

## 1. Data Summary

Following content will be defined below:

- purpose of the data collection/generation and its relation to the objectives of the project?
- types and formats of data that will be collected/generated by the project
- re-usage of any existing data, if yes how
- the origin of the data
- the expected size of the data
- data utility

### 1.1 Work Packages

ProxIMed aims to introduce proteins from sustainable sources into the Mediterranean (Med) food and feed systems to promote and establish the use of alternative proteins. Traditional alternative protein sources mainly of plant origin (lentil, faba bean (which are common in MedDiet), and chia seeds (a novel seed source introduced to the Med food chain with the recent consumer trends); protein sources defined as “Novel food” (micro-algae, insects, mycoprotein, tomato and mallow leaves, duckweed) and agro-industrial by-products (from tomato, sesame, date) are selected for this purpose. Proteins will be produced by using innovative and green processing technologies (such as microwave, microwave vacuum drying, ultrasonication, supercritical fluid extraction, cold plasma, triboelectrostatic separation, micro-fluidization and ultrafiltration) for extraction and/or concentration of the selected protein sources and for pre-treatment of by-products aiming minimal impact on nutrients. Selected proteins will then be implemented into final products according to their functionalities. Developed proteins will be adapted and introduced to Mediterranean consumers in different regions (Middle East, Northern Africa and Europe) with different products. Novel processes and the protein value chain will be assessed for environmental, economic and social impacts; adapted business models will be evaluated for alt-proteins.

Purpose of the data collection/generation: Consumer Research, Product formulation, Process Optimization and Product characterization. Data will be collected throughout the experiments, surveys and literature analysis. The main data categories are listed as follows:

- Consumer data from surveys and sensory panels
- Technofunctional properties: experiment results of ingredients and products
- Process parameters from trials

The data will be available to the users working in related areas to be used as ‘example’ data sets. New functional alternative protein ingredients and products will be formulated using above mentioned production technologies. The data that will be generated in line with the objectives of the project. The data collection and sharing will also enable researchers working in similar fields to analyse the data and use for their research. Work packages of ProxIMed can be seen as below. Except WP2 and WP7 most of the data collected in the project will be ‘experimental’. (Table 1).



Table 1. Work Packages of ProxIMed

#	WP Title	Lead Participant #	Lead Participant	Start Month	End Month
1	Management and Coordination	1	HSWT1	1	48
2	Consumers' acceptance	1	HSWT2	1	48
3	Protein production through innovative technologies	2	METU	3	42
4	Product development and sensorial evaluation	10	AINIA	12	48
5	Economic and Environmental Assessment	3	GS	1	48
6	Health, Nutrition & Safety Aspects	7	AUTh	6	48
7	Business model development	1	HSWT3	24	48
8	Dissemination and Communication Activities	2	METU	1	48

## 1.2 Types and Format of the Collected/Generated Data

The data, which will be collected/generated throughout ProxIMed, will have both quantitative and qualitative character, the former from experimental trials and the latter from consumer surveys. Following different data formats will be in use:

Textual data: XML, TXT, HTML, PDF, DOCX, DOC

Tabular data (including spreadsheets): CSV, DAT, XLSX

Databases: XML, CSV

Images: TIFF, PNG, JPEG

Audio: MPEG, WAV, MP3

Other: .RWA, .M, .OPJ

## 1.3 Re-usage of Existing Data

Mainly, new data will be generated throughout the project with experiments and trials. Literature data will also be used in WP 2 and WP5. Protein databases for proteomics will also use the data in relevant databases. Already published data or data collected by the consortium will be utilized to fulfill the objectives established in the different WPs. Collected/generated data will be used in the scientific publications.

## 1.4 Origin of the Data

Generated data is experimental data of used processes, proteins, powders, semi-finished, finished protein products or similar commercial varieties of all mentioned data. This experimental data can be physical, chemical or techno-functional properties and main WPs that will generate experimental data are WP3, 4 and 6. Additionally consumer survey and analysis results data will be generated in WP2.



### 1.5 Expected Size of the Data

Data to be generated/collected will vary greatly in size, from megabytes (MB) to gigabytes (GB). Some examples are: Trial data for extraction- 2 GB, Images for color measurements- 10 GB, Rheometer measurements- 5 GB, Particle Size Analysis- 3 GB.

### 1.6 Data Utility

Data generated in this project will be useful to the worldwide scientific community and the project teams. Especially, research groups performing studies in similar areas (extraction, protein analysis, protein solubility, product development, and scale up processes) will benefit. Generated data will as well be relevant to industry partners, companies, and relevant sectors on the production methods used in the project.

