

Disclaimer: This deliverable has been submitted to the European Commission and is currently under review. The final version after the approval may differ.

Data Management Plan update

Deliverable n°1.4 – WP n°1 – TASK n°3

Lead beneficiary: GRANT Garant (GG)

Participant beneficiaries: all partners

Delivery type (R, ETHIC, DEC, DMP)

DMP

Dissemination level (PUBlic, SENSitive)

PU

Planned due date

30/06/2026 (M18)

Version n°

01



Funded by
the European Union

D1.4 – Data Management Plan Update

Project Acronym: Project Title:	PROTEIN4IMPACT IMPACT OF ALTERNATIVE PROTEIN SOURCES TO IMPROVE NUTRITION
Grant Agreement: Project coordinator:	4101182324 CENTRALESUPELEC (CS), PIC 933906883, PLATEAU DE MOULON 3 RUEJOLIOT CURIE, GIF SUR YVETTE 91192, France
Call: Topic: Type of action: Granting Authority:	HORIZON-CL6-2024-FARM2FORK-01 HORIZON-CL6-2024-FARM2FORK-01-07 HORIZON Research and Innovation Actions European Research Executive Agency
Short summary:	This deliverable is the update of the first version of the Data Management Plan (DMP) for the PROTEIN4IMPACT project, funded under Horizon Europe. It updates and outlines the management of project resulting datasets and deliverables following FAIR principles and EU Open Science policy. The DMP ensures data findability, accessibility, interoperability, and reusability, using trusted repositories (e.g. Zenodo). Sensitive datasets are protected for commercial exploitation, while open-access outputs are licensed under Creative Commons. This document serves as a living platform for continuous updates throughout the project's lifecycle.

Table of contents

1	SUMMARY	3
2	PROTEIN4IMPACT PROJECT SUMMARY	4
3	PROTEIN4IMPACT DATA SUMMARY	5
3.1	PRINCIPLES OF THE DATA MANAGEMENT	5
3.2	OBJECTIVES OF THE DATA MANAGEMENT PLAN	5
3.3	DATA MANAGEMENT IN THE PROJECT	6
3.4	SUMMARY OF PROTEIN4IMPACT DATA AND OTHER RESEARCH OUTPUTS	7
3.4.1	DATASETS	7
3.4.2	OTHER RESEARCH OUTPUTS.....	7
4	PROTEIN4IMPACT FAIR DATA APPROACH	8
4.1	FINDABLE DATA INCLUDING PROVISIONS FOR METADATA	8
4.2	ACCESSIBLE DATA	9
4.3	INTEROPERABLE DATA.....	10
4.4	DATA RE-USE	11
5	CONCLUSION	11

List of tables

Table 1: PROTEIN4IMPACT project datasets – description.....	13
Table 2: Other research outputs - project deliverables.....	16
Table 3: Findability of PROTEIN4IMPACT project datasets including provisions for metadata	19
Table 4: Findability of other research outputs – project deliverables	23
Table 5: Interoperability of PROTEIN4IMPACT project datasets	27
Table 6: Interoperability of other research outputs – project deliverables	31
Table 7: Reusability of PROTEIN4IMPACT project datasets	36
Table 8: Reusability of other research outputs – project deliverables.....	40

1 Summary

This document presents the first update of the data management strategy for the PROTEIN4IMPACT project, a pioneering initiative supported under the Horizon Europe programme's call HORIZON-CL6-2024-FARM2FORK-01-07. The project is coordinated by CentraleSupélec (CS). The PROTEIN4IMPACT first data management plan (DMP) was developed by the task 1.3 leader GRANT Garant s.r.o. (GG) in M4, and it serves as a key strategic deliverable with public dissemination status. It acts as a blueprint for outreach and utilization of research data produced throughout the project's lifecycle. As a living document, the DMP is regularly revised and refined to accommodate the project's developments and necessary modifications. The DMP establishes a structured approach to assist the consortium in promoting research and innovation efforts, ensuring compliance with Article 17 and Annex 5 of the project's Grant Agreement. Furthermore, it defines essential guidelines and responsibilities, functioning as a reference manual for the management of the research data for members of the project consortium.

The PROTEIN4IMPACT project is dedicated to assessing the nutritional value, health implications, safety, and quality of novel protein-based foods (NFs) derived from unconventional sources while also evaluating their environmental and socio-economic effects. To ensure the project's results are effectively utilized and achieve maximum impact, a comprehensive strategy for the management of resulting project data has been established:

PROTEIN4IMPACT Data Management Plan (DMP)

The PROTEIN4IMPACT Data Management Plan (DMP), a key output of Work Package 1 (WP1) prepared by the task leader GRANT Garant (GG), ensures the responsible handling of all research data and results in alignment with the FAIR principles (Findable, Accessible, Interoperable, Reusable) and Open Access (OA) guidelines. Key elements of this deliverable and its update include:

Definition of Generated Datasets: The project will produce a diverse range of datasets and research outputs (public deliverables) across its seven work packages.

Findability: Public datasets will be primarily stored in Zenodo (within the PROTEIN4IMPACT community: <http://zenodo.org/communities/protein4impact/>), utilizing persistent identifiers (such as DOIs) and Creative Commons licenses (CC BY for data, CC0 for metadata). Confidential datasets will be managed under strict data protection protocols.

Accessibility: The consortium is committed to maximizing open access to project datasets and research outcomes, while respecting any applicable legal, ethical, intellectual property, or confidentiality constraints.

Interoperability: Data will be provided in widely used open formats (such as MS Office, CSV, FASTA, and PDF) to ensure compatibility with common tools. Proprietary formats will be supported by licensed software or Python libraries where necessary.

Reusability: Comprehensive metadata, README files, and descriptive documentation will accompany datasets, ensuring ease of access, reproducibility, and compliance with Horizon Europe requirements.

Updates: The DMP, prepared in M4, is updated in M18 and M36. The DMP provides a dynamic working framework for ongoing improvements.

The PROTEIN4IMPACT consortium is committed to transparency and open science principles while balancing ethical, legal, and commercial considerations.

2 Protein4Impact Project Summary

PROTEIN4IMPACT project is dedicated to assessing the nutritional, health, safety, and quality characteristics of novel protein foods (NFs) derived from unconventional sources, while also evaluating their environmental and socio-economic implications. To ensure a holistic approach and maximize the project's contribution to the transition toward healthier protein-based foods, PROTEIN4IMPACT focuses on utilizing selected agri-food, fisheries, and aquaculture by-products as primary raw materials. Additionally, natural protein producers—including fungi, bacteria, insects, and both micro- and macroalgae—play a key role in this alternative protein production process. Secondary by-products generated along the value chain will be repurposed for reuse within the food production cycle, materials or as a source of process energy.

The extracted proteins will undergo modification and functional enhancement, followed by sensory evaluation and testing in human food trials, as well as in aquaculture as feed components. To assess the feasibility of large-scale NF production, the project will simulate industrial-scale conditions using various methodologies and tools, including Life Cycle Assessment (LCA), Life Cycle Costing (LCC), Risk Assessment (RA), Techno-Economic Analysis (TEA), and Acceptability Optimization (AO), along with Digital Twins, real-time AI-driven monitoring, and Decision Support Systems (DSS).

At the final stage, key parameters influencing protein NF production will be optimized to enhance protein content, while impact assessments will be carried out at each phase—extraction, characterization, modification/functionalization of protein-rich fractions, and NF production—as well as across the entire integrated process. The project will also evaluate consumer acceptance and the societal relevance of NFs through sensory testing in various markets, allowing for a comparative analysis between Europe and other global regions.

By adopting the comprehensive, interdisciplinary approach proposed in PROTEIN4IMPACT, the project aims to establish a realistic and sustainable framework for the effective use of alternative proteins in the food industry.

3 Protein4Impact Data Summary

3.1 Principles of the Data Management

PROTEIN4IMPACT is a Research and Innovation Action initiative funded under the Horizon Europe programme. Over its 36-month duration, the project will generate a wide range of digital datasets from various consortium partners. Establishing a robust Data Management Plan (DMP) was a crucial initial step to ensure effective data storage, accessibility, availability, and reuse, all in alignment with the FAIR principles.

The EU Open Science policy, as outlined in Annex V of the EU Grants Annotated Grant Agreement, alongside the indicative DMP template, has provided a structured foundation for designing a comprehensive strategy for data management within PROTEIN4IMPACT.

The Data Management Plan of PROTEIN4IMPACT is designed to:

- i. Develop and maintain an up-to-date DMP, with scheduled revisions at project milestones M18 and M36.
- ii. Handle all data in accordance with FAIR principles, ensuring it is easily findable, accessible, interoperable, and reusable.
- iii. Store research data in a trusted repository (Zenodo), making it publicly available as soon as possible following scientific publication.
- iv. Guarantee open access to research data, applying the latest versions of the Creative Commons Attribution International Public License (CC BY) or the Creative Commons Public Domain Dedication (CC0). Exceptions will be granted where openness could compromise the legitimate interests of beneficiaries or conflict with obligations under the Grant Agreement (refer to Article 17) or EU competitive considerations.
- v. Provide repository-based metadata and supporting information on research outputs, including any tools or methodologies needed for data reuse and validation.
- vi. Ensure open metadata (under CC0 licensing), capturing essential details such as dataset origin, publication details (authors, title, publication date, and venue), Horizon Europe grant identifiers (project name, acronym, and number), licensing terms, and persistent identifiers for publications, authors, organizations, and grants. Where relevant, metadata will also include persistent identifiers for related research outputs or tools necessary for validating the publication's conclusions.

This structured approach of the DMP will facilitate responsible data management, ensuring transparency, accessibility, and compliance with Open Science principles throughout the PROTEIN4IMPACT project.

