19/06/2023

ENDOTARGET/

DATA MANAGEMENT PLAN

WP6, TASK D6.1

DELIVERABLE VERSION:

D6.1, V.0.1

DISSEMINATION LEVEL:

PU classified[1]

AUTHOR(S):

Author 1 Alessandro Melillo
Author 2 Marianna Vigorito
Author 3 Giovanna Scognamiglio
Author 4 Adelaide Serretiello

ENDOTARGET

[1] PU = Public - fully open
SEN = Sensitive - limited under the conditions of the Grant Agreement
EU classified – RESTREINT-UE/EU-RESTRICTED, CONFIDENTIEL-UE/EU-CONFIDENTIAL, SECRET-UE/EU-SECRET under Decision 2015/444

DOCUMENT HISTORY

PROJECT ACRONYM	ENDOTARGET
Project Title	Systemic Endotoxemia as the driver of chronic inflammation - Biomarkers and novel therapeutic targets for Arthritis
Grant Agreement Nº	101095084
Project Coordinator	HUS
Project Duration	01/01/2023 - 31/12/2026
Deliverable No.	6.1
Diss. Level	PU
Deliverable Lead	EBRIS
Status	_ Working _ Verified by other WPs/Partners x Final version
Due date	01/07/2023
Submission date	19/06/2023
Work Package	WP6
Work Package Lead	EBRIS
Contributing beneficiary(ies)	Text eingeben…
DoA	Text eingeben…

Funded by the European Union

DATE	VERSION	AUTHOR	COMMENT
06/06/2023	0.1	Melillo Alessandro, Vigorito Marianna, Scognamiglio Giovanna, Serretiello Adelaide	First draft
16/06/2023	0.2		Second draft
19/06/2023	1.0		Final version

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authorities can be held responsible for them.

ENDOTARGET is a Horizon Europe project supported by the European Commission under grant agreement No 101095084 and it is supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 22.00462.

EXECUTIVE SUMMARY

Points to be addressed:

- Explain the purpose of the data collection/generation and the relation to the objectives of the project
- Specify the types and formats of data generated/collected
- Specify if existing data is being re-used
- Specify the origin of the data
- State the expected size of the data
- Outline the data utility: to whom will it be useful
- Specify the ethical aspects

Funded by the European Union

TABLE OF CONTENT

Funded by the European Union

Funded by the European Union

LIST OF ABBREVIATIONS

ACRONYM	DESCRIPTION
AI	Artificial Intelligence
CC	Creative Commons
DB	Database
DMP	Data Management Plan
DPIA	Data protection impact assessment
DSMB	Data Safety Monitoring board
EB	Ethics board
EC	European Commission
EU	European Union
FAIR	Findable Accessible Interoperable Reusable
GA	Grant Agreement
GDPR	General Data protection regulation
HUS	Helsingin ja Uudenmaan Sairaanhoitopiirin Kuntayhtymä
IMM	Instituto De Medicina Molecular Joao Lobo Antunes
IPR	Intellectual Property Rights
IVDR	In vitro diagnostic devices regulation
MACs	Message Authentication Codes
MDR	Regulation on medical devices
DMP	Data Management Plan
OA	Osteoarthritis
PID	Persistent Identifiers
PRS	Polygenic Risk Scores
RA	Rheumatoide arthritis
RD	Rheumatic diseases
SE	Systemic Endotoxemia
SENSA	Secure sENSitive data processing plAtform
SERGAS	Servizo Galego De Saude
SIB	Sib Swiss Institute Of Bioinformatics

Funded by the European Union

SPA	Spondyloarthritis
UNICAM	Universita Degli Studi Della Campania Luigi Vanvitelli
UH	Helsingin Yliopisto
UTARTU	Tartu Ulikool

1. INTRODUCTION

The Horizon Europe Model Grant Agreement requires that a Data Management Plan ('DMP') is established and regularly updated. In completing the sections of the template the requirements for research data management of Horizon Europe as described in article 17 and analyzed in the Annotated Grant Agreement, article 17, must be addressed. Data Management Plan (DMP) of Endotarget's is to identify the project's research data and to describe how to make them findable, accessible, interoperable and re-usable (FAIR).

