ENPADASI, the European Nutrition Phenotype Assessment and Data Sharing Initiative

Program Proposal

General Information

Programme Proposal full title:
The European Nutrition Phenotype Assessment and Data Sharing Initiative

Programme Proposal acronym:
ENPADASI

Expected start:
2014-12-15

Expected end:
2016-12-18

Name of the coordinating person (Knowledge Hub Coordinator):
Jildau Bouwman

Contact details:
Netherlands Organisation for Applied Scientific Research (TNO) - Department of Microbiology and Systems Biology
jildau.bouwman@tno.nl

Information on coordinators:

Legal name of organisation: Netherlands Organisation for Applied Scientific Research (TNO) - Department of Microbiology and Systems Biology*Country:*
NL

ZIP code:
3704 HE

Town:
Zeist

Street name, number:
Utrechtseweg 48

Website:

Mrs/Mr:
Jildau Bouwman

Title:

PHD

Function:
Senior scientist Systems Biology

Phone:
+31 (0)88 866 16 78

Fax:
+31 (0)88 866 87 01

E-mail:
jildau.bouwman@tno.nl

Additional information on person (max. 1500 characters) concerning personal background, institutional funding and explain responsibilities and tasks:
TNO, The Netherlands, is an innovation driven independent research organization whose expertise and research make an important contribution to the competitiveness of companies and organizations, to the economy and to the quality of society as a whole (www.tno.nl). TNO has a track record of more than 10 years in measuring, structuring data and redirecting Systems Health, especially in the area of metabolic syndrome, T2 Diabetes and cardiovascular disease. PhD Jildau Bouwman, working at Microbiology & Systems Biology has a longstanding scientific track record in systems biology, bioinformatics nutrition and study of metabolic syndromes. Jildau Bouwman is a molecular biologist who has used informatics in biological research were needed. She is one of the driving forces behind the phenotype application and principle investigator in the Netherlands Metabolomics Centre Data support platform. She represents a partner in many (inter-)national innovation networks and public private partnerships based on the 'know-how' in consolidation and integration of big data (e.g. Cosmos, NU-AGE, EuroDISH and Food4me). Jildau Bouwman is a scientist in and team leader of the group of Systems Biology, responsible for coordinating all infrastructure activities for storage and sharing of data. As the coordinator of ENPADASI Jildau Bouwman is responsible for the overall scientific coordination and supervision of the project throughout the whole duration of the project.

Name of the deputy knowledge Hub coordinator:
Duccio Cavalieri

Contact details:
National Research Council - IBIMET
duccio.cavalieri@fmach.it

Information on coordinators:
Legal name of organisation:
National Research Council - IBIMET

Country:
IT

ZIP code:
50145

Town:
Firenze

Street name, number:
Via Giovanni Caproni, 8

Website:

Mrs/Mr:
Duccio Cavalieri

Title:
National Research Council - IBIMET

Function:

Phone:
0039-0461615153

Fax:

E-mail:
duccio.cavalieri@fmach.it

Additional information on person (max. 1500 characters) concerning personal background, institutional funding and explain responsibilities and tasks:
Duccio Cavalieri is associate researcher at IBIMET-CNR (Florence), Coordinator of the Computational Biology Department at Fondazione E. Mach (San Miichele all'Adige) and Assistant Professor at University of Florence. DC's studies deal with the interaction between genome, metagenome, nutrients and diet. The main activity is the development and application of bioinformatic tools, based on pathway and network analysis, for integrative analysis of omics data of observational nutritional studies. DC has been member of the FP7 NoE NUGO, and active in several EU funded initiatives in nutrigenomics, metagenomics, immunogenomics such as the NoE DC-Thera and the FP7 IP Sybaris.
Within ENPADASI, consortium specific concerted action aims at the development of an integrated network of computational infrastructures for nutrigenomics research. The NutriExp Directory will be an information system designed to support knowledge management and integrative data analysis for this research community and beyond, developing computational pipelines for data integration, building and sharing of methodologies and data (WP2, 3 and 4). Given the presence of internationally renowned experts in the fields of bioinformatics, nutrition, data modeling and ontology development within our Institutes, the joint institutes could play an important role to streamline bioinformatics tools enabling the development of state-of-the-art research, education and training through the network (WP4,6).

Table of contents

Abbreviations

List of abbreviations

ArrayExpress Functional genomics database
BBMRI db Biobanking and Biomolecular Resources Research Infrastructure
Bc2S Bari Computer Center for Science
BioClaims EU project on ‘BIOmarkers of Robustness of Metabolic Homeostasis for Nutrigenomics-derived Health CLAIMS Made on Food’
BioMedBridges Building data bridges from biology to medicine in Europe
BioNH Biomarkers for Nutrition and Health
BioVeL Biodiversity Virtual e-Laboratory
CC-BY Creative Common - Attribution
CONSORT Consolidated Standards of Reporting Trials
COSMOS EU coordination activity ‘COordination of Standards in MetabOlomicS’
CTMM/TraIT Center for Translational Molecular Medicine/translational Research IT
dbNP (Nutritional) Phenotype database

DEDIPAC Determinants of Diet and Physical Activity
DiOGenes Diet, Obesity and Genes
DSA Designing Services for People with Alzheimer's
DTL Dutch Techcentre for Life Sciences
ECRIN European Clinical Research Infrastructure Network
ELIXIR Building a sustainable European infrastructure for biological information, supporting life science research and its translation to medicine, agriculture, bioindustries and society
ENPADASI European Nutrition Phenotype Assessment and Data Sharing Initiative
EoI Expression of Interest
EPIC European Prospective Investigation into Cancer and Nutrition
ESFRI European Strategy Forum on Research Infrastructures
EuroDISH EU coordination activity on 'European Determinants of Intake, Status and Health'
ExCom Executive Committee
FAIRport Findable, Accessible, Interoperable and Re-usable
FAQ Frequently asked questions
Food4me EU project on 'An integrated analysis of opportunities and challenges for personalised nutrition'
GA General Assembly
GC-MS Gas Chromatography – Mass Spectrometry
GEO Gene Expression Omnibus database
IaaS Infrastructure as a Service
ISA-TAB Build around the Investigation, Study and Assay general-purpose Tabular format
ISBE Infrastructure for Systems Biology Europe
JINGO Joint Irish Nutrigenomics Organisation
JPI Joint Programming Initiative
MAQC MicroArray Quality Control
MetaboLights Database for Metabolomics experiments and derived information
Methontology Comprehensive ontology engineering methodologies
MLOSS Machine learning open source software
MLPY Machine Learning Python
NGS Next Generation Sequencing
NMC Netherlands Metabolomics Centre
Nu-AGE EU project on 'New dietary strategies addressing the specific needs of elderly population for a healthy ageing in Europe'
NuGO NutriGenomics Organization
NutriTech EU project on 'Integration of emerging technologies into nutrition research'
OBI Ontology for Biomedical Investigation
OECD Organization for Economic Co-operation and Development
OGTT Oral Glucose Tolerance Test
Ontolingua Provides a distributed collaborative environment to browse, create, edit, modify, and use ontologies.
OpenPhacts Open Pharmacological Space
OWL Web Ontology Language
PII Personal Identifiable Information
PRIDE PRoteomics IDEntification database
PRISMA guidelines Preferred Reporting Items for Systemic Reviews and Meta-Analyses
Protégé To build knowledge-based solutions in areas as diverse as biomedicine, e-commerce, and organizational modeling
RI Research Infrastructure
SB Steering Board
SEQC Sequencing Quality Control

SemSim Multiscale model-description architecture designed to facilitate the sharing, reuse, and modular construction of a wide range of biological models
SOP Standard Operating Procedure
SPIDER Supporting Public service Innovation using Design in European Regions
STROBE Strengthening the Reporting of Observational studies in Epidemiology
TIC Technology for information and Communication
WebODE Extensible ontology-engineering suite

Summary

Publishable summary (max. 3000 characters including blank spaces):

The main objective of ENPADASI project is to deliver an open access research infrastructure that will contain data from a wide variety of nutritional studies, ranging from mechanistic,/interventions to epidemiological studies including a multitude of phenotypic outcomes that will facilitate combined analyses in the future.

To fully attain both the scientific and the strategic project goals the following activities will be performed during the course of ENPADASI. Standards and minimal requirements will be developed and defined in work package 2, in order to connect existing databases and to facilitate the upload of future nutritional studies. Available studies that are not yet in a readable format will be uploaded within this WP. Furthermore, the quality of the uploaded studies will be validated and guidelines for data sharing will be drafted. Within WP3 the actual distributed infrastructure will be technically developed, tested and released. This system will connect the available databases in the consortium (e.g. the Phenotype database). Ontologies and tools for integrated analysis on multiple studies will be developed in WP4. Furthermore, identifier mappings will be developed to facilitate biological interpretation. Presently researchers are hesitant to share data as they are unaware of the sharing conditions, therefore, guidelines will be drafted for efficient and legal sharing of data, resolving ethical, data protection, intellectual property, and data sharing policy issues (work package 5). Within WP 6, nutritional researchers will be trained in standards (SOPs and ontologies) and data upload using the infrastructure. WP1 will assess methods for sustainability of the infrastructure beyond the project. In addition, this WP will be responsible for f the governance and management of the project.

The activities of the project will lead to a sustainable storage and analysis system for the nutrition research field. In order to maximize the output and impact of ENPADASI a consortium with a wide range of skills has been assembled with excellent connections to other relevant projects (e.g. ELIXIR, EuroDISH, ECRIN, DEDIPAC and BioNH).