3.2 Objectives of the Data Management Plan

The present deliverable, D1.4, constitutes an update to the initial version of the PROTEIN4IMPACT Data Management Plan (DMP), D1.1, submitted in M4. The DMP plan is designed to support the project team in effectively implementing research activities while ensuring compliance with the Grant Agreement and Open Science policy. Furthermore, it serves as a key tool for facilitating communication, disseminating project results, and fulfilling reporting obligations. The PROTEIN4IMPACT DMP has been developed as part of Work Package 1 (WP1), Task 1.3, with the following core objectives:

- ⇒ Identify and classify key research datasets generated throughout the project, linking each dataset to its corresponding research objective.

- ⇒ Ensure adherence to the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) and encourage Open Access and Open Data sharing, while considering necessary restrictions.
- ⇒ Define essential dataset details, including data formats, storage locations, access limitations for sensitive datasets (e.g., restricted access rights), and the selection of appropriate repositories for long-term storage.
- ⇒ Establish a flexible and evolving framework that allows continuous updates and improvements in data management practices over the course of the project.
- ⇒ Ensure compliance with ethical, legal, and security requirements, particularly regarding data protection, confidentiality, and intellectual property rights.
- ⇒ Implement clear guidelines for data documentation and metadata standards, ensuring datasets are well-described and easily interpretable for reuse and validation by external stakeholders.
- ⇒ Promote interoperability and standardization, ensuring that datasets are stored in widely accepted formats and are compatible with existing tools and platforms.
- ⇒ Facilitate data preservation and long-term accessibility, ensuring that project outputs remain available beyond the project's completion.

The PROTEIN4IMPACT DMP covers both digital research data and other research outputs, particularly public deliverables. The initial version of DMP (D1.1) has been developed based on dataset information collected through an internal survey of PROTEIN4IMPACT consortium members, conducted between M3 and M4 of the project. This document, the update of the DMP Plan, D1.4, reflect the necessary revisions as well as inputs and updates collected from partners in M18. The DMP Plan will be revised and expanded at M36, ensuring that data management strategies evolve in response to project developments and emerging requirements.

3.3 Data Management in the Project

Data management is a key task within WP1 (Project Management and Coordination), which is led by CS, the project coordinator. Within WP1, Task 1.3 focused on data management, is led by GG. To support effective management, PROTEIN4IMPACT use internal project repository maintained by IDENER (T1.4). The repository serves as the central internal platform for storing project documentation, datasets, deliverables, meeting materials, and collaborative working documents. Access is restricted to consortium members and is managed according to the project's data governance and confidentiality requirements.

The DMP form was prepared in coordination with the project coordinator and in close collaboration with all project team members, following a DMP template provided by EU authorities (EU Funding & Tender Portal - Reference Documents). This effort aimed to establish a solid foundation for the project's DMP, including its principles and objectives.

In addition, GG provides guidance to the entire project team throughout the data lifecycle and ensures they are informed about their relevant data management obligations. GG is also responsible for conducting periodic reviews of the DMP, working collaboratively with all project team members. All related questionnaires are securely stored internally on the project's internal repository.

3.4 Summary of PROTEIN4IMPACT data and other research outputs

The PROTEIN4IMPACT project is structured into seven work packages, five of them are dedicated to research and innovation. WP1 focuses on the dissemination, exploitation, and communication of project outcomes, as well as project management and coordination. WP7 addresses ethical aspects.

The PROTEIN4IMPACT project has identified a broad spectrum of datasets that will be generated over its course. A brief description of these datasets is provided in Tables 1 and 2 (*See Attachments*).

3.4.1 Datasets

Table 1 presents a detailed summary of the datasets produced as part of the PROTEIN4IMPACT project. These datasets primarily originate from laboratory experiments, surveys, and analytical work, and encompass a broad spectrum of research activities carried out across multiple work packages and specific tasks within the project. They serve a crucial role in advancing the development of novel foods, evaluating their environmental performance, and constructing a comprehensive methodology database aligned with the project's core objectives.

The data collected span a diverse array of formats, tailored to different research needs. These include structured numerical data from analytical measurements, textual annotations providing experimental context, high-resolution images capturing microscopic and macroscopic observations. By integrating these varied data types, the project aims to create a robust scientific foundation for novel food technologies, fostering advancements in the novel food processing technologies.

3.4.2 Other research outputs

Table 2 provides a comprehensive overview of additional research outputs produced within the PROTEIN4IMPACT project, specifically focusing on project deliverables. These deliverables encompass a variety of formal project reports, which serve as key documentation of the research progress, findings, and methodologies applied throughout the project's duration.

Once submitted to the EC, public deliverables will be uploaded to the project website and deposited in ZENODO, accompanied by a disclaimer stating that the version has been submitted to the EC. Sensitive deliverables will not be made publicly available in full. Where appropriate, metadata and/or an executive summary will be published in Zenodo, subject to confidentiality, intellectual property, and other relevant restrictions. The deposition of project outputs in Zenodo ensures their long-term accessibility and preservation.

Beyond formal reports, other research outputs may include:

- ⇒ Scientific publications in peer-reviewed journals, detailing experimental methodologies, data analysis, and key discoveries.
- ⇒ Conference presentations and posters, showcasing project advancements at national and international scientific meetings.
- ⇒ Technical guidelines and best practice documents, providing recommendations for researchers and industry stakeholders on biosensor development and environmental assessments.

- ⇒ Training materials and educational resources, designed to support knowledge transfer, including workshops, webinars, and instructional videos.
- ⇒ Software tools and computational models, developed within the project for data analysis, biosensor performance prediction, and environmental impact assessment.
- ⇒ Policy briefs and stakeholder reports, summarizing key findings and their implications for industry, regulatory bodies, and decision-makers.

By consolidating these diverse research outputs, the PROTEIN4IMPACT project aims to maximize knowledge dissemination, promote collaboration, and support future advancements.

Overview of these outputs is in Table 2 (See Attachments).

4 Protein4Impact FAIR data approach

4.1 Findable data including provisions for metadata

The PROTEIN4IMPACT project is built on a foundation of open collaboration and structured knowledge-sharing while strictly adhering to its data management policy. This policy prioritizes maximizing transparency and accessibility as early and extensively as possible while maintaining compliance with FAIR (Findable, Accessible, Interoperable, and Reusable) principles and the data management guidelines set by Horizon Europe (as specified in the Annotated Grant Agreement, particularly Article 17).

To enhance the discoverability and long-term accessibility of research data generated within the project, all datasets—along with other research outputs—will be assigned persistent identifiers (PIDs), preferably DOIs, through trusted repositories. The Zenodo platform has been designated as the primary archival repository, where the project has already established a dedicated PROTEIN4IMPACT community to store and manage datasets and related research deliverables.

All relevant datasets, except those excluded to safeguard intellectual property rights or potential patent applications, will be accompanied by comprehensive metadata frameworks. These metadata records will be made openly accessible under the CC0 public domain dedication via trusted repositories, ensuring seamless data sharing. The metadata will include:

- ⇒ Dataset descriptions, submission dates, authorship, deposition venues, and embargo periods.
- ⇒ Horizon Europe funding details, including the project's name, acronym, and grant number.
- ⇒ Licensing terms defining data usage permissions.
- ⇒ Persistent identifiers for datasets, researchers, and affiliated institutions, where applicable.
- ⇒ Links to related publications and research outputs, enabling AI-powered data harvesting and integration into broader scientific ecosystems.

To further enhance searchability and interoperability, PROTEIN4IMPACT datasets will be systematically categorized using standardized keywords. A detailed breakdown of data findability strategies, including metadata requirements, can be found in Table 3 and 4 (See Attachments).

4.2 Accessible data

As part of the **PROTEIN4IMPACT** project, an initial questionnaire was distributed in M3–M4 among project partners to assess the accessibility of datasets generated throughout the research process. The feedback collected has helped shape a data management strategy that ensures **efficient storage, accessibility, and traceability of research outputs related to novel food innovations**.

For this document, DMP update, D1.4, the follow-up questionnaire was distributed in M18 to collect updates on the datasets. All datasets produced within the project—whether **open or restricted**—will be deposited in **Zenodo**, a trusted open-access repository. This ensures that each dataset is assigned a **persistent identifier (DOI)**, making it easily citable and linkable to relevant scientific publications, supplementary datasets, or complementary research findings.

In addition to Zenodo, partners may opt out to use its **institutional repository**, as an alternative data storage platform. Each **dataset version** will be systematically tagged and accompanied by **comprehensive metadata**, with clearly defined **linkages between different dataset iterations**. To enhance dataset discoverability, the **project team will establish a standardized naming convention**, which includes:

- **A prefix** to specify whether the file is a dataset, metadata, or template.
- **A core identifier** composed of:
 - A concise, meaningful name describing the dataset or template.
 - The acronym or short name of the data-contributing partner(s).
- **A suffix** indicating the last upload date in **YYYYMMDD** format.

Before publication or other public repositories, project datasets and documentation are managed through the official PROTEIN4IMPACT repository maintained by IDENER. This repository acts as the consortium’s collaborative workspace and supports version control, secure storage, and controlled access to project data during its development and validation stages.

This structured approach ensures that datasets are **well-organized, easily searchable, and efficiently archived**. PROTEIN4IMPACT team members are strongly encouraged to utilize these repositories for data storage and archiving of project results.

Data Accessibility and Restrictions

Datasets supporting **scientific publications** will be **deposited at the time of publication**, ensuring immediate availability through **free and standardized access protocols**. While most datasets and research outputs will comply with **FAIR principles (Findable, Accessible, Interoperable, and Reusable)** and be **openly available**, some datasets will have **restricted access due to their potential for intellectual property protection and future patent applications**.

Long-Term Data Availability and Licensing

To **ensure longevity**, project datasets will remain **accessible for a minimum of five years** after the official completion of PROTEIN4IMPACT. **Publicly available datasets** will continue to be hosted on **Zenodo**, ensuring long-term preservation and accessibility.

- **Open-access datasets** will be licensed under **Creative Commons Attribution International (CC BY)**, allowing others to freely use and build upon the data while crediting the original authors.

- **Metadata** will be published under the **Creative Commons Public Domain Dedication (CC0)**, enabling unrestricted sharing and reuse.

