



REVERT

DELIVERABLE

D2.1

FAIR Data management Plan (DMP)



DOCUMENT INFORMATION

AUTHORS

Name	San Raffaele	Name	Guadagni Fiorella
Name	San Raffaele	Name	Ferroni Patrizia

Deliverable lead partner	San Raffaele
Contributing partner(s)	ULSS 4 - ProMIS, MALMO UNIV., GXP, BAM, UMEA UNIV., BioV, UCAM, IRO IASSY, SMS, LIH, IMAGO-MOL, Olomedia, UNITOV
Work Package	WP1 Project management and coordination
Deliverable type	ORDP: Open Research Data Pilot
Contractual delivery date	30.09.2020
Actual delivery date	17.12.2020
Updated version delivery date	31.12.2023
Dissemination level	Public
Version	1.4.9

ABSTRACT

A document summarizing the way in which the data will be managed inside the REVERT ecosystem, from REVERT partner and third parties. The document is aimed to describe how the project will implement the FAIR principles and how will be applied the open access paradigm or, in case of restrictions, will be explicated the reason for that choice.

STATEMENT OF ORIGINALITY

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

DISCLAIMER

The content of this deliverable represents the views of the author only and is his/her sole responsibility; it cannot be considered to reflect the views of the European Commission and/or the Consumers, Health, Agriculture and Food Executive Agency (CHAFEA) or any other body of the European Union. The European Commission and the Agency do not accept any responsibility for use that may be made of the information it contains.

TABLE OF CONTENTS

DOCUMENT INFORMATION	1
TABLE OF CONTENTS	3
LIST OF FIGURES	4
LIST OF TABLES	4
Changelog.....	5
1 Executive summary	7
2 Introduction.....	10
3 Legal framework.....	13
3.1 Principles of GDPR.....	13
3.2 GDPR and National Data Protection Law of the Countries of the Participants Institutions for what concern the “Revert” context.....	18
4 The "REVERT" context	20
4.1 Why AWS.....	21
5 Data Summary.....	23
5.1 Aim and origin of data collection/generation	23
5.2 Other data collected in the database.....	24
5.3 Other data collected in the database.....	25
5.4 Open access.....	25
5.4.1 Open Research Europe	27
6 FAIR data	29
6.1 Making data findable, including provisions for metadata.....	29
6.2 Making data openly accessible.....	29
6.3 Making data interoperable.....	30
6.4 Increase data re-use (through clarifying licences).....	30
7 Allocation of resources.....	31
8 Data security.....	32
8.1 Security measures implemented in the REVERT project.....	32
8.2 Data security requirements.....	33
9 Ethical aspects.....	34
10 Other issues	36
10.1 Specific National Regulations	36
Attachement 1 - National Data Protection Laws of the Countries of the Participant Institutions relevant within the “REVERT” context	37
End of D2.1	39

LIST OF FIGURES

Figure 1. Structure of REVERT project	10
Figure 2. Risk Assessment Methodological framework.....	11
Figure 3. Principles of Data Protection	15
Figure 4. REVERT architecture	20
Figure 5. Existing (blue) and prospective (orange) AWS Regions	21
Figure 6. Magic quadrant for cloud database management systems (from: Gartner, Magic Quadrant for Cloud Database Management Systems, Donald Feinberg, Merv Adrian, Rick Greenwald, Adam Ronthal, Henry Cook, 23 November 2020.)	22

LIST OF TABLES

Table 1. List of participants	7
Table 2. Data sets overview	23
Table 3. Data sets description and utility.....	24

Changelog

VERSION	PUBLICATION DATE	CHANGE	AUTHOR
1.0	24.07.2020	<ul style="list-style-type: none"> Initial version - Draft 	San Raffaele
1.1	05/08/2020	<ul style="list-style-type: none"> 2.1 Aim and origin of data collection/generation 3.1 Making data findable, including provisions for metadata 5 Ethical aspects 	San Raffaele
1.2	07/09/2020	<ul style="list-style-type: none"> Added IMAGO-MOL contribution: P. 5 Added IMAGO-MOL's Observation 3.1 Making data findable, including provisions for metadata 3.2 Making data openly accessible 3.3 Making data interoperable 3.4 Increase data re-use (through clarifying licences) 	San Raffaele
1.3	28/09/2020	<ul style="list-style-type: none"> Introduced c.2, 3 and 7 Modified table c. 4 and introduced c.4.1 	San Raffaele
1.4	21/10/2020	<ul style="list-style-type: none"> New version, with introduction on GDPR & Data protection topics 	San Raffaele
1.4.1	24/10/2020	<ul style="list-style-type: none"> Modified image no.1 "List of participants" p.4 	San Raffaele
1.4.2	28/10/2020	<ul style="list-style-type: none"> Modified image no.2 "Structure of REVERT project" p. 6 	San Raffaele
1.4.3	02/12/2020	<ul style="list-style-type: none"> Modified page 13 "5 The 'REVERT' context" Added p.16 "6.2 Other data collected on the database" Added p.19 "6.3.1 Open Research Europe" 	San Raffaele
1.4.4	09/12/2020	<ul style="list-style-type: none"> Modified section 7 FAIR data P.20 Modified 11.1 Specific National Regulations P.27 	San Raffaele

D2.1 FAIR Data management Plan (DMP)

1.4.5	11/12/2020	<ul style="list-style-type: none">• Document revision	San Raffaele
1.4.6	17/12/2020	<ul style="list-style-type: none">• Corrections and comments	BAM
1.4.7	18/12/2020	<ul style="list-style-type: none">• Change in document template	San Raffaele
1.4.8	23/12/2020	<ul style="list-style-type: none">• Some little corrections applied• Added “Allocation of resources”	San Raffaele
1.4.9	31/12/2023	<ul style="list-style-type: none">• Document updating	San Raffaele

1 Executive summary

THE REVERT PROJECT

The targeted therapy for advanced colorectal cancer patients (REVERT) project addresses the Topic “Systems approaches for the discovery of combinatorial therapies for complex disorders” (SC1-BHC-02-2019), which belongs to the Work Programme Part: Health, demographic change and wellbeing of the H2020 Work Programme 2018-2020.

The REVERT project is managed by a consortium that involves many excellent EU research institutes and SMEs, as well as several certified biobanks in different countries (Italy, Spain, Germany, Romania, Luxembourg, Sweden) to ensure access to a large amount of data and the involvement of numerous clinical centres.

The project sees the participation of the following organizations: The project sees the participation of the following organizations:

REVERT - targeted therapy for advanced colorectal cancer patients

List of participants

Participant No*	Name	Country
1 (Coordinator)	SAN RAFFAELE ROMA SRL (SR)	ITALY
2	AZIENDA ULSS 4 VENETO ORIENTALE (ProMIS)	ITALY
3	MALMO UNIVERSITET (MU)	SWEDEN
4	GENXPRO GMBH (GXP)	GERMANY
5	BUNDESANSTALT FUER MATERIALFORSCHUNG UND - PRUEFUNG (BAM)	GERMANY
6	UMEA UNIVERSITET (UMU)	SWEDEN
7	BioVariance GmbH (BioV)	GERMANY
8	FUNDACION UNIVERSITARIA SAN ANTONIO (UCAM)	SPAIN
9	REGIONAL INSTITUTE OF ONCOLOGY IASSY (IRO IASI)	ROMANIA
10	Instituto Murciano De Investigaciones Biosanitarias - Hospital Universitario Santa Lucia (IMIB - HGUSL)	SPAIN
11	LUXEMBOURG INSTITUTE OF HEALTH (LIH -	LUXEMBURG
12	Clusterul Regional Inovativ de Imagistica Moleculara si Structurala Nord-Est (IMAGO - MOL)	ROMANIA
13	Olomedia Srl (OLO)	ITALY
14	University of Rome Tor Vergata (UNITOV)	ITALY

Table 1. List of participants

The REVERT consortium is coordinated by IRCSS San Raffaele Pisana, a leading research institute which is devoted to fostering excellence in care and assistance for patients through innovative biomedical and pre-clinical research that can be readily transferred into clinical practice, and to the organisation and management of health services.

D2.1 FAIR Data management Plan (DMP)

The Organizations involved, for the purpose of enhancing the REVERT project, formalized a Joint Controller Agreement (JCA pursuant to art. 26, GDPR) to jointly establish the purposes and methods of processing the personal data collected, as well as define their respective responsibilities, roles and relationships based on the Consortium Agreement and this document.

The JCA provides that each Joint Controller guarantees, to the extent of its competence, to process personal data in accordance with applicable data protection regulations, storage methods and provisions inherent to IT and security incidents and through the use of IT tools that comply with the technical organizational requirements in force, as well as through its own staff, duly informed and instructed in accordance with Article 32 of the GDPR.

OBJECTIVES OF REVERT PROJECT

The REVERT project will develop an innovative decision support system based on Artificial Intelligence (AI), using the experience and data collected from the "real world" by the various hospitals operating in the EU health system.