To evaluate the actual infrastructure and to show the added value of connecting similar studies in resolving chronic diseases with lifestyle related solutions, case studies will be defined. We hypothesise that combining studies will increase the power thereby limiting the needs for new and larger intervention studies. Validation of study results in a different cohort/study will enhance the biological applicability of the conclusions. Such data comparison is instrumental to improve the interpretation and validation of results. Furthermore, it will facilitate nutrition research and thereby will increase the knowledge and understanding of how food and nutrition can improve human health.

Description of the Knowledge Hub in general

	1. Objectives and state of the art 1. a) Objectives, vision including scientific / technological challenges 2. b) State of the art 2. Scientific/ technological concept including data standards, data management, data sharing (e.g., open access) 3. Implementation/ Knowledge Hub structure/ management 4. Description of the Knowledge Hub as a whole in respect to complementarities and balance among its main themes 5. Transnational added value of the Knowledge Hub 6. Potential impact on the advancement of the research area, capacity building, plan for translation of research (suitable for ENPADASI) into public health practice or policy (in 2 years, with a perspective on a longer term) 7. Overall strategy of the work plan (including timing of different work packages and their components) including long-term strategy for sustainability 8. Dissemination/exploitation strategy (including IPR, if applicable)	
General Description	description-10-545c8ec2d42be.docx	Show Uploaded Description

A3: Description of the Knowledge Hub in general
(max. 15 pages)

1. Objectives and state of the art
a. Objectives, vision including scientific/technological challenges
This initiative will develop an open access research infrastructure (RI) for all nutritional mechanistic, intervention and epidemiological studies. For this development standardisation is crucial as combing studies depends on mapping of similar data and design. The standardisation will consider study metadata and phenotypic data (e.g. clinical data, dietary intake, lifestyle and physical activity, metabolomics, transcriptomics). For this purpose, existing data infrastructures will be connected and further developed. This will result in a system that is the leading system to share small and big data in the nutritional field. The system will include ways to perform integrated analysis on multiple studies to enable further exploitation of research data (hypothesis testing, scenario analysis and modelling), re-analysis and tackle research waste. Nutritional researchers across Europe will be trained in the usage of the system, which will maximise the impact and encourage upload of all future nutritional studies and increase quality of data collection and sharing protocols.
Nutritional mechanistic, intervention and epidemiological studies make increasing use of high throughput technologies (e.g. metabolomics, high throughput screening) yielding large datasets. The strength of these technologies is to describe the whole system (systems biology), but previously it has been difficult for researchers to map these measurements to biological processes. Part of this problem is caused by the fact that the number of parameters measured out numbers the number of subjects greatly weakening, the power to arise biological conclusions. This power may be enlarged by connecting the outcome of similar studies; however, this requires structured and standardized data storage and methods for integrated analysis.
Making research data available/accessible, after study finalization is often a prerequisite for funding, thus emphasising the need for a Research Infrastructure for deposition and future exploitation of the data. Sharing of published/unpublished data and resources is a pre-requisite for a unified solution to these issues, as it optimizes the use of existing technological RIs and implements the strength of each study through comparison of data from different experimental setups. Such data comparison is instrumental to improve the interpretation of results and creation of new working models to design further experiments. Data sharing implies not only collaboration and data exchange but also standardization of metadata and development of standardised protocols, to allow merging of datasets which can be queried as a single integrated dataset.
The European coordination activity EuroDISH has mapped the requirements for a research infrastructure

for nutritional studies and has identified some main technological issues:

• increasing use of high throughput technologies (large datasets)
• lack of metadata (including life style information, such as dietary intake and physical activity)
• hard to compare studies even with complete metadata due to unstandardized descriptions (lack of controlled vocabularies and ontologies)
• lack of comprehensive metabolite spectral databases
• difficulty to map to (bio)chemical entities between technologies (lack of identifier and identity resolution)
• little understanding of integration of complex biological processes (e.g. inflammation)
• complicated IT infrastructure, requiring modern design and project management

These technological issues will be resolved in the ENPADASI infrastructure.

b. State of the art

The (Nutritional) Phenotype database (www.dbnp.org) is an initiative of NuGO (NutriGenomics Organization) and NMC (Netherlands Metabolomics Centre), and was launched in 2007. The application is developed to store nutritional intervention studies (including those with complex design e.g. cross-over) and is meant to facilitate standardized data output and study comparisons. The system is accessible via a web interface and is built in the framework Grails and data are stored in a PostGreSQL database. Flexibility of the system is guaranteed in the system by templates. The templates make it possible to adjust the information that can be stored by adding additional fields to the database via the user interface (if the user is template administrator). The type of the field can be defined by the template administrator, making it possible to store information in text format, dates or via dropdowns. Ontologies are included to facilitate standardization, linking to the ontologies available in Bioportal (https://bioportal.bioontology.org/). For several types of data (mainly related to study design) no standards are yet available. Using the dropdown menus in the system several vocabularies specific for nutritional intervention studies have been developed. The system is secured with an authentication/authorization system. Only if you have a user account, you can include data. The person including data in the system (data owner) can give access to the data to others and can give open access to the data upon publication. The Phenotype database is built in a modular way, making it possible to link the data and design in the system to data in other databases (e.g. ArrayExpress). The system has now been fully operational for 2 years. The database currently includes 58 biological studies, mostly nutritional intervention studies. Data analysis of combinations of studies with multiple measurements and interpretation of this analysis is central to this system.

To the best of our knowledge this is only database available for depositing of nutritional studies with nutritionally relevant metadata and multi-omic data. For metabolomics data the MetaboLights system is developed this infrastructure is linked to the data in the Phenotype database via www.metabolomexchange.org.

To identify the RI needs for nutritional studies, a case study was defined within EuroDISH. The goal of the case study was to facilitate development of an infrastructure design for the European Infrastructure on Nutrition and Health Research supporting all needs for performing molecular nutrition research focusing on quantification of the health effects of nutrition which fits the goal of the ENPADASI project. This design included solutions to facilitate standardized execution of human dietary intervention studies, optimization to access to all data and information needed in nutrition and health research, a definition of the optimal IT-infrastructure, methods to harmonize and standardize all relevant mechanistic research procedures. Also considered were issues related to connecting the RI with other nutrition-health-omic infrastructures (i.e. created within EU initiatives such as ELIXIR, ECRIN, ISBE, BioMedBridges). EuroDISH has shown that a virtual cohort can be developed based on data of independent studies. The case study showed that a research infrastructure for nutritional studies should include the following aspects.

1. Study design
2. Sample collection and analysis
3. Data processing (including omics)
4. Study upload
5. Advanced data analysis and integration

The state of the art and requirements for those aspects are individually described below.

Study design

To be able to compare studies a sufficient amount of metadata on the design is required. The Nutritional Phenotype database (dbNP) was built for this purpose (www.dbnp.org). The ISA-TAB standard is designed to standardize the collection of study metadata. The Phenotype database design was based on this standard, but was insufficient to store all relevant nutritional information. Therefore the standard was extended and probably needs further developed for other studies relevant to ENPADASI (e.g. observational studies). This extension will be shared with other nutritional researchers and the ISA-TAB developers.

The mapping activity showed that many ontologies for study metadata are still lacking. For instance, ontologies for interventions, such as a high fat diet and meal challenges require standardization, as these terms are used in several studies with different interventions. Proper ontologies should be developed, preferably with partners who are experts in this subject within initiatives such as ELIXIR, FAIRport and OpenPhacts.

Thus, it is crucial to develop a common ontology for dataset structuring. Ontology building methods can be subdivided in centralized approaches and collaborative approaches. Methontology and On-T--Knowledge are examples of centralized ontology building initiatives. Consensus building initiatives include the Delphi method the Nominal Group Technique, and DILIGENT. The NeOn Methodology for building ontology networks is a scenario-based methodology that supports a knowledge reuse approach, as well as collaborative aspects of ontology development and dynamic evolution of ontology networks in distributed environments.

A number of technical solutions developed to support collaboration in ontology building activities are also worth to mention. The Ontolingua Server and WebODE represents seminal works on distributed Ontology Engineering platforms. The Web version of Protégé, the most widely used ontology editor, provides an open collaborative platform for ontology modeling and knowledge acquisition. However the collaboration facilities are limited to the concurrent editing of OWL ontologies, leaving out features such as voting and discussions. Another relevant issue regards the visual representation of ontologies, to support the users in the creation and understanding tasks. Finally, approaches based on crowd-sourcing are gaining attention also in the knowledge engineering field. For instance, crowd-sourcing through Amazon Mechanical Turk to evaluate ontology quality in representing the target domain is explored.

In the ENPADASI knowledge hub, EoI45 has a wide expertise in data and knowledge representation and management. They worked on the development of semantics-based solutions and their applications in real case scenario and developed the SemSim semantic similarity search method for the retrieval of semantically annotated digital resources. In particular, ENPADASI will develop some specifically interesting synergies with the Interomics project and the Cosys SYSBIo infrastructure.

Sample collection and analysis

Currently, as far as we know the Phenotype database is the only system that centrally stores nutritional studies in Europe (e.g. The EU projects NU-AGE, NutriTech, Bioclaims share their data). However, several other projects and applications include relevant data (e.g. the BBMRI db, ArrayExpress, study specific databases etc.). In addition, several other studies are collected in specific databases, especially for observational studies (e.g. Diogenes). In order to use all these study data standardization of terms with common ontologies is required. Part of this standardization has been developed already, but further work is needed.

Data processing (including omics)

Several of the relevant studies for the ENPADASI project will include qualitative and quantitative dietary details. The information can be best stored by linking to databases developed for this purpose (e.g. FoodDB, EPIC-Soft).

