Funded by
the European Union

D3.1
PROJECT DATA MANAGEMENT PLAN

Project	METROFOOD-EPI – *METROFOOD-RI Early Phase Implementation*
Grant Agreement n.	101130162
Call	HORIZON-INFRA-2023-DEV-01
Due date	Month 4 – April 2024
WP – Task	WP3 – Task 3.1
Lead Beneficiary	PREMOTEC GMBH
Type	DMP – Data Management Plan
Dissemination Level	PU

Funded by
the European Union

Table of contents

Funded by
the European Union

1. Executive Summary

METROFOOD-RI - Infrastructure for Promoting Metrology in Food and Nutrition (www.metrofood.eu) is a distributed RI aimed to promote scientific excellence in the field of food quality and safety. It provides high-quality metrology services in food and nutrition, comprising an important cross-section of highly interdisciplinary and interconnected fields throughout the food value chain, including agrifood, sustainable development, food safety, quality, traceability and authenticity, environmental safety, and human health. It combines a Physical-RI (P-RI) and an electronic-RI (e-RI) for open data deposition, access, and processing. The P-RI coordinates and integrates an existing network of state-of-the-art facilities including: in the "Metro" side, laboratories for the full chemical, physical-chemical and microbiological characterisation of foods and any matrix of interest in relation to the agrifood (e.g., environmental matrices from the agroecosystem of production, feeds, food contact materials, etc.), and plants for Reference Material (RM) development and production; in the "Food" side, experimental fields/farms for crop production and animal breeding, small-scale plants for food processing and storage, kitchen-labs for food preparation, and "demo" sites for direct stakeholder engagement (e.g., to run Living Labs). The e-RI consists of a service-oriented electronic architecture providing an accessible platform for sharing and integrating data, knowledge, and information on metrological tools for food analysis and for facilitating the availability and use of agrifood data to the user community. The e-RI collects, integrates, and makes the P-RI results open and interoperable, organising and complementing them with existing data and providing tools for various uses of the data, even promoting their interoperability and the integration with data arising from other existing networks and infrastructures. METROFOOD-RI users are individuals, teams or institutions who are foreseen to use the services of the RI. Four main user categories have been identified as follows: Researchers and academic communities; Policy makers/food inspection and control agencies; Food business operators (FBOs); Consumers/citizens. METROFOOD-RI is structured according to a Hub & Nodes model. The Central Hub (CH) will be the statutory seat of the ERIC and will represent the heart of the strategy, coordination, communication, and administration of METROFOOD-RI, coordinating the overall infrastructure and managing the central e-portal, which will give access to all the resources and services of the infrastructure. The CH will act as a coordinating European layer across all National Nodes (NNs), while the NNs will represent the operational sites of the infrastructure.

METROFOOD-EPI (GA 101130162) is the Horizon Europe funded project supporting the Early Phase Implementation of METROFOOD-RI. Its overarching mission is to advance the building process of METROFOOD-RI as infrastructure consolidated for its full implementation and ensure an effective start of the operational phase.

This document contains the Data Management Plan (DMP) for the METROFOOD-EPI project, addressing what and how data - including metadata - will be generated, collected, handled and preserved in repositories following the FAIR principle (findable, accessible, interoperable and re-useable) as good as possible.
According to the European Commission [1], a DMP is a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed, and/or generated by a Horizon Europe project. As part of making research data findable, accessible, interoperable, and re-usable (FAIR), a DMP should include several information. First is the handling of research data during and after the end of the project, second is what data will be collected, processed, and/or generated, third, which methodology & standards will be applied and forth whether data will be shared/made open access. Finally, it should contain information about how data will be curated, maintained and

preserved during and after the project.
The DMP for METROFOOD-EPI is related to the DMP of METROFOOD-RI as METROFOOD-EPI is the "Early Phase Implementation" project of METROFOOD-RI. However, the DMP for METROFOOD-RI is a separate document with a longer timeframe. Data which is generated during the METROFOOD-EPI project, and which will be used by METROFOOD-RI, will be included in the DMP of METROFOOD-RI. The DMP is a living document and can be updated during the project if necessary.

2. Introduction

METROFOOD-RI aims at providing high quality metrology services in food and nutrition, comprising an important cross-section of highly interdisciplinary and inter-connected fields throughout the food value chain, including agri-food, sustainable development, food safety, quality, traceability and authenticity, environmental safety, and human health.

METROFOOD-RI's mission is to enhance quality and reliability of measurement results and make available and share data, information and metrological tools, to enhance scientific excellence in the field of food quality and safety and strengthen scientific knowledge, also promoting scientific cooperation and to integration. The general objective is to enhance scientific cooperation and encourage interaction between the various stakeholders, as well as the creation of a common and shared base of data, information, and knowledge.

To serve its role, METROFOOD-RI needs to gather, process and generate various types of data. As the various types of data will accompany every single step, not only during the lifecycle of METROFOOD-RI, but also during each single development phase and operational use case, an appropriate Data Management Plan (DMP) - continuously upgraded - is crucial. This report presents the DMP for METROFOOD-EPI, with reference to METROFOOD-RI where appropriate. Where necessary, the document will differentiate between METROFOOD-RI and METROFOOD-EPI.

