

Working Package

WP no	WP name	Lead partner	Start month	End month	Deliverable no
3	Design and development	Corrado Priami <i>priami@cosbi.eu</i>	1	24	Report with the available infrastructures ,Specification of the infrastructure,Specification of the tools,Specification of the user interface,Preliminary version of the infrastructure and feedback from the users,First release of the infrastructure,Report on the testing activities,Final release,Workshop with all partners of WP2 and WP4 to share all available data and infrastructure,Hackathon with all programmers from the different consortia,Workshop with all partners of WP2 and WP4 to test the infrastructure

Detailed information on work packages

WP Leader:

Legal name of organisation: The Microsoft Research - University of Trento Centre for Computational and Systems Biology

Country: IT

ZIP code: 38068

Town: Rovereto

Street name, number: Piazza Manifattura 1

Additional (e.g., department, building...):

Website:

Mrs/Mr: Corrado Priami

Title: PROF

First name:

Last name:

Function:

Phone: +39 0464 808811

Fax:

E-mail: PROF

Additional information on person

(max. 1500 characters) concerning personal background and explain responsibilities and tasks:

Corrado Priami is professor of Computer Science at the University of Trento. The results of his PhD thesis on stochastic pi-calculus were the basis for the foundation of the Microsoft Research - University of Trento Centre for Computational and Systems Biology (COSBI), of which he is the President and CEO. Those same results are recognized as fundamental in the field of systems biology by an expanding international community, which is using them to model the behavior of biological systems (the CMSB conference is a milestone of this). He was member of the expert group on the EU 7th FP of the CRUI and has participated in many EU projects (also as coordinator) for the advancement of emerging areas of research. He serves on evaluation committees for EU projects as well as for many international funding agencies. He is a member of the Scientific committee of Fondazione Veronesi. His research covers computational methods for the modeling, analysis, and simulation of biological systems and programming languages. He published over 180 scientific papers, gave more than 50 invited talks and lectures at conferences and universities around the world, participated in many program committees for international conferences (also as chair). He founded the international conferences "Computational Methods in Systems Biology (CMSB)" and "Converging Sciences". He was a member of ISTAG-FET (Information Society Technologies Advisory Group - Future and Emerging Technologies) of the EU

Commission.

Partners Involved				
Legal name of organisation	Knowledge Hub member (main contact person/ project leader within the organisation)	Person months #	Start month	End month
Ghent University - Faculty of Medicine - Department of Public Health	De Henauw Stefaan	8	1	24
CRNH Rhône-Alpes - Bt 1A	Laville Martine	1	1	6
CRA-NUT	Giuditta Perozzi	2	1	24
Institute of Biomembranes and Bioenergetics - c/o Dipartimento di Bioscienze, Biotecnologie e Biofarmaceutica	Graziano Pesole	24	1	24
Istituto di Analisi dei Sistemi ed Informatica Consiglio Nazionale delle Ricerche	giovanni felici	3	1	24
The Microsoft Research - University of Trento Centre for Computational and Systems Biology	Corrado Priami	10.8	1	24
Alma Mater Studiorum - Università di Bologna - Dept. Scienze e Tecnologie Agroalimentari	Francesco Capozzi	5.5	1	24
National Research Council - IBIMET	Duccio Cavalieri	1.1	1	24
Politecnico di Bari – POLIBA - The Interuniversity Physics Department	Giorgio Pietro Maggi	4	1	24
Netherlands Organisation for Applied Scientific Research (TNO) - Department of Microbiology and Systems Biology	Jildau Bouwman	5.4	1	24
Max Delbrück Center for Molecular Medicine (MDC) - Molecular Epidemiology Research Group	Tobias Pischon	18	1	24
German Institute of Human Nutrition (DIFE) - Department of Epidemiology	Heiner Boeing			
Department of Nutrition and Food Sciences, University of Bonn	Ute Nöthlings			
Helmholtz Zentrum München (HMGU) - Research Group National Cohort	Jakob Linseisen			
Max Rubner-Institute (MRI), Federal Research Institute of Nutrition and Food - Department of Nutritional Behaviour	Ingrid Hoffmann			
University Heart Center, University Hospital Hamburg-Eppendorf - Preventive Medicine	Prof. Eberhard Windler, M.D.			
Bio-Competence Centre of Healthy Dairy Products (BioCC)	Andre Veskioja			
National Institute for Health Development (NIHD) - Department of Surveillance and Evaluation	Eha Nurk	8.5	1	18
University of Liège - Unit of Prof. Michèle Guillaume "Nutrition, Environment and Health" - Department of Public Health	Anne-Françoise Donneau	8	6	15
University of Copenhagen - Dept. Nutrition, Exercise and Sports	Lars Ove Dragsted	8.5	1	24
CIBER OBN - Instituto de Salud Carlos III	Dolores Corella	0.3	1	24