Metabolomics data is very relevant for nutritional studies as the dietary effect generally is on a multitude of components in the body and metabolites are close to the phenotypic outcomes. Comparison of metabolomics data between different studies is difficult as often only the relative amounts are measured. This comparison is made easier if data processing, data structure, units and naming of the molecules are similar. Studies in the Phenotype database included metabolomics identifiers to compare the different studies with each other. This solution does not include a standard pipeline for metabolomics data and precludes incorporation of metabolites that have not yet been unambiguously identified. The identification is a central bottleneck in metabolomics due to limited and fragmented availability of compounds or compound spectra in central repositories. Although metabolomics is an important platform for nutritional studies, resolving these complex problems is also an issue for other research areas. This should be resolved in collaboration with initiatives like COSMOS and ELIXIR. For transcriptomics the ArrayExpress and GEO databases are relevant and should be connected to the ENPADASI infrastructure.

EoI45 contributes to the ESFRI infrastructure for Biodiversity "Lifewatch", and leads the Italian node for Bioinformatics "ELIXIR" (http://www.elixir-europe.org/). NMR metabolomics data acquisition and their transformation in descriptors for biological functions are acquiring importance for nutritional studies. EoI45 develops algorithms for the transformation of raw data in parameters suitable for annotation and meta-analysis with other omics datasets. They recently described a foodomics protocol coupling an in vitro static simulation of digestion to a combination of omics techniques, to grant an overview of the protein digestibility of a meat-based food. They proposed the use of a multiscale approach to investigate type-2 diabetes.

Other types of datasets, like next generation sequencing NGS data, proteomics, light/electron microscopy, imaging, epigenetics and genomics also require standardized processing and storage to make cross-study comparison possible. The nutritional community should link to other projects for storage of these data structures.

Study upload

Several (ongoing) studies are uploaded in the Phenotype database and other nutritional studies are available in study specific databases. These specific databases are rarely standardized and lack study metadata. In addition, many valuable nutritional studies are unavailable in a machine-readable and standardized way (e.g. several independent excel sheets). ENPADASI will create standards and an infrastructure to facilitate sharing study data.

Advanced data analysis and integration

Major efforts for standardization have been achieved in NuGO, the FP6 Network of Excellence "NUtriGenomics Organization", a major collaborating network in molecular nutritional sciences. NuGO developed and implemented the NuGO Information Network, a distributed system for data exchange based on standard web technology, as a tool for data management and computing infrastructure that supports collaboration between nutrigenomics researchers. This resource does benefit from several efforts for standardizing ontologies in the biomedical field (e.g. OBI (ontology for biomedical investigation) and ISA-TAB for the annotation metadata). Identifier mappings were developed for omics tools in the BridgeDB framework (www.bridgedb.org). Semantic Web technologies for interoperability are developed by several groups within ELIXIR. This work will be continued within the FAIRport initiative whose goal is to make information/data Findable, Accessible, Interoperable and Re-usable (www.datafairport.org).

Metabolomics data are often expressed in relative measures (not absolute values). Within the EuroDISH case study, statistical standards were developed to make the cross-study comparison possible. The

queries developed in this case study were focused on oral glucose tolerance tests (OGTT) and metabolomics data and used data from the Phenotype database. For other research questions specific statistical solutions and queries will be needed, including queries to other data resources. Some of these questions may require modelling. The nutritional community should link to specific projects for this type (e.g. ISBE).
EoI45 has excess to a high-performance cluster to run analysis and many proprietary software tools to help researchers extract knowledge from big data.
In NUGO, a pathway analysis tool EuGene has been developed then improved as Pathway Processor 2.0, a pathway web-application designed to analyze high-throughput datasets, (e.g. for analysis of microarray and next-generation sequencing). The tool can perform two different types of analysis: the first covers the traditional Fisher's Test used by Pathway Processor and topology-aware analyses and the second is a new pathway-based method to investigate differences between phenotypes of interest (e.g. dietary regime, metabolic syndrome).
Modeling and predictive machine learning are the ultimate goal of integrative analysis. MLPY Python package for machine learning one of the most downloaded open source systems from the European MLOSS platform for predictive machine learning and most recently Nettools - published as a R package- , and the web interface ReNette have been developed for complex network analysis
Several of the ENPADASI partners have access to distributed infrastructures based both on grid and cloud paradigms. These can be used for data storage and complex analysis (e.g. infrastructures of the ReCaS project, the Cloud test bed, IaaS and PaaS of the PRISMA project, the EGI Cloud federated infrastructure, and the BioVeL,service). High performance computing platforms such as TeraByte, RAM and the computational infrastructure of the Bari Computer Center for Science (Bc2S), are available as well to the ENPADASI partners.

2. Scientific/ technological concept including data standards, data management, data sharing (e.g., open access)
Within Europe several consortia and foundations are coordinating data stewardship and management activities, such as ELIXIR and EURODISH. Several of the ENPADASI partners are connected to those initiatives. ELIXIR is building a European bioinformatics infrastructure. EURODISH is building a food and health research infrastructure. The (Nutritional) Phenotype Database was an initiative of NuGO and the Netherlands Metabolomics Centre (NMC). This database is part of the ICT infrastructure of EURODISH and is one of the projects in the nutritional & health sector of DTL/Data (the Dutch node of ELIXIR).
The Phenotype database will be used as one of the data storage solutions in this project, as it is optimized for data comparisons. This system is based on the ISA-TAB metadata standard. The sustainability, standardization and development of the system will be guaranteed though interactions with EURODISH and ELIXIR. The Phenotype Database will provide the data to the health related scientific community. The data management solutions adheres to the data stewardship requirements as described by DTL/Data (as Dutch node of ELIXIR DTL/Data focusses on interoperability), which includes a description on structuring, annotating, storing, sharing and using of data.

Data Management
Phenotype database (www.dbnp.org) is secured with an authentication/authorization system (developed as part of the Grails framework). All other databases that will become part of the ENPADASI infrastructure, should have a save authentication/authorization system. Those systems will be standardized and connected, ELIXIR has several partners that have developed systems for this purpose. Standardization is key in this project. The nutritional community has started this task by storing studies in the Phenotype database (e.g. NU-AGE, Food4me, NutriTech, EuroDISH, Bioclaims). The system is made flexible by templates, which make it possible to add information to the system where necessary. Initially the study will only be open to the owner of the study (and the administrator of the system). The original owner of the data has management rights and will be control who has access to the data. The study can be made open access upon publication of the results (under the CC-BY license). Data from

spreadsheets (xls, xlsx and csv) can be uploaded in the system. To prevent double work an exchange format between the Phenotype database and other relevant databases will be developed (based on the ISA-TAB standard). RAW data (up to 200 TB over the whole project including NGS and imaging) will not be stored in these systems, but in off-line solutions with minors to prevent data loss, till finalization of processing (draft of the paper ready). Standardized pipelines will be used for data processing, these pipelines will be made public where possible through ArrayAnalysis.org (www.arrayanalysis.org). The amount of data storage and calculation power needed for the nutritional community will be analysed in the sustainability tasks of WP1. To ensure that there is no data loss a backup plan will be implemented: Both for data and metadata a backups will be made regularly another location to prevent data loss (details of requirements will be collected in WP1 and WP5).

Metadata management
The Phenotype Database includes a system to describe the metadata in detail (such as study design, demographics, SOPs etc), which can be stored in the same database as the data (but also in other modules of the federated system). The description of the study can include any required detail on methodology, definitions, procedures, used SOPs, units and will be standardized through ontologies. Parts of the SOPs which are not relevant for data comparisons can be stored in linked files (any format), but essential parts will be made machine-readable. Links between SOPs, data, people, publications and projects will be preserved by the structure of the system. Modeling scripts used for data analysis can be stored and will be made accessible via a dashboard. In guidance with EU ethical Regulation's all data to be included and used of the case study will be anonymised.

Curation support
Data will be delivered by the partners and curated by a dedicated data manager (gatekeeper) of the project. The data manager will guard completeness of the captured data. Only the final version of the data (ready for publication) will be stored in the ENPADASI infrastructure. SOPs files and models on the dashboard will include a version number.

Data analysis support
Data analysis scripts and queries will be delivered by the bioinformaticians of the project. Models (e.g. scripts) can made available via a dashboard. They will be made accessible via Github upon publication. For scripts that are generally useful, simulation apps will be included in the ENPADASI infrastructure. A direct link between the automated scripts and the data infrastructure will make direct modelling possible.

Publishing support
Export will facilitate publication in community repositories (GEO, PRIDE, etc.) where needed (direct publication of data and metadata in the Phenotype database is possible as well). Publications can be linked to the data via a direct link to PubMed in the Phenotype database. Data can be made accessible to other project members were needed.