## 2. FAIR data

### 2.1 Making data findable, including provisions for metadata

#### Are the data produced in ProxIMed discoverable with metadata?

The data generated in ProxIMed will be available to third parties through scientific publications. **Open access journals** will be selected for publishing so the results will be indexed in main peer-reviewed literature databases (SCOPUS, PUBMED, Web of Science) as well as Open Research platform of the European Commission (<https://open-research-europe.ec.europa.eu/>). Moreover, when scientific output dissemination is performed through Social Media, links to the open data set will also be provided. Our published dataset will be shared in the publicly accessible disciplinary repository **Zenodo** using descriptive metadata as required/provided by that repository. A **'Data Repository'** link will also be available in the web of page of the project.

#### Are the data produced and/or used in ProxIMed identifiable and locatable by means of a standard identification mechanism?

Data produced/used in ProxIMed will be findable and identifiable, whether in Scientific Publications or Zenodo, through a globally unique and eternally persistent identifier (e.g. DOI). Thereby dataset of the project will get a unique Digital Object Identifier (DOI) provided by the publishers or other identifier (accession number or handle number) provided by the repository services.

#### What naming conventions do you follow?

Data will refer to the figures and table numbers used in the publications so it will be easy to identify the data. Datasets and files in the repository will be uniquely identifiable and versioned by using a name convention consisting of project name, dataset name, clear version numbers, method used, ID, place and date.

#### Will search keywords be provided that optimize possibilities for re-use?

Keywords of the publications will be relevant to the goals and defined objectives of the project so they will be provided to optimize the re-use.

#### What metadata will be created?

Minimum requirements for data generated/collected in ProxIMed is given in Table 2.

Table 2. Metadata requirements of ProxIMed



NAME	DEFINITION	TYPE
Experimenter/ Au-thor/Owner	The person who creates the data	Person
Content	Description of the content (Named according to naming convention)	Text
Date	Start, end of the data collection time and date	Date/Time
Comments	Necessary description about the dataset	Text
Keywords	Relevant keywords for the dataset	Text
File	Location of the data set (URL, DOI, ...)	URL, DOI, ...

## 2.2 Making data openly accessible

All data that result in scientific publications, and is considered necessary for third parties to be able to reproduce and validate results will be openly accessible. No data-relevant provisions are made in the consortium agreement. Datasets that are presented in publications or as supplementary materials in publications will be available in the Web Moodle Platform of ProxIMed which will be accessed through the project web page. Users need to register themselves to the platform so that identity can be registered. Data not published anywhere will be deposited to Zenodo (<https://zenodo.org/communities/proximed>) and also to Web Moodle Platform of the project.

All data is presented in digital formats that can be read by most operating systems and standard software tools. For data requiring specialized software, links are provided to open access tools required to open the data. The data, associated metadata, documentation and codes (if available) are all stored in Zenodo's data repository. Zenodo has no restrictions on usage and authorizations. The data is made available to any user following a simple registration process. If it turns out that an institutional repository of a beneficiary is used, the consortium will ensure that any user can access the data. The consortium partners will opt for public domain licenses, as these allow a high degree of reusability and do not restrict access and use of the data by third parties. In addition to Zenodo, a 'Data Repository' tab has been added to the project's website, providing links to the project data's Web Moodle platform.

## 2.3 Making data interoperable

All partners will create or enable copies of datasets and results in formats that are compatible with most common and open source software applications to enable data sharing and reuse between researchers, institutions, organizations and countries to the greatest extent possible. These formats enable the transfer of information and the recombination with different datasets from different sources. The data formats used in this project are based on a formal, accessible, common and generally applicable language to enable interdisciplinary interoperability. In the unlikely event that unusual or project-specific ontologies or vocabularies need to be used. We will provide mappings to more general standard vocabularies for all unusual data types to maximize interoperability. All assumptions made and the mapping steps to more common vocabularies will be documented in a readme file and stored in the project's Web Moodle Platform or in the data repository (Zenodo).

Following formats will be used for the data acquired:

- .txt/.csv/.xml/. opj/.dat/.raw (for data)
- .docx/.doc/.pdf/.ppt (for reports/presentations)
- .jpg/.png/.gif/.tiff/. fig (for graphics)



## 2.4 Increase data re-use (through clarifying licences)

In Zenodo, the data can be licensed under Creative Commons CC International. The open access approach is only extended to the scientific community for non-confidential data. For confidential data or data used for a patent application, the IPR strategy defined in the consortium agreement applies.