The overall aim of METROFOOD-EPI is to bring METROFOOD-RI ESFRI project to the level of a legally, financially, and technically mature pan-European research infrastructure, ready-to-implementation, and ready-to-operation (ERIC legal status). The overall objective is to develop the organizational, operational, and strategic framework of METROFOOD-RI. In particular, the objectives are:

i) to organise the legal entity that will manage the future RI
ii) to define the operation and the operational standards at the level of the whole RI and for the National Nodes, as well as the role of the RI as a service-oriented organisation
iii) to define the long-term activities for the future RI and update the Strategic Research & Innovation Agenda (SRIA), in response to current and future challenges in the agri-food sector and for the society

METROFOOD-EPI is a continuation of METROFOOD-PP ("Preparatory Phase"), established with a main goal of beginning the operational phase in the implementation of METROFOOD-RI and addressing issues and bottlenecks identified in the final evaluation report of METROFOOD-PP.

The DMP has been defined to guarantee compliance and sustainability of the project with well-established requirements of the scientific data generated, collected and/or handled as well as with all data collected and handled for the planning, management, implementation and operation of the project.
The DMP addresses data and metadata to be stored, data archiving and preservation (long-term storage), policies for access and sharing. Details on what data are generated and collected, whether and how they will be exploited or made accessible for verification, use and re-use, and how they will be curated and preserved are given, with the aim to define a clear strategy for knowledge management and protection (IPRs). Different levels of data access are defined for different user groups, and strategies for ensuring IPRs of data owners and confidentiality of results in case of hazard identification are also included.
Furthermore, the first criteria towards the provision of a data quality assessment system to classify data entering the METROFOOD-EPI repositories are defined. This standardisation provides added

value because data retrieval and integration internally and externally to the infrastructure will be harmonised and standardised, ensuring and promoting interoperability and enabling users to access, use, mine, exploit and reproduce METROFOOD-EPI data.
This document is based on the EU Commission documents 'Guidelines on FAIR Data Management in Horizon 2020' [1] and 'Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020' [2] The document is aligned and should be considered in combination with the Grant Agreement (No. 101130162) and the Consortium Agreement of METROFOOD-EPI.
It has been designed as DMP for METROFOOD-EPI as "Early Phase Implementation" of METROFOOD-RI. The plans and the criteria for data management in the subsequent steps of the infrastructure (Implementation, Operation) are also considered. This DMP is intended to be a living document in which information will be made available on a finer level of granularity through updates as the implementation of the infrastructure progresses and when significant changes occur.

3. Data summary

3.1. Dataset and Formats

Different types of datasets will be generated and collected. Beside the use case datasets from WP4, some datasets are referred to aspects related to the business plan, financial documents, and the internal organization of METROFOOD-RI. The identified datasets for the METROFOOD-EPI project are listed in the following four tables. The formats of the datasets are listed in the column 'Format', while the column 'Access Mode' gives for each dataset an overview if the dataset will be public, protected to certain user groups or private to the consortium or owner. In certain cases, the term "mostly public" is used to express that some partners will not publish certain entries for various reasons. The default setting for data involved in the electronic infrastructure is "open access", but some data needs to be opted out. This mainly concerns data related to the business plan and the internal organisation. Details are described on section 4.3. The column 'Acquisition Type' indicates if the dataset will be generated by METROFOOD-EPI or will be collected from other sources.

Table 1 - e-RI datasets

Dataset	Description	Format	Access Mode	Acquisition Type
Information about METROFOOD-RI (incl. contacts)	General information about the METROFOOD-RI on the website	HTML	Public	Generated and Collected
Information about users		Database	Private	Generated and Collected
Analytes description, analytical techniques, determination, and literature	Information on parameters/analytes, analytical techniques, and literature (GAMA Wiki, a wiki created in different former projects including PRO-METROFOOD and METROFOOD-PP).	JSON (web service), HTML	Public	Collected
Available data on food contamination and composition for all food groups	A dataset containing globally available data on food contamination and composition.	JSON (web service), HTML and CSV	Public	Generated and Collected
Thesauri	A dataset containing controlled vocabularies for analytes, units, analytical techniques and so on, that should be used for data management to harmonise data format.	JSON (web service), HTML and CSV	Public	Generated and Collected
Physical infrastructure data	A dataset comprising all data related to the facilities offering a service within the Physical-RI from project partners (Beneficiaries and Third Parties).	Word, Excel, PDF, and HTML	Mostly public	Generated and Collected
ICT infrastructure data	A dataset with information about electronic facilities and available databases from	Word, Excel, and PDF	Mostly public	Generated and Collected

	project partners (Beneficiaries and Third Parties).			
Information about Services and associated data (incl. service chart)	Information about services that can be offered by project partners (Beneficiaries and Third Parties).	Word, Excel, PDF, and database	Mostly public	Collected from partners

