PROJECT DELIVERABLE REPORT

DELIVERABLE NUMBER	D5.2
TITLE	DATA PROTECTION
AUTHOR(S)	W.PINXTEN, L.STEVNER, M.LAVILLE, M.PINART, C.ROUSSEAU, F.MATTIVI, A.HODGE, A.ASSPOLLU, R.CANALI & M.ALLIGIER, J. BOUWMAN
WORK PACKAGE	WP 5
TASK	TASK 5.2
WP LEADER	M.LAVILLE
BENEFICIARIES CONTRIBUTING TO THE DELIVERABLE	CRNH, TNO, UHASSELT, UCPH, MCD, FEM, ULG, BIOCC, AND CRA-NUT
STATUS – VERSION	FINAL - VERSION 1.0
DELIVERY DATE (MONTH)	M24
SUBMISSION DATE	M30
DISSEMINATION LEVEL – SECURITY*	PU
DELIVERABLE TYPE**	O

* Security: PU – *Public*; PP – *Restricted to other programme participants (including JPI Services)*: RE – *Restricted to a group specified by the consortium (including JPI Services)*; CO – *Confidential, only for members of the consortium (including JPI Services)*

** Type: R – *Report*; P – *Prototype*; D – *Demonstrator*; - O - *Other*

CONTENTS

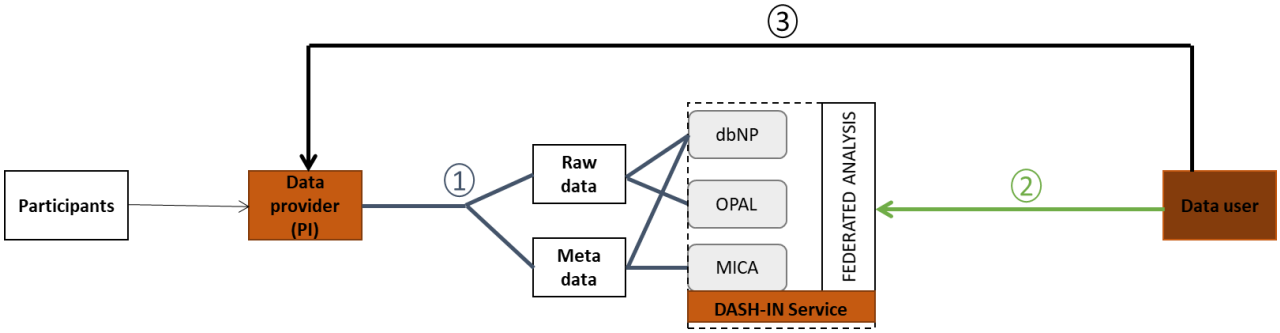
1- INTRODUCTION

ENPADASI aims to develop a database (WP3) based on the reuse of data obtained from different nutritional & clinical studies in Europe. However, the sharing of raw, meta or aggregated data raises several regulatory questions. Therefore, the scope of the WP5 is to define general rules required to share and reuse data in accordance with legal and ethical aspects of the EU participating countries and with respect to national policies. Four different topics are concerned in the WP5: ethics, data protection, data sharing polices and intellectual property.

Therefore, the deliverables of work package 5 aim to provide a set of rules and tools applicable to secondary use of nutritional data in the framework of ENPADASI project. Indeed, all the data users/providers have an obligation to operate in conformity with the requirements of their institution, and fulfill all necessary regulatory and ethical requirements imposed by their own national legislation.

Through several conference calls and one meeting in Paris, the consortium of the WP5, which gathers together several ethical & data protection experts, has identified the main ethical and data protection requirements to data sharing and has also proposed solutions and tools in order to help future data providers and users of the ENPADASI infrastructure to share and reuse their data in accordance with the current legislation.

