D1.7

Data Management Plan

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[1] management of these research data sets.[2]

Delivery Date	2nd July 2019
Work Package	WP1
Task	T1.4
Dissemination Level	Public
Type of Deliverable	Report
Lead partner	UNIVIE

[1] Findable, accessible, interoperable and reusable.

[2] EUROPEAN COMMISSION Directorate-General for Research & InnovationH2020 Programme, Guidelines on FAIR Data Management in Horizon 2020

InteropEHRate project is co-funded by the European Union (EU) Horizon 2020 program under Grant number 826106.

This document has been produced in the context of the InteropEHRate Project which is co-funded by the European Commission (grant agreement n° 826106). All information provided in this document is provided "as is" and no guarantee or warranty is given that the information is fit for any particular purpose.

CONTRIBUTORS

	Name	Partner
Contributors	Nikolaus Forgó	UNIVIE
	Katerina Polychronopoulos	UNIVIE
	Marie-Catherine Wagner	UNIVIE
	Felix Zopf	UNIVIE
	Vincent 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
Reviewers	Laura Pucci	ENG
	Adrian Bradu	SIVECO

LOGTABLE

Version	Date	Change	Author	Partner
0.1	17-05-19	First draft of ToC	Katerina Polychronopoulos Marie-Catherine Wagner Felix Zopf	UNIVIE
0.2	22-05-2019	Update of the ToC	Katerina Polychronopoulos Marie-Catherine Wagner	UNIVIE
0.3	23-05-2019	Technical Descriptions	Katerina Polychronopoulos Marie-Catherine Wagner	UNIVIE
0.4	27-05-2019	Ethical and Data Protection Aspects	Katerina Polychronopoulos Marie-Catherine Wagner	UNIVIE
0.5	31-05-2019	Integration of Information received by Partners through the Questionnaire	Katerina Polychronopoulos Marie-Catherine Wagner	UNIVIE

0.6	05-06-19	Review	Nikolaus Forgó Katerina Polychronopoulos Marie-Catherine Wagner	UNIVIE
0.7	11-06-19	First version for internal review	Nikolaus Forgó Katerina Polychronopoulos Marie-Catherine Wagner	UNIVIE
1.0	18-06-19	First internal review	Laura Pucci	ENG
1.1	20-06-19	Second internal review	Adrian Bradu	SIVECO
1.3	22-06-19	Integration of review comments	Nikolaus Forgó Katerina Polychronopoulos Marie-Catherine Wagner	UNIVIE
1.4	22-06-19	Review of section 5.2.4	Thanos Kiourtis	UPRC
1.5	23-06-19	Review of section 5.2.2	Vincent Keunen	A7
1.6	24-06-19	Review of section 5.2.5	Salima Houta	FRAU
1.7	25-06-19	Review of section 5.2.1	Gábor Bella	UNITN
1.8	26-06-19	Review of section 5.2.6	Sofianna Menesidou	UBITECH
1.9	26-06-19	Review of sections 2 and 5.3	Francesco Torelli	ENG
1.10	27-06-19	Integration of technical partners' review	Nikolaus Forgó Katerina Polychronopoulos Marie-Catherine Wagner	UNIVIE
Vfinal	02-07-2019	Final quality check and final version for submission	Laura Pucci	ENG

ACRONYMS

Acronym	Description
A7	A7 Software
CHU	Centre Hospitalier Universitaire De Liège
DMP	Data Management Plan
EC	European Commission
EFN	Fédération Européenne Des Associations Infirmières Aisbl
EHR	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[3]
FHIR	Fast Healthcare Interoperability Resources
FMTG	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
SCUBA	Spitalul Clinic De Urgenta Bagdasar-Arseni
SIVECO	SIVECO Romania SA
S-EHR	Smart Electronic Health Record
UBITECH	UBITECH Limited
UNITN	Universita Degli Studi Di Trento
UPRC	University Of Piraeus Research Center

[3] Guidelines on FAIR Data Management in Horizon 2020 [http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf].

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

1.1. Scope of the document

This Data Management Plan (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 and will outline:

- 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 generated data regarding their privacy and confidentiality, ensure that the legal and ethical standards for data generation, use, storage, and share are applied throughout the project and in line with the overall-management of the project, as foreseen in grant agreement and consortium agreement, and that appropriate technical standards are applied for data representation. Another purpose of the DMP is to ensure that InteropEHRate activities are compliant with the H2020 Open Access policy and the recommendations of the Open Research Data pilot. The DMP will address measures the Project partners forming the Consortium will employ in order to cope with legal, ethical and privacy concerning personal data usage and to assure the application of relevant national or EU regulations, primarily the GDPR.

Throughout this DMP, the organisations participating in the InteropEHRate project will be referred to as “partners” and all partners collectively as the “Consortium”.

All InteropEHRate partners commit to respect the policies outlined in this DMP and ensure that all data are created, managed and stored according to 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 Grand Agreement have to assume responsibility for the quality control of the data generated or processed during the work on that specific task.

Information to produce this deliverable was gathered by the InteropEHRate Consortium through a questionnaire (see ANNEX 1 which follows the “Guidelines on FAIR Data Management in Horizon 2020”: The different questions concerning various aspects of data management were taken from these guidelines and were arranged according to the template provided by Horizon 2020. The structure of the data management plan is also based on this template. The different tables represent mappings of this FAIR Data Management model. The DMP reflects a current picture of the Project based on the answers from each partner.

1.2. Intended Audience

The InteropEHRate Project is a participant in the Open Research Data Pilot of the European Commission (ORD pilot), which enables open access and reuse of research data generated by Horizon 2020 projects and will follow a balanced strategy between openness of data and protection of scientific information, commercialization and intellectual property rights (IPR), privacy concerns, security as well as data management and preservation questions. Regarding open data, the InteropEHRate platform will retain compatibility with standards, best practices, and guidelines for working with open data.

