Deliverable 6.1

FOODSAFER

Project Management and Pathway to impact and participation

#EVERYBODY'S BUSINESS

Deliverable D6.1

Data Management Plan

Work package number and title	*WP6: Project Management and Pathway to impact and participation*
Lead-beneficiary	*IRIS Technology Solutions*
Work package Leader	*Austrian Competence Centre for Feed and Food Quality, Safety and Innovation (FFoQSI)*
Contributors to deliverable	*IRIS TECHNOLOGY SOLUTIONS* *FFoQSI- Austrian Competence Centre for Feed and Food Quality, Safety and Innovation* *FSAI- Food Safety Authority of Ireland* *PIWET- National Veterinary Research Institute* *BARILLA* *BIOMIN* *BIOSENSE INSTITUTE* *BFR- German Federal Institute for Risk Assessment* *GHENT UNIVERSITY* *WAGENINGEN UNIVERSITY & RESEARCH* *AUSTRIAN INSTITUTE OF TECHNOLOGY*
Relevant Task	*Subtask 6.4.2 Data Management*
Participants	*All participants*
Dissemination Level	*Public*
Due Date (month)	*M06*

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101060698

Action Number: 101060698

Action Acronym: FOODSAFER
Action title: A JOINED-UP APPROACH TO THE IDENTIFICATION, ASSESSMENT AND MANAGEMENT OF EMERGING FOOD SAFETY HAZARDS AND ASSOCATED RISKS

Date: 08.05.2023

DMP version: V1.2

Prepared by: IRIS Technology Solutions

History of changes		
Date	**Version**	**Changes**
11.01.2023	V1.0	First template
09.02.2023	V1.1	Introduction of Tables for gathering partner input
08.05.2023	V1.2	First complete iteration

Contents

Introduction

Data Management is an integral part of the research process for our project. Indeed, the need for good research data management practices is becoming more recognized as a critical part of research[1]. The European Commission, established a pilot project under the Horizon 2020 program, called Open Research Data Pilot (ORD pilot)[2]. In Horizon Europe the creation of a Data Management Plan is a mandatory process in all research projects.

A data management plan (DMP) is necessary for several reasons, as it can help:

- To ensure that research data is properly organized, preserved, and accessible for the long-term. This is important for reproducibility and transparency in research, and it also helps to prevent data loss.
- With compliance during the project with Horizon Europe requirements for data sharing and preservation, and to ensure that we have a robust approach to managing and sharing the data generated by our project.
- To minimize legal and ethical issues related to data collection, storage, and sharing. For example, providing guidance in the protection sensitive or confidential data, and to obtain informed consent from research participants.
- To make the most of the data genetated by facilitating data reuse and collaboration. With a well-designed DMP, we can more easily share our data with others and incorporate data from other sources into our own work.

For the purpose of drawing up our DMP, we have based our methodology around the recommended template provided by the European Commission (EC) in order to manage the process and to arrive at a robust plan.

Our starting point for this plan is the information provided in the Description of the Action (DoA), whereby we provided a table (see Table 1 below which has been adapted from the outline Data Management Plan (DMP) provided at proposal stage) as an overview of how data/ research outputs will be managed in line with the FAIR principles (Findable, Accessible, Interoperable, Reusable).

Table 1- Summary of data management and management of other research outputs in keeping with FAIR principles

Types of data/research outputs and their estimated size:	Description of the types and formats of data (experimental, observational, images, text, numerical) the project will generate/collect (e.g. indicators of stress (measurable factors) and its combination and provenance of existing data; The origin of the data; Expected size of the data; From whom it might be useful ('data utility'). The data used in food safety ranges from unstructured to highly structured data and are stored as documents in various formats (e.g. txt, JSON) and databases (NoSQL, SQL, MongoDB database- a flexible relational database), as well as personal data of e.g. participants in the virtual Living Labs (which will be GDPR compliant).
Findability of	Make the data readily findable through trusted repositories, and we will define

[1] Johanne Medina, Abdul Wahab Ziaullah, Heesoo Park, Ivano E. Castelli, Arif Shaon, Halima Bensmail, and Fedwa El-Mellouhi (2022). Accelerating the adoption of research data management strategies. Matter 5, 3614–3642, November 2, 2022 3639