The project will use an information system, the REVERT PLATFORM, that will collect and harmonize the data and will consist of two main components: the REVERT-DataBase (REVERT-DB) and the REVERT-AI (REVERT software systems).

The REVERT software systems shall guarantee data integrity and privacy management, in compliance with the EU GDPR (EU Reg. 2016/679), with the beneficiaries' national regulations and with the EU Charter of Fundamental Rights.

The REVERT project has the specific objective of addressing the challenge of understanding, at system level, the pathophysiology of metastatic colorectal cancer (mCRC) in patients who respond well or poorly to therapies, in order to design an optimal strategy for mCRC, with therapeutic interventions adapted to the patient's condition and diagnostic characteristics.

REVERT aims to select optimal therapy by implementing predictive models (REVERT-AI) based on a large database of cases (REVERT-DB). The project will thus use an information system, the REVERT PLATFORM (RP). The RP will incorporate and harmonize data from the European BioBank Network as well as from local clinical databases.

Colorectal cancer (CRC) is among the most frequent causes of cancer-related deaths [<https://seer.cancer.gov/statfacts/>]. Around 50% of CRC patients with local or regional disease will develop distant metastases, while almost 21% of CRC patients present metastases already at the time of diagnosis, with a 5-year survival rate of 13.8% [<https://seer.cancer.gov/statfacts/>].

The specific objectives of REVERT are:

- To build the REVERT-DataBase (RDB) in order to re-analyse and characterise the pathophysiology of mCRC and to investigate the causes of positive or negative responses to treatments based on established therapeutic interventions in patients with unresectable mCRC. RDB is built upon a large number of standardized biobank samples with related structured data, and on clinical databases from several major European clinical centres;
- To build a sophisticated computational framework based on AI to predict patient responses to combinatorial therapies for mCRC care, based on the analysis of new, potentially prognostic biomarkers (e.g., gene mutations, epigenetic changes, gene expression profiling signatures) as molecular predictors of therapeutic response, treatment resistance or disease outcome, in comparison with established therapeutic interventions;
- To assess the significance of biomarkers and molecular predictors of therapeutic response or disease outcome in patients with mCRC using an innovative AI-model, feeding on data from large European clinical databases and biobanks;

D2.1 FAIR Data management Plan (DMP)

- To screen and characterize molecular mechanisms of already approved drugs as potential novel candidates for combinational therapy to effectively target metastatic cancer by using patient tumour-derived organoid models;
- To validate the health, economic and social impact of the model in preclinical/clinical studies across Europe and
- To build an EU-wide network among SMEs, Research Institutions, Clinical Centres and Biobanks focused on R&D in the field of AI-Health for the synergistic and accelerated development of personalised medicine.

2 Introduction

This Data Management Plan (hereinafter also "this Document" or "DMP") illustrates the research area, the objectives to be achieved and the methodologies to be followed in analysing the issues in the data management related to the REVERT Project.

Before presenting and discussing the DMP, we would like to make a brief statement on the COVID-19 pandemic that has appeared in Europe approximately as of February 2020. On behalf of the REVERT consortium, we would like to thank all healthcare professionals for the valuable work that they have done and continue to do in the medical field around the world, offering best possible care to the people.

The Data Management Plan is structured as follows:

- **Legal Framework:** main principles of GDPR and relevant National Data Protection Law in the context of REVERT
- **The "REVERT" context:** introduction to the REVERT context
- **Data Summary:** description of the aim and the origin of data
- **FAIR data:** description of how the FAIR principles are applied to REVERT
- **Allocation of resources:** analysis of the costs to make data FAIR and responsibilities for data management within the project
- **Data security:** description of main data security requirements and measures
- **Ethical aspects:** description of ethical or legal issues
- **Other Issues:** description of main issues regarding IT & privacy

The methods for achieving the objectives of REVERT embrace several strategies, activities and outputs that will be delivered over the four years' timeline of the project.

The key features and bodies of the management structure of the REVERT project are illustrated in Figure 1.

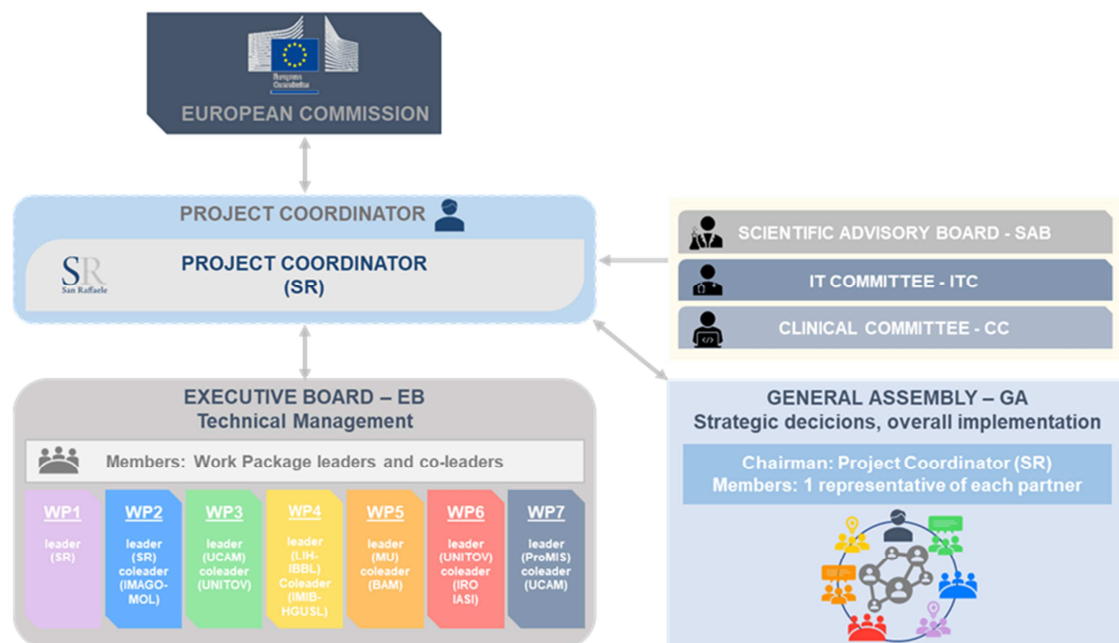


Figure 1. Structure of REVERT project

In the initial phase, a risk assessment and mitigation plan has been drafted, pinpointing various potential risks associated with software implementation and data management, which will be integrated into the next version of this DMP. Any risks pertaining to the main IT and legal aspects will be managed as outlined in Figure 2, i.e., will be identified, analysed, evaluated and mitigated in the best possible way as detailed in the Risk Management Plan:



Figure 2. Risk Assessment Methodological framework

Currently, the context in which the new artificial intelligence technologies will be used and that relies prominently on the databases attached to this project has been established (Context Establishment), allowing to draw conclusions on the requirements on the Data Protection Regulation for all relevant aspects of the project, including

- Guidelines suitable to facilitate the analysis of the risks related to the treatment;
- Guidelines suitable to facilitate the treatment;
- Training plan with the aim of promoting knowledge of the principles set out in the data protection legislation and
- Report with the technical and organizational measures adapted or being implemented by each participant to protect the personal data processed.

Moreover, any issues that arise during the next phases and that call for modifications of the IT and Data protection aspects will be included in revised Data Management Plans, this process being an ongoing one.

In connection with WP6 “Prospective study to evaluate the association between individual profiles and clinical outcomes in 1st line metastatic colorectal cancer (mCRC) patients”, the document “PR Critical risk” has been updated by considering among the potential regulatory risks that have emerged also that relating to “Potential regulatory delays” according to which it should be noted that Regulation EU No 2017/745 (Medical Device Regulation – MDR Regulation) which considers AI software used in the medical field as a “class IIa medical device” was not yet in force at the time the project was presented and the GA signature.

Based on this regulation, medical devices do not follow the phase classification of pharmacological studies but are in an initial pre-marketing (pre-registration) or post-marketing process.

After an extensive discussion with the Italian Regulatory Authorities (together with the Spanish and Romanian counterparts) and the reclassification request to the UNITOV Ethics Committee (Lead Beneficiary WP6), followed by the request for acceptance by of the IRBs of all partners involved in the pilot clinical study, according to the EU Regulation, the study previously classified as a “Pilot Phase II Prospective Clinical Multicentre Pharmacology Study” was reclassified as a “Pilot Prospective Clinical Medical Device, Non-Profit, Interventional, Pre-Commercial - Early Phase”.

However, the reclassified protocol was not changed in any way other than study reclassification.

Furthermore, the same document confirmed the delays in the project's development caused by the COVID-19 pandemic.

D2.1 FAIR Data management Plan (DMP)

With reference to the risk assessment plan (D.1.2), the work envisaged in this Deliverable identified, among the relevant risks, those related to data protection and IT.

In the context of the risks identified in that document, it emerged the need to conduct a Data Protection Impact Assessment (“DPIA”) on the personal data resulting from the REVERT project in order to highlight the process of describing the processing, assessing its necessity and proportionality and managing any risks to the rights and freedoms of individuals arising from the processing.