Health Research Institute Hospital La Fe - Endocrinology, Nutrition and Dietetic Unit	Jose M. Soriano	1	18	24
Gent University	Carl Lachat	3	1	18
Technische Universität München - Molecular Nutrition Unit	Kurt Gedrich	2	1	18
Total		114.1		

Description of work package:

General Description description-10-545c8ec2d42be.docx [Show Uploaded Description](#)

Work package 3

WP name: Design and development

WP leader: Prof. Dr. Corrado Priami (Italy)

Description of work package:

1. Scope of work package

The purpose of this WP is to connect and further develop data infrastructures (inventory of relevant systems in task 3.1) for enhancing data sharing and exchange for systems nutrition research in order to facilitate research and reuse of data. This work will result in an infrastructure that is scalable and sustainable over time. The work of EuroDISH, where the needs for research infrastructures are mapped, and the inventory of WP2 will be used as a basis for the design and the focus will lie on the development of data solutions of nutrition specific data problems. The data solution will facilitate both data storage and analysis (link with WP2 and WP4). The initial work of the WP will include development of a functional and technical design (task 3.2 for the database and 3.3 for the analysis tools), which will be based on the work of WP2, WP3 and WP4. The functional design document will include sketches of the screens of the user interface. The usability of the user interface will be analysed by a user survey (task 3.4), which will be filled out by all partners of WP2. The functional and technical design document will form the basis for the implementation of a first version of the data infrastructure that will be available to the systems nutrition community (task 3.5), which will be tested in task 3.6 by all partners of WP2. The results will be shared with WP4 for further development of the analysis tools and bugs and feature requests from these tests will be used for further development of the infrastructure (task 3.7).

Deliverables

- D3.1.1** Report with the available infrastructures (including DBs and tools) at M3
- D3.2.1** Specification of the infrastructure at M9 (possibly resulting a scientific paper to be published)
- D3.3.1** Specification of the tools at M12 (possibly resulting a scientific paper to be published)
- D3.4.1** Specification of the user interface at M9 (possibly resulting a scientific paper to be published)
- D3.5.1** Preliminary version of the infrastructure and feedback from the users at M12
- D3.5.2** First release of the infrastructure at M15
- D3.6.1** Report on the testing activities at M18
- D3.7.1** Final release at M24

Milestones

- MS3.1** Workshop with all partners of WP2 and WP4 to share all available data and infrastructure at M3

MS3.2 Hackathon with all programmers from the different consortia at M3

MS3.3 Workshop with all partners of WP2 and WP4 to test the infrastructure at M18

2. Concept and objectives

a. Objectives, vision including scientific/ technological challenges

The main objective of this work package is to further develop an infrastructure that allows the exchange of data in the field of systems nutrition at European level. A design phase in such a development is a fundamental step to move to the implementation of the infrastructure in such a way that different teams can work simultaneously (using github as sharing and versioning systems for communication and to prevent legacy problems). The main challenge is to identify the requirements of different groups (bioinformaticians, nutritionists, clinicians) in order to capture all the relevant information needed to capture experimental and observational studies in nutrition. From a technological perspective, we have to integrate many existing databases that store data in different format and that are geographically located in different countries. The challenges we face here are the typical ones for the development of distributed system/ federated database, which require communication between the systems using web services (such as the Phenotype database). An additional difficulty is that the users of the infrastructures may have a limited computing literacy so that the user interface is the bottleneck for the adoption of the interface. This is why we have a specific task for a user survey on the sketches of the interface.

State of the art

Biological data is heterogeneous in terms of format, protocols of transmission and interpretation. Providing an integrated framework to deal with the complexity of multi-level analyses is a strategic issue from both the computational and the algorithmic design point of view. The development and the implementation of user friendly solutions is a necessary step towards the efficient and widespread use of these instruments. The ENPADASI services should aim to overcome existing limitations due to hardware resources and algorithmic limitations and will make them suitable to run in a shared environment such as the ENPADASI federation of institutions and research centres. An example is the (Nutritional) Phenotype database (www.dbnp.org) was 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 with complex design (including 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. 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 open the data to the world (e.g. upon publication). The Phenotype database is built in a modular way (e.g. a study design, metabolomics module; all modules are connected via REST APIs), making it possible to link the data and design in the system to data in other databases (e.g. ArrayExpress or other instances of the Phenotype database) and has been tested to work on a cloud solution. TNO was in the lead for the functional requirements and responsible for steering the development work. The system has now been fully operational for 2 years (at four different locations). One of the instances is publicly available (studies.dbnp.org), on which daily and weekly back-ups are made. The database currently includes 58 biological studies, mostly nutritional intervention studies. The system is connected to other databases containing metabolomics data on www.metabolomexchange.org and is also used in the biomedical research area (CTMM/TraIT).