[2] European Commission. Open research data pilot in horizon 2020. https://ec.europa.eu/research/participants/docs/h2020-fundingguide/cross-cutting-issues/open-accessdatamanagement/data-management_en.htm

data/research outputs:	types of persistent and unique identifiers (e.g. digital object identifiers).
Accessibility of data/research outputs:	Make the data openly accessible in keeping with robust IPR management (data that is linked to IP will be available after IP protection mechanisms are in place), with provisions for providing controlled access to restricted data for verification purposes.
Interoperability of data/research outputs:	Ensure the interoperability of the data through standards, format and vocabularies for data and metadata, e.g. the standards developed by European and international standardization organizations like CEN and ISO), Working Groups of the Global Harmonization Initiative (GHI), as well as CODEX Alimentarius (the latter is a member of our Advisory Board).
Reusability of data/research outputs:	Increase data re-use (through clarifying licenses). Where relevant we will allow data sharing and reuse through Creative Commons and Open Data Commons licenses. Use of tools/ software/ models for data generation and validation/interpretation /re-use.
Curation and storage/preservation costs:	Through *Subtask 6.4.2* Data Management work is performed under the leadership of an experienced professional, Prof. David Rodríguez-Lázaro[3]- who has the expertise, skills and competence to perform this role of data management and quality assurance. Moreover, all partners have allocated resources (personnel involvement and costs for making data FAIR) in their budgets to participate in Data Management under the guidance of the Data Manager. In relation to long-term preservation of the data we have budgeted a reasonable amount for this to contribute to long term preservation.
Data Security:	Data recovery as well as secure storage and transfer of sensitive data.
Regulatory and Ethical aspects:	Ensure full compliance with GDPR (General Data Protection Regulation) for any personal data gathered, as relevant for our Community registrations to the Digital Hub. Scientific integrity will be respected during data collection, data sharing & collaboration amongst the different actors in & out the consortium.

The present Data Management Plan (Deliverable 6.1) builds on this outline, and serves as a living document that will be updated as the implementation of the project progresses and when significant changes occur.

Over the course of the first 6 months of the project, we have worked on creating an agile process that would enable and facilitate the participation of all partners. This collaborative process has culminated in this first version of the DMP, and which will be continually updated and iterated upon as the project evolves and develops, to ensure that we give due consideration during our project to the aspects of **creation, storage, sharing, discovery and destruction of data** created over the course of the project, and that we have a robust approach in relation to processes and procedures for handling and managing data and in keeping with **current legislative requirements and best practice guidelines**, including those relating to **GDPR compliance**. Moreover, the goal is to ultimately ensure that FOODSAFER contributed to sharing research data and other outputs following 'FAIR' data management principles (in accordance with the guidance from the EC): **Findable, Accessible, Interoperable and Reusable**.

Methodology

This present DMP (*D6.1 Data Management Plan*) is a consortium effort whereby all partners have been involved, and each partner has completed the table provided in an Excel template that was circulated by the Data Manager (IRIS[4]) and it will be continually updated as part of the ongoing efforts in *Subtask 6.4.2 Data Management*. As such, each partner has been tasked with providing an overview of the data that will be created or collected within as part of their work streams. The FOODSAFER DMP essentially specifies what data is generated by the project (purpose of the data collection/generation and its relation to the objectives of

[3] This role has since been taken over by the Digital Team at IRIS
[4] IRIS has agreed to take over the role of Data Management from partner UBU

the project), whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved to ensure our data/research outputs are findable, accessible, interoperable and reusable.

Based on the data management template provided by the European Commission[2], the FoodSafeR data management team worked on creating a table for facilitating the process of gathering input from each partner.

The table (which was circulated as an Excel file to all partners in Febraury 2023), along with a simple guidelines document, contained the following comprehensive sections:

PRELIMINARY INFORMATION:

- Partner responsible for the data
- WPs or Tasks in which the data is generated
- WPs or Tasks in which the data will be used

GENERAL INFORMATION REGARDING DATA & DATASETS:

- Dataset name
- Data Set Description & relation to project (purpose of the dataset, procedures for collection, storage, protection, preservation, destruction, etc.)
- Data source (generated, collected or reuse of existing data).
- Is the data generated from a task in the project or an external source ?
- Type of data (numerical, categorical, textual, spectral, images, videos, combination of previous types)
- Is the data real, synthetic or hybrid ?
- Category type (e.g. experiment data, publication, personal data, environmental data, etc.)
- Metadata & Standards (Which ones will be used? Will a new one be created?)
- Does documentation exist for the data? (description of inputs/outputs, taxonomy for categories, etc.)