This analysis and activity were also addressed in the General Assembly and in the periodic reports sent to the European Commission.

In particular, the DPIA clarified that for the identification of risks in the field of Artificial Intelligence, the Revert project Partners, following the development of AI, analysed and considered as reference, as discussed in the IT Committee, the main topics of the AI assessment checklist called ALTAI, established by the independent High Level Expert Group on Artificial Intelligence (AI HLEG), set up by the European Commission. In particular, the following topics were considered:

- algorithm training will be carried out on the dataset in order to reduce the error rate and/or technical faults;
- self and/or external audits will be carried out periodically to identify potential risk profiles, ethics and the use of AI algorithms in order to update the REVERT project;
- in order to ensure the controlled development of AI, actions are planned to reduce the risk of discrimination (e.g., race and/or gender); ensure the professional training of operators;
- the algorithms are limited to understanding the factors that influence the course of the disease, in order to verify whether one of the treatments under investigation produces better results than the others or not, therefore, they do not aim to identify/track patients at risk.

Taking into account the purposes for which the AI algorithms will be used (which will have the purpose to predict the impact of the drugs on the disease of mCRC contracted by the patient), the measures adopted, the AWS cloud platform used and the aggregation of the data being published, the DPIA considered that the processing operations envisaged by the Revert project present a comprehensively medium-low degree of risk for the rights and freedoms of the data subjects involved in the processing itself, based on the evidence emerging from the risk analysis.

Therefore, it has been possible to proceed with the implementation of the project and the start of the processing without further technical and organizational measures, but it is suggested to implement some measures.

3 Legal framework

With respect to the legal framework, the focus lies on the main principles of GDPR and the relevant National Data Protection Laws, also applicable at the time of the collection of the clinical study, in the context of REVERT.

3.1 Principles of GDPR

The European Regulation 679/2016 (also called “GDPR”, “regulation 2016/679” or “GDPR Regulation”) approved by the European Parliament came into force on May 24, 2016 and is fully applicable in all Member States since May 25, 2018, has introduced significant changes with respect to each national legislation framework on data privacy and personal data processing within the EU.

The European Parliament therefore wanted to protect the rights of European citizens through a series of indications that aim to form a common basis for all the countries joining the Union. The GDPR in fact has the objective of harmonizing the privacy rules of the various states and is aimed at facilitating the development of a digital single market through the creation and promotion of new services, applications, platforms and software.

The regulation relates to the protection of individuals with regard to the processing and free circulation of personal data.

The GDPR arises from specific needs, as indicated by the EU Commission itself, of legal certainty, harmonization and greater simplicity of the rules regarding the transfer of personal data from the EU to other parts of the world. It is also a necessary and urgent response to the challenges posed by technological developments and new models of economic growth, taking into account the needs for the protection of personal data increasingly felt by EU citizens.

The regulation, however, leaves some spaces of autonomy that remain with the individual Member States in disciplining, in a more specific way than the GDPR, some aspects not included in the competence of the EU based on the principle of attribution. This circumstance could give rise to conflicts between the various national supervisory authorities who find themselves specifically regulating and applying the provisions of the GDPR at national level.

The regulation bases all the information on the concept of personal data.

The definition of personal data is contained in Article 4 of the GDPR:

- *“personal data”: means any information relating to an identified or identifiable natural person (‘data subject’); 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;*

Article 4 also contains other important definitions applicable to the reference context:

- *‘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;*
- *“consent” of the data subject means any freely given, specific, informed and unambiguous indication of the data subject’s wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her;*

D2.1 FAIR Data management Plan (DMP)

- *“personal data breach” means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed;*
- *“genetic data”: personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question;*
- *“biometric data”: personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data;*
- *“data concerning health”: personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status.*

Article 9 of the GDPR highlights a further classification of **special categories of** personal data, or so-called "particular data", previously called "sensitive data" and states that:

- *“Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited”.*

With reference to the definition of data concerning health, it is important to note that Recital 35 of GDPR provides that:

“Personal data concerning health should include all data pertaining to the health status of a data subject which reveal information relating to the past, current or future physical or mental health status of the data subject. This includes information about the natural person collected in the course of the registration for, or the provision of, health care services as referred to in Directive 2011/24/EU of the European Parliament and of the Council to that natural person; a number, symbol or particular assigned to a natural person to uniquely identify the natural person for health purposes; information derived from the testing or examination of a body part or bodily substance, including from genetic data and biological samples; and any information on, for example, a disease, disability, disease risk, medical history, clinical treatment or the physiological or biomedical state of the data subject independent of its source, for example from a physician or other health professional, a hospital, a medical device or an in vitro diagnostic test”.

Paragraph 2 lists all the cases relating to the non-application of Article 9 par.1:

- *(a) “The data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject”;*
- *(g) “processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject”;*
- *(h) “processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of Union or Member State law or pursuant to contract with a health professional and subject to the conditions and safeguards referred to in paragraph 3”;*
- *(j) “processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the*

D2.1 FAIR Data management Plan (DMP)

right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject”.

In order to address the issues relating to data protection expressively with concern to the REVERT project, it is necessary to take into account the following principles laid down in the GDPR:

- I. Pursuant to art. 5 of the GDPR, the principles governing the processing of personal data are:

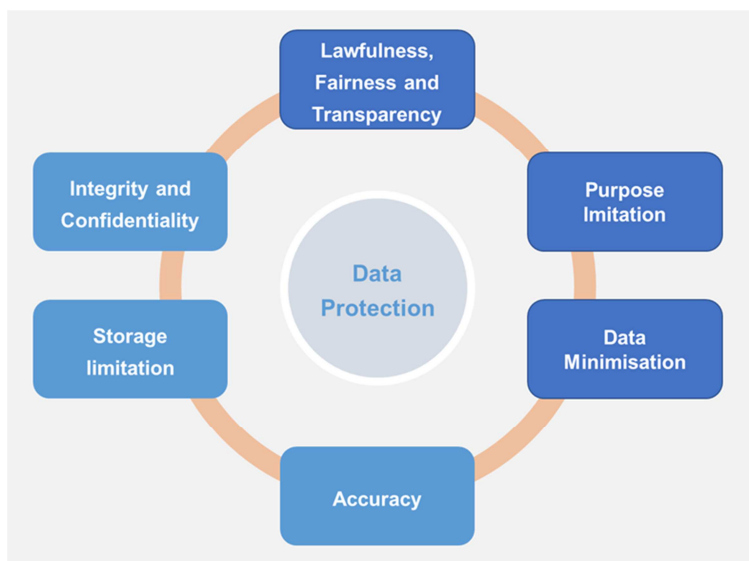


Figure 3. Principles of Data Protection

In particular, in accordance with art. 5 regarding the “Principles relating to processing of personal data”, personal data shall be:

- a. processed lawfully, fairly and in a transparent manner in relation to the data subject (‘lawfulness, fairness and transparency’);
- b. collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes (‘purpose limitation’);
- c. adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (‘data minimisation’);
- d. accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay (‘accuracy’);
- e. kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject (‘storage limitation’);
- f. processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures (‘integrity and confidentiality’).”

D2.1 FAIR Data management Plan (DMP)

Pursuant to paragraph 2 of art. 5, the “controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 (‘accountability’)”.

The "Principle of Accountability" was therefore introduced and any party concerned with data processing must demonstrate the adoption of an overall process of organizational, procedural and technical measures for the protection of personal data to the Data Controller.

- II. In accordance with article 6 “Processing shall be lawful only if and to the extent that at least one of the following applies:
 - a. the data subject has given consent to the processing of his or her personal data for one or more specific purposes;
 - b. processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;
 - c. processing is necessary for compliance with a legal obligation to which the controller is subject;
 - d. processing is necessary in order to protect the vital interests of the data subject or of another natural person;
 - e. processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
 - f. processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child”.
- III. Pursuant article 7, it is important to note that regarding the Conditions for consent:
 1. Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data;
 2. If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding;
 3. The data subject shall have the right to withdraw his or her consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent;
 4. When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract”.
- IV. Pursuant to articles 13 and 14 of the GDPR, transparent information, communications and methods must be provided also for the exercise of the rights of the data subject (communications referred to in articles 15 to 22 and article 34 of the GDPR and specific national applications);
- V. In Article 25 “Data protection by design and by default”, the GDPR deals with the requirements on privacy “by Default” and “by Design”:
 - With reference to "Privacy by Default", the Data Controller is required to put in place measures to ensure that, by default, only the personal data necessary for each specific processing purpose are processed. This obligation applies to the amount of data collected, the scope of processing, the retention period and accessibility. In compliance with this logic, the Data Controller must ensure that, by default, personal data are made accessible to a defined number of people (in other words, it adopts solutions that allow for the tracking of the subjects who acquire the data)

D2.1 FAIR Data management Plan (DMP)

- With reference to "Privacy by Design", when determining the means of processing, the Data Controller, taking into account the state of the art and the implementation costs, as well as the nature, scope of application, context, purposes and of the risks for the rights and freedoms of the data subjects deriving from the processing, is required to implement adequate technical and organizational measures, aimed at implementing the principles of data protection (lawfulness, correctness, transparency - legitimacy - minimization - accuracy - conservation - integrity and confidentiality).
- VI. It should be taken into account that according to Article 22, Automated individual decision-making, including profiling, "the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her";
- VII. Moreover, in accordance with article 32, the EU Regulation states that the Data Controllers (and Processors) must implement appropriate technical and organizational measures to guarantee a level of security adequate to the risks for the protection of the personal data collected, taking into account the state of the art and the costs of implementing the measures.