1.1 Purpose of the data collection/generation and ist relation to the objetives of the project

The purpose of data collection/ generation will be:

• Analyze lifestyle factors and gut microbiota composition and to measure biomarkers of intestinal permeability in large population cohorts to study their association with markers of systemic inflammation and SE and with risk of RA, OA and SpA.

• Determine the level of SE in healthy subjects and in patients with RA, SpA and OA as well as in high-risk pre disease cohorts to study the association of SE with the risk of RA, SpA and OA.

• Identify high-risk individuals by performing targeted proteomic and metabolomic analysis to discover novel biomarkers and factors contributing to increased intestinal permeability and SE and to study their association with risk of OA, SpA and RA and in transition from pre disease to disease.

• Determine genetic profiles of large cohorts of people and to create polygenic risk scores (PRS) for intestinal permeability and SE and study their association with risk of RA, SpA and OA and the risk of transition from health to disease.

Funded by the European Union

• Provide new recommendations that could lead to a change in current clinical practices, including systematically screening of rheumatic patients in predisease phase for intestinal abnormalities.

1.2 Types and formats of data to be generated or collected by the project

Structured and unstructured data can be of two types: anamnestic and data research.

Structured data:

- Data from paper/online questionnaires: lifestyle, health, dietary habits
- Clinical evaluation data: physiological, pathophysiology, pathogenesis sociodemographic and anthropometric, evaluation of the response to drug treatment, analysis of biological material (blood, stool, synovial fluid, tissues).

Raw Data:

- molecular data: on genome (part of/ all), transcriptome, analysis of inflammatory networks, analysis of proteomic profiles, microbiome, transcriptomic analysis, epigenomic analysis, pharmacogenetics, hormonal analysis and related immune-profiling.
- Non molecular data: AI for the extraction of biomarkers.

Anamnestic data:

- Medical history
- Patient surveys
- Clinical data
- Biobank

Data research:

- Multi omics analysis data
- non-biological and biological biomarkers are processed through artificial intelligence to generate predictive tools.

Funded by the European Union

1.3 Re-use of existing data

Data from recruitment and research activities can be reused for future studies. To this end, such data will be protected by intellectual property and stored in specific databases in accordance with the principles of the GDPR. The origin of the data comes both from the recruitment of patients, and from the involvement of pre-existing biobanks which data will be used for the set goals. Patients will be recruited in several EU countries. Results and analysis will be used to form the predictive AI tool that is one of the project's objectives.

1.4 The expected size of the data

The size will be assessed during the project and will depend on the scope and nature of the data made available. The data minimization rule will be observed by identifying the minimum set of information necessary for the project objectives.

1.5 Data utility

The project aims to allow third parties to access, extract, exploit, reproduce and disseminate this research data by attaching Creative Commons licenses to the data repository, in particular the following are the reference groups:

- Patient and patient's family
- Healthcare providers
- Healthcare system
- Researchers and pharma companies
- General public
- ENDOTARGET Consortium
- European Commission services and European Agencies
- EU National Bodies
- Standardization bodies
- Stakeholders

Funded by the European Union

2. FAIR DATA

In general terms, our research data will be 'FAIR' that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard or implementation-solution.

2.1 Making data findable, including provisions for metadata

The beneficiaries must process personal data under the Agreement in compliance with the applicable EU, international and national law on data protection (in particular, Regulation 2016/679). They must ensure that personal data is:

- processed lawfully, fairly and in a transparent manner in relation to the data subjects
- collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes
- adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed
- accurate and, where necessary, kept up to date
- kept in a form that allows the traceability of the persons concerned only to authorized operators for a period not exceeding that necessary for the purposes for which the data are processed
- processed in a manner that ensures appropriate security of the data

2.1.1 Standards for metadata creation

We will develop/implement bioinformatics pipelines in order to standardize a process of categorization and correlation of structured and unstructured data, starting from the collecting data, towards information extraction.

The AI algorithms will be applied to develop predictive tools and a repository will be customized and used.

Funded by the European Union

2.1.2 Use of persistent and unique identifiers

The assignment and management of persistent identifiers (PID) to the data will be evaluated during the project, starting from a database team verification and an evaluation of available PID services.

2.1.3 Naming conventions

The naming rules will be clearly and comprehensively defined in order to clarify the interpretation. The data structure will be then harmonized in a common tabular format to allow data processing in the same way for each cohort. Genomic, metabolomic, and metagenomic data will be aligned to standard formats and nomenclature as required to enable comparison between cohorts, and only data from standardized questionnaires will be included.