2- DATA FLOW WITHIN THE ENPADASI INFRASTRUCTURE AND IDENTIFICATION OF THE DATA PROTECTION & ETHICAL ISSUES



Overview of the database architecture/data flow

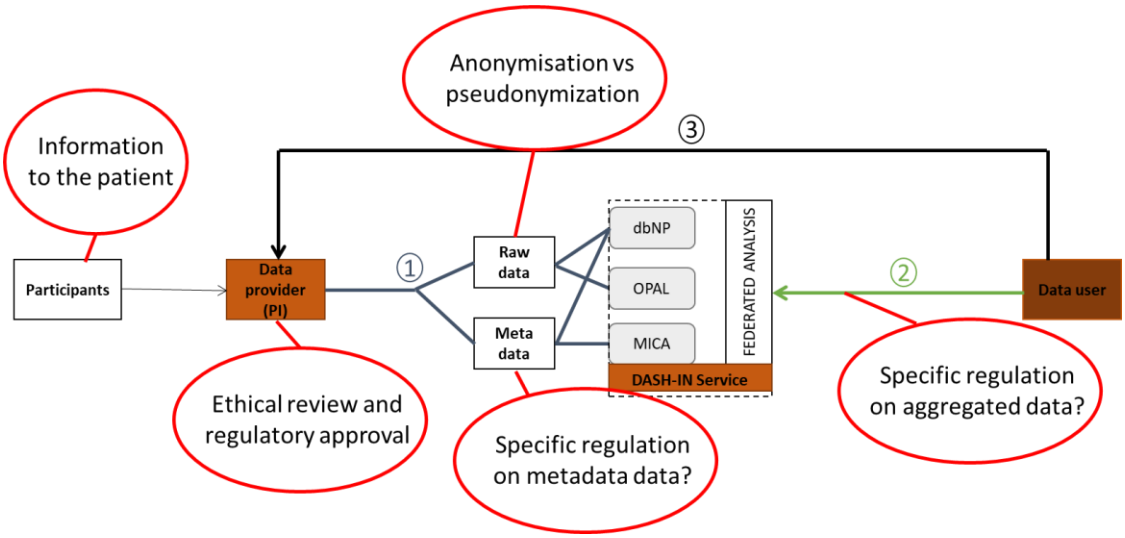
1. It will be mandatory for the Data provider (see for definitions in the definitions paragraph) to upload his metadata either in the MICA server or in the Phenotype database (www.dbnp.org). The data provider also has the possibility to upload his raw data on the Phenotype database or in an OPAL service. The metadata will be accessible within the ENPADASI consortium, thus each partner will be able to see which kind of study/data could be re-used. The data provider has also the possibility to upload the raw data of clinical studies (both interventional and observational) on the Phenotype

database or the OPAL system. But contrary to the meta data, the raw data will not necessarily be disclosed to the ENPADASI consortium. Each data provider will have the choice to restrict (even to only one data user) or open access to its raw data.

2. The data user can query the MICA server to identify studies of interest according to his scientific hypothesis. Thanks to the FEDERATED ANALYSIS tool, it will be possible to conduct a preliminary statistical analysis in order to validate or disprove the scientific hypothesis. The results of these statistical analyses are named aggregated data. The use of the FEDERATED analysis tool solves several ethical and data protection issues, allowing the future data user to combine the data and obtain statistical results without access to raw data, encountering ethics-related data-sharing concerns.

3. However, in order to go further and to publish the research from the combined analysis, access to the raw data will be mandatory. Two possibilities: the data user will contact the data provider directly, or the data user can have an access by the dbNP or OPAL server. The raw data access will raise several ethical and data protection issues that should be solved before use

3-DATA PROTECTION ISSUES RAISED BY THE DIFFERENT DATA SHARING



The scheme above sets out the main ethical (see deliverable 5.1) and data protection issues facing the future data users and providers and to which we will respond in the present document. Ethical and data protection issues are tightly linked.