c. Scientific/ technological concept

The use of distributed infrastructures, such as local computing clusters or the geographically distributed Computing Grid facilities, has been the first answer to the needs of a continuously growing community of biological data analysts, with interesting recent developments. The next step which naturally arises is the involvement of Cloud technologies. In fact, combination of workflow approach and cloud computing paradigms could facilitate big data access and processing, addressing high computation and complex storage issues coming from actual research. It is also necessary to provide technical capabilities to federate genetic and clinical data currently stored in hospitals, research archives and databases while guaranteeing strong protection for sensitive information. The implementation of such ideas within a shared platform will allow the ENPADASI Project to exploit large-scale initiatives and create data acquisition standards that could later be used by institutions and research centres to provide a common "language" and a fruitful interaction. In particular, the present proposal plans to cover a crucial aspects: 1) implementing new elastic and customisable data workflows, providing automatic biological data processing services, through cloud computing solutions while dealing with volatile requirements in terms of data format and privacy. This WP will create an infrastructure that can enhance the research in systems nutrition at EU level. The scientific and technological development follows three main steps: design, user survey and development. The design will provide a non-ambiguous specification document shared between the partners to avoid subsequent changes in the development that would make the project fail due to time constraints. The user survey is needed to identify a transparent and unique interface that hides the technical detail from the users of the infrastructure still letting them access the data they need for their studies. The implementation is then the translation in software of the specification documents produced in the previous two activities.

3. Management

The WP will be managed by the WP leader in close collaboration with leaders of the individual tasks. Since collaboration has to take place continuously over the time span of the project much of the management will be performed using teleconferences with involved partners. Annual meetings will be conducted as part of the annual project meetings. Each partner will be responsible for their own reporting to their funders, but the work input from each partner will be made visible as a reference for all reporting activities.

4. Potential impact on the advancement of the research area, capacity building, plan for translation of research into public health practice or policy (in 2 years, with a perspective on a longer term)

The availability of a European infrastructure for nutrition research will enhance the capacity of small labs or purely biological labs 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. The Phenotype database, developed by the nutritional community, is implemented in the Dutch initiative CTMM/TraIT), which focuses on medical translational research.

5. Overall strategy of the work plan

The WP is organized in 7 tasks, each of which has a task leader responsible for coordinating the activities of the task. The WP leader will oversee the task leaders and keep strict connection with the other WP leaders to ensure the needed level of information flow and collaboration to deliver an infrastructure that is then used by the community. The survey on the design of the interface, the specification of data infrastructure and tools and the testing phase with the end-users are all preventive actions to limit the risk of delivering software that is not used.

We list below the main tasks with their schedule in months and the relation with the other activities in the project (distinguishing between input and output), the risks, and the deliverables of each task with the due date.

Task 3.1: Inventory of the technical structure of currently available databases & tools - Eol45 [lead], Eol50, 51, 71, 74, 75 – Month 1-6

This task will consolidate the mapping work of EuroDISH and investigate the available data resources available within and relevant to this consortium (e.g. Phenotype database, www.metabolomexchange.org) with input from WP2 task. 2.1 These data resources will be integrated in the one infrastructure of the project which will be based on the FAIRport concept (see Fig 2) and connected to relevant pipelines. The resulting infrastructure will at least include an atlas with the main features (input from WP 2 task 2.2) of each resource, access policies, where the data are collected, data representation and available pipelines. The automated querying systems should recognize the access level of the study. We will also consider and reuse what has been done in other applicative domains (e.g. through ELIXIR) to prevent double work and to favor the integration of data from different research efforts areas through computational infrastructures. Particularly relevant for the design of the infrastructure will be the selection and management of metadata associated with molecular and clinical data in relation to nutrition research (e.g. Individual phenotype, geographic origin, diet, physiological or pathological state, etc.).

INPUT from: WP2

The inventory on data availability within the consortium and the minimal requirements in terms of metadata and quality will complement the activities of this task

OUTPUT to: WP2, WP4

The outcome of this survey will inform and complement the activities of task 3.2, WP2 and WP4.