The GDPR already indicates the main measures that allow to mitigate the risks arising from the processing of personal data:

- a) pseudonymisation and encryption of personal data;
- b) the ability to ensure the confidentiality, integrity, availability, resilience of the systems and services used;
- c) the ability to promptly restore availability and access to personal data in the event of a physical or technical accident;
- d) procedures for testing and regularly verifying the effectiveness of measures to ensure the safety of treatment.

With reference to security measures, it is essential that technical and organizational measures suitable to guarantee a level of security adequate to the risks for the protection of personal data processed are put into place.

In this regard, for example, pseudonymization of data on the RDB is ensured, so that data are entered into the platform pseudonymised in order not to be traced back to specific individuals and only the uploading centre can trace them back to the data subjects.

- VIII. With reference to potential aspects of a transfer of personal data to third countries or international organisations, Chapter V, GDPR, Article 44 specifies the general principle for such a transfer. In particular, some aspects concerning transfer to third countries could become a potential risk without adequate guarantees.
- IX. Article 68 and following, regulate the figure and role of the European Data Protection Board;
- X. Article 89 regarding the "Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes"
 - 1. *Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner*

D2.1 FAIR Data management Plan (DMP)

2. *Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes*
 3. *Where personal data are processed for archiving purposes in the public interest, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18, 19, 20 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes*
 4. *Where processing referred to in paragraphs 2 and 3 serves at the same time another purpose, the derogations shall apply only to processing for the purposes referred to in those paragraphs”.*
- Finally, it should be noted that the regulation lists criteria the violation of which entails severe administrative penalties.

Furthermore, the EDPB in the "Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research", adopted on 2 February 2021, specified that in the field of scientific research *“informed consent to participate in the medical research project is a necessary requirement, with some exceptions for situations where consent cannot be given (incapacitated individuals, emergency situations etc.). However, such consent can and should be distinguished from ‘consent as a legal basis for processing of personal data’ in Article 6(1)(a) of the GDPR. Taking into consideration that Article 6 (1) GDPR provides for legal bases other than consent and Article 9 (2) GDPR provides for exemptions other than explicit consent, it is foreseeable and not incompatible (with ethical standards) that the other legal grounds can be relied on for the processing health data for scientific research purposes. However, when relying on another legal basis in Article 6 other than consent and one of the other exemptions in Article 9 (2) GDPR, the ‘ethical’ requirement of informed consent for participation in the medical research project will still have to be met. In the GDPR-framework, this can be perceived as one of such additional safeguards as foreseen in Article 89(1) GDPR that should be in place when processing personal data for scientific research purposes.”*

GDPR and National Data Protection Laws of the Countries of the Participant Institutions relevant within the “REVERT” context.

The GDPR recognizes the possibility for Member States to maintain or introduce additional conditions, including limitations, for the processing of genetic data, biometric data or data relating to health.

An overview of the respective national regulations with relevance to the project and in compliance with the GDPR are listed in **Attachment 1**, having a focus on the processing of health data and research.

3.2 GDPR and National Data Protection Law of the Countries of the Participants Institutions for what concern the “Revert” context

The GDPR recognizes the possibility for Member States to maintain or introduce additional conditions, including limitations, for the processing of genetic data, biometric data or data relating to health.

D2.1 FAIR Data management Plan (DMP)

An overview of the respective national regulations with relevance to the project and in compliance with the GDPR are listed in Attachment 1, having a focus on the processing of health data and research.

Hosted in Italy, the REVERT DataBase (RDB) follows the rules required by the Italian and EU Regulations.

Furthermore, with regard to the European Health Data Space, as mentioned in section 5.4 of this Document, it should be noted that in the “Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space¹”, the European Commission carried out and published, on 12 February 2021, an “Assessment of the EU Member States’ rules on health data in the light of GDPR” in relation to the specific national legislation of all European countries including those of reference for REVERT project.

The purpose of the assessment is to examine and present the EU Member States’ rules governing the processing of health data in light of the GDPR, with the objective of highlighting possible differences and identifying elements that might affect the cross-border exchange of health data in the EU and examining the potential for EU level action to support health data use and re-use.

¹ For more informations and updates: https://health.ec.europa.eu/health-data-space_en

4 The "REVERT" context

The REVERT project aims to build a database (RDB) that collects the clinical data of patients from different biobanks located in several European countries as described above.

The format with which the necessary clinical data will be collected will be established by the San Raffaele team and proposed to all biobanks. The current plan is to use either a CSV (Comma Separated Values) or a JSON (JavaScript Object Notation) file, containing all the information necessary for processing.

Since the data of different biobanks can vary with respect to the measurement units that parameters are stored with or to the formulae with which secondary data are calculated, the data have to be harmonized upon import to ensure that the corresponding entries of all the records have the same meaning across all institutions that provide the samples the RDB will be built upon.

Once a common format has been established, automated routines using machine learning and artificial intelligence algorithms will be devised for uploading the data to the RDB, eventually unifying the processing irrespective of the team entering the data.

All the processing will be executed exclusively on an AWS platform. The REVERT partners that will work with the biobank and clinically archived data will use the respective AWS space and/or service in order to process all the data within the secure AWS environment, avoiding any external data transfer. The only data that can be downloaded to external IT infrastructures will be the final results of the processed data. This workflow ensures to control and trace when and by whom data have been accessed and exported. All the risks concerning IT security will be mitigated by using an Amazon cloud platform that is certified according to GDPR and characterized by high-level security standards. Figure 4 shows the diagram of the REVERT architecture.

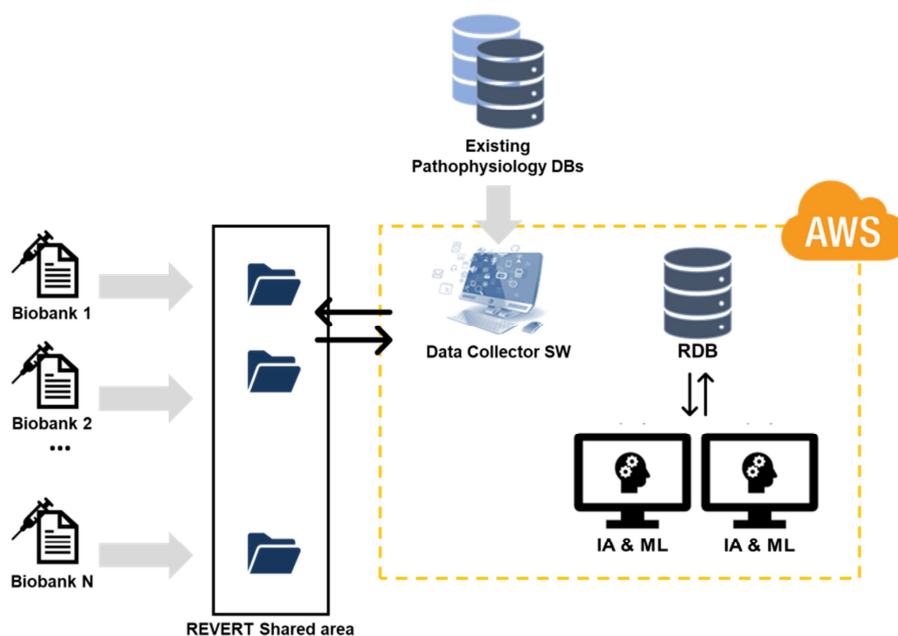


Figure 4. REVERT architecture

4.1 Why AWS

Amazon Web Services (AWS) is Amazon's cloud platform and is perhaps the most complete and most widely used in the world, offering more than 175 comprehensive data center services globally. AWS offers significantly more services and features within the services themselves than any other cloud provider, including computing power, storage and database infrastructures, machine learning and artificial intelligence options, data lakes, analytics and interfacing with the Internet of Things.

AWS also has the most advanced features within its services. For example, AWS offers a wide range of databases tailored to different types of applications, allowing to choose the best tool for an intended application thus providing a cost- and performance-optimized solution.

AWS has the largest and most dynamic community, with millions of active customers and tens of thousands of partners around the world, guaranteeing rapid and comprehensive support as well as offering solutions developed specifically for use on the Amazon cloud.