DATA FEATURES:

- Formats that will be used/generated/collected? (e.g.csv, .xlsx, .txt, .xml, .json, .html, .jpeg, .doc…)
- Requirements of Methods/Software for data access
- Size estimation (Mb/Gb min/max) if a dataset has been collected

ACCESSIBILITY OF DATA & DATASETS:

- What is the access level? (Public / Restricted)
- Will the data/metadata be stored in a trusted repository (e.g. Zenodo)? Or others (e.g. Google Drive, Teams folder, sharepoint, etc.)?
- Provide a justification if some data/metadata cannot be shared.
- Mention any Ethical/Legal issues that impact on sharing

HOW/WHICH DATA IS GENERATED:

- Methodology applied
- Data from sensors or industrial processes?
- Variables (initial list of inputs and outputs or KPIs)
- Expected data quality and coverage (high, medium, low)
- Any special information to comment in relation to data

TIMINGS

- When the data will be generated (time range in project, eg. M6-M12)
- When the data will be needed (time range in project, eg. M6-M18)
- How long will the data remain available and findable?

OTHER

- How will the data be licensed to permit the widest re-use possible?
- Are there any associated costs?

All completed tables have been complied into a single repository, which will not be referenced to, implemented and updated over the remainer of the project. These tables are presented in the Annex.

The FoodSafeR Open Digital Hub

The FoodSafeR Open Digital Hub will offer a range of features and tools to support professionals from the food safety system in the identification of emerging food hazards. In relation to access to data and information, our goal is to provide easy access to relevant data and information related to emerging food hazards, including databases of scientific research, and other relevant sources of information and data, which are currently being explored, such as:

- Public health and safety data (data sources related to public health and safety, such as disease outbreaks, foodborne illness reports, and recalls of contaminated food products).
- Environmental data (data sources related to environmental factors that affect food safety, such as weather patterns, pollution levels, and water quality).
- Supply chain data (data sources related to the supply chain of food products, such as production and distribution records, logistics data, and traceability information).
- Consumer data (e.g. social media analytics…)
- Regulatory data: data sources related to food safety regulations and standards, such as the European Union's Rapid Alert System for Food and Feed (RASFF) and the European Food Safety Authority's (EFSA) scientific opinions.

Connecting to these types of data sources would allow FoodSafeR to provide a more comprehensive and up-to-date view of emerging food hazards and associated risks, and ultimately helping risk managers and risk assessors from the EU food safety system.

Consultation with the FoodSafeR Advisory Board has to date resulted in the following list of suggested data and information sources that could be connected to the FoodSafer Hub (and which are currently in the process of being analysed):

- EU Food Safety Forum organized by FS4EU
- EFSA PRIMo
- EFSA consumption database
- Combase
- MicroHibro
- Epipulse data on foodborne outbreaks
- RASFF (Rapid Alert System for Food and Feed)
- EFSA/ECDC WGS platforms
- European Open Science Cloud
- European Health Information Gateway
- EFSA evaluations
- Food safety agencies data
- FAO has a wide range of well-researched publications on food commodities and associated risks, and on food foresight
- USDA-ARS-Food Safety Research Information Office based at the National Library
- USDA-ARS ComBase
- Interagency Risk Assessment Consortium (IRAC) a network of U.S. federal agencies with responsibilities or interests in developing food-safety risk-assessment tools or conducting or using food-safety risk assessments
- Government of Canada Open Data Portal
- CAN LINE
- WHO GEMS food database
- JEMRA
- JECFA
- JMPR

- FAO scientific advice
- WHO's Epidemic Intelligence System
- ProMed curated horizon scan
- Other commercially available horizon scan platforms
- Integration of existing databases of chemical hazards such as the NORMAN suspect exchange list.
- WHO FOSCOLLAB (the data sources are connected to evaluations of the JEFCA, JMPR GEMS food databases)

In relation to data privacy and security aspects for the FoodSafeR Open Digital Hub, the following strategy is being designed:

- Platform hosted on Google Cloud in Germany: Cloud SQL customer data is encrypted when stored in Databases, temporary files and backups.
- Secure Data Transfer protocol: to ensure the encryption of data during transit (using HTTPS or SSL/TLS protocols to secure data transfer).
- Access Control: password policies and role-based access control
- Data Encryption: Data at rest encrypted using robust encryption algorithms, and private keys kept safe and secured.
- Monitoring and Alerting: continuous monitoring of the system for any suspicious activities or potential security breaches.
- Incident Response Plan: In case of a security breach, the platform will have an incident response plan that outlines the steps to be taken to mitigate the damage and restore normal operations.
- Data Backup and Recovery: Data regularly backed up (daily) & a disaster recovery plan in place in case of any data loss or system failure.
- Data Privacy: robust data privacy and security measures to protect users' personal information and ensure compliance with relevant data protection laws (GDPR); privacy policy and cookies policy

Further actions

Our aim is also to desposit FoodSafeR research data in OpenAIRE's trusted repository hosted by CERN: **Zenodo** (https://zenodo.org).

Zendod is an open repository for all scholarship, enabling researchers from all disciplines to share and preserve their research outputs, regardless of size or format. As it is free to upload and free to access, it will make our data and scientiic outputs citable, shareable and discoverable for the long term. The intention is to create and curate our own community for the project. In Zenodo, all uploads get a Digital Object Identifier (DOI) to make them easily and uniquely citeable. All open content is harvestable via OAI-PMH by third parties. The use of Zenodo will make our data and scientific outputs citable, shareable and discoverable for the long term.

Data is stored in the CERN Data Center, where it is stored safely for the future in same cloud infrastructure as research data from CERN's Large Hadron Collider and using CERN's battle-tested repository software Invenio, which is used by some of the world's largest repositories such as INSPIRE HEP and CERN Document Server.

In FOODSAFER we will ensure full compliance with GDPR (General Data Protection Regulation) for all personal data gathered during the implementation of the project activities, such as survey's (WP1, WP5), personal data that may be gathered as part of the registration process to the FOODSAFER Open Digital Hub, as well as in the user profiles section (WP4), dissemination activities and management files records (WP6). Moreover, scientific integrity will be respected during data collection, data sharing & collaboration amongst the different actors in & out the consortium across all WPs. For any research performed with individuals over the course of the project (surveys, etc.) an informed consent process will be followed.

Conclusions

This document has described the FOODSAFER Data Management Plan to be used in the project. It includes a detailed overview of the data that will be created, processed or utilized within the FOODSAFER project with details on the type and nature of the data involved in keeping with its relation of the FOODSAFER data to the project objectives. The structured approaches of FOODSAFER in ensuring that data management adheres to the FAIR data principles defined by the EC are also provided. An updated version of this DMP will be prepared to coincide with the mid-term of the project

Annex: Partners contributions to FoodSafeR data management

DATA MANAGEMENT TABLES COMPLETED BY EACH PARTNER

STATUS AT M6

Project Acronym	FOODSAFER
Project Title	A JOINED-UP APPROACH TO THE IDENTIFICATION, ASSESSMENT AND MANAGEMENT OF EMERGING FOOD SAFETY HAZARDS AND ASSOCIATED RISKS
GA nº	101060698
Coordinator	Linked to D6.1 Data Management Plan
Related Work Package	WP6- Project Management and Pathway to impact and participation
Related task	Subtask 6.4.2 Data Management-
Author	IRIS
Contributors	All

Data is a set of values or observations that can be processed, analyzed, and used to draw conclusions or support decision making. It can come in many forms, such as numbers, text, images, or sound.

A **dataset** is a collection of data that is organized and formatted in a specific way, so that it can be processed and analyzed. Datasets can come in various forms, such as tables, spreadsheets, or databases.

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. It can include descriptive information such as title, abstract, author, and keywords, as well as administrative information such as creation date, file type, and size.