Ethical, legal, and intellectual property considerations surrounding **data sharing and commercialization** will be carefully managed to **balance openness with innovation protection**, ensuring that the project's novel food research can be effectively leveraged by both academic and industry stakeholders.

4.3 Interoperable data

Interoperability is a fundamental principle within the PROTEIN4IMPACT project, ensuring that datasets related to novel food innovation can be seamlessly integrated, shared, and utilized across different systems, disciplines, and research initiatives. To maximize interoperability, the project adheres to structured data standardization practices, metadata frameworks, and persistent identifiers, enabling smooth data exchange among various research communities, stakeholders, and digital platforms.

All datasets produced within the project—whether openly accessible or restricted—will be deposited in Zenodo, a widely recognized open-access repository. Zenodo provides each dataset with a Digital Object Identifier (DOI), ensuring it remains permanently citable and traceable while being linked to relevant scientific publications, supplementary datasets, or complementary research findings. This approach facilitates cross-referencing and supports broader scientific engagement with the project's outputs.

Additionally, the partners may choose to complement Zenodo with their institutional repository, which allows for version control and enhanced dataset management. Within such institutional repository, each dataset version will be:

- Systematically tagged, ensuring clear differentiation between data iterations.
- Accompanied by comprehensive metadata, improving discoverability and contextual understanding.
- Linked to previous versions, enabling researchers to track dataset evolution over time.
- To further support data interoperability, the opting out team will establish a structured naming convention, ensuring consistency and clarity across different datasets. This convention consists of:
 - **A prefix** specifying whether the file represents a dataset, metadata, or a template.
 - **A core identifier**, composed of:
 - **A concise**, descriptive name of the dataset/template.
 - **The acronym** or short name of the data-contributing partner(s) (PFASTwin as the default for templates).
 - **A suffix** indicating the date of the last upload in YYYYMMDD format.

By implementing these interoperability measures, PROTEIN4IMPACT enhances the searchability, accessibility, and usability of its datasets, ensuring they can be easily integrated into broader scientific and industrial applications. Researchers, policymakers, and food innovators will be able to seamlessly access and analyze the data, fostering cross-sector collaboration in novel food development.

4.4 Data Re-use

One of the key objectives of the **PROTEIN4IMPACT** project is to ensure that research data generated within the framework of **novel food innovation** is not only accessible but also effectively **reusable** by diverse stakeholders, including **scientists, industry partners, regulatory bodies, and policymakers**. Reusability is crucial for **accelerating innovation, reducing redundant research efforts, and enhancing scientific transparency**.

To facilitate reusability, datasets supporting **scientific publications** will be **deposited at the time of publication**, ensuring immediate availability through **free and standardized access protocols**. The majority of research outputs align with **FAIR principles (Findable, Accessible, Interoperable, and Reusable)** and are openly available. However, certain datasets with potential for **patent applications and commercialization** will have restricted access.

Additionally, the project promotes **responsible data-sharing practices** by employing:

- **Creative Commons Attribution International (CC BY) licenses** for open-access datasets, allowing others to **reuse, modify, and redistribute** the data, provided that proper credit is given.
- **Creative Commons Public Domain Dedication (CC0) for metadata**, ensuring unrestricted availability and enabling **metadata harvesting by AI-driven tools** for enhanced searchability and integration.

Ethical, legal, and commercial considerations will be carefully managed to **balance open science principles with the need for intellectual property protection**. By enabling **broad data reusability**, the **PROTEIN4IMPACT** project supports **scientific progress, regulatory advancements, and industrial applications**, contributing to the development of **next-generation sustainable food solutions**.

5 Conclusion

The Data Management Plan (DMP) for the **PROTEIN4IMPACT** project has been designed to ensure that all research data generated throughout the project's lifecycle is handled in a responsible, transparent, and sustainable manner. By adhering to the FAIR principles (Findable, Accessible, Interoperable, and Reusable) and complying with Horizon Europe requirements, the project guarantees that its datasets contribute meaningfully to scientific progress, regulatory frameworks, and industrial innovation in novel food development.

Through the ZENODO repository, supplemented eventually by institutional archives, **PROTEIN4IMPACT** ensures that all datasets are securely stored, well-documented, and easily retrievable. The use of persistent identifiers (DOIs), standardized metadata frameworks, and structured naming conventions enhances interoperability and data discoverability, facilitating seamless integration across academic, industrial, and policy-making sectors.

The project actively promotes open data practices, making the majority of its datasets publicly accessible while responsibly managing intellectual property and ethical considerations for restricted data. Long-term preservation strategies guarantee that research outputs remain available for at least five years after project completion, ensuring ongoing impact and reusability in future food innovation initiatives.

By implementing this comprehensive Data Management Plan, **PROTEIN4IMPACT** strengthens collaborative research efforts, accelerates scientific discoveries, and supports sustainable food solutions. The project's commitment to data openness, integrity, and usability will contribute to a more transparent and efficient research ecosystem, ultimately benefiting academia, industry, and society at large.

6. Attachments

Table 1: PROTEIN4IMPACT project datasets – description.....	13
Table 2: Other research outputs - project deliverables.....	16
Table 3: Findability of PROTEIN4IMPACT project datasets including provisions for metadata	19
Table 4: Findability of other research outputs – project deliverables	23
Table 5: Interoperability of PROTEIN4IMPACT project datasets	27
Table 6: Interoperability of other research outputs – project deliverables	31
Table 7: Reusability of PROTEIN4IMPACT project datasets	36
Table 8: Reusability of other research outputs – project deliverables.....	40

Table 1: PROTEIN4IMPACT project datasets – description

Dataset No	Name of the dataset	WP/Task	Leading partner	Data description
ER 1.3	Project Handbook	WP1	CS	Gathers all information in one document. Provides detailed guidance, management processes, and procedures for effective project execution and collaboration.
ER 1.4	List of past and ongoing EU projects	WP1	CS	Made to contact each coordinator for encouraging the establishment of a collaborative link
ER 1.5	Mapping and database of the relevant stakeholders	WP1	CS	Guide engagement strategies, ensure relevant involvement, and maximize project impact.
ER 1.6	Platform and dietary data hub sharing	WP1	IDENER	Creation of a centralized digital repository that not only stores all official consortium documents in one place but also enables seamless online collaboration. This system ensures structured, easy, and immediate access to essential documents for all partners. Additionally, it fosters real-time teamwork, streamlining workflows and improving efficiency in document management and decision-making.
ER 2.2	Bacterial fermentation process for protein from waste	WP2	DTU	Membrane aerated bioreactor for safe and resource efficient aerobic fermentation of hydrogenotrophic bacteria for microbial protein
ER 2.3	Elaboration of experimental diets for rearing larvae	WP2	ENEA	Rearing larvae using agricultural by-products represent a way to reduce waste and provide valuable source of protein
ER 2.4	Collection fish by-products samples and defatting process	WP2	CNR	The use of fish by-products as valuable source of protein is a way to reduce waste production
ER 2.5	Solid submerged fermentation bioprocess method	WP2	UMINHO	Solid submerged fermentation bioprocess to obtain food-grade mycoprotein from food industry by-products
ER 2.6	Protein concentrates obtained by extrusion	WP2	CS	Using extrusion to modify biomass in order to recuperate the protein fraction

Dataset No	Name of the dataset	WP/Task	Leading partner	Data description
ER 2.7	Protein derived from macro- and microalgae	WP2	ENEA	Macro- and microalgae protein improvement
ER 3.2	Alternative proteins (soy, algae) with improved aroma and technofunctionality	WP3	UOH	Document describing fermentation of soy proteins and algae with basidiomycetes for improved sensorial attributes and technofunctionality
ER 3.5	Human acceptability of alternative proteins	WP3	UOH	Information on sensory profiles and human acceptability of several alternative proteins from WP2 and 3
ER 4.1	Online service for New Foods protein production simulations using AI Digital Twin of realistic industrial infrastructure of typical food processors	WP4	GOLEM	Innovation online digital service for simulation of different protein production and sales scenarios and in-depth holistic assessment of industrial results and impacts by stakeholders using automated dashboards, interactive analytics reports.
ER 4.2	Open cyber-physical knowledgebase model of realistic food processor enterprise	WP4	GOLEM	The open model will allow easy customisation to optional protein production scenarios, usage of various KPIs and sustainability conditions to provide detailed insights into different approaches and conditions, including financial viability, quality management and environmental impacts
ER 5.2	Behavioral intervention methodology	WP5	UNIWA	Task 5.2 Behavioural intervention to enhance adoption of alternative proteins on the diet. Participant: UNIWA (M17-M20) A case-study was launched aiming to develop an intervention tool (based on the Transtheoretical model of behavior change) to promote greater utilization of alternative proteins. 2 target groups were employed as such: a) healthcare professionals (from a University Hospital in Athens, Greece), b) healthcare students (from UNIWA) since they are critical components in all primary healthcare national systems, supporting healthy diets. The intervention programme for each group comprised 15 group sessions (of 5–10 people) lasting 90 min each session, which will be held for 3 months. The specific approaches were as follows: 1) Development of a Dietary Interactive Behavioural Intervention (DIBI) for the above target groups. The intervention used the Transtheoretical Model of behaviour change (TTM) in conjunction with Motivational interviewing (MI), aiming in supporting the target group find motivation and commitment to change their dietary choices. 2) Through the intervention, dynamic education aiming the empowerment of these target groups, concerning dietary choices

Dataset No	Name of the dataset	WP/Task	Leading partner	Data description
				related to alternative proteins offered in interactive work sessions. The whole intervention was evaluated through scientifically accepted instruments. Relevant assessment (before/after the intervention programme) protocols were utilized.
ER 6.1	Risk assessment report	WP6	IDENER	Planning, monitoring, and execution of the environmental, health, and safety Risk Assessment (RA) and TEA.
ER 6.2	Recommendations for executing an SSbD framework	WP6	In.BIO	Project overall impact optimized, results shared broadly, exploitation plans in place, and recommendations for executing an SSbD framework disseminated