This DMP is a public document. The deliverable will be available for everybody, but it will mainly be interesting for health institutions (hospitals, clinic centres), HC 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.

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 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 (if any)

Not applicable.

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”. [4]

The ISO defines it as “recorded information” (ISO 22005:2007) 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) [5].

The InteropEHRate Data Management follows the “Guidelines on FAIR Data Management in Horizon 2020”[6], 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, which 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 re-useable, for example by researchers interested in using the data in future medical research.

These principles provide guidance for scientific data management and are relevant to all stakeholders in research projects. They 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 the following model:

RESEARCH DATA LIFECYCLE[7]

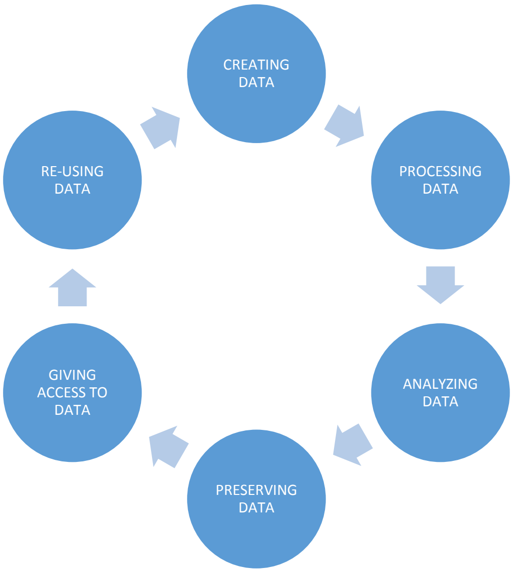

Figure 1 - Research data lifecycle

[4] https://www.merriam-webster.com/dictionary/data
[5] https://www.iso.org
[6] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
[7] Adapted from UK Data Archive : https://ukdataservice.ac.uk/media/187718/dmplanningdm24apr2013.pdf

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

Processing data: entering, transcribing, checking, validating and cleaning data, anonymizing data, describing data, manage and store 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 platforms 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”[8] .

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[9]					
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 are 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.

[8] https://www.iso.org/obp/ui#search

[9] Content adapted from “THE SIX PRIMARY DIMENSIONS FOR DATA QUALITY ASSESSMENT”, DAMA, UK [https://www.whitepapers.em360tech.com/wp-content/files_mf/1407250286DAMAUKDQDimensionsWhitePaperR37.pdf]

2.3. Collection of Information for the InteropEHRate Data Management Plan

The Information for this deliverable was gathered by the InteropEHRate Consortium through a questionnaire (see ANNEX 1), which is based on the "Guidelines on FAIR Data Management" and corresponds to a template associated with them.[10] Considering the different aspects of these recommendations the project management and the partners agreed on procedures how to map the specific details of the InteropEHRate project to the general structure provided by Horizon 2020. After various modifications the final version of the questionnaire (ANNEX 1) was sent out to the partners by UNIVIE.

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	-	ALL
Information on Technical Processes	Cross-Border Electronic Health Records (EHRs) (Section 5.2.1)	UNITN
	Smart Electronic Health Records (S-EHR) Mobile App (Section 5.2.2)	A7
	Health Care Professional (HCP) App (Section 5.2.3)	SIVECO
	Interoperability (Section 5.2.4)	UPRC
	Interoperability Profile and Standardization (Section 5.2.5)	FRAU
	Health Record Security and Privacy (Section 5.2.6)	UBITECH
Project in Pilots	-	CHU FTGM
Data Protection and Ethical Aspects	-	CHU, EFN, EHTEL, FTGM, HYG, UNIVIE

The information provided was transferred to the tables presented in the following sections. UNIVIE produced the corresponding descriptions of the technical processes on the basis of the Grant agreement in close cooperation with the responsible partners. Apart from the – in the Grant agreement -scheduled obligatory review of the Deliverable, the texts were reviewed by the respective partners and some of them additionally evaluated by the project coordinator. The DMP reflects a current picture of the Project and its effective and continued implementation will be part of a holistic and long-term data strategy.

[10] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

3. PROJECT INFORMATION

The InteropEHRate Project was conceived in reaction to an anticipated near future in which most citizens will own their own mobile health repositories, called Smart EHRs (S-EHRs). Through these S-EHRs citizens will, manage a wide range of personal health data on smart devices, regardless of whether the data were produced by health professionals, by any sensor or device or by the citizen himself/herself.

Currently, citizens' health data are stored in different IT systems scattered among several hospitals and healthcare providers. In order to better support the continuity of care, several European countries are adopting national or regional Electronic Health Records (EHRs), i.e. interoperability infrastructures that connect the EHRs of different health providers (e.g. hospitals, clinical analysis laboratories), in order to realize virtual or centralized national repositories of citizens' health records. In this way, healthcare operators have centralized access to medical records produced by different public or private providers. The national EHRs of different countries operate in different legal and standardization contexts, have different levels of maturity, and use different technologies, which makes it very difficult to exploit the available health data for healthcare and research across different countries. With the initiative called "eHealth Digital Service Infrastructure" (eHDSI)[11], the EU is starting to experiment a solution for the exchange of ePrescriptions and Patient Summaries among different countries using a common data format. This initiative follows a top-down approach, in the sense that the interaction between healthcare providers of different countries is mediated by so called National Contact Points (NCPs), each one on top of the health care providers of a single country. NCPs are provided at national level and currently support only the exchange of the aforementioned kind of documents between healthcare providers connected to NCPs. The need of a national coordination can make it difficult to quickly support new use cases.

A key goal of the InteropEHRate Project is to integrate the current interoperability solutions with a new one, based on a bottom-up approach that does not require the coordination at national level and that leaves more control of health data to the citizen and more freedom of innovation to software developers. The new solutions specified by the InteropEHRate Project will allow citizens to carry their health data with them in digital form, and manage it easily even when abroad - with benefits such as: they can obtain health data from a foreign healthcare provider directly in their own language and are able to share their personal health data directly with the Health Care Professional (HPC) of a specific healthcare organization. The same solution will allow the HPCs to obtain a translated version of a citizen's health data produced in another country directly from the citizen.