Open access
All partners generally support open access policies for non-commercial data and several have already shared data sets (e.g. via the Phenotype database). The drive of the community towards open access and the increased demand on data availability by leading journals creates a need for an infrastructure that can contain complex nutritional data and allow the researchers to meet these new demands. The direction towards more openness and data sharing clashes to some extent with demands on intellectual property rights and the trend towards more extended commercialization of university research. Good governance and overview of national legislative obstacles to data sharing is important for ENDAPASI to be sustainable and well accepted. The advantage of sharing data through standardized infrastructures over sharing data through scientific publication is that individual data and phenotypic (individual) data are also shared, whereas publications on omics data often only include the (average) omics data itself. By sharing all data after project terminations these additional data can be mined and studies can more

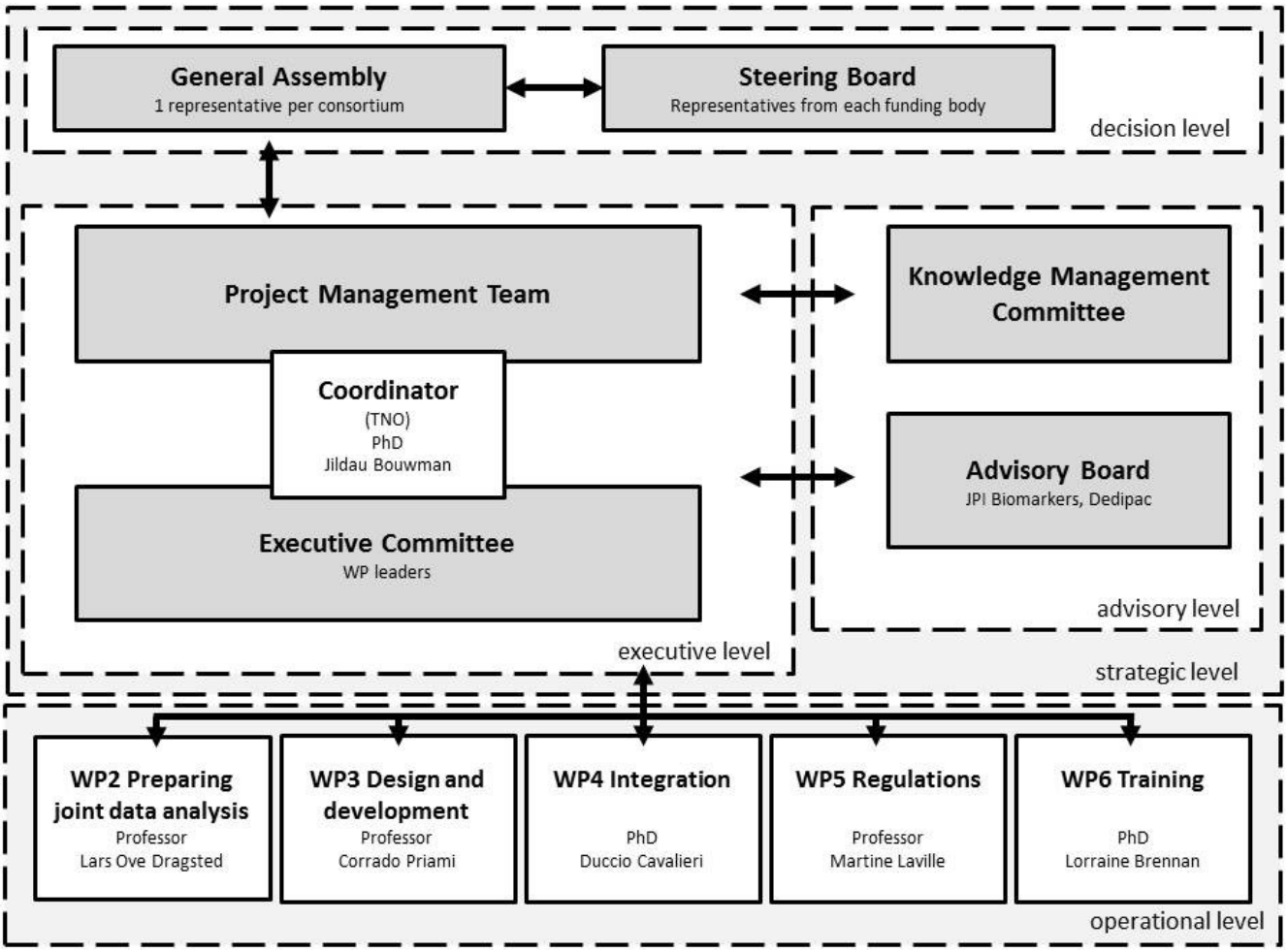
easily be combined for improved statistical analysis and can diminish the amount of animal studies needed for research.

3. Implementation/Knowledge Hub structure/management

Management structure

ENPADASI is a large, ambitious, multidisciplinary collaborative project with 16 consortia containing 51 organizations from 9 different countries. The management structure and procedures are therefore very important to ensure the smooth execution of the work plan and to facilitate collaboration between partners so that both scientific and strategic goals can be fully attained. The execution and management of the project will be regulated by the Grant Agreement as well as by the Consortium Agreement. A pragmatic organizational structure has been adopted, with the allocation of clear responsibilities.
The governance of the project entails 4 organizational levels (see figure Management structure of the ENPADASI project). The ultimate decision-making will be the responsibility of the General Assembly (GA) together with the Steering Board (SB).The executive function will be carried out by the Coordinator together with the Project Management Team (PMT) and the Executive Committee (ExCom). The advisory function will be carried out by the Knowledge Management Committee (KMC) and the Advisory Board (AB). Finally, the operational level will be ensured by the WP leaders and members.

Management structure of the ENPADASI project



Project Coordinator

TNO (The Netherlands) will coordinate the project. TNO will have full responsibility for the overall management of the project. Jildau Bouwman, PhD was elected as project coordinator during the preparatory meeting. She will be the contact person to the ExCom. The Project Coordinator will be

assisted by the deputy coordinator Duccio Cavalieri, PhD from IBIMET-CNR (Italy).
A Consortium Agreement (CA) will be drafted and agreed upon by all partners and will cover all aspects of the governance of the project including rights and responsibilities.
The major responsibilities of the coordinator, on behalf of the beneficiaries, include:
- To supervise the project in all its scientific, technological, legal, ethical aspects and supervise the preparation of deliverables and reports and ensure their appropriate quality;
- To act as intermediary between the beneficiaries and the call secretariat;
- To organise and ensure appropriate communication among partners, and chair the meetings of the governing bodies.

Project Management Team (PMT)
The project coordinator (Jildau Bouwman, PhD) will be the chairman of the Project Management Team (PMT). She will be assisted in the PMT by a Project Manager with professional and accredited project management training and skills (Annelies Dijk, PhD, funded by the project), a Financial Officer, embedded in a professional project management structure with all accountancy tools necessary, and a juridical office backup for legal issues, among others contracts and agreements. A secretary will assist the PMT with administrative tasks.
The PMT will ensure the total content and quality of the project. The PMT will be in charge of the following tasks:
- Preparation of the Consortium Agreement (to be signed by all parties at the start of the project)
- To assist the coordinator in the day-by-day management of the project, including monitoring and evaluation of overall progress and timely delivery of the deliverables, adapting the planning where necessary and appropriate, ensuring prompt delivery of all data identified as deliverables in the Grant Agreement or requested by the call secretariat for reviews and audits and taking measures in the framework of controls/audit procedures;
- Invite parties to take a seat in the Advisory Board and being the contact point to the Advisory Board;
- To organise the procedures for internal communication within the consortium as well as for collection of reports and deliverables;
- To support the organization, preparation, and follow-up of periodical meetings;
- To support the coordinator and the partners in the interaction with the call secretariat.

Executive Committee (ExCom)
The Executive Committee will be the main executive body of the project and will be composed of the six leaders of the Work Packages, to assure the smooth progression of the complex project activities. The ExCom is chaired by the Project Coordinator. The ExCom will facilitate the interdisciplinary focus of the project, collaborating with the Coordinator in ensuring the adherence to the project work plan and time-tables as well as in reviewing the project deliverables and reports. The specific tasks include:
- Coordinate the integration of the WPs;
- Planning of long-term detailed work plans and technical activities, and monitoring, guiding and controlling project progress and output on the basis of the half-yearly reports prepared by the PMT;
- Proposing the adjustment of scientific and technical roadmaps (also as a result of external developments);
- Ensuring proper interactions and communication between Work Packages and participants;
- Monitor the attainment of all milestones and deliverables and take corrective measures;
- Making proposals as to the use and dissemination of knowledge;
- Managing the connection of the project with other relevant and synergic networks/projects;
- Seeking new, external members into the Advisory Boards foreseen.
The ExCom will meet on a regular basis, but at least twice a year. In order to ensure cost-effectiveness, ExCom meetings will be coupled with the GA meeting whenever possible.

General Assembly (GA)

The General Assembly is chaired by the Coordinator and is one of the decision making bodies of the ENPADASI consortium. The GA will meet twice a year. The GA will be composed of one representative per consortium/EoI. Each Consortium/EoI will have one vote, while the consortia/EoIs with a contribution above 100 kEuro will have two votes. The GA will work on consensual decisions as much as possible, and will resort to voting only if this is unavoidable. The GA shall not deliberate and decide validly unless two-third (2/3) of its members are present or represented. The parties agree to abide by all decisions of the GA. Veto rights are determined in detail in the Consortium Agreement.
The GA will be responsible for the strategic orientation of the project, the overall direction of activities, and their re-orientation whenever necessary. In particular:
- Decisions concerning the work plan and its major changes;
- Overall monitoring of the project progress, quality and output, on the basis of half-yearly progress reports provided by the PMT;
- Monitoring and assuring adequate exchange of information between different WPs;
- Monitoring dissemination and exploitation;
- Approving the progress and management reports;
- Review and/or amendment of the terms of the Grant Agreement and Consortium Agreement;
- Decisions concerning possible premature completion/termination of the project;
- Amicably settling and disputes arising from the project implementation;
- Decisions regarding the entry of new parties as well as exclusion/withdrawal of parties and approval of the settlement on the conditions of the accession/withdrawal of such a party;
- Declaration of a party to be a Defaulting party;
- Remedies to be performed by a Defaulting party;
- Termination of a Defaulting party's participation in the consortium and measures relating hereto;
- Proposal for a change of the coordinator, suspension of all or part of the project or termination of the project and the CA.
- Ensure safeguarding privacy and ethics issues related to the knowledge hub and data sharing
- Safeguard IP rights of the ENPADASI infrastructure and knowledge developed
To ensure the relevance of the work plan with respect to the advancement of the project as well as external changes, the GA will:
- Analyse all relevant information provided by the ExCom;
- Take into account analysis on the evolution of the context in which the project is carried out, notably, scientific, legal, societal, and economical aspects.
Roles and procedures of the GA will be finely detailed in the Consortium Agreement.

