



## PROJECT DELIVERABLE REPORT

DELIVERABLE NUMBER AND TITLE	D4.1
TITLE	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
AUTHOR(S)	DUCCIO CAVALIERI, FRANCESCO VITALI, DAMARIZ RIVERO, ROSARIO LOMBARDO, FULVIO MATTIVI, PIETRO FRANCESCHI, FRANCESCO CAPOZZI, ALESSANDRA BORDONI UNIBO, ALESSIA TRIMIGNO, GIOVANNI FELICI, FRANCESCO TAGLINO, CARL LACHAT, CHEN YANG, NATHALIE DE COCK, LARS OVE DRAGSTED, JILDAU BOUWMAN
WORK PACKAGE	WP 4
TASK	TASK 4.1
WP LEADER	D. CAVALIERI
BENEFICIARIES CONTRIBUTING TO THE DELIVERABLE	CNR-IBIMET, UNIFI, COSBI, FONDAZIONE EDMUND MACH, UNIBO, CNR-IASI, UGENT, TNO, UNIVERSITY OF COPENHAGEN
STATUS – VERSION	FINAL - VERSION 3.0
DELIVERY DATE (MONTH)	M12
SUBMISSION DATE	M12

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

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



ENPADASI

European Nutrition Phenotype Assessment and Data Sharing Initiative

JOINT PROGRAMMING INITIATIVE – A HEALTHY DIET FOR A HEALTHY LIFE EUROPEAN NUTRITION PHENOTYPE ASSESSMENT AND DATA SHARING INITIATIVE

DISSEMINATION LEVEL – SECURITY*	PU
DELIVERABLE TYPE**	R

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

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



JOINT PROGRAMMING INITIATIVE – A HEALTHY DIET FOR A HEALTHY LIFE EUROPEAN NUTRITION PHENOTYPE ASSESSMENT AND DATA SHARING INITIATIVE

## CONTENTS

INTRODUCTION .....	4
METHODS .....	5
Collection of terms .....	5
Ontology development: from a common vocabulary to classes and subclasses .....	5
RESULTS AND DISCUSSION .....	8
FUTURE DEVELOPMENT .....	9
REFERENCES.....	10



## INTRODUCTION

With the great progress in analytical procedures in the past decades (i.e. the advent of *-omics* technologies), scientific community has faced a great development. Researchers in the biomedical domain in particular, are able to generate huge amounts of data almost on a daily basis. Considering its multidisciplinary nature, biomedical data are produced from a variety of scientific disciplines, and are thus traditionally stored in heterogeneous formats. Extracting new knowledge from those data can be a daunting task, and there is a clear need for facilitating (and ideally automating) the integrated use of such resources. (Carmen Legaz Garcia et al. 2016, Lamy 2016).

The necessity to structure and formalize this knowledge adequately led to the development of ontologies. As a result, hundreds of ontologies have been developed, and are playing a critical role in biomedical data and knowledge representation, integration, sharing and analysis. For example the Ontology for Biomedical Investigations (OBI) has been used for representation of a wide range of investigations (Ong and He, 2016). Ontologies can link knowledge together and produce new knowledge using a reasoner (Lamy 2016). Biomedical ontologies are consensus-based controlled biomedical vocabularies of terms and relations with associated definitions, which are logically formulated to promote automated reasoning (Cimino JJ, Zhu X 2006). They consist of defined concepts, typically structured within trees or networks, where the concept represents the nodes which are interconnected by semantic relationships (i.e. is-a, part-of) (Groß et al 2015). It is important to note, that in the biomedical domain the term “ontology” is often not used in the sense of formal, axiom-based ontologies but instead for a wide spectrum of simpler terminologies including a.o. thesauri, taxonomies and is-a-hierarchies. (Groß et al 2015). A formal ontology is a set of terms and relations that represent entities in a specific domain and how they relate to each other.