PRELIMINARY INFO			GENERAL INFO REGARDING DATA & DATASETS							
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Dataset name	Data Set Description & relation to project: purpose of the dataset, procedures for collection, storage, protection, preservation, destruction, etc)	Data source (generated, collected or reuse of existing data). Is the data generated from a task in the project or an external source ?	Type of data (numerical, categorical, textual, spectral, images, videos, combination of previous types)	Is the data real, synthetic or hybrid ?	Category type (e.g. experiment data, publication, personal data, environmental data, etc.)	Metadata & Standards (Which ones will be used? Will a new one be created?)	Does documentation exist for the data? (description of inputs/outputs, taxonomy for categories, etc.)
FFoQSI	WP2.3	WP2.3	Listeria WGS	Strains and whole genome sequence data from *Listeria monocytogenes* isolated from food processing facilities across partner countries.	Internally generated	numerical, categorical, textual, images	Real	Experimental Data		Yes
FFoQSI	WP2.3	WP2.3	Facility Microbiome	Amplicon sequence data of microbial communities co-occurring with *Listeria monocytogenes* in food processing facilities across partner countries.	Internally generated	numerical, categorical, textual, images	Real	Experimental Data		Yes
FFoQSI	WP2.3	WP2.3	Biofilm Transcriptome	Transcriptomic data generated from biofilm experiments.	Internally generated	numerical, categorical, textual, images	Real	Experimental Data		Yes

PRELIMINARY INFO			DATA FEATURES			ACCESSIBILITY OF DATA & DATASETS			
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Formats that will be used/generated/collected? (e.g.csv, .xlsx, .txt, .xml, .json, .html, .jpeg, .doc…)	Requirements of Methods/Software for data access	Size estimation (Mb/Gb min/max) if a dataset has been collected	What is the access level? (Public / Restricted)	Will the data/metadata be stored in a trusted repository (e.g. Zenodo)? Or others (e.g. Google Drive, Teams folder, sharepoint, etc.)?	Provide a justification if some data/metadata cannot be shared.	Mention any Ethical/Legal issues that impact on sharing
FFoQSI	WP2.3	WP2.3	.fasta, .csv, .jpeg, .doc, .xlsx	Microsoft Office, open source programs for data analysis	Gb	Public upon publication	Deposited in NCBI sequence repository	Confidentiality of sensitive information, including facility names where samples were collected	
FFoQSI	WP2.3	WP2.3	.fasta, .csv, .jpeg, .doc, .xlsx	Microsoft Office, open source programs for data analysis	Gb	Public upon publication	Deposited in NCBI sequence repository	Confidentiality of sensitive information, including facility names where samples were collected	
FFoQSI	WP2.3	WP2.3	.fasta, .csv, .jpeg, .doc, .xlsx	Microsoft Office, open source programs for data analysis	Gb	Public upon publication	Deposited in NCBI sequence repository		

PRELIMINARY INFO			HOW/WHICH DATA IS GENERATED					TIMINGS			OTHER	
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	lology applied	Data from sensors or industrial processes?	Variables (initial list of inputs and outputs or KPIs)	Expected data quality and coverage (high, medium, low)	Any special information to comment in relation to data	When the data will be generated (time range in project, eg. M6-M12)	When the data will be needed (time range in project, eg. M6-M18)	How long will the data remain available and findable?	How will the data be licensed to permit the widest re-use possible?	Are there any associated costs?
FFoQSI	WP2.3	WP2.3	Facility swab sampling, Microbiological enrichment culture, DNA extraction, Whole Genome Sequencing			High		M6-M36	M6-M48	Permanently	No specific restriction. Data will be deposited in repositories for public access.	Within the budget of the tasks/resources in the project.
FFoQSI	WP2.3	WP2.3	Facility swab sampling, DNA extraction, 16S/ITS amplicon sequencing			High		M6-M36	M6-M48	Permanently	No specific restriction. Data will be deposited in repositories for public access.	Within the budget of the tasks/resources in the project.
FFoQSI	WP2.3	WP2.3	Facility swab sampling, RNA extraction, sequencing			High		M8-M40	M8-M48	Permanently	No specific restriction. Data will be deposited in repositories for public access.	Within the budget of the tasks/resources in the project.

PRELIMINARY INFO			GENERAL INFO REGARDING DATA & DATASETS							
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Dataset name	Data Set Description & relation to project: purpose of the dataset, procedures for collection, storage, protection, preservation, destruction, etc)	Data source (generated, collected or reuse of existing data). Is the data generated from a task in the project or an external source ?	Type of data (numerical, categorical, textual, spectral, images, videos, combination of previous types)	Is the data real, synthetic or hybrid ?	Category type (e.g. experiment data, publication, personal data, environmental data, etc.)	Metadata & Standards (Which ones will be used? Will a new one be created?)	Does documentation exist for the data? (description of inputs/outputs, taxonomy for categories, etc.)
AIT	Task 2.1	WP4 & WP5		overview of microbiological and production data collected within the task	generated	Combination of textual and numerical variables		experimental data	we will define a metadata structure and SOP in the WP (SOPs will be partially based on existing standards like ISO norms)	not yet
AIT	Task 1.1	Task 1.1 & 1.3		Overview of trends, drivers and barriers in the food system	generated	combination of textual and numerical variables		Desk research	metadata structure will be developed	Not yet, glossary will be developed