Steering Board (SB)
The Steering Committee is composed of representatives from each funding body that have signed the ENPADASI memorandum of Understanding (MoU). The members of the SB will be responsible for:
- Approval of Joint Action procedures;
- Dissemination of the call text in their country;
- Eligibility check of the EoIs of new parties;
- Appointment of an international joint panel of experts for the peer review of the programme proposal;
- Funding decisions
- Safeguarding ethical and privacy issues related to the development of future use of personal data
- Implementation of monitoring the Networking Programme.

Advisory Board (AB)
An Advisory Board (AB) will be established shortly after the start of the project to ensure active involvement of key stakeholders, end-users and scientific experts. Specifically, the JPI BioNH and Dedipac consortia will closely interact. The AB will meet annually, and serve as a discussion platform to consult in different aspects of the project. Also, the Advisory Board will provide guidance and inputs to the design and monitoring of project activities and will propose routes for dissemination, education and

exploitation. The AB will advise the PMT on following issues:
- Scientific aspects of the project;
- External developments relevant to achieving progress towards the impact of the call;
- Maximising exploitation benefits;
- Advise with respect to dissemination, education and exploitation of the project results.
Members of the AB will have access to the project documents and outputs upon signature of a non-disclosure agreement (NDA). They will be invited to (selected) ExCom or annual meetings and will draw up an advisory report after each consortium meeting. The AB will be assisted in this latter task by the PMT.

Knowledge Management Committee (KMC)
The Knowledge Management Committee will oversee the handling and publication of any knowledge generated by the project and will ensure proper IP protection. The Committee will thus implement those parts of the CA covering intellectual property. The KMC will see it that:
- Pre-existing know-how is identified and included in the Consortium Agreement;
- Merger of pre-existing know-how and know-how generated in ENPADASI is avoided;
- IP issues and related legal affairs are handled correctly;
- IP is identified and appropriately protected at an early stage;
- Project results are disseminated at a proper time (in relation to IP protection).
- Advice the GA on ethical and privacy issues of the data in ENPADASI
The KMC will be composed of three members of the Executive Committee and will report to the ExCom, at least during each ExCom meeting. The KMC will be assisted by the PMT.

Work Packages (WP)
Each WP will be led by a WP leader (WPL). The WP leaders have been selected for their known experience in the areas of each specific WP and for their management and organizational skills (Table 1). The WP leaders will be responsible for day-to-day management and (scientific) coordination of tasks and activities in their WP. All WP leaders have a seat in the Executive Committee to ensure direct communication lines with the researchers, and representation of all ENPADASI research lines at the management level. The key responsibilities of the WP leaders include:
- Detailed planning of the tasks within their Work Package;
- Facilitating communication and information flows within the WPs;
- Monitoring the technical progress against Milestones and the Deliverable deadlines and assuring the (scientific) quality thereof;
- Preparing WP reports and deliverables and presenting them to the Project Coordinator;
- Organizing and chairing WP meetings, and reporting the progress to the PMT and the General Assembly (if the latter requires more detailed information on some issues);
- Collecting inputs, ideas, opportunities arising within the WP activities as well as any possible problems or conflicts (e.g. delays, misperformance, etc.) and bring them to the ExCom; on the other hand, translate the decisions of GA, SB and ExCom into clear and operative tasks for the WP members;
- Discussing approach and methodologies with the scientists;
- Solving problems within the WP;
- Dissemination of the results generated within their WP;
- Initiating contact with the PMT, minimal on a bimonthly basis.

Table 1. WP leaders.

WP no	WP leader	EoI	Organisation
1	Dr. Jildau Bouwman	50	Netherlands Organisation for Applied Scientific Research (TNO)
2	Prof. Dr. Lars Ove Dragsted	71	University of Copenhagen

3	Prof. Dr. Corrado Priami	45	The Microsoft Research - University of Trento Centre for Computational and Systems Biology
4	Dr. Duccio Cavalieri	45	IBIMET-CNR
5	Prof. Dr. Martine Laville	37	CRNH Rhône-Alpes
6	Dr. Lorraine Brennan	65	UCD

Communication and project meetings

Meetings and interpersonal interactions will be organised and arranged. Telephone- and videoconferencing networks will be used were possible, in order to minimise unnecessary or repeated travelling. This will also optimise overall cost-effectiveness, in all the activities necessary for ensuring the smooth running and timely delivery of the project. In addition, a project website will be developed with functionalities for document exchange, consortium and WP-specific email lists, web-based conferencing, calendar functions, methods portal and data exchange. The ENPADASI project website will consist of two parts. The first part will be 'open access' and will be used for informing and communication with stakeholders outside the consortium. The other part will be username and password restricted to the consortium members and the external advisory board. This area will include relevant documents, such as administrative documents, descriptions of all WPs, documentation of draft program progress, a contact list, meeting agendas and minutes, and web conferencing facilitation. This area will also be used for the management, storage and exchange of the project's research results.

Ultimate consideration will be paid to the need for face-to-face meetings, interactions, and communication which will ensure the interpersonal contacts within the project necessary for the successful and optimal impact of trans-national activities. This is particularly important given the broad geographical sweep of ENPADASI partner organizations, their cross-disciplinary diversity and the importance to the overall success and impact of the project of optimizing scientific/industrial contacts and opportunities. An overview of the different types of project meetings is provided in the following table.

Table 2. Overview of ENPADASI meetings.

Type of meeting	Frequency	Scope	Participants
Plenary Group meetings	Kick-off + month 12 and 24 (face-t--face)	Monitoring, coordination and integration of scientific and technical WP activities	All members
GA meetings	Kick-off + month 6, 12, 18 and 24 (also via Telephone- and videoconferencing networks)	Coordination and decision making	GA members
ExCom meetings	Every 3 months (also via Telephone- and videoconferencing network)	Scientific and technical coordination and monitoring	ExCom members
WP meetings	Every 3 months and whenever necessity should arise (mainly via Telephone- and videoconferencing network)	Coordination and monitoring of WP activities	WP members
AB meetings	Every 6 months and whenever necessity should arise (mainly via Telephone- and videoconferencing network)	Scientific and technical advice and guidance, coordination and monitoring of dissemination, education and exploitation	AB members

Progress evaluation, reporting, risk assessment and control

The PMT will provide planning and control tools to facilitate the monitoring of activities and to ensure minimum reporting effort while maintaining effectiveness for control purposes. In addition to the reports, a short tabular progress report on each WP activity will be presented every 6 months in order to allow

evaluation by the ExCom. When milestones are reached, or deliverables are ready, the relevant reporting will be assessed for quality by the coordinator and ExCom members. The PMT will work closely with the call secretariat staff and administrations of the partners to keep administrative procedures to a minimum so that resources can be targeted on the agreed deliverables and milestones. Criteria for assessing the achievement of each milestone will be identified.
Risk assessment, covering all ENPADASI activities, will be implemented to produce a timely response to issues unforeseen in the planning phase. An integrative approach will be used to execute, monitor, and evaluate the progress within and between WPs and tasks. Each task has clearly specified deliverables and milestones. WP leaders are responsible for planning, implementation and control of tasks within their WP. For each task, a consistent set of planning documentation will be available as part of the ENPADASI project management, including resources involved, deliverables, milestones and expected results.
Taken together, the progress will be monitored by timeline-based regular reporting and meetings of the PMT, ExCom and GA. Reports will be communicated between partners by e-mail and by the project intranet. The monitoring of the project at different levels is outlined below:
- The secured intranet section of the website will serve as the main platform for direct exchange of information between researchers within and between WPs. This limited-access portion of the website will also be used as a site to distribute documents of general use of the consortium;
- The WP leaders will be responsible for the work within their respective WP. Participants in the WPs will inform the WP leaders when problems are encountered. He/she will take appropriate measures. Major problems will be discussed in (teleconference) meetings with the PMT;
- Meetings with the ExCom will be organised every three months, and additionally at the moment interaction between WP's is required;
- Half-yearly progress reports will be prepared by the PMT on the basis of WP and ExCom input, using the project intranet;
- On basis of half-yearly progress reports, the PMT will monitor progress of the project, identify major bottlenecks and, together with the ExCom, find and implement solution for these problems. Where needed, adaptations to the project plan will be made with the aim to ensure the delivery of the project results. Major adaptations need to be approved by the GA.

Conflict resolution
In case of conflict, the WP leaders will have first responsibility to find an amicable solution, if appropriate in consultation with the project coordinator. In case the conflict cannot be resolved, the WP leader will put the issue forward to the ExCom. If the ExCom is unable to find an amicable solution, the issue will be voted on by the GA.
In case this happens, the issue and its outcome will be reported to the call secretariat.
All disputes arising out of or in connection to the Consortium Agreement, which cannot be solved amicably, shall be finally settled under the Rules of Arbitration of the International Chamber of Commerce by one or more arbitrators appointed in accordance with the said Rules.

4. Description of the Knowledge Hub as a whole in respect to complementarities and balance among its main themes
TNO is a not-for-profit organisation. The Microbiology and Systems Biology group focuses on improving health by life-style interventions. TNO is elected to be the coordinator of ENPADASI (WP1, all tasks). Furthermore, TNO will contribute to all other WPs either in-kind, in cash or both. In particular, TNO will share datasets, will contribute to the inventory of available databases and tools and the establishments of requirements of the infrastructure and tools. In addition, TNO will contribute to definition of a common language and querying data and data integration, respectively. TNO will be task leader of task 6.5 'Online FAQ and help desk'.
EoI45 involves 8 Italian institutes whose activities are dedicated to the development and application of bioinformatic tools for the integration of omics data (genome, metagenome, metabolome) with nutritional

studies information. EoI45 will coordinate the activities of WP3 (leader C. Priami) and WP4 (leader D. Cavalieri) in which consortium specific action aims to develop a network of computational infrastructures as well as pipelines for data integration in nutrigenomics studies and is elected as deputy for the ENPADASI consortium

CRA-NUT is the former National Research Institute on Food & Nutrition (INRAN). CRA-NUT coordinates the JINN-DAT Consortium in ENPADASI (EoI41), which proposes to represent a national hub for data collection, integration and harmonization. CRA-NUT contribution to the ENPADASI KH consists of transcriptomic datasets representing the output of two-hit challenge studies applying secondary inflammatory hits to nutritionally stressed cells, to reveal underlying mechanisms that are not phenotypically evident following nutritional challenge alone.

