**Project Number:** 1836

**Project Acronym:** InnoSol4Med

**Project title:** Innovative sustainable solutions for ready-to-eat traditional Mediterranean products and non-conventional healthy foods

# DATA MANAGEMENT PLAN

## 1. Data Summary

The purpose of the data collection/generation is to support the research objectives and goals of the project. Specifically, it is intended to:

- Provide empirical evidence and data-driven insights to validate research hypotheses.

- Facilitate the exploration of innovative matrices (essential oils, natural compounds/extracts) from sustainable sources (agro-food by-products and medicinal plants) and autochthonous microbial strains, as new functional ingredients and biotechnological solutions for the improvement of quality, safety, nutritional value and functionality of traditional foods and development of new food prototypes in synergy with non-thermal innovative technological solutions (ozone, nanoemulsions, fermentation).

- Enable the development of new and tailored business models and awareness campaigns that can support the adoption and exploitation of the innovative solutions developed in the project through measuring consumer attitudes and preferences for low-processed and healthy ready-to-eat foods (RTE).

The data collected or generated for this project originates from various sources, including:

- Laboratory research activities,

- Pilot scale productions in collaboration with company partners

- Consumer attitudes and preferences surveys.

The project will generate and collect various types and formats of data, including:

- Structured data from surveys and questionnaires (consumer attitudes and preferences for the solutions and food models developed in the project),

- Textual data from scientific literature and reports,

- Experimental data measured during analyses (characterization of natural bioactive compounds, extracts, and essential oils; *in vitro* investigations of biological activities of the extracts, essential oils and bioactive compounds; isolation and characterization of autochthonous lactic acid bacteria (LAB) from naturally fermented matrices; *in vitro* screening for efficient folate-producing LAB strains of dairy origin; effects of non-invasive innovative technological solutions (ozone and nanoemulsions) on the stability and availability of the functional ingredients/bioactive compounds *in vitro*; use of *in vitro* digestion methods to assess the functional properties of the selected innovative ingredients; laboratory and pilot scale Mediterranean food prototype preparation),

- Observational data (sensory analyses).

Existing data will be utilized when applicable and relevant to the project. This re-use may involve the results of the funded projects of the Consortium partners:

- PRIMA2019 BioProMedFood (ID1467): InnoSol4Med will use knowledge of autochthonous LAB strains and potent ingredients from Med biodiversity to investigate

technological and safety issues and test them for functionality when used in non-conventional matrices;

- PRIMA-MED 2020 MED-LINKS (ID: 1591): Business models developed and tested in MED-LINKS will be used as a reference to identify product-specific business strategies proposed in InnoSol4Med;

- H2020 FF-IPM2019: InnoSol4Med will apply ozone application know-how from the agricultural sector and investigate its application in the food sector;

- PRIMA2019 BIOfreshCloud know-how on extending the shelf life of Med fresh produce using sustainable preservation technologies and predictive modeling will be used.

- H2020-HEALTH/0795 PlasticsFatE Methodology know-how will be used for ingredients/foods in InnoSol4Med.

The expected size of the data will vary depending on the type and format.

The data generated and collected in this project may be valuable and useful to several stakeholders, including:

- Fellow researchers and students in the biotechnology field.

- Policymakers and government agencies.

- Environmental organizations and NGOs.

- Private companies.

## 2. FAIR data

### 2. 1. Making data findable, including provisions for metadata

We are committed to making project data discoverable, identifiable, and locatable. We will assign persistent and unique identifiers, such as Digital Object Identifiers (DOIs) and URN-NBN persistent identifiers, to datasets generated during the project. These identifiers will ensure that the data can be reliably located and cited, contributing to data findability.

Project data will be assigned with metadata as required by selected repository.

We adhere to established naming conventions to ensure consistency and clarity in dataset names. Naming conventions will follow the format e.g., "ProjectAbbreviation_DatasetName_Date" to provide meaningful and standardized names for datasets.

Folders will be structured hierarchically.

Comprehensive search keywords and tags associated with each dataset will be provided to optimize discoverability and re-use. These keywords will be selected to accurately reflect the content and subject matter of the data, making it easier for other researchers to find and utilize the datasets.

Clear version numbers for datasets will be provided to track changes and updates over time. Version numbers will follow a standardized format, such as "v1.0," "v2.0," etc., and will be documented alongside each dataset.

Metadata will be created to describe each dataset comprehensively (administrative, descriptive and/or structural metadata). This metadata will include (depending on the type of dataset):

- Dataset title and description + permanent identifier,

- Authorship and contributor information,

- Date of creation and update,

- Data format and file structure,

- Data collection methods and instruments,

- Data processing and analysis procedures,

- Spatial and temporal coverage,

- Keywords and subject classifications,

- Licensing and access information,

- Persistent identifiers (e.g., DOIs),

- Version history.