The proposed solution will also change the relationship between citizens and medical research. Scientists will be able to obtain health data directly from citizens, using transparent mechanisms for communicating the purpose of the research. Citizens will be able to donate their health data (produced by institutions or by the citizen themselves) securely and apply specific constraints on the usage of their data.

This will be obtained thanks to a new kind of product and three related communication protocols that will be openly specified by the InteropEHRate project:

[11] https://ec.europa.eu/cefdigital/wiki/display/EHOPERATIONS/eHDSI+Mission

- Smart EHRs (S-EHRs): a new model of secure mobile applications for the storage, control (i.e. data and access right management), anonymization and exchange of health data on smart devices (i.e. iOS and Android smartphones or tablets), without the obligation to store data in the cloud. The project will specify the characteristics that any secure S-EHR mobile application must have in order to be considered trustable and usable by citizens. For those citizens that want to store their health data also in a cloud, the project will specify what characteristics a storage system (called S-EHR Cloud) must have in order to be considered secure by the citizens and to interact with the S-EHR mobile applications.

- Healthcare Interoperability Protocols: The project will produce the open specification of two protocols for the exchange of health data between S-EHRs of citizens and information systems of healthcare organizations. A first protocol, called D2D (Device to Device) protocol, will allow the exchange of health data without the internet, but using short range communication technologies (e.g. Bluetooth). A second protocol, called R2D (Remote to Device) protocol, will allow the exchange of health data at distance, by means of the Internet. The R2D protocol will also allow to exchange health data with an S-EHR Cloud.

- Research Interoperability Protocol: The project will also produce the open specification ~~of~~ for a third communication protocol for the exchange of health data between any citizen using an S-EHR and research centres. It will allow scientists to engage voluntary citizens at cross-national levels in new research trials and retrospective studies, and will allow citizens to easily and securely donate health data, in pseudonymized or anonymized form.

The Project will realize a prototype of S-EHR and prototypes of libraries and components (exploitable by healthcare organizations and research centres to extend their information systems) implementing the specified protocols. The full set of software prototypes produced by the project is called “InteropEHRate Framework”.

The developed prototypes will be periodically assessed and finally validated in pilots at clinical sites, by final users (e.g. patients, researchers and health-care professionals).

However, the primary goal of this project is not t the development of new software, but the drafting of new concepts that might develop into standards which eHealth will have to fulfil, no matter who develops and offers it. The development of software within the project is needed to check that those specifications are actually implementable. Similarly the execution of Pilots (i.e. experimentations with real data and real users during the third year) is needed to check if the new software satisfies these open specifications.

4. SYNOPSIS: "FAIR" INTEROPEHRATE DATA SETS

A variety of data will be produced by the InteropEHRate Project. Most of this data falls into one of the following three major groups of data: administrative data (personal[12] and non-personal data), technical data (software code and design), 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[13] filled out by relevant partners of the InteropEHRate Consortium (A7; CHU; EFN; EHTEL; ENG; FRAU; FTGM; HYG; SIVECO; 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, Mailing lists, Partner Contact Details, Meeting Minutes and Information Technical Data: Software Code, Software Design Patient Data in Pilots: Generated Clinical Data (*i.e.,* randomly generated, fake data) and Collected Real Clinical Data (detailed in chapter 5.3 Project in Pilots)
Data Format	Administrative Data: PDF, ZIP, Google Docs, Google Sheets, Google Slides 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	Too early to say, as major pieces of software code are still being developed. Administrative Data currently is likely not larger than a few Gigabytes. Patient data: No generated/patient data has been generated/collected
Software tools for creating/processing /visualizing data	Administrative Data: Google Drive Technical Data: IDE Patient Data in Pilots: DICOMM VIEWER, HCP App and clinical research related software
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: Not likely only generated clinical data
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

[12] All partners of the project agreed to have their personal data processed for administrative purposes for the lifetime of the project.
[13] See in annex 1.

	ORGANIZATION, DOCUMENTATION AND METADATA OF DATA INTENDED TO BE PUBLISHED
Standards for Documentation and Metadata	Administrative Data: No particular standardization protocols observed. Technical Data: No Patient Data in Pilots: No
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: None identified at this stage.
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 defined prior to pilot start.
Directory and File Naming Convention	Administrative Data: Defined in the InteropEHRate Project deliverable D1.5 Technical Data: Java naming convention Patient Data in Pilots: Will be defined prior to pilot start.

	DATA ACCESS
Risks	Administrative Data: unauthorized access Technical Data: Stealing of GitLab credentials and access to source 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 GoogleDrive terms of usage Technical Data: ISO 27001, Private GitLab repository (User and Password protection) Patient Data in Pilots: Avoidance by using generated patient and clinical data
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
Procedures to Follow a Data Breach	Administrative Data: Will be defined at a later stage of the Project and documented in the next version of the DMP. Technical Data: Will be defined at a later stage of the Project and documented in the next version of the DMP. Patient Data in Pilots: Will be defined at a later stage of the Project and documented in the next version of the DMP.