Related deliverables and milestones: D3.1.1

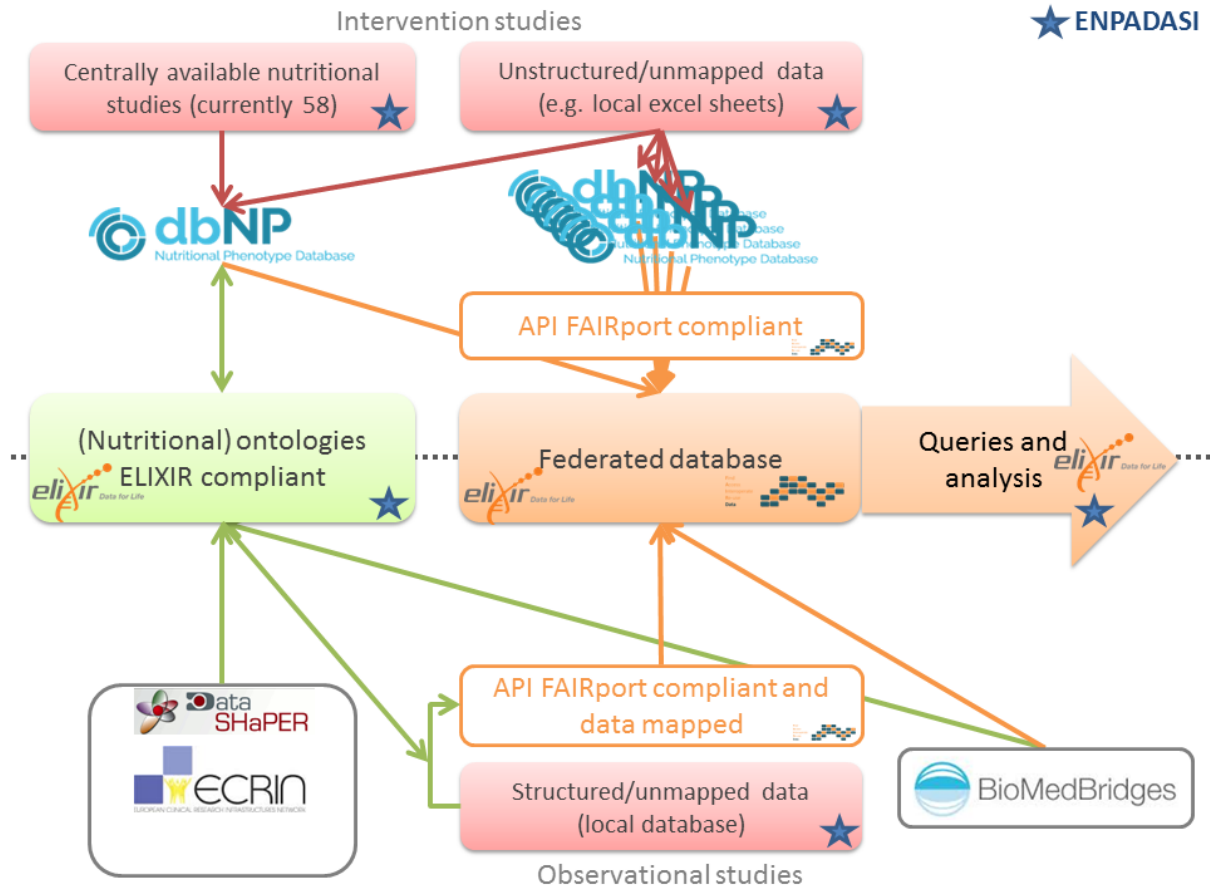


Figure 3. Database structure.

Task 3.2: Functional/technical requirement data infrastructure – Eol 45 [lead], 38, 50, 51, 71, 74, 75 – Month 1-18

This task will produce a specification of the data infrastructure that will be the basis for the implementation of the infrastructure (task 3.5) and will be based on work done earlier in the field (e.g. Phenotype database). The functional specification will be done in strict collaboration with the biologists (end-users) of the consortium and comparing our development with existing infrastructure for systems biology (e.g. SBML, OpenBEL, Combine community, etc.). The infrastructure will connect to the ongoing work in complementary efforts (JPI DEDIPAC, ELIXIR, etc.) and will integrate the relevant structured data in the consortium (depending on the selected study cases (task 2.4)). For intervention studies the infrastructure (Fig. 3) will make use of the Phenotype database (www.dbnp.org), which can store metadata on the study design (e.g. all details on the intervention that would also be needed in a publication) and measurement data (e.g. clinical chemistry, anthropometry, etc). This database makes use of templates and therefore will require limited adjustments depending on the selected study cases. Observational studies are generally stored in local databases; these will be made available by inclusion of an API, depending on the study cases. It will be essential to specify how to manage big datasets and how to integrate multi-source, multi-level 'omics' data with clinical markers (in collaboration with WP4). The identification of functional requirements will be important also for identifying the right scope of the ontologies to be developed in WP4. The requirement documents will be updated based on the test performed in task 3.6.