3.1) METADATA SHARING AND ETHICAL & DATA PROTECTION CONCERNS

Metadata (see for definitions in annex 1) is defined as data that describes other data such as descriptive data on the clinical study: title, purpose, type of population, design of the study, funding, duration of the study, place of investigation, period for the recruitment, informed consent, scientific publications… and also on the raw data collected (help to described the data collection): type of

analysis, type of technical measurements, characteristic of the recruited population (BMI, specific pathology, exclusion/inclusion criteria…)

→ Usually, these types of metadata can be shared and used without any ethical and data protection restrictions since the data does not directly concern the volunteers. However, the future data providers must ensure that by overlapping metadata it does not become possible to identify a volunteer, this may be the case for example with rare disease. Therefore, the data providers have to ensure that the metadata is totally anonymous. Although the metadata is not in the scope of data protection, it is generated by the consortium of the project and thus must meet some legal requirements (see the D5.3) in order to share and access them in a legal manner.

3.2) RAW DATA SHARING AND ETHICAL & DATA PROTECTION CONCERNS

Personal data and sensitive data (including all the health data) are the two types of raw data mainly concerned by and data protection regulation.

The EU Data Protection Directive 95/46/EC, which will shortly be replaced by the General Data Protection Regulation (which will enter into force in May 2018) provides the European regulation regarding the processing of all personal data collected from an EU citizen. Both in the Directive and the regulation, the main data protection and ethical rule in which each European must adhere to is: **data sharing may only be considered permissible if data is unlinked anonymized or the data subject has given specific consent for the use of their (personal) data for the intended use.**

The notions of anonymized data, information to the patients and the initial purpose of the study are crucial and detailed in the present document.

3.3) AGGREGATED DATA SHARING & DATA PROTECTION CONCERNS

Thanks to the FEDERATED analysis tool developed in the WP3, the future data users will have access to aggregated data without any access to the raw data stored in the different tools by the data providers. The processing (access, storage, sharing) of aggregated data is not controlled by any data protection and ethical regulations.

3.4) THE INFORMATION RELATED TO DATA PROTECTION NEEDED TO BE IMPLEMENTED ON THE DASH-IN SERVICE

We provide a list of minimum information related to data protection covering the following domains: legal, data and intellectual property. The list was developed based on the templates circulated to identify both observational and intervention/mechanistic studies, coupled with the in-depth discussions regarding legal advice on issues related to data protection and ethics during the TCs held within the Work Package 5.

The first domain requests information related to the signed informed consents (see deliverable 5.1), Ethics Committee approval, potential limitations regarding secondary use of data and Data Transfer Agreements.

The second domain requests information related to the willingness of sharing data and metadata. In addition, the list contains two questions related to anonymization or pseudnonymization of the data for those partners willing to share raw data.

Label	Unit	Comments
Legal		
Does the data provider have the agreement of ENPADASI consortium to share data?	Yes/No	
Are there limitations for a secondary use of data?	Yes/No	
If yes, please enumerate limitations for secondary use of data	Text	
Should a Data Transfer Agreement (DTA) be signed between the data provider and the data user before any re-use of data?	Yes/No	
Data		
Will you make metadata publicly available?	Yes/No	
Are data stored in the ENPADASI database?	Yes/No	
If Yes, which database?	DbNP/OPAL	
Which kind of data are available?	Data/Biosamples/other	
Is data anonymised? If yes, provide the technique	Yes/No Text	
Is data pseudononymised? If yes, provide the technique	Yes/No Text	
Embargo period after the finalization project	MM/YYYY	

Checklist of minimum information for data protection and ethics in the DASH-IN research infrastructure

4- DATA PROTECTION SOLUTIONS NEEDED

For ethical reasons the personal data should be carefully protected. This means that the data provider will take all the precautions to prevent a breach in the systems. This requires separation of personal data and raw data in separated locations (independent file locations).

Technically, this means regular updates of the systems were the data are stored, hashing of the personal data and logging of the person accessing these data. Authentication and authorization systems should be implemented that are updated regularly. This last part is not only applicable to personal data, but also for anonymous data, as only the people that have the legal right to use these data should have access.

Given the considerable diversity in national legal regulations involved, it is impossible to account for all aspects. Therefore, we emphasize that a contact with the national data protection agency is often required; The list of the main national agency is put in annex 2.