The semantic web defines a new form of web content which is meaningful to computers, where the meaning is provided by ontologies (Garcia et al. 2006). In virtues of this, ontologies play a pivotal role in the semantic web, making possible and encouraging the dissemination of machine understandable data (Jean-Baptiste LAMY, 2016; Doan et al. 2006). The fact that a machine can interpret the meaning of a content allows the use of automated reasoning, allowing the deduction of new information (del Carmen Legaz Garcia et al. 2016) at unprecedented pace, and ultimately paving the way to astonishing advances in the biomedical field. However, in order to effectively improve health care systems, it is essential to reach interoperability between different ontologies and different databases/datasets. Although benefits of semantic interoperability are clear for improving accuracy and efficiency of diagnoses and treatment (Shen and Lee 2016), mapping of ontologies or databases from different experimental and other sources remains nontrivial (Demontier and Wild 2012).

As a lot of other scientific disciplines, the nutrigenomics field faced a great development in the last decades due to the development of high-throughput *-omics* technologies. Food researchers are now able to expand the common saying “we are what we eat”, taking in consideration the crucial influence of our genome, microbiome, food composition, biotransformation and their molecular interactions, our health status, age, etc. (Cifuentes 2016). Starting from the four major types of *-omics* measurements (genomics, transcriptomics, proteomics, and metabolomics), a variety of omics subdisciplines (epigenomics, lipidomics, interactomics, metallomics, diseasomics, etc.) has emerged (Capozzi and Bordoni 2013). Thanks to the omics approach, researchers are now facing the possibility of connecting food components, foods, the diet, the individual, the health, and the diseases, but this broad vision needs not only the application of advanced technologies, but mainly the ability of looking at the problem with a different approach, a “foodomics approach” (Capozzi and Bordoni 2013).



ENPADASI, following a “foodomics approach”, aims 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. The construction of an ontology for nutritional studies is one of the aims of WP4. In this deliverable, the construction of a controlled vocabulary dedicated to nutrition studies is presented.

## METHODS

### COLLECTION OF TERMS

Relevant terms for the nutritional field and for supporting the informatics infrastructure of ENPADASI (DASH-IN), were collected from multiple sources.

1. **TEMPLATE TERMS:** Starting from the templates for nutritional studies description supplied from WP2 and WP3 (namely the “Any study template with QAT 30-06-2016.xlsx” and “DataSHIELD\_Casestudy\_OS\_Variable\_catalogue\_V4.xlsx” files), our group created a controlled vocabulary, which should be extended based on the needs of the community. Variables in the two above mentioned templates constitute a comprehensive dictionary for ENPADASI, functioning as a starting point for ontology development.
2. **METADATA SEARCH TERMS:** A second set of terms was collected exploiting the “Metadata Terms collection” quest started by Rosario Lombardo (WP3) in April 2017. This quest was originally aimed at collecting terms that the ENPADASI consortium considered relevant to perform a study search based on metadata. Most of those terms were also considered relevant for the ontology.
3. **NUTRITIONAL TERMS:** A third source of terms came from the discussions connected to the preparation of the draft of the manuscript presenting the ONs ontology. A series of general terms (i.e. diet, food, nutrient, etc.) were considered pivotal for a nutritional ontology, mainly as high-mid level terms, and were thus inserted.
4. **FOOD TERMS:** Terms related to food description were included by importing a subset of terms from an external ontology; the FOODON ontology (Griffiths et al. 2016). This solution seemed the most convenient: FOODON list all the “food” relative terms under a single class (“Food source” class) and connect most of them to the corresponding taxon by mean of an equivalency axiom “*part of some* [NCBI Taxon Entity]” and to the human consumer by equivalency axiom “*has consumer some homo sapiens*”. To preserve this organization, we imported the whole sub-classes under the “Food Source” and “Organism”.