AWS is designed to be the most flexible and secure cloud computing environment on the market to date. The core infrastructure is designed to meet security requirements for highly sensitive organizations such as military, global banking, and more, underpinned by a wide range of cloud security tools. AWS supports 90 security standards and compliance certificates, and all 117 AWS services that store customer data offer the ability to encrypt that data. Furthermore, all AWS services can be used in compliance with the GDPR. This means that, in addition to benefiting from all of the measures that AWS already takes to maintain service security, customers can deploy AWS services as a key part of their GDPR compliance plans.

AWS offers the largest global cloud infrastructure. No other cloud service provider offers the same amount of regions, each with multiple connected Availability Zones (AZs) with low latency, high throughput, and highly redundant networks. AWS operates in 77 AZs spread across 24 geographic regions worldwide, has announced plans for nine additional AZs and three additional AWS Regions in Indonesia, Japan and Spain. The AWS Region / Availability Zone model has been recognized by Gartner as the recommended approach for running enterprise applications that require high availability.

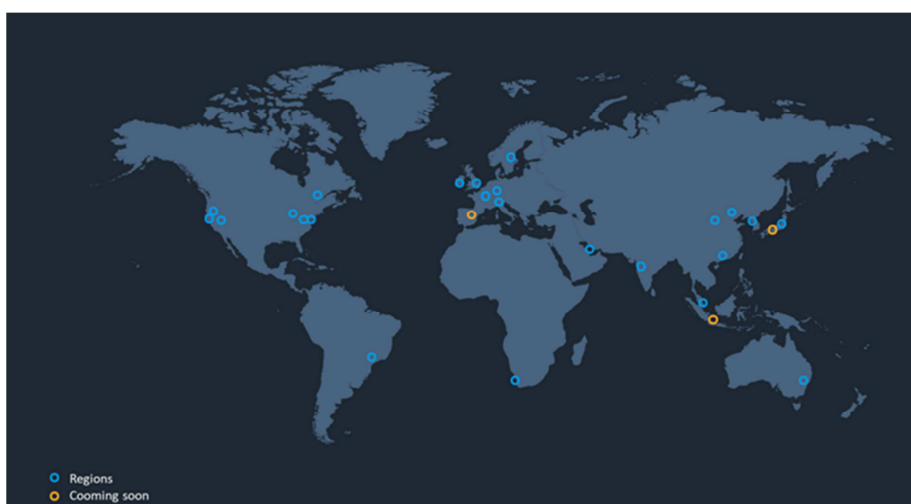


Figure 5. Existing (blue) and prospective (orange) AWS Regions

D2.1 FAIR Data management Plan (DMP)

Gartner Research, one of the leading global research and advisory firms, places AWS in the leader quadrant in the new 2020 magic quadrant for Cloud Infrastructure & Platform Services (CIPS) solutions. CIPS services, in the context of the magic quadrant, are defined as "standardized and highly automated offers, in which infrastructure resources (e.g. processing, network and storage) are complemented by integrated platform services".



Figure 6. Magic quadrant for cloud database management systems (from: Gartner, *Magic Quadrant for Cloud Database Management Systems*, Donald Feinberg, Merv Adrian, Rick Greenwald, Adam Ronthal, Henry Cook, 23 November 2020.)

5 Data Summary

The targeted therapy for advanced colorectal cancer patients (REVERT) project will develop an innovative decision support system based on AI, through the use of the REVERT PLATFORM that will collect and harmonize biobank and other clinical data.

The REVERT PLATFORM will consist of two main components: the REVERT-DataBase and the REVERT-AI.

The REVERT-DataBase (RDB) will collect data from different sources and will feed the AI algorithms. It will receive data from different biobanks located in several European countries.

5.1 Aim and origin of data collection/generation

The aim is to build a sophisticated computational framework based on AI able to predict patient responses to combinatorial therapies for mCRC care, based on the analysis of new, potentially prognostic biomarkers.

To achieve the objectives, patient data will be collected from different biobanks located in various European countries as described above.

The integrity of the data entered into the database will be ensured and the tools (Keys, Indexes, Constraints and Triggers) to implement data integrity will be developed. In addition, customized software scripts to connect different database formats and automatically collect the data will be built and ran periodically.

The BioBIM (InterInstitutional Multidisciplinary Biobank) of the SR Institute, together with all the other biobanks/databases and the SMEs, will jointly work on the infrastructural design, which will be implemented as an ICT framework for integrating data from the different databases coordinated in the project. This infrastructure will consist of an information system to extract and harmonize the data obtained from the different project partners, serving as the basis for the decision support systems to be developed. For this purpose, all available parameters will be selected and the standardization criteria – which guarantee data intercomparability across all sources and are mandatory for the successful development of predictive algorithms – will be defined.

One important task for the REVERT software system is to ensure the integrity of data entered into the database. Each data set should be identified by a unique primary key, all tables should be linked using foreign keys, and the database should not allow redundant data or untraceable data to be stored. In this context, tools (keys, indexes, constraints and triggers) which set the framework for data integrity will be developed. On the other hand, most of the partners have different database formats so that customized software scripts structuring import and connecting the databases will be generated, eventually allowing the automated collection of data.

In order to provide an overview of the different data sets that are currently and will be produced in the REVERT project, the following table shows the data type, the origin of the data and the format, in which the data will be presumably stored.

ID	DATA TYPE	ORIGIN	FORMAT
1	Biobank clinical data	Biobanks	CSV

Table 2. Data sets overview

Table 3 describes the data set and the purpose of the data collection in relation with the objectives of the project. Additionally, it shows the data utility for clarifying to whom the data might be useful.

ID	DATA TYPE	DESCRIPTION & PURPOSE	UTILITY
1	Biobank/Clinical data	Biobank data with indication of: <ul style="list-style-type: none"> - Personal data of the donor - Biological data - Process related data - Rates (part of the sample used, e.g. given to a university for research) - Pandemic data for COVID-19, when available 	Every biobank will provide clinical data to feed the RDB. The data in the RDB will be used to train the AI algorithms and to produce future predictions.

Table 3. Data sets description and utility

The data will be obtained from the biobanks/clinical archives used by the partners. In some cases, if the data cannot be transferred automatically, a special import procedure will be established.

RDB is built upon a large number of standardized biobank samples with related structured data, and clinical databases from several major European clinical centres and has a central role in the REVERT Platform as it will contain all the information to enable the REVERT AI modules and the according preclinical and clinical studies.

As already reported in the paragraph on the dissemination of the results, RDB follows the rules required by the Italian and EU Regulations and will implement the following standards:

- Standard PREanalytical Code (SPREC).
- International Statistical Classification of Diseases and Related Health Problems (ICD X)
- Entity Attribute Value (EAV)
- EAV/CR (Entity-Attribute-Value with Classes and Relationships)

5.2 Other data collected in the database

In addition to the previously identified data sets, closely linked to research on colorectal cancer, a dedicated area will be set up which, if available, will allow the collection of clinical data relating to the current pandemic situation (COVID-19).

Such data will be made available by the REVERT consortium to interested, reputable bodies for epidemiological and health surveys upon explicit request and will remain, as all data provided by biobanks and clinical archives, in the AWS infrastructure.

5.3 Other data collected in the database

The data that will feed the Revert Database (RDB) originates from different biobanks, participating in the project, present in several states of the European Union.

In order to avoid heterogeneous data formats, a standard format will be defined that all biobanks will have to respect. This will ensure that all biobanks have a "common language" and that they provide the same data using the same measurement units and calculation standards as any indices.

The uniformity of the data is an important requirement as it allows to limit errors arising from different reference systems or the use of different mathematical formulae for the calculation of the indicators that will be subsequently processed by artificial intelligence algorithms.

In this initial phase, the format that the files containing all the clinical data must adhere to will also be established (e.g. csv, JSON, txt, etc.).

What types of data will the project generate/collect?

RDB is built upon a large number of standardized biobank samples with related, structured data, and clinical databases from several major European clinical centres and has a central role in the REVERT Platform as it will contain all the information to enable the REVERT AI modules and the respective preclinical and clinical studies.

5.4 Open access

The "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020"¹ illustrates the rules on open access to scientific peer reviewed publications and research data that beneficiaries have to follow in projects funded or co-funded under Horizon 2020.

The concept of Open access "(OA)", as indicated in the aforementioned Guidelines refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable.

Moreover, always taking into consideration the Guidelines, there are the following aspects and definitions that are reported in order to illustrate the main points. In the context of research and innovation, "scientific information" can mean:

1. Peer-reviewed scientific research articles (published in scholarly journals) or
2. Research data (data underlying publications, curated data and/or raw data).

The two main routes to open access are:

- A. Self-archiving / 'green' open access;
- B. Open access publishing / 'gold' open access.