Table 2 - Use cases (WP4) datasets

Dataset	Description	Format	Access Mode	Acquisition Type
Use Case 1	Characterization of fermentation processes and their bio-based derived products from the food industries	Currently not known	Private	Generated
Use Case 2	Pilots for food production and integrated analytical services	Currently not known	Private	Generated

Table 3 - METROFOOD-RI preparation datasets

Dataset	Description	Format	Access Mode	Acquisition Type
Financial data and information	Financial data and information used for preparing the cost book and business plan for METROFOOD-RI	Word, Excel, and PDF	Private	Collected and generated
ERIC setup, statutes, and policies	Generation of EU and national regulations for the ERIC set up and ERIC statutes and policies	Word and PDF	Private	Collected and generated
Inter-Ministerial Group Members	Data about IMG members and participants to the IMG meetings	Word, Excel, and PDF	Private	Collected and generated
Stakeholders	A dataset about stakeholders of METROFOOD-RI	Word, Excel, and PDF	Private	Generated and collected
Partners	A dataset about partners of METROFOOD-RI	Word, Excel, and PDF	Private	Generated and collected
Global supporting organisations	A dataset with contacts to organizations supporting METROFOOD-RI on a global level	Word, Excel, and PDF	Public and private	Collected
IP information	Data about IP aspects for the METROFOOD-RI consortium	Word, Excel, and PDF	Private	Generated and collected
List of strengths, weaknesses, opportunities, and threats	List of the main SWOT identified in relation to METROFOOD-RI	Word and PDF	Private	Collected and generated
Dissemination materials	List of dissemination materials available for METROFOOD-RI	PDF	Public and private	Generated and collected
Scientific performance data of the partners	Data about key performance indicators from the partner institutes and applicants for joining the consortium	Word, Excel, and PDF	Private	Collected

Set of official documents, regulations, and internal directives for the HR policy	Legal documents, internal official materials of participating institutions	Word, Excel, and PDF	Private	Collected
Set of regulations and internal directions for the open access policy	Legal documents, internal official materials of participating institutions	Word, Excel, and PDF	Private	Collected
Set of regulations and internal directives for the public tenders	Legal documents, internal official materials of participating institutions	Word, Excel, and PDF	Private	Collected
The list of KPIs	The dataset on KPIs and criteria for their evaluation	EXCEL	Private	Generated and collected

Table 4 - Project internal datasets

Dataset	Description	Format	Access Mode	Acquisition Type
Project proposal, project description, financial information etc.	Project proposal, DoA, information on the SEDIA portal, financial information, internal communication etc.	Word, HTML, e-Mail etc.	Private	Collected and generated
Conferences, seminars, fairs, and other events	List of events organised by METROFOOD-RI or with participation of METROFOOD-RI project participants	Word, Excel, and PDF	Public and private	Generated and collected
Dissemination materials	List of dissemination materials available for free download from the project website	PDF	Public and private	Generated and collected
List of potential risks	List of the potential risks that may impede the successful implementation of the project and the related mitigation actions	Word and PDF	Private	Collected and generated
Report on use cases	Information about participant companies and results of the use case	Word and PDF	Private	Collected and generated
Stakeholders	A dataset about stakeholders of METROFOOD-EPI	Word, Excel, and PDF	Private	Generated and collected

3.2. External Datasets

The acquisition type cannot only be assigned to datasets but needs to be considered for every data record within a dataset. The acquisition type "Generated" indicates that the data record origins directly from METROFOOD-PP/METROFOOD-EPI/METROFOOD-RI, whereas 'Collected' means that other sources such as publications, laboratory reports, online systems or other data sources were/will be used. To integrate external datasets, an agreement on access and integration must be found with the data owners. Preferred is the Creative Commons license "CC BY 4.0", which is the broadest open access license available and is further described in section 4.5. If such an agreement could be reached, data needs to be imported into one of the databases of METROFOOD. The details of the technical approach will be described in section 4.1. As an alternative, descriptive general

information are/will be given and a direct link to the main source will be provided. Currently, data related to the RI is spread all over Europe and beyond. The goal of METROFOOD-RI is to list all these datasets together in a web application on the e-RI. Many of the consortium partners are data owners and METROFOOD-RI aims to involve more data owners into the consortium or to define cooperation agreements / Memorandums of Understanding with other data owners, so no external datasets are needed. In any case, the origin of a data record must be provided to give information on data provenance, which is also a quality indicator.

The origin and re-use are described in the following table for external scientific datasets and for the METROFOOD-RI inventory datasets.