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Re-Use of Data	Administrative Data: N/A Technical Data: Public deliverables will be published on the InteropEHRate website Patient Data in Pilots: Only anonymized results of clinical research will be available for re-use.
Organization/ Labelling of Data for Easy Identification	Administrative Data: N/A Technical Data: Not defined yet Patient Data in Pilots: Not applicable as such data will not be re-used by any third party.
Data Sharing Requirements	Administrative Data: N/A Technical Data: IDE, Standardization tools Patient Data in Pilots: Will be defined at a later stage of the Project and documented in the next version of the DMP.
Audience for Re-use	Administrative Data: N/A Technical Data: Anyone (software vendors, Healthcare organizations, research organizations, public and private institutions) Patient Data in Pilots: Anonymized data may be re-used by researchers
Restrictions on Re-Use of Data	Administrative Data: N/A Technical Data: Not for public deliverables Patient Data in Pilots: None identified so far. However, one the application is on the market, patients will be capable of selecting what data they want to share (and for what purpose(s)).
Publication	Administrative Data: N/A Technical Data: GitLab, scientific conferences, standardization workshops, InteropEHRate Website Patient Data in Pilots: Anonymized data may be re-used by researchers and may be published in scientific papers

	DATA PRESERVATION AND ARCHIVING
Archiving of Data for Preservation and Long-term Access	Administrative Data: Google Drive, EC Website and local repositories Technical Data: GitLab and back-up Patient Data in Pilots: Generated clinical and/or patient data will not be stored longer than necessary for the project.
Data Retention	Administrative Data: anticipated to be for at least 5 years Technical Data: at least 5 years Patient Data in Pilots: Will be stored in accordance with Article 17 GDPR
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.
Data Archives	Administrative Data: EC Portal Technical Data: Local repositories Patient Data in Pilots: According to national regulation
Long-term Maintenance of Data	Administrative Data: Not yet defined, will be presented in the next version of the DMP Technical Data: Not yet defined, will be presented in the next version of the DMP Patient Data in Pilots: Not yet defined, will be presented in the next version of the DMP

5. PROCESS-ORIENTED ANALYSIS OF "FAIR" DATA SETS

This section addresses the life cycle of administrative data and standards of project documentation.[14]

This chapter also provides descriptions of the major technical processes put in place for the InteropEHRate project activities 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.

The information reflects the current state of the activities and will be updated later in the project.

5.1. Administrative Data

A variety of administrative data (personal[15] and non-personal data) will be generated and collected during the course of the Project. Some examples are:

- Planning data, concepts; e.g. the coordination partners' 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;
- Data from the marketing and commercialization process.

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.

[14] InteropEHRate D1.5 - Quality Plan (www.interopehrate.eu/resources)

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

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. As mentioned before the structure of the tables follow the FAIR principle:

	DATA PRODUCTION AND STORAGE
Data Generated/Collected	Reports/Deliverables defined in the GA Templates Partner contact information Meeting/Web conference related material (participants' list, agenda, presentations)
Data Format	GoogleDrive PDF and Doc
Reproducibility	GoogleDrive maintains a history. All the deliverables will be uploaded to the Participant Portal website. 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 the GoogleDrive or offline through usual document readers.
Storage and Backup Strategy	GoogleDrive (The project management will move the project data to one of the paid versions of GoogleDrive), Project website and EC Portal. Living versions of deliverables are stored in the collaborative workspace of the project repository on Google Drive, in the related WP/Task subfolder; final 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.

	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 Google Drive is a collaborative space that is the main exchange means of documents among partners. As far as Software artefacts are concerned, a Software repository will be implemented and the Consortium will agree on rules later in the project.
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 • for confidential data (confidential deliverable, contact details of partners, etc.) the risk is also of non-authorized access. On Project repository, data is protected by user and password controls and by other specific security protocols granted by terms of usage of GoogleDrive.
Data Access	Data (on the Project repository) can be accessed only by specific persons indicated by the Partners Project Managers (including their deputies) to whom the Project Coordinator gives grant to access
Correct execution of the access process	Project Coordinator will be responsible for ensuring secure access protocols for administrative data.

	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 GA.
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[16] www.InteropEHRate.eu. Public deliverables are published on the project website as soon as they are approved by the EU Commission. They may also be published as draft documents on 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 website. The Project Coordinator will archive all deliverables and project related documentation on its internal cloud service for future audits until five years after the end of the Project.
Data Retention	At least 5 years.
File Formats	The consortium has not decided on any format yet. The Project Coordinator archives its copies of deliverables in the original format. All final documentation is in pdf format.
Data Archives	The archive location for project management data (deliverables and effort data reported in the Project Progress Reports) is the EC portal, which is considered an institutional archive that preserves the deliverables and other information submitted permanently.
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.

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

5.2. Technical Data

This section addresses the main software development data generated during the project. After a brief description of the goals and the intended products the life cycles of the data are discussed. The tables reflect the stage of the project at the time of writing this deliverable, as far as information is available and decisions on the processing of data have already been made. Some project activities have not started yet and so changes and updates are expected.

5.2.1. Cross-Border Electronic Health Records (EHRs)

The Health Records (HRs) of different countries operate in different legal and standardization contexts, have different levels of maturity, and use different technologies. So it is currently difficult to transfer medical records from country to country, sometimes even from institution to institution.

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. A fundamental characteristic of the HL7 FHIR[17] standard is its extensibility: As information that is not part of the basic model can be added to each health record as a custom structured child element, the data model can be extended with new fields with a view to reusability in different applications. 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. This profile will be based on the HL7 FHIR standard, 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. It will be a meta-profile integrating existing profiles, standards and models compatible with 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.

A multilingual knowledge model will provide interoperability for formally encoded and natural language data. Formal encodings include both local and international encoding schemes for diseases, procedures, care unit identifiers, etc. AI technologies such as machine translation and knowledge extraction will be used, where applicable for translations. The solution will also include natural language understanding services for extracting structured data from natural language text, as well as machine translation to other natural languages in order to make international data exchange easier and allow healthcare professionals to exploit data from citizens and health care operators of different countries to assure continuity of care as the citizen moves from country to country.

Software Development

During the InteropEHRate Project, such a health data integration platform is developed. 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

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

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.