¹ https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

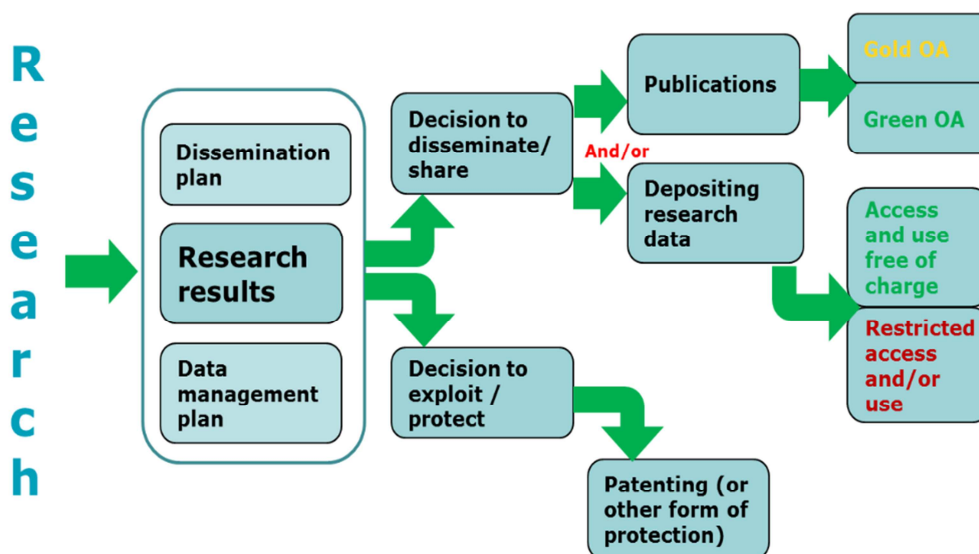
D2.1 FAIR Data management Plan (DMP)

The dissemination and communication of REVERT project results in peer-reviewed scientific publications will be also ensured through open access for any user. In particular, we will pursue immediate Gold open access to the deposited publication and the bibliographic metadata via the publisher website or PubMed Central® (PMC) as soon as possible after final peer-review and acceptance. The REVERT partners are also planning to deposit at the same time the aggregated research data needed to independently validate the results presented. These data might, indeed, represent a valuable resource to address secondary questions and, if used in combination with other study data, will facilitate meta-analyses and future research planning.

With reference to Research data, it is also necessary to illustrate that:

- Open access to research data refers to the right to access and reuse digital research data under the terms and conditions set out in the Grant Agreement;
- In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images;
- Users can normally access, mine, exploit, reproduce and disseminate openly accessible research data free of charge;
- The legal requirements for participating projects are set out in Article 29.3 of the Model Grant Agreement, included by default in the Grant Agreement (but can be removed by opting out);
- Not all data can be open.

In order to clarify the main aspects and provide evidence about the possible choices that will be adopted in terms of open access, the following graph is shown*



Graph: Open access to scientific publication and research data in the wider context of dissemination and exploitation

In this context, it is necessary to review the reasons according to which the Commission (as regards Horizon 2020) allows to opt out of open access publishing.

As indicated in the aforementioned Guidelines, “Projects can therefore opt out at any stage (either before or after signing the grant) and so free themselves retroactively from the obligations associated with the conditions if:

- Participation is incompatible with the obligation to protect results that can reasonably be expected to be commercially or industrially exploited;

D2.1 FAIR Data management Plan (DMP)

- Participation is incompatible with the need for confidentiality in connection with security issues;
- Participation is incompatible with rules on protecting personal data;
- Participation would mean that the project's main aim might not be achieved;
- The project will not generate / collect any research data;
- There are other legitimate reasons (you can enter these in a free-text box at the proposal stage).

The Commission's approach can therefore be described as "as open as possible, as closed as necessary".

During the lifetime of a project, a total opt-out is possible for any of the reasons highlighted above. In this case, Article 29.3 is removed from the Grant Agreement via an amendment".

In particular, as regards the REVERT project, the following issues will be taken into consideration during the Risk Assessment and in future updates of the Data Management Plan:

- The RDB and AI services will be open to all partners and, after project completion, will also be available to EU research institutions for future studies;
- Consistently with the fundamental principle that science should be a collaborative effort to maximize common benefits, the REVERT investigators will put efforts into data sharing taking into account, however, all the issues related to patient privacy and consent, as well as the correct use of data that must be regulated based on the EU GDPR regulation;
- We must keep in mind that, currently, informed data privacy consent forms generally do not contemplate that the data could be publicly shared with others and, therefore, patients do not explicitly opt out of sharing data publicly;
- Moreover, failure of a consent form to include an opt-in statement to publicly share data does not constitute tacit approval to share data publicly;
- The RDB could be presently used only within national and international research projects aimed to knowledge advancements in the biomedical field and not-for-profit intent, warranting the correctness of data sharing and use only for scientific purposes.

The European Commission, on 3 May 2022, presented a "Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space" to establish a European Data Space, a health-specific ecosystem comprising common structures, rules, standards and practices and a governance structure that aims to:

- empowering individuals by giving them greater digital access to and control over electronic personal health data, at national and European level, and supporting their free movement, as well as fostering a true single market for electronic health record systems, relevant medical devices and high-risk artificial intelligence systems (primary use of data);
- provide a coherent, reliable and efficient system for the use of health data for research, innovation, policy-making and regulatory activities (secondary use of data).

The Proposal will be discussed in the Council and the European Parliament in order to proceed to the subsequent approval.

5.4.1 Open Research Europe

Open Research Europe is a new publishing platform that provides an Open Science protocol to the H2020 beneficiaries. It represents a great step forward for EU R&I programme beneficiaries and research communities from all natural sciences, social sciences and humanities fields. As reported in the "Aim and Scope" section of the official website:



D2.1 FAIR Data management Plan (DMP)

“Open Research Europe publishes articles across the Natural Sciences, Engineering and Technology, Medical Sciences, Agricultural Sciences, Social Sciences and Humanities stemming from Horizon 2020 funding.

Each publication on Open Research Europe must have at least one author who has been, or still is, a recipient of a Horizon 2020 grant.

Articles must be original (not duplications). All research is welcome and will be published irrespective of the perceived level of interest or novelty; confirmatory and negative results, as well as null studies are all suitable.

All articles are published using a fully transparent, author-driven model; the authors are solely responsible for the content of their article. Invited peer review takes place openly after publication, and the authors play a crucial role in ensuring that the article is peer reviewed by independent experts in a timely manner. Articles that pass peer review will be indexed in a number of bibliographic databases and repositories, following formal approval by these services.

Open Research Europe is an Open Research platform: all articles are published open access under a CC-BY license; the publishing and peer-review processes are fully transparent; and where applicable, authors are asked to include detailed descriptions of methods and to provide full and easy access to the source data underlying the results in order to improve reproducibility.”

The screenshot shows the top navigation bar of the Open Research Europe website. It includes the European Commission logo, a search bar, and navigation links for 'How to Publish', 'About', 'My Account', and 'Sign In'. Below the navigation bar, there is a section titled 'Rapid & Transparent Publishing' with a brief description and a 'SUBMIT YOUR RESEARCH' button. The page is powered by FICOO Research.

The REVERT consortium will always consider the possibility to contribute with the obtained results, reserving the privilege to publish only those data that will not occur in privacy or legal issues.



6 FAIR data

The REVERT project, in relation to the question of keeping data open, applies the principle “as open as possible, as closed as necessary”.

Below we detail the ways in which REVERT applies the FAIR principles.

6.1 Making data findable, including provisions for metadata

The dataset, although findable, will not be publicly accessible, but a formal request for access to the data can be submitted to the “REVERT Data Access Committee”, specifically set up for this aim, which will define from time to time the ways of accessing data, the specific limitations, also regarding privacy requirements or limitations given by the consortium in a specific case.

The “REVERT Data Access Committee” will be made up of a representative from each institution that contributed to the creation of the database and will evaluate case by case the possibility of granting access to the data, for the sole purpose of scientific research.

6.2 Making data openly accessible

After a positive evaluation of the “REVERT Data Access Committee”, the methods of access to the data and the limitations which the applicant must comply with will be defined. The user will require to be authenticated on the REVERT systems, in order to log every action executed on the system and to monitor the way of using the data. It will also be necessary to sign a Data Use Agreement (DUA) that prevents the use of data for prohibited purposes.

As emerged from the Data Protection Impact Assessment, it was clarified that all the data on the RDB are pseudonymised, to minimize the risks related to privacy.

The dissemination and communication of the project results will take place through peer-reviewed scientific publications and open access for any user to the deposited publication will be ensured, as previously stated.

How will this data be exploited and/or shared/made accessible for verification?

The dissemination and communication of project results in peer-reviewed scientific publications will be also ensured through open access for any user. In particular, we will pursue immediate Gold open access to the deposited publication and the bibliographic metadata via the publisher’s website, PubMed Central® (PMC) or other open publishing tools as soon as possible after final peer-review and acceptance. The REVERT partners are also planning to deposit at the same time the aggregated research data needed to independently validate the results presented. These data might, indeed, represent a valuable resource to address secondary questions and, if used in combination with other study data, will facilitate meta-analyses and future research planning.

How will the identity of the person accessing the data be ascertained?

Each person wishing to access the data must ask to be authenticated. Such request must be discussed by the “REVERT Data Access Committee” and approved. In this way the REVERT consortium can monitor when and how data is accessed by which external user of the RDB resources.