All data obtained within the project will be made available in the last three months after the end of the project. Parts of the data may be available earlier due to journal publications. There will be no embargo period. Other scientists from the fields of physics, chemistry, statistics, food science, mathematics, consumer science, and energy research can use the data. Neighboring disciplines and interdisciplinary re-search groups may also be interested due to the interdisciplinary characteristics of ProxIMed's consortium.

Each partner who is responsible for it ensures data quality. The tools needed to describe and identify the dataset and to create the metafiles. The consortium will adhere to key data quality standards set by peer-reviewed journals to ensure the production of high quality data that can be shared with others.

As open formats are used for data archiving, the data remains reusable until the repository withdraws the data or ceases operations.

## 3. Allocation of resources

Costs for making data FAIR in ProxIMed is covered through individual institutions since all partners considered open access publication fees during project proposal stage. Data repository and Web Moodle Platform of the project are cost neutral. The implementation of the data management plan will be mainly responsibility of the Coordinator. Consortium will be updated on the plan regularly. Social Media Co-ordinator of ProxIMed will monitor the Open Access publication process and let consortium members know about the opportunities to publish their research.

Secure long-term storage will be provided by HSWT. HSWT is currently developing an archiving solution for research data, which will be available in 2024. Data will either be deposited in HSWT's own servers or cloud services of Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities (<https://www.lrz.de/english/>).

Costs of this long-term storage are secured by Institutional funds. Data will be kept for at least 10 years, following Deutsche Forschungsgemeinschaft's (DFG, German Research Foundation) recommendations on good scientific practice. The archiving period begins on the date when the results are made publicly available (<https://wissenschaftliche-integritaet.de/en/code-of-conduct/archiving/>).

## 4. Data security

All data generated as part of the project, whether research data or otherwise (including all documents generated as part of the administrative activities), will be centralized and stored in Web Moodle Platform of HSWT, Coordinator. All members of the consortium will have access to this service. Zenodo will also be used as data storage for data sets which are not published anywhere. A copy of all original data will also be stored in cloud service Sync+Share of Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities which will form the basis for data recovery and secure storage in the event of a failure of the online platform. The transfer of sensitive data subject to IPR restrictions is also enabled by the cloud service Sync+Share of Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities.



## 5. Ethical aspects

Any study conducted on human subjects as part of ProxiMed will comply with ethical principles and applicable international, European and national law. The researcher will ensure respect for human beings and human dignity, the equitable distribution of research benefits and the protection of the values, rights and interests of research participants.

Specific documentation is provided to assist ProxiMed' researchers in the study of human behaviour in ethical issues. The following standards and guidelines, norms and publications are followed in the preparation of this documentation:

- Regulation 2016/679 General Data Protection Regulation 27th of April 2016
- Ethics for researchers, Facilitating Research Excellence in FP7
- ESOMAR (e.g. ESOMAR International code on Market, Opinion, Social Research and Data Analytics 2016)
- DFG or German Society for Sociology (e.g. Ethics codex of DGS)
- ISO 9186-1 Graphical symbols – Test methods – Part 1: Method for testing comprehensibility.
- Second Edition 2014-03-15
- Guidance How to complete your ethics self-assessment. Version 4.01, October 2015. European Commission. Directorate-General for Research & Innovation. European Commission. Ethics for researchers. Facilitating Research Excellence in FP7. Directorate-General for Research and Innovation Science in society /Capacities FP7.

Every test with consumers will monitor following subtasks:

- Consumer test objectives, samples and methodology variables definition and scheduling. Sampling method by: Gender, social class and rural-urban representation and differences
- Pilot preparation: Ensuring that all the requested project material will be available for the test with consumers
- Questionnaire/ plot design
- Questionnaire programming (just for quantitative studies)
- Consumer's recruitment
- Fieldwork (carrying out the interviews)
- Statistics analysis
- Report elaboration

## 6. Other issues

All necessary procedures have been described.

## 7. Further information on DMP for consortium

The Research Data Alliance provides a [Metadata Standards Directory](#) that can be searched for discipline-specific standards and associated tools.

The [EUDAT B2SHARE](#) tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.

Useful listings of repositories include: [Registry of Research Data Repositories](#) Some repositories like [Zenodo](#), an OpenAIRE and CERN collaboration), allow researchers to deposit both publications and data, while providing tools to link them. Other useful tools include [DMP online](#) and platforms for making individual scientific observations available such as [ScienceMatters](#).