### ONTOLOGY DEVELOPMENT: FROM A COMMON VOCABULARY TO CLASSES AND SUBCLASSES

The ontology was developed following the principles indicated by the OBO Foundry (<http://obofoundry.org/>) (Smith et al 2007) and using the software Protégé (<http://protege.stanford.edu/>). In this way, if desired, it will be easier to release the ontology for public use via the OBO foundry itself. To achieve publication in fact, upon request to the OBO foundry, the ontology will be revised and evaluated according to those principles. ONS builds on the subset of the Ontology for Biomedical Investigations (OBI) (Bandrowski et al. 2016), that was

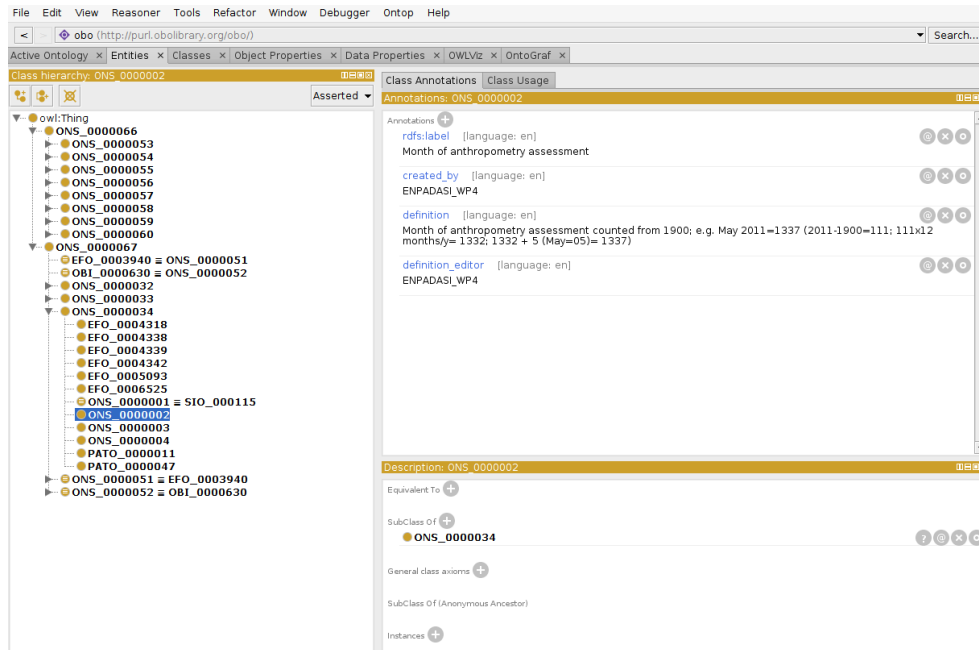


created using the ONTODOG tool (Zheng et al. 2014), and is composed of all terms relevant to nutritional investigations, and extended also in accordance to the bioinformatics infrastructure of ENPADASI. Terms related to food description, as already introduced, were also included by importing a subset of terms from the FOODON (Griffiths et al. 2016).

A key principle followed in ontology development was to avoid the definition of new terms if they (or their suitable synonyms) were present in other ontologies already in use (principle of ontology orthogonality). For each term in the ENPADASI dictionary we firstly checked its presence in other ontologies using ONTOBEE project (<http://www.ontobee.org/>) (Xiang, et al. 2011) (Figure 1), and defined it as new if no suitable matches were found. Each new term was named with the prefix ONS followed by an underscore and a sequential, 7 digits, number (e.g ONS\_0000001, ONS\_0000002, etc.). Each of them was annotated with a short text providing its definition, a label, and an indication on the editor and creator of the definition and the term (annotated as ENPADASI\_WP4) (Figure 2).