In addition, a set of API will also be defined that will allow to retrieve the research results in an aggregate way to allow research validation.

6.3 Making data interoperable

To allow interoperability of data, it must be possible to combine and compare them with different sources, both by humans and by machines. To strengthen interoperability, aggregate data supporting articles (both on Open Research Europe and/or other research portals) will be stored in a non-proprietary open file format and described using standard vocabulary (if available).

Even the pseudonymised data, present on the RDB for users who will be allowed access, will be provided with a data dictionary, in order to allow the user to understand the values stored on the RDB and apply them to future researches.

6.4 Increase data re-use (through clarifying licences)

The ultimate goal of FAIR is to foster the reuse of data. Consistently with the fundamental principle that science should be a collaborative effort to maximize common benefits, the REVERT investigators will put efforts into data sharing taking into account, however, all the issues related to patient privacy and consent, as well as the correct use of data that must comply with the EU GDPR regulations. We must keep in mind that, currently, informed consent forms generally do not contemplate that the data could be publicly shared with others and, therefore, patients do not explicitly opt out of sharing data publicly. Moreover, failure of a consent form to include an opt-in statement to publicly share data does not constitute tacit approval to share data publicly. We agree with the concerns raised in the viewpoint by M. C. Gibson (Moving From Hope to Hard Work in Data Sharing. *JAMA Cardiol.* 2018;3(9):795-796) that “data could be used in a way that was never intended by patients or researchers and could result in unforeseen damages”.

In light of these considerations, access to data (in any case in pseudonymised form) will be evaluated from time to time by the designated committee and access to them will be allowed only after signing binding agreements on the purpose of use and confidentiality of the data. In this way, re-use will be guaranteed, while maintaining control over the use of data.

Also, the inclusion of additional documentation alongside the data will ensure that the data are understandable and thus reusable. As a general rule, someone who is not familiar with the data should be able to understand what it is about by using only the metadata and documentation provided.

7 Allocation of resources

For all institutions that have planned to publish scientific works, an average of about 5,000 euros has been reserved (within the item “Other direct costs”). Further costs will be borne by the individual institutions.

The cost of the cloud platform, that will allow the execution of the project, was borne by the coordinator, for the initial 6 years (4 planned for the project and 2 further years). The renewal after this initial period will be managed by the “REVERT Data Access Committee”, a committee of which one member from each organization of the REVERT consortium will be part, with the financial participation of all the institutions involved.

A further renewal of at least 4 years is already envisaged. This cost, not included in the project costs, will be shared among all the participating institutions and always managed by the "REVERT Data Access Committee"

The responsibility for the data management of the REVERT project will be borne by each institution. Each of them will identify a representative who, together with the ICT committee and the coordinator, will take care of this aspect.

8 Data security

The REVERT project has to address privacy and information security during software design and development and will be supported by a sub-contractor highly specialized in privacy and cyber-security regulations for all participating EU countries.

The aim is to guarantee cyber-security and legally compliant archiving of data respecting specific EU and national recommendations.

The data transformation from input sources to the outputs produced has many challenges regarding the need to ensure, in every step of the process, data availability exclusively to the pre-authorized participants, and a constant monitoring of data quality. This generally implies that the risk to infringe confidentiality and corrupt the quality of data is minimized.

In depth, for software and systems involved in the project, it will be necessary to define data security requirements in terms of:

- Logistical security of data communications from sources (e.g. Biobanks) to any system involved
- Logistical security of the individual components of the systems involved
- Mapping of authorized accesses to each system involved
- Logic of functional and technical monitoring of the data to guarantee non-alteration
- Logic tracing of the data transformation
- System procedures for changes with identification of participants and tasks
- Procedures for analysing the adequacy of data procedures
- Escalation procedures for data or other security breach

8.1 Security measures implemented in the REVERT project

One of the most important activities is the risk assessment within REVERT and the definition of strategies to avoid, mitigate or remove risks.

The approach used in the REVERT project will be that of risk assessment, evaluated according to quality management models. As required by current regulations, all the most modern technologies will be used, based on GDPR compliance and ISO 27001: 2013 certification. As an example, the following technologies will be used:

1. To mitigate the possible risks of intrusion related to classical architectures, it was decided to use DOCKER containers, which exceed the limits of the current common virtualized infrastructures and allow greater control of the various compartments;
2. To mitigate the risk of dissemination of sensitive data via the Internet – as required by current regulations – all transactions will take place via an HTTPS protocol, with the acquisition of SSL certificates (with quarterly renewal), which will identify the ownership of the instance. Furthermore, VPN (LAN to LAN) will be used for the partners' network;
3. To mitigate the risk of dissemination of sensitive data due to data theft, an encryption system - independent of the DBMS of the data present in the backups - will be created; and
4. To manage the risk of unauthorized access to the system operator, a rigorous system of consent management will be introduced, able to manage in a capillary manner the privileges relating to each individual area of the system.

All the project partners have to ensure during every stage of the project that technical security measures are put in place to guarantee data security and prevent cyber threats.

D2.1 FAIR Data management Plan (DMP)

To do this, the following activities are planned during all the project execution:

- Constant vulnerability assessment, network scan, code review and risk management activities to identify vulnerabilities and remedy them;
- Draw up clear and punctual plans that make incident management and disaster recovery possible;
- Develop policies and procedures relating to IT security and the development of secure software;
- Ensuring privacy by design at every stage of the project;
- Application of international security standards;
- Use of encryption formats of the exchanged input and output flows;
- Securing the perimeter of the DataBase on all levels of the OSI stack (from the physical to the application level);
- Calculation of the hash of records (exchanged, stored) to ensure that there is no corruption; and
- Mapping and risk analysis of the entire data use cycle.

8.2 Data security requirements

The data transformation from input sources to the outputs produced has many challenges regarding the need to guarantee, in every step of the process, data availability exclusively to the pre-authorized participants, and constant monitoring of data quality. This generally implies that the risk to confidentiality and quality of the data is minimized. In depth, for software and systems involved in the project, it will be necessary to define data security requirements in terms of:

- Logistical security of data communications from sources (e.g. Biobanks) to any system involved;
- Logistical security of the individual components of the systems involved;
- Mapping of authorized accesses to each system involved;
- Logic monitoring of functional and technical data to guarantee non-alteration;
- Logic tracing of the data transformation;
- System changes procedures with identification of participants and tasks;
- Procedures for analysing the adequacy of data procedures; and
- Procedures for the data scalability or other security breach.

Privacy by Design is based on the principle of incorporation of privacy, starting from the design of the related IT support applications:

- Prevent the correction of problems assessed at the design stage to avoid the introduction of new risks Privacy as default setting. The hardware and software devices will collect only the data necessary for the pre-established purposes. Possibilities to be evaluated: use of Pkey technology, API access (row, column) and volume tracking Privacy incorporated in the project, for example, the use of pseudonymization techniques, data minimization, encryption (i.e. MD5);
- Security granted during the whole product or service lifecycle;
- Visibility and transparency of processing so that data protection is verifiable; and
- User centrality and therefore respect for rights, timely and clear answers to requests for access.

9 Ethical aspects

The REVERT project requires to set up the Ethical, Social and Humanities Support Package (ESH-SP) with the aim to ensure that all project activities comply with good practices as well as legal requirements for ethical, psychological, privacy and personal data protection issues, taking care also of gender issues and gender equality.

The ESH-SP will offer partners guidance and support regarding ethics, social and privacy issues raised by the project, as opposed to ethics and privacy issues raised by the potential impact of results of the project.

A key objective for the ESH-SP is to establish the research ethics, social and privacy framework and procedures for carrying out the demonstration cases for testing the platform and integrated solutions.

To this end, the ESH-SP will provide:

- Important information, analysis, and checklists for research ethics and data protection, as well as key sources of ethical principles and standards in relation to project objectives;
- A clear roadmap with instructions and research protocols templates for End Users needing to obtain approvals from research ethics committees or Data Protection Authorities;
- Template informed consent and information sheets for research participants; and

The REVERT Project involves human participants who are not healthy volunteers, nor vulnerable individuals (e.g. pregnant women, certain elderly populations or persons judged as lacking mental capacity) or groups or children/minors. The human participants to the REVERT studies are metastatic colorectal cancer (mCRC) patients.

Full details of the recruitment, inclusion and exclusion criteria and informed consent procedures are reported within the Clinical Trial Template. REVERT investigators will inform and obtain approval from the Institutional Review Board (IRB) for the conduct of a study at a named site, the protocol, informed consent documents and any other written information that will be provided to the patients and any advertisements that will be used. Written approval will be obtained prior to recruitment of patients into a study. Proposed amendments to the protocol and documents will be submitted to IRB.

Per GCP guidelines, the investigator will be responsible for ensuring that an annual update is provided to the IRB to facilitate continuing review of the study and that the IRB is informed about the end of the study.