2.1.4 Approach towards search keywords

For pseudo-naming, an alphanumeric code will be used to identify the registration center, the patient, the type of sample, the type of collection and destination of the sample. In any case any future reference within the project results will be anonymous or encrypted.

2.1.5 Clear versioning

The metadata will have three types of versions: draft, in review, and versions that vary depending on the change of the metadata itself. It's important to keep track of managing the metadata. There are two components to the "Configuration Management System": version control and task management. It's important to keep track of any changes or updates recorded: the version number, the author, a brief summary of that iteration's changes of the document, the date. Versions are 0.1, 0.2 etc. until such point as the document is approved, then it becomes version 1.0.

Funded by the European Union

Subsequent edited versions become 1.1, 1.2. The '.x' part of the number represents a small change. We rename the document to the next whole number if it's a major update, e.g.

Version	Date	Author	Rationale
0.1	1 March 2017	Name, Surname	First draft
0.2	15 March 2017	Name, Surname	Review by Supervisor
0.3	22 March 2017	Name, Surname	Wider review
		Name, Surname	by project team
		Name, Surname	updated (new dates)
0.5	03-apr-17	Name, Surname	Final version
		Name, Surname	for signature;
		Name, Surname	costs updated
1	14-apr-17	Name, Surname	Issued

The rationale column is for a brief description which highlights what is different about this version compared to the last one. It may have no changes beyond the version number, because the 'change' is that it has been approved and become version 1.0, the 'final' one (at least until the next update). The history of a document, a result or a delivery can be contained in an index within it or in a separate index file. The first solution is less expensive and easier to manage.

2.2 Making data openly accessible

The Open Data Management plan is the first mandatory step before creating an open data store. It will make an inventory of all the data that will be generated in the project, from clinical data to statistical analysis and bioinformatics. The nature, volume and file types containing the data will be reported. Common ways of naming and referencing the data will be chosen. The degree of data accessibility will be decided, taking into account regulatory constraints (e.g. GDPR and human personal data). On the basis of this analysis, the characteristics of the databases to be developed will be listed.

We will publish the search results /results in open access journals. Anonymised raw data will be made available in a public database if it does not violate IPR generated by consortium partners.

Funded by the European Union

Data accessibility will be ensured by 3 different security levels.

The first level will cover documents and publications and will be accessible to all, because it is necessary for dissemination and communication activities.

The second level concerns personal data (physiological, pathophysiology, pathogenesis sociodemographic and anthropometric, evaluation of the response to drug treatment). This data, which will be stored in the appropriate repository, will have access only to authorized operators through login and password.

The third level of security concerns structured research data resulting from the processing of raw data analysis. This data will be protected by encryption and will only be accessible to authorized operators.

The beneficiaries may grant their personnel access to personal data only if it is strictly necessary for implementing, managing and monitoring the Agreement. The beneficiaries must ensure that the staff is under a confidentiality obligation. The beneficiaries must inform the persons whose data are transferred to the granting authority and provide them with the Portal Privacy Statement.

An external ethics advisor was appointed to act as an independent expert, assisting the European Commission on ethical aspects of research through a monitoring activity until the end of the project. He/she offers guidance, advice, monitoring and recommendations for all project activities, maintaining an overview of all the operations involved in ENDOTARGET, thinking ahead about possible problems and how they can be addressed. In summary, the role of the Ethics Advisor is to diligently monitor the objectives, methodology and implications of the research carried out in ENDOTARGET and to ensure that all activities comply with the highest ethical standards.

Once extracted, aligned and harmonized, pseudonymized data will either be securely transferred to a dedicated project space on SENSA sensitive data server at SIB, Lausanne (sensa.sib.swiss) or if ethical or legal restrictions prevent this, data will be made available for federated analysis on local servers.

Beneficiaries must manage the digital search data generated in the action (data) in a responsible manner, in line with the FAIR principles and by depositing the data in a trusted repository as soon as possible and within the established deadlines.

Co-ownership is governed by Article 16.4 of the grant agreement and Annex 5, Properties of Results, with the following additions: unless agreed otherwise between the co-owners, each of them shall have the right to exploit the joint results as it sees fit and to grant non-exclusive licenses without obtaining any consent.