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. The tables are organized in accordance with the FAIR principles:

	DATA PRODUCTION AND STORAGE
Data generated/collected by UNITN	Source code Data concerning general medical knowledge
Data Format	Software code for the implementation of data integration/interoperability systems is written in Java. 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.
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 results will be stored by the user. Source 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	Several MBs of data are produced
Software tools for creating/processing/ visualizing data	All such tools will be provided by UNITN.
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 source code will also be used, either brought in as background or open-source third-party tools.
Storage and Backup	Versioning systems (both local and project-wide) will be used for non-personal

Strategy	data and knowledge, Google Drive for documentation data.

	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.
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	All such tools will be provided by the software developer (UNTIN).
Project and Data Identifiers	Project and data identifiers will be specified at a later stage of the project.
Community Standard for Metadata Sharing/Integration	Standard and agreed methods will be used to attach metadata to the source 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	All software tools and the entire software developing processes of partner's institution (UNITN) will be formally security- audited by a professional external company.
Data Access	The source code will be accessible on the project-wide GitLab platform (using login + password). Non-personal medical knowledge will be accessible by using a dedicated web-based GUI of the partner (protected by username + password).
Correct execution of the access process	UNITN will provide a local developer/system administrator to check the process.

	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 source code will be released in a modular form.
Data Sharing Requirements	Apart from source code, in publications UNTIN will only use a few examples of non-personal data and, if needed, generated EHR data.
Audience for Re-use	System developers interested in adopting the results of the project may reuse released source code.
Restrictions on Re-Use of Data	The license will be defined in accordance with the Consortium Agreement and other partners.
Publication	The data sets collected/generated in the course of the development of cross-border EHR software will be published on an open website 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	Maintenance beyond the project will be up to negotiation based on the long-term needs of project partners.

5.2.2. Smart Electronic Health Records (S-EHR) Mobile App

At the moment most European citizens' health data are stored in different IT systems of various hospitals and healthcare providers. However, some European countries have already introduced national or regional Electronic Health Records (EHRs), i.e. interoperability infrastructures that connect the EMRs (Electronic Medical Records) of different health providers in order to realize virtual or centralized national repositories of citizen's health records. In this way, healthcare operators have centralized access to medical records produced by different public or private providers. Due to centralized storage architectures patients, however, do not have control over their health data.

Objectives

It is the goal of the Project that most citizens will have their own mobile health repositories, called Smart EHRs (S-EHRs). Such an S-EHR can manage a wide range of personal health data on smart (mobile) devices no matter if the data was produced by health professionals, by any sensor or device or by the citizen himself/herself.

As discussed, using the S-EHR mobile app, citizens will effectively be the “hub” of their own data and- due to the FHIR standard- will be able to quickly and safely share/gather their health data, wherever they are, just bringing the smart device (e.g. a smartphone or a smart watch) next to the terminal of the healthcare provider’s record system.

An S-EHR is a kind of 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 HER, which is under the control of an institution. The S-EHRs will be integrated with legacy EHRs and EMRs by means of standardized Application Program Interfaces (APIs) allowing the import of data from EHRs or EMRs. 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 short-range D2D 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). This way of communication is similar to contactless payment.
(2) On the internet: *The remote protocol* will allow use of the S-EHR to exchange health data at distance with healthcare operators, by means of secure communications on the internet.

Software Development

A prototype of an S-EHR mobile app will be developed and implemented. The architecture for the interaction with related services and the optional cloud service will also be designed so that the S-EHR mobile app will be able to import/share data from/with EMRs and systems of research organizations, using both short-range wireless D2D (device to device) communication and remote communication protocols.

FAIR Lifecycle of Data Set for Software Coding and Development of S-EHR Application

The following tables provide additional detail regarding the characteristics of the data generated from the tasks associated with developing the S-HER software. Again, FAIR principles are addressed:

	DATA PRODUCTION AND STORAGE
Data generated/collected by A7	Code of the integration of S-EHR App by Andaman7 (A7).
Data Format	D2D Protocol Library is provided as a Java Script file.
Software Tools for creating/processing /visualizing data	Source 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 Java Script.
Best Practices/Guidelines for Data Management	The S-EHR component developed in this task is provided as a library in Kotlin (Android), what makes its integration easy in an Android app.
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 source code by an unauthorized person. GitLab platform access requires user login and password.
Data Access	Access to source code is provided via a login in GitLab on a developer account.
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	For the S-EHR app library, the one requirement is to use an IDE allowing to read and indent 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 pushed on GitLab.

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

5.2.3. Health Care Professional (HCP) App

In the current situation it is sometimes difficult or impossible for hospitals and other health care professionals to process and transfer electronic health data from or to other institutions. In order to enable interoperability in the future, the Project will provide a tool for Health Care Professionals in the form of an app.

Objectives

Interoperate is planning to provide hospitals with a secure web app (the so-called Health Care Professional Web App) which is based on FHIR standards and will enable HCPs to securely exchange health data of their EMRs with any S-EHR and to read health data stored in federated EHRs.

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 remote protocols, to:

(1) import/export data directly from/to the S-HER on the smartphone;
(2) import/export data from/to S-EHR cloud;
(3) access integrated health records imported from EHRs/ EMRs of other healthcare providers.