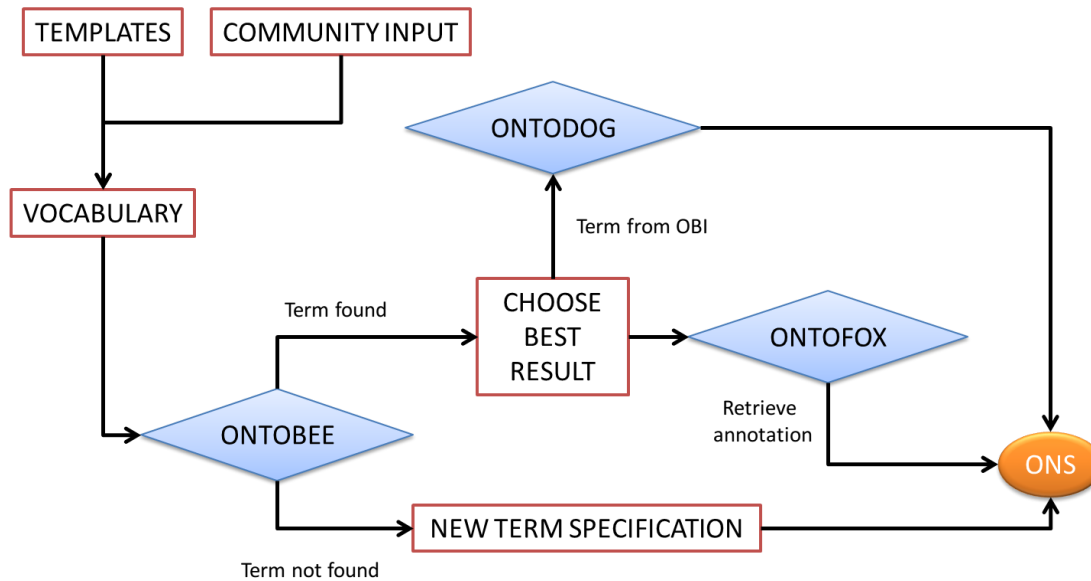
**Figure 1:** Screenshot of ONTOBEE resource. In this example, the term “principal investigator” was searched, and a list of the ontologies containing the term is returned. From this list, the most suitable term (in this case OBI\_000103) is selected.



**Figure 2:** Screenshot of Protégé program illustrating one of the new terms defined by ONS. In this example, the term “month of anthropometry assessment” from the DataSHIELD\_Casestudy\_OS\_Variable\_catalogue\_V4 template was named ONS\_0000002, and a description and label was defined.

As stated above, every term was firstly searched using ONTOBEE, and a list of suitable terms, or rather, a list of their IRI (Internationalized Resource Identifier), was created. To efficiently integrate terms from external ontologies in ONS, their label, definition, and other annotations should be retrieved in addition to their IRI. For this purpose we used ONTOFOX (<http://ontofox.hegroup.org/>) (Xiang et al. 2010) with *includeAllAnnotations* option. In this way, ONTOFOX fetches all the annotation properties from the selected terms and create a new ontology, which was imported in ONS using Protégé.

A graphical summary of the procedure followed for the construction of ONS controlled vocabulary, is presented in figure 3.



**Figure 3:** Schematic representation of the methodology used to develop the controlled vocabulary and the class-subclass structure of the ontology..

Whether it was coming from other ontologies through the ONTOBEE-ONTOFOX approach, or it was newly defined, each time a term was inserted it also was contextualized in a hierarchical schema of classes and subclasses. This basic organization of the ontology has the scope of defining the most simple type of relation in an ontology: the *is a* relation. This relation connects a class (i.e. a concept) with its super-class (i.e. a concept that is ideally 1 level higher). Importantly, a consequential assumption of this type of relation is the inheritance of properties from the parent to the child class. It follows that anything that is necessarily true of a parent class is also necessarily true of all of its child classes. For the terms that were fetched from OBI using the ONTODOG approach, no further action was needed, as ONTODOG create a proper subset of the original ontology, maintaining annotations, axioms, and obviously hierarchical organization of class.

## RESULTS AND DISCUSSION

Following, some key characteristics of the ontology our group developed:

- The ontology was named “**Ontology for Nutritional Studies**” with acronym **ONS**. (In the following we will refer to the developed ontology simply as ONS)
- ONS was made openly available by copyrighting under a Creative Commons CC-BY license version 4.0 (<https://creativecommons.org/licenses/by/4.0/>), and by hosting it in a GitHub repository at <https://github.com/FrancescoVit/Ontology-for-Nutritional-Studies>. In this way, the process of development, revision, and integration of new terms in ONS is fully traceable and public, stimulating and facilitating the development in a collaborative environment. The GitHub repository was created following <https://github.com/cmungall/ontology-starter-kit>.
- ONS was uploaded to the Bioportal repository, and is accessible at <http://bioportal.bioontology.org/ontologies/ONS>





## FUTURE DEVELOPMENT

Continuous update of the ontology will be guaranteed, and managed primarily through the Github repository, to ensure versioning and traceability.