Table 5 - Data origin for METROFOOD-EPI

Dataset	Origin	Re-use
Analytes description, analytical techniques, determination, and literature	EuroFIR, EuroFIR Nexus, EuroFIR AISBL, PRO-METROFOOD, METROFOOD-PP	GAMA Wiki will be maintained.
Available data on food contamination and composition for all food groups	Collected from different project partners and brought together in PRO-METROFOOD and METROFOOD-PP.	Data is used in a search and compare web app.
Thesauri	From different EU projects and official organisation like EFSA, EuroFIR AISBL	Will be re-used for documentation and data exchange.
Reference material data	Directly from providers, which can be consortium or external partners	A search engine will be implemented to find available reference material.
Physical infrastructure data	These datasets are collected information from consortium partners (Tasks 3.2, 3.3).	Specification for apps
ICT infrastructure data		
Service data		

3.3. Data size

The expected sizes of the datasets for METROFOOD-EPI are listed in the following table and mainly depend on the size of the generated and/or collected information. Only scientific datasets are listed while project internal or METROFOOD-RI preparation datasets are not. The expected size of a dataset was estimated by the owning database manager and where not possible by a database expert within the consortium.

Table 6 - Data sizes

Dataset	Expected size in GB
Analytes description, analytical techniques, determination, and literature	0.20
Available data on food contamination and composition for all food groups	0.50
Reference material data	0.50
Thesauri	0.10
Physical infrastructure data	0.20
ICT infrastructure data	0.10

Services data	0.10
Results from electron microscopy analyses	0.10
Results from minimizing acrylamide levels in bakery products	0.10
Reference material data	1.00

3.4. To whom it might be useful

Users were identified and categorised at the early stages of the RI, in PRO-METROFOOD and in METROFOOD-PP, as part of the business plan and Stakeholder Analysis. Users were categorised into four groups, which are food business operators, policy makers/inspection & control agencies, research/academic, and consumers/citizens. Within each category, further user groups were identified:

i) researchers from universities and research institutes, National Metrology Institutes
ii) public and private laboratories
iii) policy makers, inspection, and control agencies
iv) Food Business Operators (private resources such as food producers, and company related to the food chain production, producer associations)
v) food consumers/citizens, consumer association
vi) IT Businesses

As part of METROFOOD-EPI the list will be revised and updated. The following list gives an overview of the most important user groups.

Table 7 - Usability for user groups

User Group	Why is it useful
Researcher from Universities and research institutes, National Metrology Institutes	Researchers need guidance in performing laboratory analysis and encyclopaedia for analytes, measurements and analytical results can be of high value for them (especially young researchers need also general information as a basis for their knowledge). Analysts and technicians (Lab. workers) also need information and data on metrological tools (reference material, reference and official methods, proficiency testing). Sharing of data and information will support the definition of best practices and will promote the integration of knowledge. Duplications of efforts will be avoided.
Public and private laboratories	Laboratories need reference materials for their daily business and must be able to quickly find and order them, as well as information about reference and official methods and proficiency testing, threshold values and guidelines on metrology. If a laboratory needs cooperation with other laboratories for certain analytes, METROFOOD-RI could be the platform to search for cooperation partners and to share data, procedures, and methodologies. Duplications of efforts will be avoided.
Policy makers, Inspection, and control agencies	National and European bodies can perform food analysis by using METROFOOD-RI where experts for different sectors and application fields are cooperating and broader analyses on raw materials, food products, agroecosystems of production, food contact materials, etc.

	are enabled. The organized data and information on metrological tools could be of great interest as well. Data and information provided by the e-RI will support policy makers and inspection and control agencies in permitting more efficient food controls and improving risk assessment and risk/benefit analyses.
Food Business Operators (FBO) (Private resources such as food producers, and company related to the food chain production, Producer associations)	FBO can get and share data and information on quality & safety of raw materials and products, food authenticity and traceability, agroecosystems of production and processes characteristics. This will permit them to improve quality & safety of their products, strengthening internationalisation and promoting trade and competitiveness. They will be facilitated also in communicating the added value of their products to consumers. Industrial sector needs quickly to find laboratories, where they can get assays for nutrients, contaminants, anti-nutrients etc. They need also to know which metrological tools are available in case an industrial partner has its own laboratory facilities.
Food consumers/citizens, consumer association	Food consumers can get information on food quality & safety, food authenticity and traceability, impact of domestic preparation/storage procedures and food contact materials, changing of components during cooking (loses and formation)
IT Businesses	Software and hardware providers might be interested in developing new software and hardware to be added to the METROFOOD-RI. These new apps will allow new ways of using the available scientific datasets.

4. FAIR Data

4.1. Architecture of the e-RI

The scientific datasets will be spread across project partners and external providers while some data will be stored on the central infrastructure of METROFOOD-RI. Therefore, it is a distributed architecture, and a central platform builds the core of the infrastructure. This architecture is similar to the architecture of the P-RI with different national laboratories and a central node. The architecture also facilitates access policy and Intellectual Property Rights (IPR) by keeping data on data owners' side. For further details please consult the corresponding deliverables.