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. Again, the tables are organized in accordance with the FAIR principles:

	DATA PRODUCTION AND STORAGE
Data produced/generated by SIVECO	Data necessary for designing and implementing the HCP App solution as well as for the experimental use and testing of the HCP App. The data is randomly generated data or fake data sets compliant with HL7 FHIR format.
Data Format	These aspects will be set within analysis stage – InteropEHRate Deliverable D2.2 data format and will be compliant with HL7 FHIR format.
Reproducibility	The process of data generation is reproducible. The design of HCP App solution will ensure the accessibility and reusability of generated data.
Size of the Data	Too early to say at this stage of project implementation.
Software tools for creating/processing /visualizing data	InteropEHRate partner developing the HCP App will use tools and software for creating/processing and visualizing the data, which will be determined later in the project.
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	Specific implementation of metadata for HCP App cannot be defined at this stage of the project. These aspects will be set within analysis stage of the Project. 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.
Tools for Formatting Data	None in this stage of project implementation.
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 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[18] schemes/mechanisms for data.
Community Standard for Metadata Sharing/Integration	Too early to determine specific implementation of metadata for HCP App; these aspects will be dealt with during analysis stage. The Project's standardized metadata schemas and encoding schemes for the HCP App as well as a local repository for data and associated metadata have not yet been specified/configured. The responsible partner will use naming conventions for identifiers specific to Java technology/programming.
Automatic Creation of Metadata	Too early to determine specific implementation of metadata for HCP App; these aspects will be dealt with within analysis stage

	DATA ACCESS
Risk Management	Major risks: data breach, loss of data, security threats, weak authentication (e.g. single factor passwords). The partner responsible for the development of the HCP App (SIVECO) (certified ISO 27001) will implement security solutions compliant with ISO 27001 (SR ISO/IEC 27001:2013) (IT Security). To protect privacy, security, confidentiality, intellectual property or other rights the partner will apply the requirements and stipulations of ISO 27001. The partner responsible for the development of the HCP App (SIVECO) has prepared a formal risk assessment addressing each of the major risks to data security and potential solutions in accordance with the requirements and stipulations of ISO 27001.
Data Access	At this stage only project partners have access to the data.

[18] Regulation (910/2014) on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC (23 July 2014).

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Audience for Re-use	The HCP App will be used mainly by medical staff as a demonstrator during pilots. Too early to specify/describe accurately the audience for reuse.
Restrictions on Re-Use of Data	Too early to define restrictions; these aspects will be set within the analysis stage of the Project.
Publication	The deliverables will be published on Google Drive (common repository) and submitted to the EC.

	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 (SIVECO) will preserve data for at least five years after the end of the project
File Formats	Not yet determined
Data Archives	Too early to say in this stage of project implementation
Long-term Maintenance of Data	The Administrator of HCP App and the technical support team of the partner will maintain the data for the long-term.

5.2.4. Interoperability

In the current situation there are countless interoperability infrastructures the health care institutions of the various regions and countries rely on. There are different standards for medical records and health data. Standardization and the development of an innovative common architecture that integrates the numerous interoperability infrastructures is needed for cross-country interoperability of health care providers and patients. Common technologies and an open health platform, where software vendors, institutions and citizens of different countries may securely collaborate, are needed to enable cross-country interoperability

Objectives

As noted, 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.

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.

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.

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 Deliverables and Publications
Data Format	Software in Java code (will be updated to other programming languages - if needed.) Deliverables and Publications will be in the format of documents (google docs/ docs/ pdfs), containing textual and image data
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 responsible partner (UPRC) does not currently know the size and the growth rate of this data. This data will change as follows: Software code will change weekly.
Software tools for creating/processing /visualizing data	No
Use of pre-existing data	No pre-existing data will be used.
Storage and Backup Strategy	The data will be stored locally, and on personal cloud repositories as a backup strategy.

	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, or 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.
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 Re-use	Data are 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.

5.2.5. Interoperability Profile and Standardization

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.

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.

As far cyber risks are concerned analogous work will be done regarding the identification of conformance security levels of a standard S-EHR CLOUD, *i.e.* what are the constraints that a cloud-storage service, that stores citizens' health data, has to fulfil in order to be considered trusted and reliable.

The responsible partner (FRAU) will check and make sure that the project work is in line with the relevant standards by evaluating the various processes, fostering dialogues with relevant bodies. It is a goal of these activities to generate or contribute to representative standards and proposing them to the appropriate European or international standardization bodies: InteropEHRate plans to engage with the most important European and International standardisation organisations.

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. Again, the tables are organized in accordance with the FAIR principles:

	DATA PRODUCTION AND STORAGE
Data Generated/collected by FRAU	• Documentation of specification of interoperability profile • Technical specification of interoperability profile • Software source code • Specification of privacy and security conformance levels of the S-EHR mobile app
Data Format	• technical specification of relevant standards: HL7 Profile, HL7 Resource, Web Service • documentation of other relevant standards (e.g. hl7 FHIR international patient summary) and relevant literature for studies / related work: PDF, ZIP • software programs, modules or libraries for specification: packages or executables • Specification of S-EHR mobile privacy and security conformance levels: WORD, PDF
Reproducibility	All produced data by FRAU 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	Still unknown at this time
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	According to the storage and backup strategy of the partners.

	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. 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. We will register the Implementation Guide for the Interoperability profile in the HL7 Implementation Guide Registry.
Tools for Formatting Data	For checking 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 “uri” (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 metadata is created by the Publishing Tool of the HL7 Organization.

	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[19], ISO[20], HL7[21]): - documentation of the specification of interoperability profile - technical specification of the interoperability profile - information on balloting process - publications - software source 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, version, …
Data Sharing Requirements	Research communities usually offer templates and define requirements. Standardization tools support special machine-readable formats. The responsible partner (FRAU) uses the Implementation Guide Publishing Tool for publishing the Implementation Guide of the interoperability profile.
Audience for Re-use	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

[19] European Committee for Standardization.
[20] International Organization for Standardization.
[21] Health Level 7.

	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

5.2.6. Health Record Security and Privacy

In order to ensure security and privacy for electronic health records, a set of conformance criteria, common interaction protocols and security APIs that provide interoperability among components and applications of the InteropEHRate platform are needed.

Objectives

A sophisticated authorization engine that regulates access to devices and data integrating state of the art identity management, consent management and encryption mechanisms is required. The critical question which entity verifies the value of the attributes that are used in the policies will have to be clarified. Access/usage/revocation of the encryption/decryption keys will be granted upon proper authorization. Attribute verification is the cornerstone of consent management. InteropEHRate will support a hybrid mode of attribute-verifiability in order to satisfy diverse requirements.