INPUT from: task 3.1, WP2, WP4, WP5

The input from WP2 on which data to be stored and how data should be shared is fundamental to determine the schema of the infrastructure as well as the management of the access rights and logging systems. Ontologies and mapping of terms from WP4 is needed to design efficient storing policies and internal representation strategies that allow to easily find terms with common meaning. To determine a scalable structure with different kinds of users the input on IP policies and privacy from WP5 is needed.

OUTPUT to: T3.5, WP4

The specification will be used to drive the implementation of the infrastructure (task 3.5) and will also inform the integration activities run in WP 4.

Related deliverables and milestones: D3.2.1

Task 3.3: Functional/technical requirements tools – Eol45 [lead], 45, 50, 51, 71, 74, 75 – Month 6-24

This task will automate querying/analysis tools that are developed in WP4 in such a way that they can be easily reused on different datasets. The specification will be done in strict collaboration with the biologists (end-users) of the consortium and comparing with /reusing existing data infrastructures in our development. We will design several workflows that integrate the different data sources equipped with schedulers and querying systems for the batch processing of demanding tasks. The requirement documents will be updated based on the test performed in task 3.6.

INPUT from: The same as INPUT from in the previous task

OUTPUT to: The same as OUTPUT to in the previous task

Related deliverables and milestones: D.3.3.1

Task 3.4: User survey for usability of infrastructure - Eol45 [lead], all partners of WP2, Month 1-9

This task will design the interface (with sketches) of the common access point to the infrastructure and will execute a survey of such an interface for usability and easiness of use with end-user. The final goal of this task is to specify the interface that will be then implemented.

INPUT from: all WP2 partners (and related consortia e.g. EuroDISH and JPI DEDIPAC)

This is essential to limit the risk of having an infrastructure that is not usable.

OUTPUT to: T3.5, WP4

The design of the interface will guide the interface implementation (task 3.5) and the integration activities in WP4

Related deliverables and milestones: D3.4.1

Task 3.5: Development of infrastructure and first release – Eol 45 [lead], 50, 64 – Month 6-15

This task will implement the infrastructure starting from the specifications delivered by tasks Task 3.2, Task 3.3 and Task 3.4.

INPUT from: T3.2, T3.3, T3.4, WP4

Besides the specifications produced in this WP, also the activities in WP4 will inform development.

OUTPUT to: whole consortium

Related deliverables and milestones: D3.5.1, D3.5.2

Task 3.6: Testing the first release– Eol 45 [lead], all partners of WP2 and WP4 – Month 15-18

This task will plan testing activities to obtain feedback from the user in the consortium to improve the usability and quality of the infrastructure as well as to remove bugs. The main activity will be the identification of significant case studies: extraction of data from the developed infrastructure and their integration in relation to a particular biological state of the individual (e.g. age, disease, diet type, etc.) to answer scientific questions. We will design batch procedures to run workflows to perform a stress test of the system.

INPUT from: all WP2 and WP4 partners

This is essential to limit the risk of having an infrastructure that is not usable. Particularly relevant here will be the input from WP2 on case studies.

OUTPUT to: T3.7, WP4

The outcome of the testing and of the activities in WP4 on integration will guide the delivery of the final release of the interface

Related deliverables and milestones: D.3.6.1

Task 3.7: Refinement and final release– Eol45 [lead] – Month 18-24

This task will update the first release of the infrastructure with the main requirements obtained from the testing phase in T3.6. In

collaboration with WP1 (sustainability) a location will be selected for the production location of the infrastructure. All developed code will also be shared via github (via the phenotype foundation: <https://github.com/PhenotypeFoundation>) under an open source license (e.g. Apache license). This will facilitate reuse of the code and will make local installation of the developed tools possible.

INPUT from: T3.6

OUTPUT to: whole consortium

Related deliverables and milestones: D3.7.1

Budgetary table				
	2014	2015	2016	Total Costs
personnel [k€]				0
travel [k€]				0
consumables [k€]				0
equipment [k€]				0
dissemination [k€]				0
others [k€]				0
direct costs [k€]				0
indirect costs [k€]				0
requested funding [k€]				0
Total [k€]	0	0	0	0