Table 2: Other research outputs - project deliverables

Dataset No	Name of the dataset	WP/Task	Leading partner	Data description
ER 1.1	DMP Plan (D1.1)	WP1	GG	Strategic plan for the data management, with the aim to fulfil the KPI's
ER 1.2	PEDR Plan (D1.2)	WP1	GG	Strategic plan for the project DEC activities, with the aim to fulfil the KPI's
ER 1.7	Practice abstract batch (D1.3)	WP1	CS	Highlights research outcomes' practical applications, exploitation potential, and sectorial impact for stakeholders.
ER 1.8	Policy brief (D1.8)	WP1	CS	Document outlining key recommendations, findings, and implications related to the production, extraction, transformation and commercialization of alternative proteins and products containing alternative proteins
ER 2.1	Report on PROTEIN4IMPACT alternative protein production (D2.1)	WP2	ENEA	This dataset constitutes a mandatory project deliverable
ER 3.1	Modified, functionalized and characterized protein for their conversion into food and feed (D3.1)	WP3	UOH	Document describing all enzymatic treatments performed on the alternative proteins as well as the characterization of the enzymatic modified protein
ER 3.3	Report on pilot scale production process of selected compounds. Feasibility study (D3.3)	WP3	VRLS	Report on pilot-scale production and feasibility of three alternative-protein supplement products from Tenebrio molitor: (i) high-purity ALK-IEP protein isolate (>71 % protein, yield 40.33 %), (ii) Alcalase bioactive peptide hydrolysate (DH 36.2 %; ABTS 98.76 %; MICA 92.73 %), (iii) multiphase micropellets-in-capsule supplement (spray-dried, controlled GI release). First-in-category insect-derived multiphase peptide capsule format. Prototype products developed and presented at consortium meeting (Paris, May 2026).

Dataset No	Name of the dataset	WP/Task	Leading partner	Data description
ER 3.4	Quality assessment of novel food produced with alternative proteins (D3.4)	WP3	NTUA	This dataset constitutes a mandatory project deliverable
ER 3.6	Report on testing alternative proteins as fish feeds (D3.6)	WP3	AQB	This dataset constitutes a mandatory project deliverable
ER 4.3	Technology platform to model protein production plants (D4.1)	WP4	GOLEM	This research output is a mandatory project deliverable
ER 4.4	Digital Twins interfaces (D4.2)	WP4	GOLEM	This research output is a mandatory project deliverable
ER 4.5	Digital Twin simulator (D4.3)	WP4	GOLEM	This research output is a mandatory project deliverable
ER 5.1	Report on citizen satisfaction survey (D5.2)	WP5	UNIWA	It would be the first time that these novel foods (NFs) will be consumed by customers thus their feedback and satisfaction will be indicative for success of the project
ER 5.3	Report on data collection related to the drivers of acceptance of unconventional proteins (D5.1)	WP5	NDF	This dataset constitutes a mandatory project deliverable
ER 5.4	Report on agrifood stakeholders training activities (D5.3)	WP5	In.Bio	This dataset constitutes a mandatory project deliverable
ER 6.3	Report on energy recovery from production processes (D6.1)	WP6	In.BIO	This dataset constitutes a mandatory project deliverable
ER 6.4	Financial and environmental sustainability (LCC/LCA) and DSS (D6.2)	WP6	AgC	An innovative ex-post assessment will be used to evaluate the financial and environmental sustainability of the proposed solutions over the short and long term, using modern LCA and costing (LCA/C) theories.
ER 6.5	EOSC, dietary guidelines and regulatory framework (D6.3)	WP6	IDENER	Protein4Impact data integrated into EOSC and a comparative analysis of current European dietary guidelines and consumption habits performed.

Dataset No	Name of the dataset	WP/Task	Leading partner	Data description
ER 7.1	OEI - Requirement No. 1 (D7.1) - MoU, Conflict of interest, CV of the Ethics Advisor	WP7	CS	Document detailing the mechanisms that prevent conflicts of interest, maintaining scientific integrity, and ensuring the credibility of research findings; document detailing the role of the Ethics Advisor; Curriculum vitae of the Ethics Advisor
ER 7.2	OEI - Requirement No. 2 (7.2) - Report 1	WP7	CS	Analysis of the activities conducted for the project from an ethics point of view, mainly to ensure compliance with EU ethical standards.
ER 7.3	OEI - Requirement No. 3 (D7.3) - Report 2	WP7	CS	Analysis of the activities conducted for the project from an ethics point of view, mainly to ensure compliance with EU ethical standards.

Table 3: Findability of PROTEIN4IMPACT project datasets including provisions for metadata

Dataset No.	Will the dataset be identified by a persistent identifier?	What metadata will be created?	Will you provide keywords in the metadata? What?	In which trusted repository do you plan to publish the metadata on your dataset?	Will metadata remain available after the data is no longer available?
ER 1.3	Y	title, description, authors, project acronym, GA number, keywords, date, licence, access conditions.	Handbook	Zenodo	Yes
ER 1.3	Y	title, description, authors, project acronym, GA number, keywords, date, licence, access conditions.	EU projects, European projects, Horizon projects	Zenodo	Yes
ER 1.3	Y	title, description, authors, project acronym, GA number, keywords, date, licence, access conditions.	Stakeholders	Zenodo	Yes
ER 1.3	Y	title, description, authors, project acronym, GA number, keywords, date, licence, access conditions.	Data Platform	Zenodo	Yes
ER 2.2	Yes, all the information related to the stored dataset will be provided at the end of the project	Experimental metadata will be recorded as Excel files and stored in the shared consortium repository. Metadata will include sample identification, experimental conditions and analytical methods	Yes. Keywords will refer to the work package, task, experimental activity, sample type, analytical method, or alternative protein source investigated.	At present, no public Open Access repository has been identified for the publication of project metadata. Such decision will be taken in agreement with all Consortium partners.	yes
ER 2.3	Yes, all the information related to the stored dataset will be provided at the end of the project	Experimental metadata will be recorded as Excel files and stored in the shared consortium repository. Metadata will include sample identification, experimental conditions and analytical methods	Yes. Keywords will refer to the work package, task, experimental activity, sample type, analytical method, or alternative protein source investigated.	At present, no public Open Access repository has been identified for the publication of project metadata. Such decision will be taken in agreement	yes

Dataset No.	Will the dataset be identified by a persistent identifier?	What metadata will be created?	Will you provide keywords in the metadata? What?	In which trusted repository do you plan to publish the metadata on your dataset?	Will metadata remain available after the data is no longer available?
				with all Consortium partners.	
ER 2.4	Yes, all the information related to the stored dataset will be provided at the end of the project	Experimental metadata will be recorded as Excel files and stored in the shared consortium repository. Metadata will include sample identification, experimental conditions and analytical methods	Yes. Keywords will refer to the work package, task, experimental activity, sample type, analytical method, or alternative protein source investigated.	At present, no public Open Access repository has been identified for the publication of project metadata. Such decision will be taken in agreement with all Consortium partners.	yes
ER 2.5	Yes, all the information related to the stored dataset will be provided at the end of the project	Experimental metadata will be recorded as Excel files and stored in the shared consortium repository. Metadata will include sample identification, experimental conditions and analytical methods	Yes. Keywords will refer to the work package, task, experimental activity, sample type, analytical method, or alternative protein source investigated.	At present, no public Open Access repository has been identified for the publication of project metadata. Such decision will be taken in agreement with all Consortium partners.	yes
ER 2.6	Yes, all the information related to the stored dataset will be provided at the end of the project	Experimental metadata will be recorded as Excel files and stored in the shared consortium repository. Metadata will include sample identification, experimental conditions and analytical methods	Yes. Keywords will refer to the work package, task, experimental activity, sample type, analytical method, or alternative protein source investigated.	At present, no public Open Access repository has been identified for the publication of project metadata. Such decision will be taken in agreement with all Consortium partners.	yes
ER 2.7	Yes, all the information related to the stored	Experimental metadata will be recorded as Excel files and stored in the shared	Yes. Keywords will refer to the work package, task, experimental activity,	At present, no public Open Access repository has been	Yes.

Dataset No.	Will the dataset be identified by a persistent identifier?	What metadata will be created?	Will you provide keywords in the metadata? What?	In which trusted repository do you plan to publish the metadata on your dataset?	Will metadata remain available after the data is no longer available?
	dataset will be provided at the end of the project	consortium repository. Metadata will include sample identification, experimental conditions and analytical methods	sample type, analytical method, or alternative protein source investigated.	identified for the publication of project metadata. Such decision will be taken in agreement with all Consortium partners.	
ER 3.2	Y	authors, date of creation, methodology, variable descriptions, units of measurement	microalgae, BSG, soy protein, fermentation, basidiomycota	Zenodo	Y
ER 3.5	Y	authors, date of creation, methodology, variable descriptions, units of measurement	microalgae, BSG, soy protein, fermentation, GC-MS, volatile compounds, aroma profile, sensory evaluation, basidiomycota	Zenodo	Y
ER 4.1	DataCite DOI Persistent identifier minted automatically on deposit in Zenodo	CodeMeta (JSON-LD, Schema.org-based): software name, description, version, authors (with ORCID), licence, runtime/OS, dependencies, related publications, funding source, and development status, stored as a codemeta.json file in the source repository. A Citation File Format (CITATION.cff) file accompanies it for human- and machine-readable citation.; data import from xlsx documents, online access using Internet https;	Project-specific free-text keywords (e.g., protein production, insect/alternative protein, feedstock, bioconversion, digital twin, process simulation, sustainability KPI, Protein4Impact). DataCite DOI metadata > global PID Graph and scholarly infrastructures (DataCite, OpenAIRE, ORCID, B2FIND/EOSC). Records discoverable and indexable by humans and machines for discovery and re-use.	Zenodo	Yes
ER 4.2	DataCite DOI Persistent identifier minted	CodeMeta (JSON-LD, Schema.org-based): software name, description, version, authors (with ORCID), licence,	Project-specific free-text keywords (e.g., protein production, insect/alternative protein, feedstock,	Zenodo	Yes