## 2.2. Making data openly accessible

The data and associated metadata will be made accessible by depositing it in a coordinator's (University of Split) institutional repository that supports open access. The selected repository aligns with FAIR data principles and it is OpenAIRE and OpenDOAR compliant.

Access to the data will primarily require standard software tools and formats commonly used in the field. Comprehensive documentation will accompany the data to facilitate access. Where specialized software is necessary, we will include instructions and, whenever possible, provide open-source or publicly available software to ensure accessibility.

In case of generating data for new product technology with patent potential, data will be protected according to the intellectual property rules included in the Consortium agreement.

## 2.3. Making data interoperable

The data produced in the project will be designed to be interoperable, ensuring that they can be exchanged and reused seamlessly by researchers, institutions, organizations, and across geographical boundaries. We are committed to adhering to relevant standards and formats to maximize data interoperability and compatibility with open software applications. The goal is to enable easy re-combinations of our datasets with those from different origins.

Also, the chosen repository has OAI-PMH protocol to ensure interoperability for data and service providers.

We will follow established data and metadata standards and vocabularies to enhance data interoperability. These standards include, but are not limited to:

For data formats: .docx, .txt, .odf, .pdf, .xlsx, .csv, .ods, .tiff, .jpg, .png, .mp3, .flac, .wav, .mp4, .mpeg4, .mpeg, etc.

For metadata: Dublin Core, DataCite Metadata Schema, domain-specific metadata schemas as appropriate, etc..

We will prioritize the use of standard vocabularies for all data types present in our dataset to promote inter-disciplinary interoperability. Standardized vocabularies will facilitate data integration across different disciplines, making it easier for researchers from various domains to work with our data.

In cases where project-specific ontologies or vocabularies are necessary, we will provide mappings to more commonly used ontologies whenever possible. These mappings will serve as bridges to ensure that our data can be integrated with datasets using established ontologies and vocabularies, enhancing overall data interoperability.

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

The data will be licensed to permit the widest re-use possible by applying an open and permissive license. Specifically, we will use the Creative Commons Attribution 4.0 International (CC BY 4.0) for the data. This license allows for maximum re-use, redistribution, and adaptation of the data, provided proper attribution is given to the original source.

The data will be made available for re-use as soon as possible after the end of the project. We recognize the importance of timely data sharing. However, in exceptional cases where there is a need for temporary embargoes, such as for publication or patent purposes, embargoes will be applied for a specified period. Any embargoes will be clearly justified.

The data produced and used in the project are usable by third parties, particularly after the end of the project. We are committed to facilitating data re-use by providing open access to the data. However, in cases where specific data must be restricted, restrictions will be clearly explained (e.g. patent applications).

The data are intended to remain re-usable indefinitely. We are committed to maintaining the accessibility and reusability of the data over the long term. The data will continue to be available even after the project's conclusion, and measures will be taken to ensure that the data remains accessible and relevant for future research and applications.

The data quality assurance processes include:

- Data validation and verification procedures during data collection,

- Documentation of data processing and analysis methodologies,

- Regular data quality checks to identify and address errors or inconsistencies,

- Use of naming conventions,

- Version control to track changes and improvements in data quality,

- Procedures for metadata documentation to provide context and information about data quality,

- Collaboration with domain experts to ensure data accuracy and validity.

## 3. Allocation of resources

There is no cost for making the data of this project FAIR.

## 4. Data security

We have implemented comprehensive provisions for data security to safeguard project data throughout its lifecycle. Data is stored in encrypted, access-controlled, and redundant storage systems to protect against data loss or unauthorized access (e.g. Microsoft OneDrive, Microsoft Teams, Dropbox). Regular backups and version control mechanisms are in place to ensure data recovery in the event of accidental loss or corruption. Measures are taken to comply with data protection regulations, including anonymization and pseudonymization where applicable.

The data is safely stored in a certified repository that specializes in long-term data preservation and curation.

## 5. Ethical aspects

Informed consent for data sharing and long-term preservation will be included in our questionnaires when dealing with personal data. Participants will be provided with clear information about how their data will be used, shared, and preserved. They will be given the option to provide or withhold consent for these specific purposes. Informed consent forms an integral part of our ethical data handling practices.

## 6. Other issues

Research data management process for this research project is in compliance with University of Split's Open Science Policy.

| HISTORY OF CHANGES | | |
|---|---|---|
| **Version** | **Publication date** | **Change** |
| 1.0 | 29.09.2023 | ▪ Initial version |