4.1) DATA ACCESS COMMITTEE

A Data access committee could be put in place. The DAC seeks to ensure that the data access applicant (the User) willing to process the raw data provides adequate protection with regard to the ENPADASI Data Sharing Policy principles, the provisions of this DTA, and in compliance with their domestic law, European and International laws and relevant guidelines in the concerned field of research and presents sufficient guarantees for ensuring responsible data sharing and use. In this mission, DAC members will ascertain the adequacy and consistency of the access requests: with the applicant qualifications, authorisation and ethical approval obtained; of the methodology presented as regards to the purpose of the study; and with regard to any restriction indicated by the Provider, in particular those arising from research participants consent or program specific restrictions. While the DAC is not an Ethics Committee, it should be entitled to assess the ethical, legal issues related to an access application that would arise and undertake further evaluations, if necessary in cooperation with the Provider.

Members of ENPADASI should follow the terms and conditions of use for access to data (see annex 3) and use a Data Access form (see annex 4)

5- DEFINITIONS

Anonymous/ised data: Information which does not relate to an identified or identifiable natural person and personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable, including through cross-analysis by overlapping data.

Data Access Committee (DAC): Integral component of the DASH-IN Service for managing access to the data. The DAC is responsible for reviewing, approving or disapproving applications from potential users for a variety of restricted access cases.

Data Provider: The Provider (or 'Data Provider') is the individual researcher or investigator or body of researchers or investigators that makes data available for access and use within the context of the ENPADASI consortium and database. (It does not refer to the research participants.). The data provider should be the legal person or body that is responsible (owns) the data.

Data User: The 'Data User' is the individual researcher or investigator or body of researchers or investigators from either academia or industry that requests access to samples and/or data and use through the Data Sharing In Nutrition (DASH-IN) Service. The Data User is a 'data processor' in the meaning of the EU General Data Protection Regulation. The Data User may seek access outside of the context of the DASH-IN Service environment.

Ethics Committee: The term 'ethics committee' in this document refers to a committee which has given ethics approval for a study which has/intends to collect and use health data that will be subsequently made available by the Data Provider within the database and the DASH-IN Service. (It does not refer to the ENPADASI Data Access Committee.)

Metadata: Metadata is data describing other data. Metadata summarizes basic information about data, which can make finding and working with particular instances of data easier. Metadata can be created manually, or by automated information processing. Manual creation allow is the user to input any information they feel is relevant or needed to help describe the file, which is very relevant for example for the description of the study design. Automated metadata creation can display information such as file size, file extension, when the file was created, who created the file and can also include the logs of the machine used to generate the data. Metadata are highly aggregated and generalised data. Metadata is most often anonymised data.

Personal Data: 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.

The use of the term 'personal data' in this document covers sensitive categories of personal information as defined within the EU General Data Protection regulation data such as health data, biological and clinical data and the use of wellbeing data. Such data are particularly protected under privacy rules and secured management and access processes.

Pseudonymisation: The processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.

Raw data: Raw data refers to any data object that has not undergone thorough processing, either manually or through automated computer software. Raw data may be gathered from various processes and IT resources. Most digital equipment does not record the raw data but immediately processes it through vendor-defined algorithms into a vendor-specific primary record while discarding the original signals recorded in the equipment. In this context, such primary record files will be seen as analogous to raw data.

Raw data is primarily unstructured or unformatted repository data. It can be in the form of files, visual images, database records or any other digital data. Raw data is extracted, analysed, processed and used by humans or purpose-built software applications to draw conclusions, make projections or extract meaningful information. The processed data takes the form of information. Raw data can include personal data in the meaning of Article 4 of the General Data Protection Regulation of the European Union (Regulation (EU) 2016/679 [1]). In such a case the respect of applicable personal data protection laws will need to be ensured by the data providers and the users of the DASH-IN service.

[1] http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679

Raw data shall be pseudonymised before any exchange in order to ensure appropriate data protection.