Dataset No.	Will the dataset be identified by a persistent identifier?	What metadata will be created?	Will you provide keywords in the metadata? What?	In which trusted repository do you plan to publish the metadata on your dataset?	Will metadata remain available after the data is no longer available?
	automatically on deposit in Zenodo	runtime/OS, dependencies, related publications, funding source, and development status, stored as a codemeta.json file in the source repository. A Citation File Format (CITATION.cff) file accompanies it for human- and machine-readable citation.; data import from xlsx documents, online access using Internet https;	bioconversion, digital twin, process simulation, sustainability KPI, Protein4Impact). DataCite DOI metadata > global PID Graph and scholarly infrastructures (DataCite, OpenAIRE, ORCID, B2FIND/EOSC). Records discoverable and indexable by humans and machines for discovery and re-use.		
ER 5.2	Y - DOI)	name of the ER, authors, format, description of the ER-Metadata will include the name of the ER, authors, date, version, format, description of the ER, keywords and access conditions. Metadata will be maintained according to FAIR principles and project data management procedures.	behavioural intervention- Behavioural intervention to enhance adoption of alternative proteins on the diet.	ZENODO	YES
ER 6.1	Identifier associated to the project deliverable	Standard format use for General communication, Libraries, Research Data	Y	ZENODO	Y
ER 6.2	Identifier associated to the project deliverable	Standard format use for General communication, Libraries, Research Data	Y	ZENODO	Y

Table 4: Findability of other research outputs – project deliverables

Dataset No.	Will the dataset be identified by a persistent identifier?	What metadata will be created?	Will you provide keywords in the metadata? What?	In which trusted repository do you plan to publish the metadata on your dataset?	Will metadata remain available after the data is no longer available?
ER 1.1.	Y	title, description, authors, project acronym, GA number, keywords, date, licence, access conditions	Data Management	Zenodo	Yes
ER 1.2	Y	title, description, authors, project acronym, GA number, keywords, date, licence, access conditions	Communication, dissemination, exploitation	Zenodo	Yes
ER 1.7	Y	title, description, authors, project acronym, GA number, keywords, date, licence, access conditions	Practice abstract	Zenodo	Yes
ER 1.7	Y	title, description, authors, project acronym, GA number, keywords, date, licence, access conditions	Policy	Zenodo	Yes
ER 2.1	Yes, all the information related to the stored dataset will be provided at the end of the project	Experimental metadata will be recorded as Excel files and stored in the shared consortium repository. Metadata will include sample identification, experimental conditions and analytical methods	Yes. Keywords will refer to the work package, task, experimental activity, sample type, analytical method, or alternative protein source investigated.	At present, no public Open Access repository has been identified for the publication of project metadata. Such decision will be taken in agreement with all Consortium partners.	yes
ER 3.1	Y	authors, date of creation, methodology, variable descriptions, units of measurement	Enzymatic modification, enzymatic functionalization, peptidase, protein-glutamine glutaminase	Zenodo	Y

Dataset No.	Will the dataset be identified by a persistent identifier?	What metadata will be created?	Will you provide keywords in the metadata? What?	In which trusted repository do you plan to publish the metadata on your dataset?	Will metadata remain available after the data is no longer available?
ER 3.3	A DOI will be assigned upon deposition in Zenodo. D3.3 (public deliverable) will receive a persistent identifier through the Horizon Europe OpenAIRE portal (linked to GA No. 101182324). Author identifiers: ORCID for all VRLS authors.	DataCite Metadata Schema 4.x for dataset-level metadata. Dublin Core for report metadata. OpenAIRE Guidelines for EC-funded research. FAIR principles (Findable, Accessible, Interoperable, Reusable). Metadata fields: title, authors (ORCID), date of deposit, project (Protein4Impact, GA 101182324), funder (Horizon Europe), licence, persistent identifiers, keywords.	Yes. Keywords to be provided in metadata: insect protein, <i>Tenebrio molitor</i> , alkaline extraction, isoelectric precipitation, enzymatic hydrolysis, bioactive peptides, microencapsulation, spray drying, novel food, alternative protein, antioxidant activity, food supplement, pilot scale, feasibility study. Metadata will be provided in machine-readable format to enable OAI-PMH harvesting.	Zenodo (https://zenodo.org) — CERN-operated open-access repository endorsed by the European Commission for Horizon Europe projects. The mandatory public deliverable D3.3 will also be deposited in the OpenAIRE portal and the Protein4Impact project repository coordinated by GG.	Yes. Metadata records will remain permanently available in Zenodo and OpenAIRE even if the underlying dataset is removed or access is restricted, in compliance with FAIR principle F4.
ER 3.4	Y	Y	Y	Zenodo	Y
ER 3.6	Y (DOI)	Dataset-level (Discovery) Metadata-Data Dictionaries	Y	General Purpose Repository (e.g. Zenodo)	Y
ER 4.3	DOI	CodeMeta (JSON-LD, Schema.org-based): software name, description, version, authors (with ORCID), licence, runtime/OS, dependencies, related publications, funding source, and development status, stored as a codemeta.json file in the source repository. A Citation File Format (CITATION.cff)	Y	N	N
ER 4.4	DOI	CodeMeta (JSON-LD, Schema.org-based): software name, description, version, authors (with ORCID), licence, runtime/OS, dependencies,	Y	N	

Dataset No.	Will the dataset be identified by a persistent identifier?	What metadata will be created?	Will you provide keywords in the metadata? What?	In which trusted repository do you plan to publish the metadata on your dataset?	Will metadata remain available after the data is no longer available?
		related publications, funding source, and development status, stored as a codemeta.json file in the source repository. A Citation File Format (CITATION.cff)			
ER 4.5	DataCite DOI Persistent identifier minted automatically on deposit in Zenodo	CodeMeta (JSON-LD, Schema.org-based): software name, description, version, authors (with ORCID), licence, runtime/OS, dependencies, related publications, funding source, and development status, stored as a codemeta.json file in the source repository. A Citation File Format (CITATION.cff) file accompanies it for human- and machine-readable citation.; data import from xlsx documents, online access using Internet https;	Project-specific free-text keywords (e.g., protein production, insect/alternative protein, feedstock, bioconversion, digital twin, process simulation, sustainability KPI, Protein4Impact). DataCite DOI metadata > global PID Graph and scholarly infrastructures (DataCite, OpenAIRE, ORCID, B2FIND/EOSC). Records discoverable and indexable by humans and machines for discovery and re-use.	Zenodo	Yes
ER 5.1	DOI/repository identifier to be assigned upon deposit in Zenodo	Metadata: title, description, authors, project acronym, GA number, keywords, date, licence, access conditions.		Yes. Keywords: alternative proteins, consumer acceptance, citizen satisfaction, trust, sustainable diets, Protein4Impact. Metadata harvestable and indexable.	Zenodo
ER 5.3	Y-DOI	name of the ER, authors, format, description of the ER	unconventional proteins acceptance-Systematic review will be conducted providing an up to date, in-depth overview of drivers of acceptance of a	ZENODO	YES

Dataset No.	Will the dataset be identified by a persistent identifier?	What metadata will be created?	Will you provide keywords in the metadata? What?	In which trusted repository do you plan to publish the metadata on your dataset?	Will metadata remain available after the data is no longer available?
			wide range of alternative proteins		
ER 5.4	The dataset will not be assigned a persistent identifier	Basic descriptive metadata will be created	Keywords will be included	N/A	N/A
ER 6.3	Identifier associated to the project deliverable	Standard format use for General communication, Libraries, Research Data	Y	ZENODO	Y
ER 6.4	Identifier associated to the project deliverable	Standard format use for General communication, Libraries, Research Data	Y	ZENODO	Y
ER 6.5	Identifier associated to the project deliverable	Standard format use for General communication, Libraries, Research Data	Y	ZENODO	Y
ER 7.1	Identifier associated to the project deliverable	title, description, authors, project acronym, GA number, keywords, date, licence, access conditions	Y	ZENODO	Y
ER 7.2	Identifier associated to the project deliverable	title, description, authors, project acronym, GA number, keywords, date, licence, access conditions	Y	ZENODO	Y
ER 7.3	Identifier associated to the project deliverable	title, description, authors, project acronym, GA number, keywords, date, licence, access conditions	Y	ZENODO	Y

Table 5: Interoperability of PROTEIN4IMPACT project datasets

Dataset No.	Standards, formats or methodologies	Ontologies or vocabularies	Data format(s)	Software tools needed to access the data	Reference to other data
ER 1.3	FAIR	No specific domain ontologies or controlled vocabularies are used. Key definition and vocabulary is explained in the document if necessary.	PDF	PDF viewer	N
ER 1.4	FAIR	No specific domain ontologies or controlled vocabularies are used. Key definition and vocabulary is explained in the document if necessary.	PDF	PDF viewer	N
ER 1.5	FAIR	No specific domain ontologies or controlled vocabularies are used. Key definition and vocabulary is explained in the document if necessary.	PDF	PDF viewer	N
ER 1.6	FAIR	No specific domain ontologies or controlled vocabularies are used. Key definition and vocabulary is explained in the document if necessary.	PDF	PDF viewer	N
ER 2.2	Datasets and metadata will be recorded as excel files providing information that follows the recognized FAIR principles (Findable, Accessible, Interoperable, and Reusable).	No specific ontology or controlled vocabulary will be adopted across the project. Scientific names, analytical parameters, units of measurement will follow commonly accepted terminology	Data and metadata will be stored, shared, and archived as Microsoft Excel files (.xlsx).	The data can be accessed, viewed, and analysed using Microsoft Excel or any compatible software capable of reading .xlsx files.	The datasets generated within the project will be linked to the corresponding scientific publications (when available) and DOIs number will be provided.
ER 2.2	Datasets and metadata will be recorded as excel files providing information that follows the recognized FAIR	No specific ontology or controlled vocabulary will be adopted across the project. Scientific names, analytical	Data and metadata will be stored, shared, and archived as Microsoft Excel files (.xlsx).	The data can be accessed, viewed, and analysed using Microsoft Excel or any compatible	The datasets generated within the project will be linked to the corresponding scientific