The JINGO consortium represents a total of 4 universities in Ireland who have extensive experience in performing nutritional studies ranging from intervention's to national food consumption studies. The JINGO consortium brings a wealth of knowhow in relation to nutrition studies and will take the lead in the Training WP.

The FiND consortium unites high quality research groups from different Flemish academic Institutions. The consortium connects expertise, experience and data (from fundamental, clinical and epidemiological research) within a broad research framework aimed at better understanding the biological and environmental (including lifestyle) determinants of chronic diseases and healthy ageing. This consortium covers all the required expertise for the tasks elaborated in this Knowledge Hub.

NEH-ULg consortium consists of two Belgian organizations. The NEH-ULg team has a good IT knowledge in the development and management of databases and web-based applications and has developed specific expertise in population surveys, being in charge of the health interview surveys and of the food consumption survey.

INDIANA is a multidisciplinary consortium with 8 researchers from 4 Flemish Universities. INDIANA will contribute with knowledge on semantics, data harmonization, analysis, knowledge on nutritional requirements, assessment and harmonization protocols. In addition, the consortium has experience in database development, data quality issues related to the detection and handling of duplicate and non-structured data and integration of consumer behavior in the assessment and standardization of data related to dietary habits.

The Health Research Institute La Fe is a non-profit organisation that carries out the scientific and research policy of the "La Fe" Health Department. The IIS La Fe comprises multidisciplinary and complementary research groups with consolidated expertise in cellular and molecular biology, cell culture, functional analysis imaging, "omics" and clinical research.

The MNES (Metadatabase for Nutritional Epidemiological Studies) consortium consists of six partner institutions from Germany, each conducting or planning large scale population-based studies with a focus on nutritional epidemiological aspects and/or diet-related chronic diseases. The consortium will contribute the identification of observational studies on nutritional epidemiological aspects for integration and collection of metadata for these studies as well as the identification of minimal requirements for observational study data, and the definition of quality criteria of observational studies.

The Molecular Nutrition Unit (MNU) is part of the research department Nutrition and Food Sciences at the TUM School of Life Sciences Weihenstephan of the Technische Universität München (TUM) in Munich Germany. TUM will provide raw data sets from some of the human studies carried out with well characterized and phenotyped healthy and prediabetic volunteers and will to be involved in investigating and optimizing the usability of the user-interface of the ENPADASI platform.

The Estonian consortium consists of two organisations: the National Institute for Health Development (NIHD) and the Bio Competence Centre of Healthy Dairy Products LLC (BioCC). The consortium will actively contribute in WP2 and WP5 by standardizing methods and data, defining the standardized phenotypic information with regard to diet, physical activity levels, other lifestyle factors and all biological, clinical and physiological measurements, and elaborating guidelines and regulations.

CIBERDEM is the Diatebes Research Area of the Centro de Investigación Biomédica en Red (CIBER). The work of CIBERDEM will consist of reviewing and searching collections of data for data integration at

national level and identifying the minimal requirements for studies and validating the studies quality. In addition, it will contribute to defining the common language, the ontologies for nutritional studies and will also contribute to the integration guidelines.
The Spanish consortium CIBEROBN is a biomedical institution of excellence integrating different research groups focused on nutrition, obesity and cardiovascular diseases. CIBEROBN will participate in WP1 task 2 (Governance & sustainability) and contribute to WP2 (sharing existing data), providing expertise in Epidemiology, Nutrition, Genetics, Disease Phenotypes, Clinical, "omics", environmental exposure. Finally it will provide expertise in nutritional ontologies and computer science and information systems expertise.
The Department of Nutrition, Exercise and Sports (NEXS) within the University of Copenhagen (UCph) has a large capacity for conducting human studies and has facilities and trained staff for quality control, data hub management as well as internationally leading nutritional researchers to support high-quality data flows, data security and data interpretation. NEXS also has a large study database to support ENPADASI; the data include many different study- and data types, including matrix-structured data. NEXS will be involved in the developmental phase of the ENDAPASI infrastructure. NEXS will act as WP leader (Lars Ove Dragsted) for WP2.
The FORCE-EISBM consortium consists of two different partner institutions from France. It has cutting-edge expertise in both basic sciences on adipose tissue and energy metabolism, facilities and know-how in clinical research in nutrition, physical activity and bariatric surgery and the clinical organization for patient recruitment as Integrated Centres for Obesity (ICO). The will consortium lead WP5. The objective is to make data sharing possible with respect of the policy, ethical, privacy, IP aspects taking into account the differences between the national regulations and the new regulation that are in discussion at the EU level.

5. Transnational added value of the Knowledge Hub
None of the partners or countries have all the knowledge, tools and/or data that are essential to develop an effective open access research infrastructure for the nutritional community.

6. Potential impact on the advancement of the research area, capacity building, plan for translation of research (suitable for ENPADASI) into public health practice or policy (in 2 years, with a perspective on a longer term)
This initiative will develop an open access research infrastructure (RI) for all nutritional mechanistic, intervention and epidemiological studies. For this development standardisation is crucial as combining studies depends on mapping of similar data and design. In two year this work will lead to a fully functional prototype of an infrastructure system that allows generalized data sharing and our impact will be to demonstrate the possibilities of fulfilment of these needs and of a much better use of nutritional study data for future analyses across data sets and study types. The work classifying nutritional study types and unifying the use of terms from vocabularies generated by others will have a large impact on future data sharing also outside of the ENPADASI consortium by linking to initiatives such as ELIXIR. There is also a significant impact by the contributions to visualize the potentials of data sharing in the form of case studies and training materials that will impact on a number of nutrition researchers who will be informed also of their own gains from data sharing. No doubt this infrastructure will need further refinement and continuous updates to become sustainable and the prospects and future needs to sustain the infrastructure will be described as part of WP1 by the end of the two-year project period. Building a common language for federating the nutritional databases and parallel interrogation of all the publically funded resources present within the EU will permit a change of paradigm in the way scientific datasets are used by the lawmakers, medical doctors, funding institutions, etc. Our project will grant intelligent access and categorization of big data, including negative results and unpublished information, hat are currently most often unstructured and poorly accessible. The potential outcome will be easier access to medically and nutritionally relevant datasets and exploitation of previously unreleased datasets as well as published ones to the benefit of the wider EU based scientific community. On the longer term,

this infrastructure will help to form better and more informed hypotheses in future studies and make it possible to actually test some new hypotheses without having to conduct additional studies. The availability of a European infrastructure for nutrition research will enhance the capacity of biologists/nutritionists and clinicians to carry out high-impact research. It will constitute an atlas of findings in the systems nutrition area that will be useful to improve the health system. This possibility is even increased by the presence in the consortium of clinicians that will also help structuring the infrastructure in such a way that it is suitable to support translational medicine. Translation of research in health practice or policy will be considered in the sustainability plan. In the future, the infrastructure may be expanded to other European countries and could be considered as or may be linked to the data sharing infrastructure in related fields.

7. Overall strategy of the work plan (including timing of different work packages and their components) including long-term strategy for sustainability

The work of the proposed project will be integrated in the following 6 Work Packages (WP). ENPADASI will resolve several of the technological challenges. Work package 2 will facilitate and stimulate data sharing by defining standards and minimal requirements. Work package 5 will limit the thresholds to share data by resolving and mapping ethical, IP and personal data collection issues. Work package 4 will develop relevant ontologies and mappings to facilitate data integration and scripts for actual data integration to show the added value of data sharing. Work package 3 will develop the integrated infrastructure and work package 6 will take care of the training in the use of the infrastructure. WP1 will develop plans for long-term sustainability.

The planning of the WPs and their component is show below (gant chart and Fig. 1) and is described in detail in WP1.