To create a really complete nutritional ontology, we strongly believe that a broader community involvement is a pivotal prerequisite. Admittedly, the development of the version of the ontology herein presented, lacked in this aspect. This has been principally caused by development issues for a rather simple motivation: we believe it is more convenient, and we believe that it is easier, to involve large number of researchers in this effort when a first version of an ontology is actually released. When the ontology is out, in fact, a researcher can rather easily browse the classes and their definition, both by online repositories or, more effectively, by actually downloading the raw file and browsing it with Protégè.

Moreover, during the development of ONS, and through the discussion on the words (i.e. the classes) to be inserted and their definition, carried out primarily within the WP4, it came out how lots of terms did not well fit with a nutritionist vision. But not only this, it was also apparent that, at present, a broad consensus in the nutritionist community on terms definition is lacking. We believe that nutritionist among all other nutritional researcher, could be considered as a primary target of ONS, and that they could greatly benefit from this standardization effort.

To foster, facilitate, and allow community involvement, as a future development of ONS, we are beginning to design a resource, which will likely be in the form of a Wiki, to collect contribution from expert in all the aspects of nutritional sciences on all part of the ontology (term proposal, term definition, annotations, properties connecting various classe, etc) with the objective of continuously improve the ontology and its representation of nutritional sciences.



## REFERENCES

- Bandrowski, Anita, et al. "The Ontology for Biomedical Investigations." PloS one 11.4 (2016): e0154556.
- Capozzi, Francesco, and Alessandra Bordonì. "Foodomics: a new comprehensive approach to food and nutrition." Genes & nutrition 8.1 (2013): 1-4.
- Cifuentes, Alejandro. "Foodomics: Still a long way to go... and we keep going!." Electrophoresis 37 (2016): 1747.
- Cimino, J. J., and X. Zhu. "The practical impact of ontologies on biomedical informatics." Yearb Med Inform 2006 (2006): 124-35.
- del Carmen Legaz-García, María, et al. "Generation of open biomedical datasets through ontology-driven transformation and integration processes." Journal of biomedical semantics 7.1 (2016): 1.
- Doan, AnHai, et al. "Learning to map between ontologies on the semantic web." Proceedings of the 11th international conference on World Wide Web. ACM, 2002.
- Dumontier, Michel, and David J. Wild. "Linked Data in Drug Discovery." IEEE Internet Computing 16.6 (2012).
- Griffiths, E., Dooley, D., Buttigieg, P.L., Hoehndorf, R., Brinkman, F., and Hsiao, W. (2016). FoodOn: A global farm-to-fork food ontology. ICBO Conference. Corvallis, OR, USA [www.foodon.org](http://www.foodon.org)
- Groß, Anika, Cédric Pruski, and Erhard Rahm. "Evolution of biomedical ontologies and mappings: Overview of recent approaches." Computational and Structural Biotechnology Journal (2016).
- LAMY, Jean-Baptiste. "Ontology-Oriented Programming for Biomedical Informatics." Transforming Healthcare with the Internet of Things: Proceedings of the EFMI Special Topic Conference 2016. Vol. 221. IOS Press, 2016.
- Ong, Edison, and Yongqun He. "Community-based Ontology Development, Annotation and Discussion with MediaWiki extension Ontokiwi and Ontokiwi-based Ontobedia." AMIA Summits on Translational Science Proceedings 2016 (2016): 65.
- Shen, Feichen, and Yugyung Lee. "Knowledge Discovery from Biomedical Ontologies in Cross Domains." PloS one 11.8 (2016): e0160005.
- Smith, Barry, et al. "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration." Nature biotechnology 25.11 (2007): 1251-1255.
- Xiang, Zuoshuang, et al. "OntoFox: web-based support for ontology reuse." BMC research notes 3.1 (2010): 1.
- Xiang, Zuoshuang, et al. "Ontobee: A Linked Data Server and Browser for Ontology Terms." ICBO. 2011.
- Zheng, J., Xiang, Z., Stoeckert, C. J., & He, Y. (2014). Ontodog: a web-based ontology community view generation tool. Bioinformatics, btu008