PRELIMINARY INFO			DATA FEATURES			ACCESSIBILITY OF DATA & DATASETS			
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Formats that will be used/generated/collected? (e.g.csv, .xlsx, .txt, .xml, .json, .html, .jpeg, .doc...)	Requirements of Methods/Software for data access	Size estimation (Mb/Gb min/max) if a dataset has been collected	What is the access level? (Public / Restricted)	Will the data/metadata be stored in a trusted repository (e.g. Zenodo)? Or others (e.g. Google Drive, Teams folder, sharepoint, etc.)?	Provide a justification if some data/metadata cannot be shared.	Mention any Ethical/Legal issues that impact on sharing
AIT	Task 2.1	WP4 & WP5	vision is to have an overview in an Excel table format; individual assays will generate different data formats as listed below	Microsoft Office or similar	kb range for overview file(s)	TBD	For the start at the project's sharepoint	needs to be kept confidential till publication, with publication it will have to be made public anyyway	none known; however if we end up sampling within production facilities complanies might request that some data is kept confidential
AIT	Task 1.1	Task 1.1 & 1.3	.xlsx	Microsoft Office or similar		TBD	project sharepoint	research in progress - public availabilty upon publication	none

PRELIMINARY INFO			HOW/WHICH DATA IS GENERATED					TIMINGS			OTHER	
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	dology applied	Data from sensors or industrial processes?	Variables (initial list of inputs and outputs or KPIs)	Expected data quality and coverage (high, medium, low)	Any special information to comment in relation to data	When the data will be generated (time range in project, eg. M6-M12)	When the data will be needed (time range in project, eg. M6-M18)	How long will the data remain available and findable?	How will the data be licensed to permit the widest re-use possible?	Are there any associated costs?
AIT	Task 2.1	WP4 & WP5	TBD as in depends on the range of assays that we still have to define. We'll definitely use standard microbiological methods, but to some extend also chemical assays and molecular-biological assayssuch as qPCR or sequencing	TBD		medium		M6 - M36				
AIT	Task 1.1	Task 1.1 & 1.3	literature review, internal discussion	no	not applicable	medium		M2-M15	M6-48			

PRELIMINARY INFO			GENERAL INFO REGARDING DATA & DATASETS							
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Dataset name	Data Set Description & relation to project: purpose of the dataset, procedures for collection, storage, protection, preservation, destruction, etc)	Data source (generated, collected or reuse of existing data). Is the data generated from a task in the project or an external source ?	Type of data (numerical, categorical, textual, spectral, images, videos, combination of previous types)	Is the data real, synthetic or hybrid ?	Category type (e.g. experiment data, publication, personal data, environmental data, etc.)	Metadata & Standards (Which ones will be used? Will a new one be created?)	Does documentation exist for the data? (description of inputs/outputs, taxonomy for categories, etc.)
WU	WP 1	WP 1, WP 2 and WP 3	Expert input drivers	In WP1 workshop expert ellicitation via a survey will be used to parameterize drivers of change and weights of for each parameter. Will be used to make driver prioritization. In WP2 & 3 the prioritization results will be used	Expert, generated	Numerical, categorical and textual	real	Expert data		not yet
WU	WP 5 task 1	WP 5 task 1	Resilience expert input	Expert input will be gathered with a survey to parameterize resilience and weigh their factors. This will allow for ranking of food management strategies based on resilience	Expert, generated	Numerical, categorical and textual	real	Expert data		not yet