Funded by the European Union

As indicated in Section 1.3 ee) or ff) of the Framework on State Aid for Research and Development and Innovation (EU 2014/C 198/01; hereinafter the "Framework") the partners, both scientific and industrial, are joint owners of the Results. They shall, where requested, verify their contributions, document the results and, where required in the light of the Framework, provide fair and reasonable compensation to those bodies. However, each of the co-owners has the right to use their Results in joint ownership for research and teaching activities not generating revenue on a royalty-free basis.

Sufficient and legal measures will be put in place to show that data is handled appropriately. Clinical partners will provide a contact person to help facilitate access to data.

Metadata of deposited publications must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent, in line with the FAIR principles (in particular machine actionable) and provide information at least about the following: publication (author(s), title, date of publication, publication venue); Horizon Europe or Euratom funding; grant project name, acronym and number; licensing terms; persistent identifiers for the publication, the authors involved in the action and, if possible, for their organizations and the grant. Where applicable, the metadata must include persistent identifiers for any research output or any other tools and instruments needed to validate the conclusions of the publication. Only publication fees in full open access venues for peer-reviewed scientific publications are eligible for reimbursement.

The data will be open and accessible for no more than 10 years from the end of the project activities.

2.2.1 Openly accessible via ENDOTARGET consortium, European Commission services and European Agencies; EU National Bodies; The general public

This project will have to comply with different regulations and standards. Some of these, such as the GDPR, are at the European level. However, in some cases we will also have to comply with local laws. In addition, the decision support instrument may, where appropriate, fall within the scope of the EU Regulation on in vitro devices

Funded by the European Union

(IVDR) or the Regulation on medical devices (MDR). The overall compliance of the project with all relevant regulations identified in this task will be assessed. It will also report on the final status of the project's GDPR, including the preparation of the Data Protection Impact Assessment (DPIA). This work will feed into a final report on the overall compliance of the project in all relevant regulations, laws and standards.

2.3 Making data interoperable

Within the consortium, genetic, omic, immune-data are exported in open, document tabular format accessible to humans and standard software. A common vocabulary and code lists of predefined values for harmonizing the descriptions of Human Biomonitoring metadata and data are under definition and will be defined in the course of the project, using when it's possible the code lists and their values, as defined in the INSPIRE implementing rules on metadata (Commission Regulation (EC) No 1205/2008): https://inspire.ec.europa.eu/metadata-codelist.

2.4 Increase data re-use (through clarifying licenses)

Each partner must give to the other participants access to the background as data, know-how, or information including any rights such as intellectual property rights.

All research results will be shared in Fair way, according to the European Community Orientation. The partners will retain their copyright, and for the publication and dissemination of the project results will grant common licenses in the manner and within the limits established by the European Commission in the Horizon 2020 program, as well as in the art.16.3 of ENDOTARGET Grant Agreement (use for its own purposes, distribution to the public, editing or redrafting, translation, storage, archiving, third parties' rights). If materials or documents are subject to moral rights or third party rights (including intellectual property rights or rights of natural persons in their image), the beneficiaries must ensure that they comply with their obligations under the Agreement in particular, by obtaining the necessary licenses and authorisations from the rights holders concerned.

Funded by the European Union

3. OTHER RESEARCH OUTPUTS

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

4. RESPONSIBILITIES FOR DATA MANAGEMENT

All the ENDOTARGET partners will be responsible for the data management of the project.

HUS will oversee the following functions:

- study synopsis;
- study protocol;
- amendments;
- provision of legal representation for EU and non-EU clinical sites;
- project overview;
- production of study reports;
- collecting, reviewing to verify consistency and submitting reports, other deliverables (including financial statements and related certification) and specific requested documents to the Granting Authority;

Funded by the European Union

- preparing the meetings, proposing decisions and preparing the agenda of General Assembly meetings, chairing the meetings, preparing the minutes of the meetings and monitoring the implementation of decisions taken at meetings
- development of newsletters;

As for the data management aspects, when registering, patients will be pseudonymised through the GDPR rules. Only the authorized researchers will have access to the sensitive data of the patient.

In accordance with the guidelines of the Horizon Europe programme, research data will be stored in a dedicated archive.

Once the project is finished, clean and definitive databases will be blocked and stored for audit and archiving via DB blocking procedures. This data will be kept for the expected period and the logs will be kept according to the storage procedures.