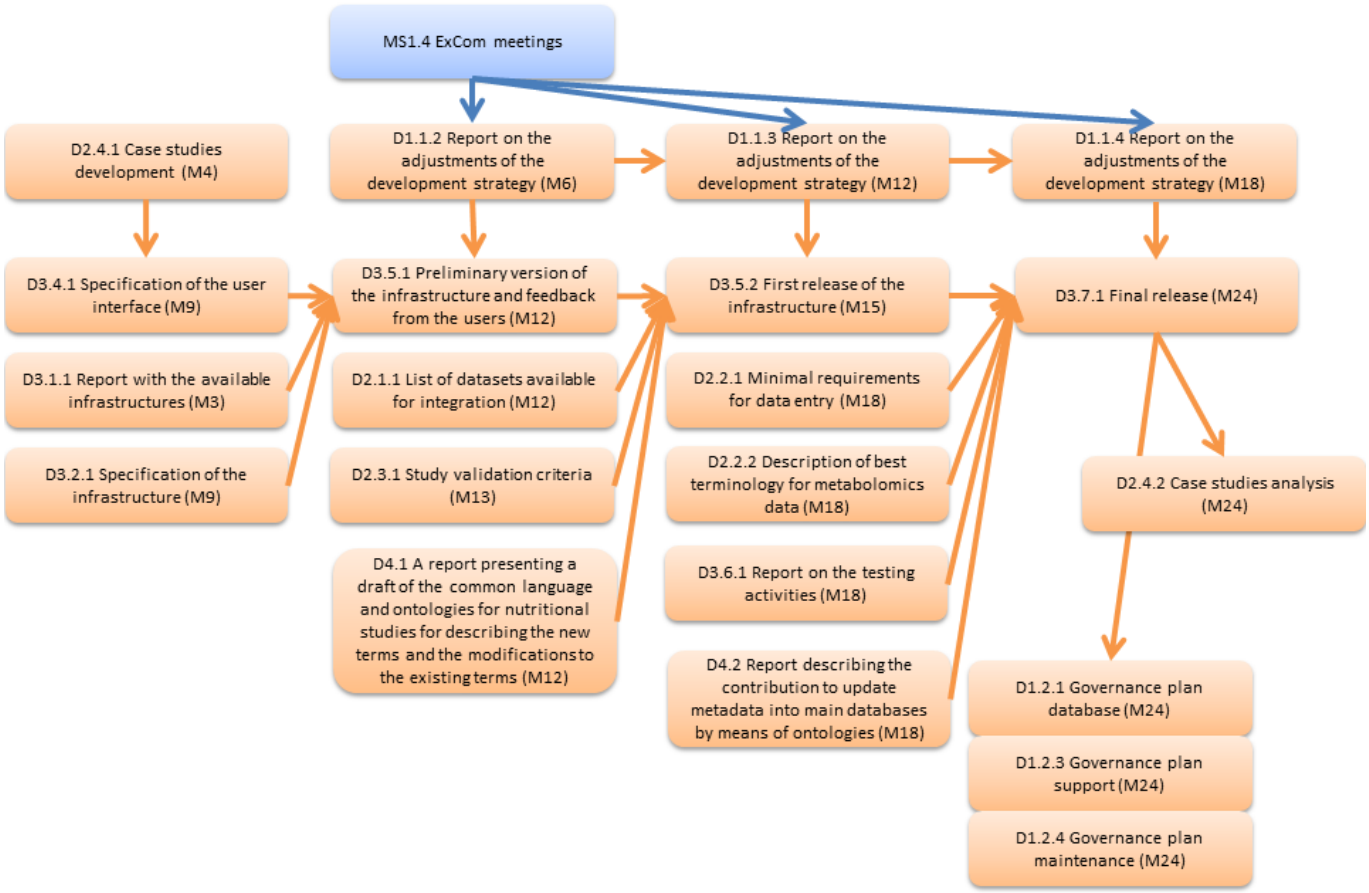


Figure 1. Relations between the database related deliverables. Other deliverables in the project have no relation or only one with another deliverable

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Activity		year 1												year 2											
WP1		WP 1 Management, coordination, governance and sustainability																							
T1.1.1	Overall coordination	1																							24
T1.1.2	Project boards, internal communication and sharing	1																							24
T1.1.3	Organisation of kick-off and periodical meetings	1																							24
T1.1.4	Legal, administrative management	1																							24
T1.2.1	Database governance													13											24
T1.2.2	Analysis governance													13											24
T1.2.3	Governance training and support													13											24
T1.2.4	Maintenance													13											24
WP2		WP 2 Preparing joint data analysis																							
T2.1	Collection of data for sharing	1																	18						
T2.2	Minimal requirements for study data			3																					24
T2.3	Validation of study quality	1											12												
T2.4	Case studies with existing data												12												24
T2.5	Guidelines for data sharing																16								24
WP3		WP 3 Design and development																							
T3.1	Inventory of the technical structure of currently available datasets & tools	1					6																		
T3.2	Functional/technical requirement data infrastructure	1																	18						
T3.3	Functional/technical requirements tools						6																		24
T3.4	User survey for usability of infrastructure	1					6																		
T3.5	Development of infrastructure and first release						6									15									
T3.6	Testing the first release															15			18						
T3.7	Refinement and final release																		18						24
WP4		WP 4 Integration																							
T4.1	Define a common language and building ontologies for nutritional studies	1											12												
T4.2	Mapping of terms; metadata integration	1																							24
T4.3	Integration of ontologies and pathways	1																							24
T4.4.2	Querying data and Data integration; semantics	1																							24
T4.4.1	Querying data and Data integration; algorithms	1																							24
T4.5	Integration of guidelines	1																	18						
WP5		WP 5 Regulations																							
T5.1.1	Ethics; Mapping of the national ethics guidelines and procedures	1																			20				
T5.1.2	Ethics; Track, share, and take into account European recommendations	1																			20				
T5.2.1	Data protection; Mapping of the national regulations	1																			20				
T5.2.2	Data protection; Mapping of potential EU regulations	1																			20				
T5.3.1	Intellectual property; Mapping of the national IP rules	1																			20				
T5.3.2	Intellectual property; Mapping of the rules established at the EU level for IP in data sharing	1																			20				
T5.4.1	Data sharing policies; Mapping of the data sharing policies established at the EU level by different consortia and by the JPI task force												12								20				
T5.4.2	Data sharing policies; Definition of general policies of ENPADASI												12								20				
T5.5	General guidelines about regulations																					21			24
WP6		WP 6 Training																							
T6.1	Providing training in ethics, privacy and IP	1																							24
T6.2	Resource of Standard Operation Procedures (SOPs) for data collection	1																	18						
T6.3	Training in ontology													13											24
T6.4	Training in Database Interface								8																24
T6.5	Development of online FAQ and help desk													13											24

8. Dissemination/exploitation strategy (including IPR, if applicable)

Before the start of the project a Consortium Agreement (CA) detailing all aspects concerning IPR-related matters will be agreed among the partners. In particular, the CA will address consistently with the Grant Agreement provisions, all issues related to intellectual property rights in the context of the project, including items such as confidentiality, liability, publication and exploitation of the project results, use of the project's database, rules for arbitration and recourse. This will also include the management of partners' background IP.

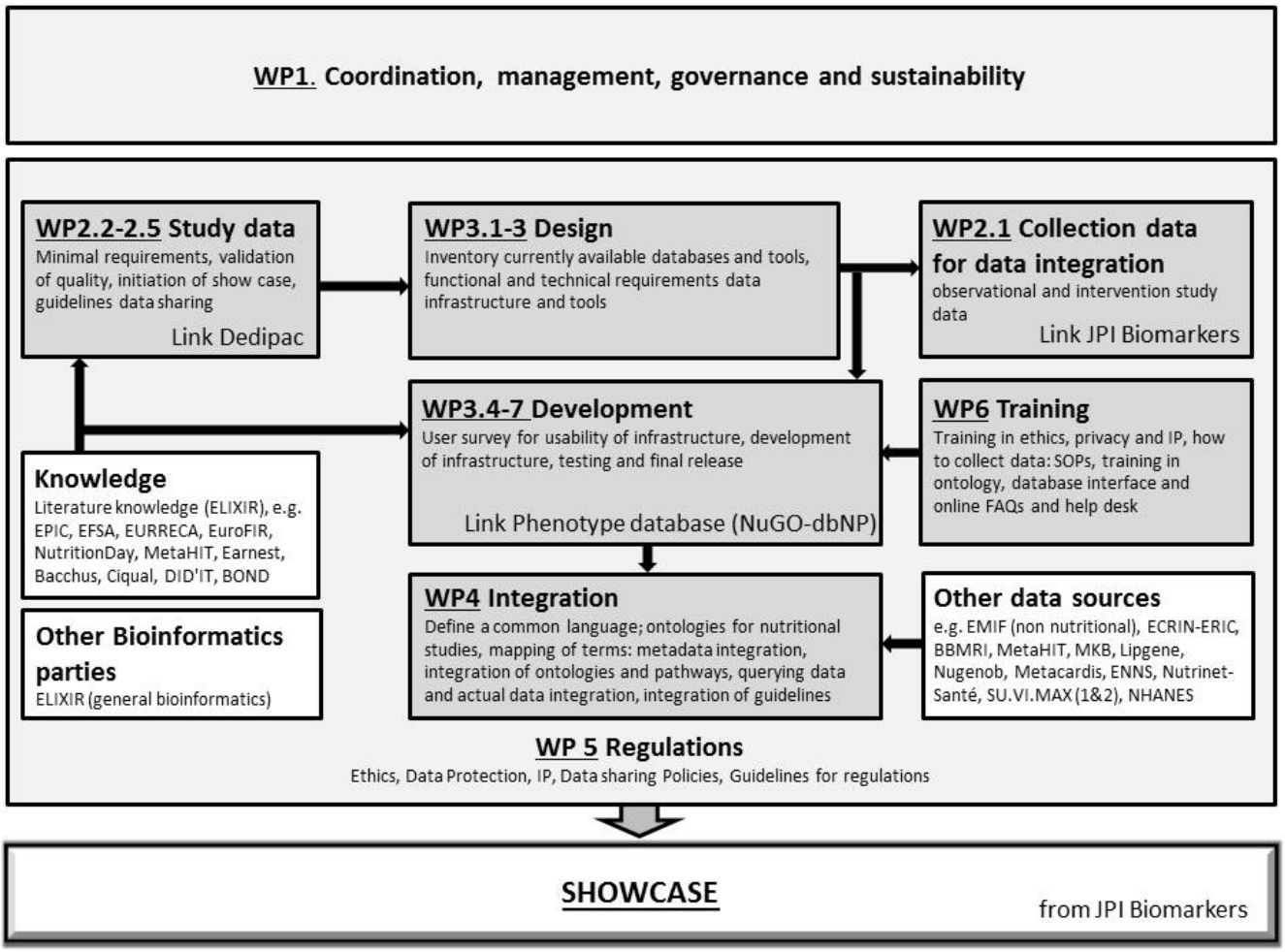
Foreground knowledge shall be the property of the Party carrying out the work generating that Foreground. If the Parties jointly generate the Foreground, they shall have joint ownership on basis of the ratio number of person months used to generate that Foreground. The joint owners will seek to agree between them for obtaining and/or maintaining such shared rights on a case-by-case basis. The Background of each party will be lay down in an Annex to the Consortium Agreement, at the start of the project. To prevent misunderstanding no list of 'Background excluded' will be used. All Background that is not mentioned in the list 'Background included' is automatically excluded.