4.2. Making Data findable, including Provisions for Metadata

Data models for several datasets were assembled in PRO-METROFOOD and METROFOOD-PP, and further elaborated in other projects. Where standard data models exist, these models are used while for missing standards new data models were generated. The data models do not only define entities and their relation but also define the attributes for the entities (metadata) and the thesauri that should be used. Among the data model standards are:

i) Standard Sample Description current version 2.0 from the European Food Safety Authority (EFSA) [4]
ii) BS EN 16104:2012 [5]
iii) EuroFIR proposal for structure and detail of a EuroFIR Standard on food composition data, which is based on BS EN 16104:2012 [6]

The entity attributes include value documentation as well as identification and description systems such as food groups or FoodEx2 for food items defined by EFSA and used by FAO/WHO. The attributes also include a unique identifier (primary key) which can be an auto-increment integer, or a Global Unique Identifier (GUID) offered by many database management systems. The Web Platform will have search facilities including identifiers and metadata.

Controlled vocabularies (thesauri) are defined for certain attributes in the data models describing available terms that must be used. Such an approach will eliminate potential errors while entering data into the system if users can only select an entry from the thesaurus. Thesauri also help in data exchange: both sides use the same pre-defined thesauri and are able to understand and interpret data. The general rule of thumb should be "to use as much thesauri as possible and useful". This will not only increase data quality and decrease the probability of potential errors but will also facilitate interoperability of the datasets. As thesauri evolve over time, each of them should also have a version number for identification.

Versioning is another concept that must be used for identification if datasets are evolving over time. For most of the datasets such as reference material or analytical results, a concept of versioning is not necessary as these data represents basic analysed data. However, when a need for versioning will come up, these datasets will also be versioned.

4.3. Making Data Openly Accessible

4.3.1. Open Access

Datasets that are generated and collected in this project are all used for METROFOOD-RI and the datasets which are open to the public will be made publicly available over different apps, wikis or repositories of the METROFOOD e-RI. The tables in section 3.1 shows that most scientific datasets

will be open access. The METROFOOD e-RI also guarantees the sustainability of data access.
In the following sub-sections, more details about open access are described.

4.3.2. Access Restrictions

The access modes to datasets are given in the tables under section 3.1. The tables show that METROFOOD-EPI and METROFOOD-RI follow the EU commission approach "as open as possible, as closed as necessary". Expressed in terms of a publication, this would mean that data should have Green Open Access. However, there are some datasets that will be partly or entirely opted out. The reasons are described in the following tables.

Table 8 - Access restrictions

Dataset	Reason
Results from electron microscopy analyses	A call will be launched for transnational access which is a feasibility study to test such remote services where users should submit a food sample and then can use the facility over remote tools. Depending on the conditions of how the food arrives at the facility and how good the remote analysis works, data can be published. But as this is only a feasibility study, several steps can go wrong and useless data can be produced, the data is classified as private.
Results from minimizing the acrylamide level in bakery products	The same argument applies here. It is only a feasibility study; several steps can go wrong and useless data can be produced. Therefore, the data are classified as private.
Physical infrastructure data	The goal is to collect data about physical facilities that can be publicly available. However, it may happen that some institutes deliver data which are not relevant to METROFOOD or even confidential. Such data will be excluded from publishing.
ICT infrastructure data	The goal is to collect data about physical facilities that can be publicly available. However, it can happen that some institutes deliver data, which are not relevant to METROFOOD or even confidential. Such data will be excluded from publishing.
Service data	The goal is to collect data about physical facilities that can be publicly available. However, it may happen that some institutes deliver data which are not relevant to METROFOOD or even confidential. Such data will be excluded from publishing.
Datasets for the preparation of METROFOOD-RI and internal project datasets	This is private information to the project and METROFOOD-RI to support their business activities.

The distributed architecture of the e-RI allows the data owner to restrict access to certain data if there are legal or internal restrictions. It allows them also to decide on embargo periods if they want to first

publish data before making them available online. The general policy is that data should be made publicly available as soon as possible.

If a dataset that is integrated in the central web application, has access restrictions for certain data and user groups, these requirements need to be reflected in the central web application by using user authentication and authorisation. Private datasets will also require user authentication and authorisation and can be stored in separate places.

4.3.3. Governance

A data management committee is needed to maintain this DMP, audit activities around data management and propose solutions for data management issues to the General Assembly. The committee will elaborate the details, including licensing, in agreement with data owners and bring a proposal to the consortium where final decision is taken. In particular, it needs to be checked if data owners have a procedure for data management that must be followed. The committee is a group of persons with knowledge in metrology, food and ICT and has no formal decision-making authority, and it may also be considered as a working group.

As planned in Task 3.1 and according to Milestone MS1 (Data Management Committee & Ethical Board established – Due date: Month1, January 2023 during the kick-off meeting (Freising, 24 January 2024) the following persons have been elected to the data management committee by the GA: WP3 leader Karl Presser (PMT), Francesca Zinni (ENEA), Matěj Božik (CZU), Montserrat Saladié (EUT), Sidney Tomé (INSA), Barbara Koroušić Seljak (JSI) and Denisa Duta (IBA). The consortium already agreed on the datasets in the grant agreement.

In general, each data provider is responsible for the data management of their own data. The data management committee is collaborating with WP3 and supports all data providers on the implementation for the e-RI. But it also collaborates with all other WPs where data is produced.