The final reconciled and clean database once locked, will be transferred to the statistical study for data analysis according to the approved statistical analysis plan. The statistician will prepare an integrated report at the end of the study. The results of the study will be presented to the EC and the Ethics Committees, and then disseminated in scientific journals and scientific congresses.

HUS, as Coordinator of the project, is responsible that all legal arrangements within the project consortium and between individual partners are in place.

5. DATA SECURITY

With respect to Privacy and Data Protection, the EU-legislation - the General Data Protection Regulation - imposes several new obligations upon the consortium partners being data processors. Moreover, several new rights are granted to data subjects and significant fines are introduced in case of a data breach.

Apart from this legislation, the consortium partners regard privacy and data protection as a fundamental principle and hence apply a strict policy on this matter.

All Parties are required to keep appropriate documentary evidence of data generation and handling.

Funded by the European Union

5.1 Data confidentiality

Partners must keep confidential all data, documents or other material that is identified as sensitive.

Beneficiaries may disclose sensitive information to their staff or other involved participants in the action only if it is necessary to know it in order to implement the project and/or are bound by an obligation of confidentiality.

The beneficiaries may disclose sensitive information to their personnel or other participants involved in the action only if they: need to know it in order to implement the Agreement and are bound by an obligation of confidentiality.

EU institutions may disclose sensitive information to its staff and other EU institutions and bodies.

It may also disclose sensitive information to third parties if it is necessary to implement the agreement or to safeguard the EU's financial interests and the recipients of the information are bound by confidentiality.

5.2 Data integrity

Integrity Authentication and correctness are security features provided primarily by cryptography such as digital signatures (and Message Authentication Codes - MACs). They enable the recipient to use digital information by verifying the authenticity of its origin (authentication) and checking that the digital information is intact (integrity). In addition to providing authentication and integrity, digital signatures also provide a key feature of cryptography called non-repudiation. Non-repudiation also prevents the sender from denying that they have sent the information.

5.3 Data availability

Business continuity and data availability shall be ensured by Ethics Advisor. For the duration of the projects, procedures and physical and logical checks will be applied to ensure the security of electronic records in accordance with the GDPR, and applicable laws and regulations. In addition, a DPIA (Data Privacy Impact Assessment) process, developed on software, has been set up to manage data privacy. Other

Funded by the European Union

procedures will then be carried out as required by the GDPR (e.g. a log of processing activities, data breach notification procedures, etc.).

It may be necessary to produce or oversee the creation of several GDPR documents, including data processing agreements, registers of processing activities and security risk assessments.

6. ETHICAL ASPECTS

Research under the ENDOTARGET project will respect the highest ethical standards. A good balance will always be struck between the research objectives and the means used to achieve them. All research conducted by partners requires ethical approval through local or national ethics committees responsible for all aspects of research work where approval for specific activities on a legal and ethical basis is required. It is essential that each partner has all the necessary authorizations regarding ethics or confidentiality when material is exchanged between partners. It is essential to collect settlement information in each partner country, establish a mechanism to ensure that all medical procedures and protocols are duly authorized and monitored and that protocols are established to ensure patient confidentiality, the use and storage of data meet strict security criteria.

The Consortium has an Ethics Board (EB) that consist of one representative of each Party participating in clinical trials (HUS, UTARTU, SERGAS, UH, UNICAM, iMM).

The EB is a consultative body established to review and assess the research activities carried out under Project to ensure compliance with ethical principles. In addition, the project has also an external ethics advisor who will be taking care of EU data protection and largescale processing of genetic data.

The process of completing the composition of the Data and Safety Monitoring Board (DSMB) is on going. The DSMB is an independent function in charge of monitoring each of the proposed studies, in which an Ethics Expert is appointed to monitor the ethical issues involved in the project and to produce periodic reports on particularly relevant ethical issues. All approvals/mandatory documents from the relevant national, local ethics committees or competent bodies, such as data protection authorities,

Funded by the European Union

before the start of the relevant activities shall be kept on file and provided to the granting authority upon request if not already provided as requested deliverables.

The transfer of data on human subjects to the ENDOTARGET repository is considered only when: informed consent, ethical approval and - where appropriate - the approval by local data protection authorities concerns the purpose that the data are intended to be used within ENDOTARGET and allow the transfer of individual or aggregated data to the repository.

Funded by the European Union