Sensitive data: data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, genetic data, data concerning health or data concerning a natural person's sex life or sexual orientation.

6- ANNEX

ANNEX 1: DIRECTORY OF THE MAIN DATA PROTECTION AGENCIES

1) Belgium
The Commission for the Protection of Privacy (CPP) / "The Privacy Commission"
Phone: +32 (0)2 274 48 79
E-Mail: commission@privacycommission.be
Web address: https://www.privacycommission.be/en
Additional information: Clinical Trials must comply with data protection rules as specified in the Belgian Data Protection Act and its implementing Royal Decrees provided on the CPP website (https://www.privacycommission.be/en/privacy-act-and-implementing-decrees)

2) Denmark
Danish Data Protection Agency
Phone: +45 3319 3200
Web address: http://www.datatilsynet.dk/english/
Additional information: If information from patients' records is to be used – in register research projects without the use of biological material – an application for such approval shall be submitted to the DKMA.

3) France
Commission Nationale de l'Informatique et des Libertés - CNIL / National Committee for Data Protection
Phone: +33 153 73 22 22
Web address: https://www.cnil.fr/
Additional information: The trial must be submitted to the committee for data protection (Commission Nationale de l'Informatique et des Libertés - CNIL) assessing the storage and to the Comité Consultatif sur le Traitement de l'Information en Matière de Recherche dans le Domaine de la Santé (CCTIRS) assessing the content of information collected. A simplified process is possible if a reference methodology is used.

4) Germany
Die Bundesbeauftragte für den Datenschutz und die Informationsfreiheit
Phone: +49 (0)228-997799-0
E-Mail: poststelle@bfdi.bund.de
Web address: https://www.bfdi.bund.de/DE/Home/home_node.html

Additional information: Further details concerning data protection are listed in section 40 subsection 2a AMG and the Bundesdatenschutzgesetz BDSG) - German Data Protection Act (see also the unofficial English version published in the Federal Law Gazette)

5) Ireland
Office of the Data Protection Commissioner
Web address: http://www.dataprotection.ie/

6) Italy
Italian Data Protection Authority DPA - Garante per la Protezione dei Dati Personali
Phone: +39 06 6967 71
E-Mail: rp@gpdp.it
Web address: http://www.garanteprivacy.it
Additional information: The rights of the subject to physical and mental integrity, to privacy ant to the protection to the personal date are regulated and safeguarded by the Data Protection Code - Legislative Decree no. 196/2003, superseding the Data Protection Act 1996 (Statute no. 675 of 31 Dec 1996).

7) Netherlands
Dutch Data Protection Authority (DPA) / College Bescherming Persoonsgegevens (CBG)
Phone: (+31) - (0)70 - 888 85 00
Web address: http://www.cbpweb.nl./en/
Additional information: Dutch Personal Data Protection Act (unofficial English translation) / Wet Bescherming Persoonsgegevens WBP lays down the main rules for handling and protecting personal data. It has been translated into a code of conduct known as the Use of Data in Health Research (Code Goed Gedrag)

8) Spain
AGPD (Agencia Española de Protección de Datos) C/Jorge Juan, 6
Phone: +34 901 100 099

9) Estonia
Estonian Data Protection Inspectorate
Phone: +372 627 4135

10) Czech Republic
The Office for Personal Data Protection
Phone: +420 234 665 111
E-mail: posta@uoou.cz
Web address: https://www.uoou.cz/en/

ANNEX 2: TERMS AND CONDITIONS OF USE FOR ACCESS TO DATA

Dear User.

You are accessing to the ENPADASI database.

By using the Platform and its related data and tools you recognize to have read, and understood and therefore to comply with these rules.

1. General Principles

The use of this open access Platform and related content including the use of the Federated Access and Analysis Tool, shall respect applicable laws and best practices in your country as well as the following terms and conditions of use. You will be accountable regarding any misuse of the data and tools accessed.

Data access request and use shall be adequate, relevant and limited to what necessary in relation to the purposes for which they are processed ('data minimisation').