During the project implementation, the ExCom will be in charge of monitoring the produced foreground IP and to identify any specific issues that requires (prior) agreement among or between partners. Possible patentable results will be sent to the project coordinator as well as to the IP officer of the partner that has generated the specific knowledge. TNO has ample expertise in managing the IP matters within its organization, ranging from confidential disclosure agreements to patent applications and the generation of spin-off companies.

The ExCom will establish a Knowledge Management Committee that will specifically execute all monitoring, judgment and valorisation issues. The Knowledge Management Committee will oversee the handling and publication of any knowledge generated by the project and will ensure proper IP protection

implementing those parts of the CA covering intellectual property.
If considered appropriate, additional collaborators will be identified, contacted, and invited to talk to the ENPADASI partners related to exploitation of IP during the course of the project. TNO will advise the ExCom on legal issues related to IP management and exploitation.
Any issues regarding conflicts and IPR protection that could not be removed bilaterally within WPs, will be solved according to CA. This will be addressed by defining clearly in the CA the responsibilities and procedures related to the definition of ownership of results as well as access rights. Decisions related to IPR issues will be taken by the General Assembly (GA).

Pert Chart


WP1. Coordination, management, governance and sustainability
WP2.2-2.5 Study data
Minimal requirements, validation of quality, initiation of show case, guidelines data sharing
Link Dedipac
WP3.1-3 Design
Inventory currently available databases and tools, functional and technical requirements data infrastructure and tools
WP2.1 Collection data for data integration
observational and intervention study data
Link JPI Biomarkers
WP3.4-7 Development
User survey for usability of infrastructure, development of infrastructure, testing and final release
Link Phenotype database (NuGO-dbNP)
WP6 Training
Training in ethics, privacy and IP, how to collect data: SOPs, training in ontology, database interface and online FAQs and help desk
Knowledge
Literature knowledge (ELIXIR), e.g. EPIC, EFSA, EURRECA, EuroFIR, NutritionDay, MetaHIT, Earnest, Bacchus, Ciqual, DID'IT, BOND
Other Bioinformatics parties
ELIXIR (general bioinformatics)
WP4 Integration
Define a common language; ontologies for nutritional studies, mapping of terms: metadata integration, integration of ontologies and pathways, querying data and actual data integration, integration of guidelines
Other data sources
e.g. EMIF (non nutritional), ECRIN-ERIC, BBMRI, MetaHIT, MKB, Lipgene, Nugenob, Metacardis, ENNS, Nutrinet-Santé, SU.VI.MAX (1&2), NHANES
WP 5 Regulations
Ethics, Data Protection, IP, Data sharing Policies, Guidelines for regulations
SHOWCASE
from JPI Biomarkers

Gantt Chart

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Activity		year 1												year 2											
WP1		WP 1 Management, coordination, governance and sustainability																							
T1.1.1	Overall coordination	1																							24
T1.1.2	Project boards, internal communication and sharing	1																							24
T1.1.3	Organisation of kick-off and periodical meetings	1																							24
T1.1.4	Legal, administrative management	1																							24
T1.2.1	Database governance													13											24
T1.2.2	Analysis governance													13											24
T1.2.3	Governance training and support													13											24
T1.2.4	Maintenance													13											24
WP2		WP 2 Preparing joint data analysis																							
T2.1	Collection of data for sharing	1																	18						
T2.2	Minimal requirements for study data			3																					24
T2.3	Validation of study quality	1											12												
T2.4	Case studies with existing data												12												24
T2.5	Guidelines for data sharing																16								24
WP3		WP 3 Design and development																							
T3.1	Inventory of the technical structure of currently available datasets & tools	1					6																		
T3.2	Functional/technical requirement data infrastructure	1																	18						
T3.3	Functional/technical requirements tools						6																		24
T3.4	User survey for usability of infrastructure	1					6																		
T3.5	Development of infrastructure and first release						6									15									
T3.6	Testing the first release															15			18						
T3.7	Refinement and final release																		18						24
WP4		WP 4 Integration																							
T4.1	Define a common language and building ontologies for nutritional studies	1											12												
T4.2	Mapping of terms; metadata integration	1																							24
T4.3	Integration of ontologies and pathways	1																							24
T4.4.2	Querying data and Data integration; semantics	1																							24
T4.4.1	Querying data and Data integration; algorithms	1																							24
T4.5	Integration of guidelines	1																	18						
WP5		WP 5 Regulations																							
T5.1.1	Ethics; Mapping of the national ethics guidelines and procedures	1																			20				
T5.1.2	Ethics; Track, share, and take into account European recommendations	1																			20				
T5.2.1	Data protection; Mapping of the national regulations	1																			20				
T5.2.2	Data protection; Mapping of potential EU regulations	1																			20				
T5.3.1	Intellectual property; Mapping of the national IP rules	1																			20				
T5.3.2	Intellectual property; Mapping of the rules established at the EU level for IP in data sharing	1																			20				
T5.4.1	Data sharing policies; Mapping of the data sharing policies established at the EU level by different consortia and by the JPI task force												12								20				
T5.4.2	Data sharing policies; Definition of general policies of ENPADASI												12								20				
T5.5	General guidelines about regulations																					21			24
WP6		WP 6 Training																							
T6.1	Providing training in ethics, privacy and IP	1																							24
T6.2	Resource of Standard Operation Procedures (SOPs) for data collection	1																	18						
T6.3	Training in ontology													13											24
T6.4	Training in Database Interface								8																24
T6.5	Development of online FAQ and help desk													13											24

Deliverables					
Deliverable no	Deliverable name	WP no	Date	Nature	Dissemination level (public, confidential, …)
D1.1.1	Template for reporting by each partner to the Coordinator and then to the EC and possible updates		M3	R	PU
D.1.1.2	Report on the adjustments of the development strategy		M6	R	PU
D1.1.3	Report on the adjustments of the development strategy		M12	R	PU
D1.1.4	Report on the adjustments of the development strategy		M18	R	PU
D1.2.1	Governance plan database		M24	R	PU
D.1.2.2	Governance plan analysis pipeline		M24	R	PU
D.1.2.3	Governance plan support		M24	R	PU
D1.2.4	Governance plan maintenance		M24	R	PU
MS1.1	Advisory Board established		M3	R	PU
MS1.2	Knowledge Management Committee established		M3	R	PU
MS1.3	First Plenary Group meeting		M3	Symposium programme	PU
MS1.4	ExCom meetings (every 6 months)		M6, M12, M18	R	PU
MS1.5	Second Plenary Group meeting		M12	Symposium programme	PU
MS1.6	Finalization meeting (Plenary Group)		M24	Symposium programme	PU
D2.1.1	List of datasets available for integration	Preparing joint data analysis	M12	O, list	PU
D2.2.1	Minimal requirements for data entry	Preparing joint data analysis	M18	R	PU
D2.2.2	Description of best terminology for metabolomics data	Preparing joint data analysis	M18	R	PU
D2.3.1	Study validation criteria	Preparing joint data analysis	M13	O, manuscript	CO, until final publication
D2.4.1	Case studies development	Preparing joint data analysis	M24	R	PU
D2.5.1	Guidelines for nutritional data sharing	Preparing joint data analysis	M24	O, guidelines	PU
D3.1.1	Report with the available infrastructures	Design and development	M3	R	PU
D3.2.1	Specification of the infrastructure	Design and development	M6	O, manuscript	CO, until final publication
D3.3.1	Specification of the tools	Design and development	M12	O, manuscript	CO, until final publication
D3.4.1	Specification of the user interface	Design and development	M6	O, manuscript	CO, until final publication

D3.5.1	Preliminary version of the infrastructure and feedback from the users	Design and development	M12	P	PU
D3.5.2	First release of the infrastructure	Design and development	M15	P	PU
D3.6.1	Report on the testing activities	Design and development	M18	R	CO
D3.7.1	Final release	Design and development	M24	D	PU
MS3.1	Workshop with all partners of WP2 and WP4 to share all available data and infrastructure	Design and development	M3	Workshop programme	PU
MS3.2	Hackathon with all programmers from the different consortia	Design and development	M3	Hackalthon programme	PU
MS3.3	Workshop with all partners of WP2 and WP4 to test the infrastructure	Design and development	M18	Workshop programme	PU
D4.1	A report presenting a draft of the common language and ontologies for nutritional studies for describing the new terms and the modifications to the existing terms	Integration	M12	R	PU
D4.2	Report describing the contribution to the update metadata into main databases by means of ontologies	Integration	M18	R	PU
D4.3	Report describing the update of Nutrition Pathways Data Model	Integration	M24	R	CO, until final publication
D4.4	A report on protocols for querying data, data integration and usage of pathway tools	Integration	M24	R	CO, until final publication
D4.5	Guidelines for study reporting and study design	Integration	M18	O, guidelines	PU
MS4.1	Organization of an Ontology meeting with the community of domain experts and ontology experts	Integration	M6	Meeting programme	PU
MS4.2	One day workshop on metadata integration	Integration	M12	Workshop programme	PU
MS4.3	Data model definition	Integration	M24	R	PU
MS4.4	Setup of dedicated website for web-based resource for data querying	Integration	M24	website	PU
MS4.5	Integration of guidelines for study reporting and study design into the ENPADASI website	Integration	M18	R	PU
D5.1	Ethics; Guidelines	Regulations	M24	O, guidelines	PU
D5.2	Data protection; Guidelines	Regulations	M24	O, guidelines	PU
D5.3	Intellectual property; IP guidelines	Regulations	M24	O, guidelines	PU