Dataset No.	Standards, formats or methodologies	Ontologies or vocabularies	Data format(s)	Software tools needed to access the data	Reference to other data
	principles (Findable, Accessible, Interoperable, and Reusable).	parameters, units of measurement will follow commonly accepted terminology		software capable of reading .xlsx files.	publications (when available) and DOIs number will be provided.
ER 2.2	Datasets and metadata will be recorded as excel files providing information that follows the recognized FAIR principles (Findable, Accessible, Interoperable, and Reusable).	No specific ontology or controlled vocabulary will be adopted across the project. Scientific names, analytical parameters, units of measurement will follow commonly accepted terminology	Data and metadata will be stored, shared, and archived as Microsoft Excel files (.xlsx).	The data can be accessed, viewed, and analysed using Microsoft Excel or any compatible software capable of reading .xlsx files.	The datasets generated within the project will be linked to the corresponding scientific publications (when available) and DOIs number will be provided.
ER 2.2	Datasets and metadata will be recorded as excel files providing information that follows the recognized FAIR principles (Findable, Accessible, Interoperable, and Reusable).	No specific ontology or controlled vocabulary will be adopted across the project. Scientific names, analytical parameters, units of measurement will follow commonly accepted terminology	Data and metadata will be stored, shared, and archived as Microsoft Excel files (.xlsx).	The data can be accessed, viewed, and analysed using Microsoft Excel or any compatible software capable of reading .xlsx files.	The datasets generated within the project will be linked to the corresponding scientific publications (when available) and DOIs number will be provided.
ER 2.2	Datasets and metadata will be recorded as excel files providing information that follows the recognized FAIR principles (Findable, Accessible, Interoperable, and Reusable).	No specific ontology or controlled vocabulary will be adopted across the project. Scientific names, analytical parameters, units of measurement will follow commonly accepted terminology	Data and metadata will be stored, shared, and archived as Microsoft Excel files (.xlsx).	The data can be accessed, viewed, and analysed using Microsoft Excel or any compatible software capable of reading .xlsx files.	The datasets generated within the project will be linked to the corresponding scientific publications (when available) and DOIs number will be provided.
ER 2.2	Datasets and metadata will be recorded as excel files providing information that follows the recognized FAIR principles (Findable, Accessible, Interoperable, and Reusable).	No specific ontology or controlled vocabulary will be adopted across the project. Scientific names, analytical parameters, units of measurement will follow commonly accepted terminology	Data and metadata will be stored, shared, and archived as Microsoft Excel files (.xlsx).	The data can be accessed, viewed, and analysed using Microsoft Excel or any compatible software capable of reading .xlsx files.	The datasets generated within the project will be linked to the corresponding scientific publications (when available) and DOIs number will be provided.

Dataset No.	Standards, formats or methodologies	Ontologies or vocabularies	Data format(s)	Software tools needed to access the data	Reference to other data
ER 3.2	ISO, FAIR	No specific controlled vocabulary or ontology will be formally adopted across all datasets. To ensure consistency, terminology will be standardised using internationally recognised scientific nomenclature, relevant ISO standards where applicable, and terminology commonly used in the fields of biotechnology, food science, and microalgae research.	xls, xlsx; doc, docx; jpg	MS Office; Default image browser	N
ER 3.2	ISO, FAIR	No specific controlled vocabulary or ontology will be formally adopted across all datasets. To ensure consistency, terminology will be standardised using internationally recognised scientific nomenclature, relevant ISO standards where applicable, and terminology commonly used in the fields of biotechnology, food science, and microalgae research.	xls, xlsx; doc, docx; jpg	MS Office; Default image browser	N
ER 4.1	FAIR principles (F-A-I-R), CodeMeta, W3C PROV-O, JSON, ISO 8601	AGROVOC, FoodOn, ENVO, UO / QUDT	xlsx; ods; pdf; html; xml; json; jpg; png;	Web browser	N
ER 4.1	FAIR principles (F-A-I-R), CodeMeta, W3C PROV-O, JSON, ISO 8601	AGROVOC, FoodOn, ENVO, UO / QUDT	xlsx; ods; pdf; html; xml; json; jpg; png;	Web browser	N
ER 5.2	Data collection and reporting will follow the approved study protocol and relevant reporting guidelines for	Standardized terminology and controlled vocabularies commonly used in public health, nutrition, behavioral sciences, and epidemiology will be applied where appropriate. Keywords and variable	Data will be stored in CSV (open), XLSX (proprietary), PDF/A (open), and DOCX (proprietary) formats. Where applicable,	Data will be accessible using standard software tools, including Microsoft Excel or LibreOffice Calc for tabular data,	The dataset is primarily project-generated and does not directly reuse existing datasets. However, the study design, variables, and

Dataset No.	Standards, formats or methodologies	Ontologies or vocabularies	Data format(s)	Software tools needed to access the data	Reference to other data
	behavioral intervention research, where applicable.	names will be defined consistently across project outputs and datasets. For literature-based data, terminology used in established scientific databases and indexing systems (e.g., MeSH terms where applicable) will be followed. Where no specific ontology or controlled vocabulary exists, a project-specific data dictionary and standardized variable definitions will be used to ensure consistency and facilitate reuse	statistical analysis files may be stored in SPSS (.sav) format. Open and widely adopted formats will be preferred for long-term preservation and reuse.	SPSS/R/Stata for statistical analyses, and PDF readers and word-processing software for reports and documentation.	assessment instruments may build upon previously published research and validated measurement tools, which are referenced in the study protocol and related deliverables.
ER 6.1	FAIR	MS office build-in vocabularies	DOC, pdf	MS office, pdf viewer, web browser	
ER 6.2	FAIR	MS office build-in vocabularies	DOC, pdf	MS office, pdf viewer, web browser	

Table 6: Interoperability of other research outputs – project deliverables

Dataset No.	Standards, formats or methodologies	Ontologies or vocabularies	Data format(s)	Software tools needed to access the data	Reference to other data
ER 1.1	The questionnaire is based on Horizon Europe Data Management Plan Template	No specific domain ontologies or controlled vocabularies are used. Key terms are explained the document if necessary.	PDF	PDF viewer	N
ER 1.2	The questionnaire is based on Horizon Europe Data Management Plan Template	No specific domain ontologies or controlled vocabularies are used. Key definition and vocabulary is explained in the document if necessary.	PDF	PDF viewer	N
ER 1.7	FAIR	No specific domain ontologies or controlled vocabularies are used. Key definition and vocabulary is explained in the document if necessary.	PDF	PDF viewer	N
ER 1.8	FAIR	No specific domain ontologies or controlled vocabularies are used. Key definition and vocabulary is explained in the document if necessary.	PDF	PDF viewer	N
ER 2.1	Datasets and metadata will be recorded as excel files providing information that follows the recognized FAIR principles (Findable, Accessible,	No specific ontology or controlled vocabulary will be adopted across the project. Scientific names, analytical parameters, units of measurement will follow commonly accepted terminology	Data and metadata will be stored, shared, and archived as Microsoft Excel files (.xlsx).	The data can be accessed, viewed, and analysed using Microsoft Excel or any compatible software capable of reading .xlsx files.	The datasets generated within the project will be linked to the corresponding scientific publications (when available)

Dataset No.	Standards, formats or methodologies	Ontologies or vocabularies	Data format(s)	Software tools needed to access the data	Reference to other data
	Interoperable, and Reusable).				and DOIs number will be provided.
ER 3.1	FAIR	No specific controlled vocabulary or ontology will be formally adopted across all datasets. To ensure consistency, terminology will be standardised by using internationally recognised scientific nomenclature, consistent naming of samples, enzymes, substrates, and analytical parameters, as well as terminology commonly used in the fields of biotechnology, food science, and protein research.	xls, xlsx; doc, docx; tiff	MS Office; Default image browser	N
ER 3.3	Laboratory methods: standard analytical protocols for protein content determination (Kjeldahl method), moisture and ash (gravimetric methods), antioxidant activity (ABTS radical scavenging assay and metal ion chelating activity assay), degree of hydrolysis (pH-stat method). FTIR-ATR spectroscopy for	EU Novel Food Regulation terminology (Reg. (EU) 2015/2283 and implementing regulations including Reg. (EU) 2025/89 for T. molitor authorisation). EU food supplement legislation terminology (Dir. 2002/46/EC). EU nutrition and health claims terminology (Reg. (EC) No 1924/2006). Standard scientific nomenclature for species (Tenebrio molitor) and chemical compounds.	Open formats preferred: CSV/TSV for tabular numerical data; PDF/A for reports and deliverables; PNG/TIFF for images (SEM micrographs, FTIR spectra, prototype photos); XLSX for raw experimental data tables. Where proprietary instrument formats are used, data are exported to open formats for archiving and sharing.	Standard office and data analysis software (spreadsheet and word-processing applications) for tabular data and report preparation. Instrument-specific software for FTIR and SEM data acquisition — raw data exported to open formats. No specialist proprietary software required to access the archived open-format datasets.	Builds on: WP2 upstream material characterisation data (alternative protein sources: T. molitor, P. ostreatus, D. salina, fish by-products). Linked to: WP6 Techno-Economic Assessment (In.Bio, D6.2) and Life-Cycle Costing