Any breaches of data security or confidentiality must be reported immediately to the DASH-IN Service without delay (see the data breach notification form).

Examples of data security breaches include (but are not limited to):

- Access by any unauthorised person (i.e. has not signed Terms and Conditions of Use for the relevant data set);
- Sharing ENPADASI data with unauthorised persons;
- Failing to ensure sufficient data protection; such as a loss of login details
- Erasure, loss or alteration of the data

The user is responsible to comply with any other notification obligation regarding authorities, employer etc. as fixed in applicable laws and professional regulations.

2. Ethical guidelines in accessing to the ENPADASI databases

Researcher(s) requesting data must act in the respect of confidentiality and any other restrictions imposed by the Data Provider for the use of the data that will be provided by ENPADASI. Any data provided remains the property of the initial Data Provider and shall be used only for the purposes for which data access was granted.

For accessing to further detailed datasets, you need to fill in the Data Access Form *(insert link toward online version of the form)*. Data will have to be used only for the purposes of the research described in the Data Access Form and in the respect of applicable laws and regulations.

Access fees may apply to some, but not all, data access requests:

a. Approved "open access data" requests will be provided at no charge.

b. Other data access requests may require a processing fee, which will be explained to the applicant after the request is reviewed.

Data must be stored in a secure location with access limited to research team members only.

Research data accessed can be used for academic or practitioner educational purposes. Professors may request data to use in teaching undergraduate and graduate classes. Doctoral students can request data for developing their dissertations.

Academic researchers are encouraged to produce peer-reviewed articles as a way to increase the body of knowledge about nutritional health research within the academic community.

Research will not be used to denigrate ENPADASI Infrastructure or its components, or in legal proceedings.

The cooperation of ENPADASI will be acknowledged in all research produced as a result of such cooperation (see below, point 5).

For reports written in English, the researcher will email a one- to two-page executive summary of the report to xxxxxxx@xxxxx.xxx within one (1) month of publication.

If the report is published in a language other than English, the researcher will provide a one-paragraph summary in English within one (1) month of publication.

The use of the data set for commercial purposes or marketing is forbidden.

Communications and reports based on ENPADASI data must focus on benefiting the public interest and the health improvement of population. They must include a disclaimer stating that the views in the report reflect the researchers' opinions and are not intended to represent the position or policies of The IIA or The IIARF.

3. Data access modalities

All the data available in open access are contained in anonymous form in the databases and are accessed through the DASH-IN Service and the Federated Access and Analysis Tool.

Raw data access, including finest and potentially personal data strictly protected and confidential, is subject beforehand to the fulfilment of a special Data Access Form *(insert the link towards online form if any)* and authorisation.

4. Open data/access licensing

The open access to databases is ensured by the DASH-IN Service and includes the use of the Federated Access and Analysis Tool (FAAT).

Databases are made available through an Open Access license (Creative Common - CC-BY 4.0), if this is compatible with the initial terms of access of the database.

The CC-BY-NC 4.0 Licence - allow the user to:

- **Share** — copy and redistribute the material in any medium or format
- **Adapt** — remix, transform, and build upon the material.
- You may not use the material for commercial purposes.
- The licensor cannot revoke these freedoms as long as you follow the license terms.

5. Return of data

Any data generated through research permitted by the access to ENPADASI databases, must be returned to the source database to encourage ongoing use by the research community. The user is required to provide ENPADASI with a copy of all data collected and/or generated, to be archived for future use.

The Data User returns, at a time not before the date of publication of a paper that describes the results of any analyses of the accessed data, the following information/results for a publication via the ENPADASI and Providers websites:

- General information about the analysis performed to inform the public.
- Summary data about the study results.
- ….

The Data Users shall provide a copy of any report of its results that derive from the use of the database to ENPADASI in any format (e.g. paper journal, on-line report, meeting abstract).

Notices required under these Terms and Conditions of Use will be in writing and will be delivered by email to the addressees specified to you by ENPADASI (by email or postal sending).