4.3.4. Long-term Preservation

METROFOOD-RI plans to maintain its own e-RI, which is sustainable and where data, metadata and public documents can be deposited and where the redundant deposition of information on other repositories will be avoided. Some available certified repositories where a cooperation or integration can be possible, will be considered in the early implementation phase of METROFOOD-RI (METROFOOD-EPI). The following list contains the evaluated repositories:

i) Metadata Directory of the Research Data Alliance, which allows to deposit metadata.
ii) EUDAT platform, which allows to share and store research data, metadata and publications and provides some tools to select the appropriate data license.
iii) Zenodo, which is OpenAIRE and CERN collaboration and allows to share and store research data and publications.

Private documents will be stored on a shared drive on the central node or on a cloud service where access will be restricted.

After the EPI project, all datasets will remain on the METROFOOD e-RI as METROFOOD-RI continues

towards the implementation and operation phases. The long-term preservation is therefore given by METROFOOD-RI. However, if METROFOOD-RI does not continue for some reason, the generated and publicly available datasets will be uploaded and deposited on Zenodo. The upload for FP7 and H2020 funded projects is free and a GitHub or ORCID login is needed.

4.3.5. Software for Data Access

Applications, wikis, or repositories are used to allow users and machines to access the data generated or collected during the project. The apps, wikis or repositories are either already existing and provided by project partners or will be developed in the project. Most important is that these tools provide user interfaces for machine-to-human interactions and APIs (Application Programming Interfaces) for machine-to-machine interactions which makes the data openly accessible. If these tools are not in place on the data owner side, data can alternatively be provided in files for download. All tools will have short manuals which are defined as deliverables and APIs will be documented. Data provider tools are either own developments or developed by third parties. The consortium strongly recommends adjusting these tools to make data FAIR and to have the tools documented.

All apps that will be developed in the project, should be at least open source to the consortium and to METROFOOD-RI. The consortium agreement states for innovations that are done by several partners that they should find an agreement on the exploitation. The project partners are encouraged to agree on open-source solutions and follow the FAIR principle. All apps should have a license that allows for later continuation and modification for METROFOOD-RI and proposed licenses are MIT, GNU GPL, or Apache License 2.0.
In order to facilitate the access to source code to all contributors, the source code should be deposited in one of the well-known and available repositories/tools with open access such as: GitLab (https://gitlab.com/), GitHub (https://github.com/) or Bitbucket (https://bitbucket.org/) where GitLab and GitHub are currently the preferred options.

Following the open-source strategy of the European Commission, the software development within the METROFOOD-EPI project should use open-source technology and components, preferably licensed under MIT, GNU GPL, Apache License 2.0 etc., so that continuation of that technology and component is foreseen. However, it is possible that not all the required software components can be implemented using only open-source libraries. In such a case, the corresponding software components need to be excluded if the software is going to be made open source. Partners who bring their software or system into the project must decide on their own if the software or system will be open source.

4.4. Long-term Preservation

Data produced in METROFOOD-EPI and overall, in METROFOOD-RI will be made interoperable, allowing data exchange, use and re-use between Consortium partners and between external users of the e-RI. The general policy is to adhere as much as possible to standards for formats, be compliant with available (open) software applications and thereby facilitating re-combinations with different datasets from different origins.

For data exchange within and outside METROFOOD-EPI and METROFOOD-RI, a format will be defined that is based on the data model in section 4.2. The exchange format can easily be derived from

the data model and JSON will be used as default format. Experiences in former projects shows that XML as exchange format is going to be large and they are difficult to handle [7], [8]. As JSON is a general-purpose file format, it can be particularly useful for third party tools, which therefore will be able to access data from METROFOOD-RI in a unified way. Each data provider needs to follow the thesauri and exchange format to be compliant to the METROFOOD-RI and get involved in the e-RI.

The data model is based on three different existing models as described in section 4.2 and the thesauri will also be based on these existing models as well as on other official vocabularies. The thesaurus of EFSA is taken for contaminants while for others such as metrological concepts, units and matrix units, the International Vocabulary of Metrology (VIM) [9] together with the ones from EuroFIR and TDS-Exposure will be used, which are based on ISO definitions. Where necessary the existing thesauri will be extended.

A data tagging concept will be used to enrich data with semantical meaning. The above-mentioned existing models and standards, exchange format definition, thesauri, and the data tagging concept can be considered as ontology. METROFOOD-EPI and METROFOOD-RI will start working on these concepts and will consider an ontology if the concepts are proved to be insufficient.

4.5. Increase Data Re-use (through clarifying Licences)

Public datasets in the METROFOOD-EPI project should be licensed under the Creative Commons license 'CC BY 4.0' to maximise usability. This license allows users to share and adapt data even for commercial use. The licensor cannot revoke these freedoms as long as the user follows the license term. The same standard should be followed in METROFOOD-RI expect for restricted datasets where other licenses will be used. As not only generated data but also external data will be among the datasets, the data access committee needs to check with data owners if their data can run under this license or find a solution otherwise.