Dataset No.	Standards, formats or methodologies	Ontologies or vocabularies	Data format(s)	Software tools needed to access the data	Reference to other data
	structural characterisation. SEM for particle morphology. Spray-drying process documented in D3.3. Good laboratory practice principles applied throughout.				(AgC, D6.3) which will use D3.3 pilot yield and cost data as inputs. Related publications: peer-reviewed articles on T. molitor protein hydrolysates (to be cited with DOI upon publication). D3.3 cross-references D1.5 (DEC Plan).
ER 3.4	FAIR	N/A	xlsx	Excel	N/A
ER 3.6	ISO, FAIR, Internal SOPs and GMPs	Life Sciences / Biomedical	xlsx; pdf; docx	MS office; PDF viewer	N
ER 4.3	FAIR principles (F-A-I-R), CodeMeta, W3C PROV-O, JSON, ISO 8601	AGROVOC, FoodOn, ENVO, UO / QUDT	xlsx; ods; pdf; html; xml; json; jpg; png;	Web browser	N
ER 4.4	W3C	AGROVOC, FoodOn, ENVO, UO / QUDT	xlsx; ods; pdf; html; xml; json; jpg; png;	Web browser	N
ER 4.5	FAIR principles (F-A-I-R), CodeMeta, W3C	AGROVOC, FoodOn, ENVO, UO / QUDT	xlsx; ods; pdf; html; xml; json; jpg; png;	Web browser	N

Dataset No.	Standards, formats or methodologies	Ontologies or vocabularies	Data format(s)	Software tools needed to access the data	Reference to other data
	PROV-O, JSON, ISO 8601				
ER 5.1	FAIR principles, Horizon Europe DMP, GDPR-compliant survey methodologies.	Food systems and consumer behaviour terminology; controlled keywords used consistently.	CSV, XLSX, PDF/A (open/preferred formats).	Excel, LibreOffice, CSV-compatible software, PDF reader.	Related WP5 outputs and supporting literature references
ER 5.3	The review methodology follows the PRISMA guidelines for systematic reviews and evidence synthesis	Standardized terminology and controlled vocabularies commonly used in public health, nutrition, behavioral sciences, and epidemiology will be applied where appropriate. Keywords and variable names will be defined consistently across project outputs and datasets. For literature-based data, terminology used in established scientific databases and indexing systems (e.g., MeSH terms where applicable) will be followed. Where no specific ontology or controlled vocabulary exists, a project-specific data dictionary and standardized variable definitions will be used to ensure consistency and facilitate reuse	Data will be stored in CSV (open), XLSX (proprietary), PDF/A (open), and DOCX (proprietary) formats. Where applicable, statistical analysis files may be stored in SPSS (.sav) format. Open and widely adopted formats will be preferred for long-term preservation and reuse.	Data will be accessible using standard software tools, including Microsoft Excel or LibreOffice Calc for tabular data, SPSS/R/Stata for statistical analyses, and PDF readers and word-processing software for reports and documentation.	The dataset is based on data extracted from published scientific literature identified through systematic searches of bibliographic databases (e.g., PubMed, Scopus, Web of Science). References to the original publications, including DOIs where available, are maintained in the review database and reported in the

Dataset No.	Standards, formats or methodologies	Ontologies or vocabularies	Data format(s)	Software tools needed to access the data	Reference to other data
					corresponding deliverable and publications.
ER 5.4	No specific international or domain standard applies, as D5.3 is a qualitative reporting deliverable documenting training activities carried out with agrifood stakeholders	No formal ontology or controlled vocabulary is used	The report will be delivered in PDF format (open, non-proprietary). Supporting materials (e.g. training attendance lists, survey results, presentation slides) may additionally be provided in open formats such as .xlsx and .pptx.	Standard office software is sufficient	D5.3 builds upon activities and outputs produced within WP5 (stakeholder engagement and training) and is linked to other WP5 deliverables. It does not reference or reuse external datasets.
ER 6.3	FAIR	MS office build-in vocabularies	DOC, pdf	MS office, pdf viewer, web browser	
ER 6.4	FAIR	MS office build-in vocabularies	DOC, pdf	MS office, pdf viewer, web browser	Ecoinvent
ER 6.5	FAIR	MS office build-in vocabularies	DOC, pdf	MS office, pdf viewer, web browser	
ER 7.1	FAIR	n/a	PDF	PDF viewer	n/a
ER 7.2	FAIR	n/a	PDF	PDF viewer	n/a
ER 7.3	FAIR	n/a	PDF	PDF viewer	n/a

Table 7: Reusability of PROTEIN4IMPACT project datasets

Dataset No.	Validation of data analysis and facilitation of data re-use	Data availability in the public domain to permit the widest re-use possible	Relevant data quality assurance processes	Usability of the third parties
ER 1.3	N/A	Dataset will be available on ZENODO.	The project coordinator (CS) write the document based on the GA and project policies. It is accessible to all partners for review and updates.	No.
ER 1.4	N/A	Dataset will be available on ZENODO.	All partners provide dataset information through a questionnaire. The project coordinator (CS) reviews all inputs, identifies missing or inconsistent information, and requests corrections to ensure completeness, and accuracy. It is accessible to all partners for review and updates.	Yes
ER 1.5	N/A	Dataset will be available on ZENODO.	All partners provide dataset information through a questionnaire. The project coordinator (CS) reviews all inputs, identifies missing or inconsistent information, and requests corrections to ensure completeness, and accuracy. It is accessible to all partners for review and updates.	Yes
ER 1.6	N/A	Dataset will be available on ZENODO.	All partners provide dataset information through a questionnaire. The project coordinator (CS) and task leader (IDENER) reviews all inputs, identifies missing or inconsistent information, and requests corrections to ensure completeness, and accuracy. It is accessible to all partners for review and updates.	Yes
ER 2.2	The information required to validate data analysis and facilitate data re-use will be included in the excel files containing the datasets and metadata.	At this stage, no decision has been made regarding the public release of the datasets.	Data quality assurance will follow the FAIR principles (Findable, Accessible, Interoperable, and Reusable).	The use of project data by third parties after the end of the project will be evaluated by the consortium before the end of the project.
ER 2.3	The information required to validate data analysis and facilitate data	At this stage, no decision has been made	Data quality assurance will follow the FAIR principles (Findable, Accessible, Interoperable, and Reusable).	The use of project data by third parties after the end of the

Dataset No.	Validation of data analysis and facilitation of data re-use	Data availability in the public domain to permit the widest re-use possible	Relevant data quality assurance processes	Usability of the third parties
	re-use will be included in the excel files containing the datasets and metadata.	regarding the public release of the datasets.		project will be evaluated by the consortium before the end of the project.
ER 2.4	The information required to validate data analysis and facilitate data re-use will be included in the excel files containing the datasets and metadata.	At this stage, no decision has been made regarding the public release of the datasets.	Data quality assurance will follow the FAIR principles (Findable, Accessible, Interoperable, and Reusable).	The use of project data by third parties after the end of the project will be evaluated by the consortium before the end of the project.
ER 2.5	The information required to validate data analysis and facilitate data re-use will be included in the excel files containing the datasets and metadata.	At this stage, no decision has been made regarding the public release of the datasets.	Data quality assurance will follow the FAIR principles (Findable, Accessible, Interoperable, and Reusable).	The use of project data by third parties after the end of the project will be evaluated by the consortium before the end of the project.
ER 2.6	The information required to validate data analysis and facilitate data re-use will be included in the excel files containing the datasets and metadata.	At this stage, no decision has been made regarding the public release of the datasets.	Data quality assurance will follow the FAIR principles (Findable, Accessible, Interoperable, and Reusable).	The use of project data by third parties after the end of the project will be evaluated by the consortium before the end of the project.
ER 2.7	The information required to validate data analysis and facilitate data re-use will be included in the excel files containing the datasets and metadata.	At this stage, no decision has been made regarding the public release of the datasets.	Data quality assurance will follow the FAIR principles (Findable, Accessible, Interoperable, and Reusable).	The use of project data by third parties after the end of the project will be evaluated by the consortium before the end of the project.

Dataset No.	Validation of data analysis and facilitation of data re-use	Data availability in the public domain to permit the widest re-use possible	Relevant data quality assurance processes	Usability of the third parties
ER 3.2	N/A	N	Data quality will be ensured through standardized experimental procedures and quality control measures. All experimental measurements will be conducted in at least triplicate to ensure accuracy and reproducibility. Data will be reviewed for completeness, consistency, and potential	N
ER 3.5	N/A	N	Data quality will be ensured through standardized experimental procedures and quality control measures. All experimental measurements will be conducted in at least triplicate to ensure accuracy and reproducibility. Data will be reviewed for completeness, consistency, and potential	N
ER 4.1	On-line help assistance, workflow presentation files	Yes, after end user accepting license agreements	Workflow: Sources data estimation > Calculation of KPIs > Calculation of product sustainability > Assessment of a protein production sustainability for a whole plant under realistic assumptions of typical food manufacturing processes and customer demand	Yes
ER 4.2	On-line help assistance, workflow presentation files	Yes, after end user accepting license agreements	Workflow: Sources data estimation > Calculation of KPIs > Calculation of product sustainability > Assessment of a protein production sustainability for a whole plant under realistic assumptions of typical food manufacturing processes and customer demand	Yes
ER 5.2	Data reuse will be facilitated through the provision of README files, metadata, variable descriptions, coding manuals, and methodological documentation. Relevant protocols,	Y - ER will be published on ZENODO- Data will be shared in accordance with ethical, legal, and GDPR requirements. Anonymized and/or aggregated data may be made	Data quality will be ensured through standardized protocols, staff training, validation checks, data cleaning procedures, and documentation of all processing steps. Completeness, consistency, and accuracy of the data will be regularly monitored	Y- The data produced in the project will be reusable by third parties after the end of the project, where ethical,

Dataset No.	Validation of data analysis and facilitation of data re-use	Data availability in the public domain to permit the widest re-use possible	Relevant data quality assurance processes	Usability of the third parties
	deliverables, and publications will be referenced to support interpretation, validation, and reuse of the data.	available where feasible, while access to sensitive data may be restricted. Metadata and documentation will remain publicly available to maximize discoverability and reuse. Appropriate reuse licenses will be applied where possible, and data provenance will be fully documented.		legal, confidentiality, and intellectual property considerations permit. Appropriate metadata, documentation, and supporting materials will be provided to facilitate understanding and reuse. Personal or sensitive data will be anonymized, aggregated, or access-restricted in accordance with GDPR and ethical requirements. Data sharing and reuse conditions will be clearly defined through appropriate licenses and repository access policies.
ER 6.1	Associated deliverable	Sensitive	Peer-review by project partners and coordinator team	Y
ER 6.2	Associated deliverable	Sensitive	Peer-review by project partners and coordinator team	Y