PRELIMINARY INFO			DATA FEATURES			ACCESSIBILITY OF DATA & DATASETS			
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Formats that will be used/generated/collected? (e.g.csv, .xlsx, .txt, .xml, .json, .html, .jpeg, .doc…)	Requirements of Methods/Software for data access	Size estimation (Mb/Gb min/max) if a dataset has been collected	What is the access level? (Public / Restricted)	Will the data/metadata be stored in a trusted repository (e.g. Zenodo)? Or others (e.g. Google Drive, Teams folder, sharepoint, etc.)?	Provide a justification if some data/metadata cannot be shared.	Mention any Ethical/Legal issues that impact on sharing
WU	WP 1	WP 1, WP 2 and WP 3	csv, xlsx, doc, pdf, jpeg/png	Microsoft office	<1gb	TBD	Sharepoint	Restricted until publication. For some information sensitive or privacy (GDPR) aspects	Privacy breach or IP
WU	WP 5 task 1	WP 5 task 1	csv, xlsx, doc, pdf, jpeg/png	Microsoft office	<1gb	TBD	Sharepoint	Restricted until publication. For some information sensitive or privacy (GDPR) aspects	Privacy breach or IP

PRELIMINARY INFO			HOW/WHICH DATA IS GENERATED					TIMINGS			OTHER	
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	dology applied	Data from sensors or industrial processes?	Variables (initial list of inputs and outputs or KPIs)	Expected data quality and coverage (high, medium, low)	Any special information to comment in relation to data	When the data will be generated (time range in project, eg. M6-M12)	When the data will be needed (time range in project, eg. M6-M18)	How long will the data remain available and findable?	How will the data be licensed to permit the widest re-use possible?	Are there any associated costs?
WU	WP 1	WP 1, WP 2 and WP 3	MCDA	No	TBD	Medium	No	M6-M12	M8-end	At minimum to the end of FoodSafeR project	TBD	Likely no, but TBD
WU	WP 5 task 1	WP 5 task 1	MCDA	No	TBD	Medium	No	M19-M30	M19-end	At minimum to the end of FoodSafeR project	TBD	Likely no, but TBD

GHENT UNIVERSITY

PRELIMINARY INFO			GENERAL INFO REGARDING DATA & DATASETS							
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Dataset name	Data Set Description & relation to project: purpose of the dataset, procedures for collection, storage, protection, preservation, destruction, etc)	Data source (generated, collected or reuse of existing data). Is the data generated from a task in the project or an external source ?	Type of data (numerical, categorical, textual, spectral, images, videos, combination of previous types)	Is the data real, synthetic or hybrid ?	Category type (e.g. experiment data, publication, personal data, environmental data, etc.)	Metadata & Standards (Which ones will be used? Will a new one be created?)	Does documentation exist for the data? (description of inputs/outputs, taxonomy for categories, etc.)
UG	WP1; T1, 2	all WP	Hazard-perspective-microbiological_T1.1	In WP1, T1.1 a list of drivers is established based on a literature review, considering microbiological hazards.	collected	textual		publication	-	-
UG	WP1	WP 1, 2, 3	Expert input drivers (refer to input WU)							-
UG	WP2; T2	WP 2		Overview of microbiological data regarding different food categories; refer to SOP sampling and analysis	collected	numerical, images, and textual		experimental	Standardised & harmonised metadata collection will be established	-
UG	WP 5; T2	WP 1, 5	Living-Lab-indicators	In Task 5.2, proposed indicators for the assessment of food safety management systems in FBOs will be tailored based on a living lab (expert input).	expert, generated	numerical, categorical, textual		Expert data	Standardised & harmonised metadata collection will be established	-

PRELIMINARY INFO			DATA FEATURES			ACCESSIBILITY OF DATA & DATASETS			
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Formats that will be used/generated/collected? (e.g.csv, .xlsx, .txt, .xml, .json, .html, .jpeg, .doc…)	Requirements of Methods/Software for data access	Size estimation (Mb/Gb min/max) if a dataset has been collected	What is the access level? (Public / Restricted)	Will the data/metadata be stored in a trusted repository (e.g. Zenodo)? Or others (e.g. Google Drive, Teams folder, sharepoint, etc.)?	Provide a justification if some data/metadata cannot be shared.	Mention any Ethical/Legal issues that impact on sharing
UG	WP1; T1, 2	all WP	›cx, .pdf	Microsoft office	< 1 Gb	Restricted for publication purposes	Sharepoint/Onedrive	Restricted for publication purposes	
UG	WP1	WP 1, 2, 3							
UG	WP2; T2	WP 2	.xlsx, .cxv, .docx, .pdf, .png/.jpeg	Microsoft office	< 10 Gb	Restricted for publication purposes	Sharepoint/Onedrive	Restricted for publication purposes	
UG	WP 5; T2	WP 1, 5	›cx, .pdf	Microsoft office	< 1 Gb	Restricted for publication purposes	Sharepoint/Onedrive	Restricted for publication purposes	