In certain limited cases, the secure erasure of the original data provided through the DASH-IN service could be explicitly required by the DASH-IN managers.

6. Intellectual property rights and publication

Any publication using and permitted by the data is must be made in open access.

The users, on the occasion of the publication of their research results, will acknowledge the database by using the following quote: …….

The above quote can be completed by specific quotations as indicated during the process for accessing raw data. They shall be used as an independent complement of the above mentioned general quotation.

If relevant, we recommend the use of the CoBRA guidelines (citation of Bioresources in journal articles), in order to facilitate citation in the articles of datasets used in scientific research.

7. Auditing

On reasonable notice, and in order to confirm or investigate compliance with the provisions of these Terms and Condition of Use, ENPADASI may itself or via appropriate third parties:

- choose to inspect the premises and other relevant facilities of the Data User, in order to review the security, storage or other arrangements for the data ;
- request additional information about the data accesses (e.g. statistics) and related Approved Research Project and/or about the data processing progresses as it can be reasonably required to the User.
- the Data User will not bear the costs of such audits unless a data default within the procedures and processes is discovered, in which case the User will be obliged to reimburse the reasonable costs to ENPADASI and any relevant third parties.

ANNEX 3: DATA ACCESS FORM

By requesting access to the ENPADASI databases, you can have access to different databases, each one integrating different type of data (raw data, metadata…). Access to data requires to fill an access form in which is described the project and the reasons why you need access to particular datasets.

A wide range of resources is available through access to ENPADASI platform. All the data are obtained from different authorised nutritional & clinical studies in Europe.

Please use this form to request access to data managed by the ENPADASI Infrastructure. Completed forms should be emailed to xxxxxx@xxxxx.xx

Name and Scope of Project	
Name of Project	
Geographic Scope (country or regions to be covered)	

Lead Researcher—Name and Affiliations	
First Name/Given Name	
Family Name/Last Name/Surname	
Job Title/Designation	
Certifications	
Organization Affiliation	
Industry of Organization	
IIA Institute Affiliation (if applicable)	
Country of Residence	

Lead Researcher—Contact Information	
Email Address	
Telephone Number	

Complete Mailing Address	

Lead Researcher—Qualifications	
Experience Related to Nutritional Studies	
Main Previous Activities in the Field	
Main Previous Publications in the Field	
Academic Degrees	
Data Analysis Experience/Skills	

Other Researchers—Name and Qualifications	
Name, Qualifications, Tasks, Location	
Name, Qualifications, Tasks, Location	

Project Overview	
Start Date (DD/MM/YYYY)	
Expected Completion Date (DD/MM/YYYY)	
Access period whished (if different from the ones of the project) (DD/MM/YYYY)	
One-Paragraph Description of Project Objectives (50 to 150 words)	

Maximum number of persons expected to have access to the data	
Categories of recipients and location (Organisation, Country, Tasks)	
Primary Methods of Distribution (for example, download from institute website)	
Publication Planned (Yes or No)	
Do you confirm that the project has received any necessary authorisations / ethical approvals required under applicable law? *If yes, please, provide GA number or reference to the obtained official documents* *If no, please justify and detail the authorisations/approval to be obtained*	

Type of Data Requested	
Kind of datasets / ref	*Processing Purposes and Methodology, Processing Location and Responsible Person if different from the applicant*

Upon approval, the applicant will be granted access to the data, either directly through the database or with a contact with the relevant Data Providers where the specificity of the data requires specific contractual agreements or authorisations. In this latter case a contact from ENPADASI will inform and guide you.

An amendment to the original access form must be completed if any of the following changes occur during the research using accessed data concerned by your approved request:

- Significant extension of the project's scope ;
- End date ;
- New researchers need to access the data ;
- Change in institution ;
- If any additional data are required

This amendment shall be further detailed and notified to your ENPADASI contact person without delay for approval and guidance.

Proposals for access may be refused. Reasons for refusal will be notified to the access applicant.