The REVERT project complies with:

- WMA Declaration of Helsinki;
- Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine (Oviedo, 4 April 1997) (Oviedo Bioethics Convention);
- EU Directive 2005/28/EC of 8 April 2005 laying down principles and detailed guidelines for good clinical practice as regards investigational medicinal products for human use as well as the requirements for authorization of the manufacturing or importation of such products (OJ L 91, 9.4.2005, p. 13);
- EU Regulation No 536/2014 of the European Parliament and of the Council on clinical trials on medicinal products for human use, repealing Directive 2001/20/EC (OJ L 158, 27.5.2014); and
- Directive 2001/20/EC (Clinical Trials Directive), Regulation EU No 536/2014 (Clinical Trials Regulation), Regulation EU No 2017/745 (Medical Device Regulation) or Regulation EU No 2017/746 (In-Vitro Diagnostic Medical Devices Regulation)

D2.1 FAIR Data management Plan (DMP)

In particular, Regulation EU No 2017/745 (Medical Device Regulation – MDR Regulation) led to the reclassification of the REVERT pilot clinical trial. Based on this regulation, medical devices do not follow the phase classification of pharmacological studies but are in an initial pre-marketing (pre-registration) or post-marketing process.

Specifically, the study previously classified as a “Pilot Phase II Prospective Clinical Multicentre Pharmacology Study” was reclassified as a “Pilot Prospective Clinical Medical Device, Non-Profit, Interventional, Pre-Commercial - Early Phase.

All REVERT investigators ensure respect for patients and for human dignity and fair distribution of the benefits and burden of research, and that they will protect the values, rights and interests of the research participants, whose participation will be entirely voluntary, who will provide a fully informed consent that has been approved by IRB.

All REVERT investigators ensure that the research methodologies do not result in discriminatory practices or unfair treatment.

Informed consent form and detailed information sheets:

- Are written in a language and in terms they can fully understand;
- Describe the aims, methods and implications of the research, the nature of the participation and any benefits, risks or discomfort that might ensue;
- Explicitly state that participation is voluntary and that anyone has the right to refuse to participate and to withdraw their participation, samples or data at any time — without any consequences; and
- State how biological samples and data will be collected, protected during the project and either destroyed or reused subsequently.

10 Other issues

Consistently with the fundamental principle that science should be a collaborative effort to maximize common benefits, the REVERT investigators will put efforts into data sharing taking into account, however, all the issues related to patient privacy and consent, as well as the correct use of data that must be regulated based on the EU GDPR regulation.

The REVERT DB (RDB) will implement the following standards briefly summarized:

- Standard PReanalytical Code (SPREC);
- International Statistical Classification of Diseases and Related Health Problems (ICD X);
- Entity Attribute Value (EAV); and
- EAV/CR (Entity-Attribute-Value with Classes and Relationships).

The Data Governance & Privacy has to be in line with the needs required by the activities to be implemented:

- Protect data during storage and transit through the network;
- Ensure risk awareness through constant log analysis and monitoring of actions and roles on various network file systems;
- Allow preventive, corrective actions in real time against vulnerabilities or detected incidents that could represent a danger to the data;
- Provide assessment tools for the effectiveness of security policies; and
- Implement data recovery mechanisms to restore access to data and systems when an incident affects availability.
- Furthermore, based on regulatory provisions, the data collected for the purposes of scientific research and public interest will be protected to minimize the data collected through effective pseudonymisation techniques (digital health);
- Automatic analysis of any non-anonymized data; and
- Data Breach notification procedure.

10.1 Specific National Regulations

During the execution of the project, any critical issues arising from the application of specific regulations enacted in individual countries of partners operating in the REVERT consortium will be assessed and managed.

Attachement 1 - National Data Protection Laws of the Countries of the Participant Institutions relevant within the “REVERT” context

COUNTRY	PARTICIPANT ORGANISATION NAME	DATA PROTECTION NATIONAL LAWS	DATA PROTECTION AUTHORITY
ITALY	<ul style="list-style-type: none"> ▪ (COORDINATOR) SAN RAFFAELE ROMA SRL (SR) ▪ AZIENDA ULSS 4 VENETO ORIENTALE (PROMIS) ▪ OLOMEDIA SRL (OLO) ▪ UNIVERSITY OF ROME TOR VERGATA (UNITOV) 	<p>PERSONAL DATA PROTECTION CODE - LEGISLATIVE DECREE NO. 196/03 AND SUBSEQUENT AMENDMENTS;</p> <p>LEGISLATIVE DECREE 10 AUGUST 2018, N.101 - PROVISIONS FOR THE ADAPTATION OF NATIONAL LEGISLATION TO THE PROVISIONS OF REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, OF 27 APRIL 2016, RELATING TO THE PROTECTION OF INDIVIDUALS WITH REGARD TO PROCESSING OF PERSONAL DATA, AS WELL AS THE FREE CIRCULATION OF SUCH DATA AND REPEALING DIRECTIVE 95/46 / EC (GENERAL DATA PROTECTION REGULATION)</p> <p>https://www.gazzettaufficiale.it/anteprima/codici/datiPersonali</p>	<p>https://www.garanteprivacy.it/</p>
GERMANY	<ul style="list-style-type: none"> ▪ GENXPRO GMBH (GXP) ▪ BUNDESANSTALT FUER MATERIALFORSCHUNG UND -PRUEFUNG (BAM) ▪ BIOVARIANCE GMBH (BIOV) 	<p>FEDERAL DATA PROTECTION ACT (BDSG) –</p> <p>FEDERAL DATA PROTECTION ACT OF 30 JUNE 2017, AS LAST AMENDED BY ARTICLE 10 OF THE ACT OF 23 JUNE 2021</p> <p>(BUNDESDATENSCHUTZGESETZ, NEUFASSUNG 2018 - “BDSG”)</p> <p>https://www.gesetze-im-internet.de/englisch_bdsge/englisch_bdsge.html</p>	<p>https://www.bfdi.bund.de/DE/Home/home_node.html</p>
SWEDEN	<ul style="list-style-type: none"> ▪ UMEA UNIVERSITET (UMU) ▪ MALMO UNIVERSITET (MU) 	<p>THE SWEDISH DATA PROTECTION ACT (2018:218) (SWE. LAG (2018:218) -THE “DATA PROTECTION ACT”</p> <p>THE SWEDISH DATA PROTECTION REGULATION (2018:219) (SWE. FÖRORDNING (2018:219) - THE “DATA PROTECTION REGULATION”</p> <p>https://www.riksdagen.se/en/documents-and-laws/</p> <p>https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/lag-2018218-med-kompletterande-bestammelser_sfs-2018-218</p>	<p>https://www.imy.se/</p>

D2.1 FAIR Data management Plan (DMP)

COUNTRY	PARTICIPANT ORGANISATION NAME	DATA PROTECTION NATIONAL LAWS	DATA PROTECTION AUTHORITY
SPAIN	<ul style="list-style-type: none"> FUNDACION UNIVERSITARIA SAN ANTONIO (UCAM) SERVICIO MURCIANO DE SALUD INSTITUTO MURCIANO DE INVESTIGACIONES BIOSANITARIAS - HOSPITAL UNIVERSITARIO SANTA LUCIA (IMIB - HGUSL) 	<p>ORGANIC LAW 3/2018, OF 5 DECEMBER, ON THE PROTECTION OF PERSONAL DATA AND GUARANTEE OF DIGITAL RIGHTS - THE "DATA PROTECTION ACT" ("LEY ORGÁNICA 3/2018, DE 5 DE DICIEMBRE, DE PROTECCIÓN DE DATOS PERSONALES Y GARANTÍA DE LOS DERECHOS DIGITALES")</p> <p>https://www.boe.es/buscar/act.php?id=boe-a-2018-16673</p>	<p>https://www.aepd.es/es</p>
ROMANIA	<ul style="list-style-type: none"> REGIONAL INSTITUTE OF ONCOLOGY IASSY (IRO IASI) CLUSTERUL REGIONAL INOVATIV DE IMAGISTICA MOLECULARA SI STRUCTURALA NORD-EST (IMAGO-MOL) 	<p>LAW NO. 190/2018 ON IMPLEMENTING MEASURES TO REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL OF 27 APRIL 2016 ON THE PROTECTION OF NATURAL PERSONS WITH REGARD TO THE PROCESSING OF PERSONAL DATA AND ON THE FREE MOVEMENT OF SUCH DATA AND REPEALING DIRECTIVE 95/46/EC (GENERAL DATA PROTECTION REGULATION)</p> <p>https://www.dataprotection.ro/index.jsp?page=legislatie_primara&lang=ro</p>	<p>https://www.dataprotection.ro/</p>
LUXEMBOURG	<ul style="list-style-type: none"> LUXEMBOURG INSTITUTE OF HEALTH (LIHIBBL) 	<p>LAW OF 1 AUGUST 2018 ORGANISING THE NATIONAL COMMISSION ON DATA PROTECTION AND IMPLEMENTING THE GDPR (THE "DATA PROTECTION LAW")</p> <p>http://data.legilux.public.lu/eli/etat/leg/loi/2018/08/01/a686/jo</p>	<p>https://cnpd.public.lu/fr.html</p>

End of D2.1