Software Development

The project will develop and implement reusable components of protocols and APIs considering the most advanced crypto-primitives as well as the most up to date standards. Security libraries will be designed and implemented for mobile operating systems (Android and iOS) and for HCP's web application to support all the InteropEHRate scenarios.

Software Protocols for identity management (IDM), for device identification will be designed and their relationships with CEF eID will be defined for both remote and D2D protocols. The Project will also develop software for components and the functional primitives regarding identity management, encryption and electronic signing considering the eIDAS Regulation. The capability of seamless integration with open solutions (e.g. OpenID Connect), will be a primary concern since it will broaden the applicability of the developed solution. 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/EMR/ Cloud services. Emphasis will be given to the adoption of protocols, regarding hashing, encryption, signing, that are considered battle-tested and unbroken.

Cross-border consent management and decentralized authorization mechanisms will be developed and implemented by the Consortium.

Moreover, the full security layer of the InteropEHRate platform will be implemented, integrated and tested as well. Emphasis will be laid on the delivery of sound and usable APIs that can be used by third party libraries. A crucial aspect of this task is the definition of end-to-end usage scenarios that prove the robustness of developed mechanisms. The development of the software for a secure access is the main process in this task.

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

The following tables describe additional characteristics of this secure access software in accordance with FAIR principles:

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

	ORGANIZATION, DOCUMENTATION AND METADATA OF DATA INTENDED TO BE PUBLISHED
Tools for Formatting Data	For implementation purposes the responsible partner will use IDE tool to write the source 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 source code.
Automatic Creation of Metadata	Configuration files of the source 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)
Data Access	Source code should be available only to InteropEHRate partners during the project lifetime.

	DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION
Organization/ Labelling of Data for Easy Identification	Not applicable at this stage. Information will be provided in the next version of the DMP.
Restrictions on Re-Use of Data	No restriction for the consortium members.
Publication	Not applicable at this stage. Information will be provided in the next version of the DMP.

	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	Source 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.

5.3. Project Pilots

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 trying the application from final users: citizens. This will be done by experimenting with the application's usage in at least three different international scenarios requiring the cross-border exchange of health data to fulfil the requirements of the three kinds of users – citizens, Health Care professionals and researchers. These tasks are referred to as "pilots". The sites for the pilots are the hospital partners included in the project Consortium.

Validation Process

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 to have access, even in case of not responding patient but with patient's prior consent, to emergency health data.
- 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.

During the software development phases, the Project partners intend to utilize only randomly generated/fake data sets, in order to perform test of correctness of the software and to show the running software to selected final users that will provide early feedback.

During the Pilots volunteer citizens/patients will provide their health data.

Project Pilots will begin in year 4 of the Project lifetime. Project partners have not decided yet, how health data will exactly be collected and processed during the Pilots. However, each volunteer will provide informed consent before his or her data is accessed.

The Grant Agreement also provides guidelines with respect to who these volunteers may be. The Grant Agreement states that only adult populations will be eligible, in particular men and women aged 18 through 70 years, except pregnant women. No children will be asked to join the study. Exclusion criteria are: (1) subjects with mental illnesses not able to express informed consent; (2) subjects aged <18; and (3) subjects aged >70. Eligible subjects will be informed about the objectives of the study and procedural details of the investigation. The Project Ethical Committee will draft a GDPR-compliant informed consent form, providing participating candidates of the research goals, possible adverse events, and possibilities to refuse participation or withdraw from the research at any time without consequences, etc.

Additional information about the Pilots execution will be provided in the second iteration of this DMP.

Lifecycle of “FAIR" Data Sets: Patient Data in Pilots

Data Collection	Hospitals will collect Clinical/Health patient produced by the HCPs of the hospital or imported from patient’s S-EHR (and S-EHR Cloud) using D2D and R2D protocols and will eventually store them within the hospital IT systems.
Data Format	Data collected through Hospital EHR and paper documents/form, in proprietary and HL7-V3-RIM format. Data collected through S-EHR will be in pdf and HL7-FHIR format.
Reproducibility	Reproducibility of patient data will be only possible if the data controllers of the clinical data provide the data again.
Size of the Data	Some GB
Storage and Backup Strategy	Google Drive will be used for documentation purposes. However, no personal/sensitive data will be stored on Google Drive. Further, by design, no personal/clinical data will be stored long-term at technical Project partners. Hospitals will apply their own backup strategies with respect to each data type in accordance with the hospital’s policy and national regulations.
Risk Management	Project partners responsible for coding the security and authentication part of the application will assure that patient data cannot be accessed by non-authorized parties through the testing of their encryption and other security software. Thus, when installed and used properly, the risk of a security breach of the tools, used in this project, is minimal to non-existent. In order to protect patient privacy, hospitals conduct risk assessments to identify technical, organizational and legal measures that must be followed. Actual patient data, if ever used, will only be processed during the short period of time of the use case scenario execution. Afterwards, the data will be deleted from the servers.
Access and Re-Use	No third party will have access to clinical/patient data. Every employee of a hospital has to sign a confidentiality agreement in order to be allowed to access clinical data. This is either done along with the employment contract in the case of a nurse or doctor or it is done on a case by case basis for other employees. Every access to clinical data will be monitored. However, no clinical/patient data of this project will be re-used.

6. 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.[22] Article 8(1) of the Charter of Fundamental Rights of the European Union (the 'Charter') and Article 16(1) of the Treaty on the Functioning of the European Union (TFEU) grant everyone the right to the protection of personal data concerning him or her 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).

Wherever personal data - information relating to an identified or identifiable natural person is processed[23] 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.[24]

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"[25]. It will be ensured 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.

Personal Data

During the lifetime of the InteropEHRate project two categories of personal and sensitive data are/may be generated or collected[26]:

(1) "Data related to stakeholders"(project partners and individuals working for or with Project partners and employees of the partner institutions, participants to the project meetings/web conferences, etc.): Information on these data subjects such as contact details (like e-mails and names), their signatures, authorship of deliverables, etc. is collected and processed by all partners of the Project.