Table 8: Reusability of other research outputs – project deliverables

Dataset No.	Validation of data analysis and facilitation of data re-use	Data availability in the public domain to permit the widest re-use possible	Relevant data quality assurance processes	Usability of the third parties
ER 1.1	N/A	Dataset will be available on ZENODO.	All partners provide dataset information through a questionnaire. The Data Manager (GG) reviews all inputs, identifies missing or inconsistent information, and requests corrections to ensure completeness, accuracy, and FAIR compliance. Before any dataset is accepted or shared, provided information are checked for consistency and quality.	No
ER 1.2	N/A	Dataset will be available on ZENODO.	All partners provide dataset information through a questionnaire. The Data Manager (GG) reviews all inputs, identifies missing or inconsistent information, and requests corrections to ensure completeness, and accuracy. An internal tool (the Release Log) for tracking and monitoring all DEC activities is in place. It is accessible to all partners for review and updates.	No
ER 1.7	N/A	Dataset will be available on ZENODO.	All partners provide dataset information through a questionnaire. The project coordinator (CS) reviews all inputs, identifies missing or inconsistent information, and requests corrections to ensure completeness, and accuracy. It is accessible to all partners for review and updates.	Yes
ER 1.8	N/A	Dataset will be available on ZENODO.	All partners provide dataset information through a questionnaire. The project coordinator (CS) reviews all inputs, identifies missing or inconsistent information, and requests corrections to ensure completeness, and accuracy. It is accessible to all partners for review and updates.	Yes

Dataset No.	Validation of data analysis and facilitation of data re-use	Data availability in the public domain to permit the widest re-use possible	Relevant data quality assurance processes	Usability of the third parties
ER 2.1	The information required to validate data analysis and facilitate data re-use will be included in the excel files containing the datasets and metadata.	At this stage, no decision has been made regarding the public release of the datasets.	Data quality assurance will follow the FAIR principles (Findable, Accessible, Interoperable, and Reusable).	The use of project data by third parties after the end of the project will be evaluated by the consortium before the end of the project.
ER 3.1	N/A	N	Data quality will be ensured through standardized experimental procedures and quality control measures. All experiments will be conducted at least in triplicate. Appropriate controls and blanks will be included to account for background signals, matrix effects, and non-enzymatic reactions. Data will be reviewed for completeness, consistency, and potential outliers.	N
ER 3.3	D3.3 will include full methodology appendix covering extraction conditions, enzyme specifications, analytical protocols, instrument settings and statistical methods. Raw experimental datasets will be deposited in Zenodo alongside the report. README files will describe file structure, variable names and units. Triplicate measurements and standard deviations reported for all key parameters. Independent cross-validation of protein content with NTUA.	D3.3 is a public (PU) deliverable: the full report will be openly available under CC BY 4.0 licence via the OpenAIRE portal and Zenodo. Supporting datasets will be deposited under CC BY 4.0 or CC0 in Zenodo. Provenance documented through DataCite metadata. Any data with IP sensitivity (patent-relevant process parameters) will be released under CC BY 4.0 with a 12-month embargo, after	All analytical measurements performed in triplicate; results reported as mean $\pm$ standard deviation. Statistical analysis of significant differences performed. Instrument calibration records maintained for all analytical equipment. Cross-laboratory validation: protein content cross-checked with NTUA. Data recording in laboratory notebooks (paper and digital); reviewed by second VRLS researcher before entry into project database.	Yes. The D3.3 feasibility study and supporting datasets are designed for reuse by: (i) food ingredient companies evaluating insect-protein processing feasibility; (ii) academic researchers benchmarking T. molitor protein isolation and encapsulation methods; (iii) regulatory bodies (EFSA) assessing novel-food safety dossiers; (iv) future Horizon Europe or national research projects building on Protein4Impact outputs.

Dataset No.	Validation of data analysis and facilitation of data re-use	Data availability in the public domain to permit the widest re-use possible	Relevant data quality assurance processes	Usability of the third parties
		patent filing, as permitted under the Grant Agreement.		Open-format data archiving ensures accessibility after project end.
ER 3.4	N/A	Y	N/A	Y
ER 3.5	N/A	N	Data quality will be ensured through standardized experimental procedures and quality control measures. All experimental measurements will be conducted in at least triplicate to ensure accuracy and reproducibility. Data will be reviewed for completeness, consistency, and potential	N
ER 3.6	Peer-reviewed paper will serve as the high-level conceptual guide, while the repository (like Zenodo or Figshare) will hold execution files.	Yes, data will be made available to permit the widest reuse possible, in compliance with Grant Agreement. However, to protect the research team's priority, dataset will be placed under a temporary embargo and will be released following the publication .	During data collection, QA will be maintained through adherence to SOPs. To preserve data integrity, the original raw files will remain unaltered and stored as read-only.	After publication, data's longevity is guaranteed past the end of the project lifecycle. By depositing the data package into a repository (e.g., Zenodo). The repository will maintain a Digital Object Identifier (DOI),
ER 4.3	On-line help assistance, workflow presentation files	Yes, after end user accepting license agreements	Workflow: Source data estimation > Calculation of KPIs > Calculation of production sustainability > Assessment of a protein production sustainability for a whole plant under realistic assumptions of typical food manufacturing processes and customer demand	Yes

Dataset No.	Validation of data analysis and facilitation of data re-use	Data availability in the public domain to permit the widest re-use possible	Relevant data quality assurance processes	Usability of the third parties
ER 4.4	On-line help assistance, workflow presentation files	Y	N/A	Yes
ER 4.5	On-line help assistance, workflow presentation files	Y	N/A	Yes
ER 5.1	Metadata records, questionnaires, variable descriptions, methodological documentation and data processing protocols	Yes, subject to GDPR, anonymisation and ethical requirements	Survey validation, consistency checks, data cleaning, duplicate removal, GDPR compliance and internal quality controls	Yes. Anonymized datasets and supporting documentation will remain accessible after project completion, subject to legal and ethical constraints, enabling future research, policy development, education and innovation activities.
ER 5.3	n/a	Y - ER will be publish on ZENODO	Predefined inclusion criteria and standardized data extraction procedures will be applied to ensure data reliability and reproducibility	y
ER 5.4	As D5.3 is a descriptive reporting deliverable, formal data validation procedures are not applicable. The report will include a methodology section describing how training activities were organised, who participated, and how outcomes were assessed. A summary of results from participant surveys and feedback will be included to allow internal review and quality assessment.	The dataset will be made freely available in the public domain. Even, it is an internal project deliverable submitted to the European Commission and shared within the consortium, it can constitute part of broader disseminations. The provenance of the data is documented within the report itself (partner	Quality assurance for D5.3 follows the internal review process. The deliverable will be reviewed by the WP5 leader and the project coordinator before submission.	D5.3 is an internal reporting deliverable specific to the Protein4Impact project. It documents training activities carried out within the project scope and the results from dissemination activities are encouraged to be reused by third parties after the project ends. The methodological approaches and lessons learned described in the report may, however, inform future training initiatives by

Dataset No.	Validation of data analysis and facilitation of data re-use	Data availability in the public domain to permit the widest re-use possible	Relevant data quality assurance processes	Usability of the third parties
		contributions, dates of activities, methods used).		consortium partners in their respective organisations.
ER 6.3	Associated deliverable	Sensitive	Peer-review by project partners and coordinator team	Y
ER 6.4	Foreground dataset inventory	Sensitive	Peer-review by project partners and coordinator team	Y
ER 6.5	Associated deliverable		Peer-review by project partners and coordinator team	Y
ER 7.1	All partners provide dataset information through a questionnaire. The project coordinator (CS) reviews all inputs, identifies missing or inconsistent information, and requests corrections to ensure completeness, and accuracy.	Sensitive	Peer-review by project partners and coordinator team	N
ER 7.2	All partners provide dataset information through a questionnaire. The project coordinator (CS) reviews all inputs, identifies missing or inconsistent information, and requests corrections to ensure completeness, and accuracy.	Sensitive	Peer-review by project partners and coordinator team	N
ER 7.2	All partners provide dataset information through a questionnaire. The project coordinator (CS) reviews all inputs, identifies missing or inconsistent information, and requests corrections to ensure completeness, and accuracy.	Sensitive	Peer-review by project partners and the coordinator team	N

7. List of Tables

Table 1: PROTEIN4IMPACT project datasets – description	13
Table 2: Other research outputs - project deliverables	16
Table 3: Findability of PROTEIN4IMPACT project datasets including provisions for metadata	19
Table 4: Findability of other research outputs – project deliverables	23
Table 5: Interoperability of PROTEIN4IMPACT project datasets	27
Table 6: Interoperability of other research outputs – project deliverables	31
Table 7: Reusability of PROTEIN4IMPACT project datasets	36
Table 8: Reusability of other research outputs – project deliverables.....	40

8. List of Abbreviations

Abbreviation	Abbreviation for
CC0	Creative Commons Public Domain Dedication
CC BY	Attribution International Public License
DMP	Data Management Plan
OA	Open Access

9. Project Consortium



<https://www.centralesupelec.fr>



<https://www.dtu.dk>



<https://katalanotis.org.cy>



<https://www.uni-hohenheim.de>



<http://www.enea.it>



<https://nordicdf.eu>



<https://www.venusroses-labsolutions.eu>



<https://www.cartif.es>



<http://www.grant-garant.cz>



<https://www.uniwa.gr>



<https://golem.at>



<https://www.cnr.it>



<https://www.ntua.gr>



<https://www.aquabt.com>



<http://www.agriclima.eu/>



<https://www.uminho.pt>



<http://www.idener.es>



Innovazione e Bioeconomia