As a rule, all data should be available as soon as possible. If a dataset was generated within METROFOOD-EPI, the authors can claim for an embargo until the dataset is published. An embargo should be justified to data management committee and an agreement for the embargo period must be found between the authors and the committee. A period between 6 and 12 months is considered to be normal. If METROFOOD-EPI is not the owner of generated data, the owner decides on the embargo period.

For the METROFOOD-EPI and METROFOOD-RI datasets, a Creative Commons data license will be used. Creative Commons has the advantage that besides the normal license, a non-lawyer and a machine-readable version of the license are available. The non-lawyer version is a simplified version that summarises the content of the license in an understandable manner. The machine-readable license is targeted on automated data exchange.

4.6. Data Quality Assurance

Data quality assurance is a multilevel approach that consists of data prevention, data evaluation and data analysis [10].
The quality prevention part checks each new data record at entry time into the system and can give data quality feedback as soon as possible to the user. This part is well known in many applications

where users are not able to proceed before they have entered all mandatory data or certain data in a certain format. Once data have entered the database, they need to be checked by a second person before used and published. A third step of automated data quality analysis is needed to backup human data checks and to check requirements that cannot be checked at entry time. This is done by an on-going automated routine that is checking data in the system and reports issues back to super-users. The combination of these three steps maximises data quality assurance while moving data quality checking from humans to computers.
The set of automated checks needs to be defined. Examples of such checks are duplication checks, range checks, mandatory information checks. As the architecture of the e-RI is distributed, data provenance is getting more important. The above-mentioned approach should not only be applied at each data owner side but also overarching data quality checks should be performed. With increasing size of the datasets this task can be time consuming.

In addition, organisational procedures must be introduced to define how data are stored, accessible, manipulated, collected, and published. This is done with a set of Standard Operating Procedures (SOPs) describing each step of data management.

5. Allocation of Resources

The costs to make the available datasets FAIR can be evaluated as a part of the costs allocated for the implementation of data publishing tools. The grant funding covers these costs. The apps that are planned in this project are designed to make data publicly available and therefore the costs to make data FAIR were included during the proposal.

Concerning METROFOOD-RI, the costs for software and hardware maintenance must be included in the cost for the RI. METROFOOD-RI will generate income through membership and paid services of the physical and electronical infrastructure, as well as projects and conferences. These incomes need to cover expenses including the ones for making data FAIR and long- term preservation. The Partners involved in the e-RI will provide further support as in-kind contribution. Details will be covered in the business plan and in the cost book and business plan.

6. Data Security

Proper solutions for secure and efficient access and exchange of data will be implemented, as they are essential for building an e-RI. As METROFOOD-EPI has a distributed architecture, data providers have the control over their data access. There are several security steps that should be taken for the data providing systems:

i) Access to restricted data or modification of data is only allowed to authenticated and authorised users.
ii) The communication over the internet should use HTTPS (SSL) and secured interfaces such as OAuth to get to the level of Internet banking.
iii) Physical servers should be in access restricted and access-controlled server rooms and use mirrored hard drive setups.
iv) For each database, a backup should be done every night and keep monthly backups for a longer time. The preferred solution is to have daily backups for the last 30 days, monthly backups for the last 12 months and yearly backup for the last 5 years. The server where the backup files are stored should follow the same security steps.
v) Software server such as database or web servers will also need authentication and authorisation for users and are access restricted to administrators.
vi) All software used like DBMS will also have authentication and authorisation for users and are access restricted to administrators.
vii) User password should not be stored plaintext in the database but encrypted by MD5, SHA family or other asymmetric or one-way encryption.
viii) Firewalls are used to protect ports of the servers.
ix) Intrusion detection or intrusion prevention system should be used.
x) Systems should use a logging mechanism to register what, how and when data were modified.

Providers of sensitive data will maybe have additional requirements that will be agreed with the data management committee and will be implemented on a case-by-case basis.

7. Ethical Aspects

Ethics issues on data could concern specific types of data, that can be classified as “personal data” and/or can be referred to “human participation”. This is for example for food consumption data (that are “personal data”) and concerns further processing of previously collected personal data (secondary use). In fact, food consumption data could be processed and combined with food analyses data. Any used food consumption data will be collected under the Declaration of Helsinki (version applicable at the time of data collection) and national legislation regarding medical-ethical studies. Any used food consumption data will be anonymized before integration into the e-RI. All the necessary procedures to protect these data and their confidentiality will be put in place. Moreover, issues on both “personal data” and “human participation” are related to relations with users and stakeholders and their direct involvement in the activities related with planning and strategy definition (e.g. stakeholder analysis, user strategy and access policy), that started in the “Early Phase” (PRO-METROFOOD project) and “Preparatory Phase” (METROFOOD-PP project) and continue in the next phases. This refers to the definition of the “User Database” and to all the activities foreseeing surveys and/or focus groups. Any personal data and information, as well as contact details of the eventual direct interlocutors, will be handled (stored, protected and maintained) according to EU data protection rules (as regards name, affiliation, contact details, etc. for all participants in surveys, workshops, etc.) and anonymised for any further utilisation. METROFOOD-EPI partners will not make public any information that could be directly related to an individual.
Ethical guidelines for METROFOOD-EPI are further described in D1.1.