[22] Article 8 EU Charter of Fundamental Rights.

[23] '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)

[24] GDPR Art 9(2)(a).

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

[26] As discussed earlier in this document, the DMP addresses only information collected or generated during the lifetime of the Project (and retained after the end of the Project). It does not currently address information collected or generated once the application becomes available to the market (to patients).

(2) “Patient’s health data in pilots” (sensitive data): At this stage of the project it is has not been decided if fake or real data, contributed by volunteer citizens, will be used for validating the functionality of the InteropEHRate platform. That is why the Project Consortium has not yet determined the procedures and criteria that will be used to identify/recruit research participants. Possibly the procedure will be determined individually by each hospital partner that wishes to use volunteers’ EHRs after consultation with and approval from the Project Ethics Committee.

The project will have two phases: a software development phase (using only fake data) during first three years, followed by experimentation (i.e. Pilot deployment and execution) phase during the fourth (half) year Technical partner will never have access to any real data. Fake data will be used by technical partners on their installations of the software within their premises and/or on the infrastructure offered by BYTE/SILO. Hospitals, however, might have to store real data on their IT systems.

As already discussed, the Project research will not include any specific clinical activity related to the InteropEHRate project. No medical services will be administered based on the patient data received, if any. In the pilot/scenario testing, data donations from patients, if needed at all, will be used only for research purposes and the application will only be tested for its capability of access to donated data, without visualization or extraction or transferring of such data to clinical researcher or healthcare personnel or any other third parties. No processing of personal and/or sensitive data is envisioned to occur during the Project outside of the pilot scenarios.

All information collected falls under the special category of personal data definition (Article 9 GDPR). Personal data will be stored in accordance to relevant national and international legislation and good practice. Sensitive data will be stored only in high-security facilities.

Informed Consent

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

Whenever personal data is collected from research participants, the participants’ informed consent must be sought by means of a procedure that meets the minimum standards of the GDPR. This requires consent to be given by a clear affirmative act. For consent to data processing to be ‘informed’, the data subject must be provided with detailed information about the envisaged data processing in an intelligible and easily accessible form, using clear and plain language.

The InteropEHRate partner will explain to research participants what the research is about, what their participation in the project will entail and in what risks they may be involved. The partner will give information as to whether data will be shared with or transferred to third parties and for what purposes and how long the data will be retained before they are destroyed. The participating patients will be also be informed about the right to withdraw consent or access to their data. They will also be told the procedures to follow should they wish to do so. They will also receive information on their right to lodge a complaint with a supervisory authority. The data subjects must also be made aware if data are to be used for any other purposes, and if it is shared with research partners or transferred to organisations outside the EU (see article 13 GDPR).

Only after making sure that participants have fully understood this information, the InteropEHRate partner will seek permission to include the patient in the project. Records documenting the informed consent procedure, including the information sheets and consent forms provided to research participants, will be kept.

The consent process(es) and the information provided to the data subjects will cover all the data-processing activities related to their participation in the InteropEHRate research.

If in the course of the InteropEHRate research project any significant changes to methodology or processing arrangements that have a bearing on the data subjects' rights or the use of their data, should occur, the data subjects will be made aware of the intended changes, and their express consent further use of the data will be sought.

If the Consortium makes a decision in favour of real patient data in pilots, the informed and explicit consent materials will be drafted at Consortium level within the next several months with input from the Ethics Committee.

Privacy by Design

To innovate ethically and responsibly, researchers and developers apply the concept of 'privacy by design', which provides a framework for focusing the design of systems, databases and processes on respect for data subjects' fundamental rights. A wider concept of 'data protection by design', now included in the GDPR, requires the implementation of appropriate technical and organisational measures to give effect to the GDPR's core data-protection principles.[27] Data protection by design is one of the best ways to address the ethics concerns that arise within a research Project.

Minimisation of data is essential in this respect.

Data processing must be lawful, fair and transparent. It should involve only data that are necessary and proportionate to achieve the specific task or purpose for which they are collected[28]

The partners of the InteropEHRate project will only collect data that are needed to meet the research objectives. Data sharing will be limited to a subset of health-related information, strictly necessary for scenario testing.

[27] Articles 5 and 25 GDPR.

[28] Article 5(1) GDPR.

Whenever personal data is collected, there are both ethical and legal obligations to ensure that participants' information is properly protected. This is fundamental to safeguarding their rights and freedoms, and minimising the ethics risks related to the data processing. In the InteropEHRate project data security is provided on all levels. The tables in Section 5 give details on the architecture.

Only authorized users will have access to digital information. The Project will adopt recommendation and standards provided by ENISA[29]. It is the goal of all project partners to mitigate the risk for all participating patients. That is why the use of generated clinical data is considered for the first test phase.

Deletion and Archiving of Data

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

Reuse of Data

A potential later use of the InteropEHRate platform may permit medical researchers to use data sets for the purpose of conducting medical research. The procedure for this possibility has not yet been addressed by the Project partners. As a result of this effort, ethical and legal considerations may arise with respect to large scale or big data processing and they will be discussed in a later version of this DMP. Specifically, Project partners will consider how to best provide researchers access to EHR data that may be collected via the application in an anonymized (aggregated) way. It is too early to discuss specific implementation of data models for the HCP application, however.

Publication of Results

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

Integration of Future Aspects

The DMP is a living document and further considerations will be made, especially with respect to donated health records and potential for the application to be used as a research platform, in later iterations of this document.

[29] https://www.enisa.europa.eu/

ANNEX 1

Data Management Plan Questionnaire

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:[30] 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?[31]

[30] 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

[31] Deliverables are not included in this question.

- How will you collect the data? In what formats?

- How will you trace the collected data?
  - *For example, how do you trace the provenance of the data collected? Or 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.