The project will also follow the EU legislation: All data will be collected and handled (stored, protected and maintained) according to the Data Protection Directive (EC Directive 95/46) [11] on the protection of individuals with regard to the processing of personal data and on the free movement of such data, updated on 25 January 2012 (European Commission draft European General Data Protection Regulation that will supersede the Data Protection Directive) and the National Laws. If necessary, informed consent procedures will be put in place. Also, the WMA Declaration of Helsinki on ethical principles for medical research involving human subjects, including research on identifiable human material and data [12] will be incorporated.

8. References

[1] European Commission, Directorate-General for Research & Innovation (2016), Guidelines on FAIR Data Management in Horizon 2020, Version 3.0, 26.07.2016
[2] European Commission Directorate-General for Research & Innovation (2017), Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, Version 3.2, 21 March 2017
[3] Presser K. et al. (2012), Software Requirement Specification of FoodCASE-Risk, TDS-Exposure project, grant no. 289108, Deliverable 6.2
[4] EFSA (European Food Safety Authority) (2013), Standard Sample Description ver. 2.0, EFSA Journal 2013, 11(10):3424, 114 pp., doi: 10.2903/j.efsa.2013.3424
[5] BS EN 16104:2012 (2013), Food data. Structure and interchange format
[6] Becker W. et al (2008), Proposal for structure and detail of a EuroFIR Standard on food composition data, II. Technical Annex, EuroFIR TECHNICAL REPORT 1.2
[7] Møller A. et al (2012), EuroFIR Web Services - EuroFIR Food Data Transport Package, Version 1.4, EuroFIR Nexus Technical Report D2.1
[8] Pakkala H. et al (2009), EuroFIR Web Services – Specification of Request-Response Message Exchange Patterns, Version 1.2, EuroFIR Technical Report D1.8.29
[9] Joint Committee for Guides in Metrology (2012), International vocabulary of metrology – Basic and general concepts and associated terms (VIM), JCGM 200:2012
[10] Presser K. (2012), A Requirement-Oriented Data Quality Model and Framework of a Food Composition Database System, Ph.D. thesis ETH Zurich, doi: http://dx.doi.org/10.3929/ethz-a-007605248
[11] European Commission (1995), Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, and amendments
[12] The World Medical Association (WMA) (2013), Declaration of Helsinki - Ethical principles for medical research involving human subjects, including research on identifiable human material and data. First version, 1964. Last update, Oct. 2013

List of abbreviations

API – Application Programming Interface
CC – Creative Commons
CH – Central Hub
CSV – Comma-separated values
CZU - Ceska Zemedelska Univerzita v Praze
D – Deliverable
DBMS - Database management system
DMP – Data Management Plan
DoA - Description of the action
EFSA – European Food Safety Authority
ENEA - Agenzia Nazionale per le Nuove Tecnologie, L'energia e lo Sviluppo Economico Sostenibile
E-RI – Electronic Research Infrastructure
ERIC - European Research Infrastructure Consortium
ESFRI - European Strategy Forum on Research Infrastructures
EU – European Union
EUT - Fundacio Eurecat
FAIR - findable, accessible, interoperable, and re-usable
FAO – Food and Agriculture Organization
FBO – Food business operator
FP7 – Framework Programme 7
GA – Grant Agreement
GB – Gigabytes
GNU GPL – GNU General Public License
GUID – Global Unique Identifier
H2020 – Horizon 2020 Framework Programme
HTML - HyperText Markup Language
HTTPS – Hypertext Transfer Protocol Secure
HR – Human resources
IBA - Institutul National de Cercetare-Dezvoltare Pentru Bioresurse Alimentare
ICT – Information and communications technology
INSA - Instituto Nacional de Saude Dr. Ricardo Jorge
IP – Intellectual Property
IPRs – Intellectual Property Rights
ISO - International Organization for Standardization
IT – Information Technology
JSI - Institut Jozef Stefan
JSON – JavaScript Object Notation
KPI – Key performance indicator
MD5 - Message Digest Algorithm 5
METROFOOD-EPI – METROFOOD-RI Early Phase Implementation
METROFOOD-PP – METROFOOD-RI Preparatory Phase
METROFOOD-RI – METROFOOD Research Infrastructure
MIT – Massachusetts Institute of Technology
MS – Milestone
NN – National Node
OAuth – Open Authorization
P-RI – Physical Research Infrastructure
PDF – Portable Document Format
PMT - Premotec
PU – Public
RI – Research Infrastructure

RM – Reference Material
SEDIA – Single Electronic Data Interchange Area
SHA - Secure Hash Algorithm
SOP – Standard Operating Procedures
SRIA – Strategic Research & Innovation Agenda
SSL – Secure Sockets Layer
SWOT – Strengths, weaknesses, opportunities and threats,
TDS – Total Diet Studies
VIM – International Vocabulary of Metrology
WHO - World Health Organization
WMA – World Medical Association
WP – Work Package
XML - Extensible Markup Language