PRELIMINARY INFO			HOW/WHICH DATA IS GENERATED					TIMINGS			OTHER	
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	dology applied	Data from sensors or industrial processes?	Variables (initial list of inputs and outputs or KPIs)	Expected data quality and coverage (high, medium, low)	Any special information to comment in relation to data	When the data will be generated (time range in project, eg. M6-M12)	When the data will be needed (time range in project, eg. M6-M18)	How long will the data remain available and findable?	How will the data be licensed to permit the widest re-use possible?	Are there any associated costs?
UG	WP1; T1, 2	all WP	view	No	-		-	M1-M10	M6 - end	At minimum to the end of FoodSafeR project	TBD	Not foreseen atm
UG	WP1	WP 1, 2, 3										
UG	WP2; T2	WP 2	Microbiological analysis, SOP for sampling and analysis	Microbiological analysis, SOP for sampling and analysis	TBD		-	M1-M24	full project	At minimum to the end of FoodSafeR project	TBD	sampling, analysis, etc.
UG	WP 5; T2	WP 1, 5		TBD	TBD		-	M24-M43	M24-end	At minimum to the end of FoodSafeR project	TBD	not foreseen atm

Bundesinstitut für Risikobewertung

PRELIMINARY INFO			GENERAL INFO REGARDING DATA & DATASETS							
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Dataset name	Data Set Description & relation to project: purpose of the dataset, procedures for collection, storage, protection, preservation, destruction, etc)	Data source (generated, collected or reuse of existing data). Is the data generated from a task in the project or an external source ?	Type of data (numerical, categorical, textual, spectral, images, videos, combination of previous types)	Is the data real, synthetic or hybrid ?	Category type (e.g. experiment data, publication, personal data, environmental data, etc.)	Metadata & Standards (Which ones will be used? Will a new one be created?)	Does documentation exist for the data? (description of inputs/outputs, taxonomy for categories, etc.)
BfR	WP2.5	WP2	microbial and molecular analysis	Data on antimicorbial resistant bacteria and their molecular characterization, including related resistance genes	generated	numerical, textural, images etc.	real	experimental data	N.A.	not yet, will be created; harmonized protocols will be used
BfR	WP2.5	WP2	next generation sequencing	sequencing of aquired resistance determinants or metagenomic sequences	generated	sequences	real	experimental data	N.A.	not yet, will be created; including interpretation of sequencing data
BfR	WP2.1	WP2	software code	programming code to develop specific functions in the FoodChain-Lab web application	generated	software code	real	software code	N.A.	not yet, will be created

Bundesinstitut für Risikobewertung

PRELIMINARY INFO			DATA FEATURES			ACCESSIBILITY OF DATA & DATASETS			
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Formats that will be used/generated/collected? (e.g.csv, .xlsx, .txt, .xml, .json, .html, .jpeg, .doc...)	Requirements of Methods/Software for data access	Size estimation (Mb/Gb min/max) if a dataset has been collected	What is the access level? (Public / Restricted)	Will the data/metadata be stored in a trusted repository (e.g. Zenodo)? Or others (e.g. Google Drive, Teams folder, sharepoint, etc.)?	Provide a justification if some data/metadata cannot be shared.	Mention any Ethical/Legal issues that impact on sharing
BfR	WP2.5	WP2	xlsx, .txt, .jpeg, tif, .docx	office program	< 1 GB	restricted until publication, public afterwards	data will be stored on internal servers at BfR	N.A.	data on aquaponic farms might be problematic because of company secrets or privacy
BfR	WP2.5	WP2	.fastq	open source programs or web browser based programs for data analysis avaiable	500 Mb per raw read sequence	restricted until publication, public afterwards	data will be stored on internal servers at BfR	raw reads too big to be shared easily; assembled sequences can be shared	N.A.
BfR	WP2.1	WP2	angular.js	web browser	<10 MB	open-source	GitHub	N.A.	N.A.

Bundesinstitut für Risikobewertung

PRELIMINARY INFO			HOW/WHICH DATA IS GENERATED					TIMINGS			OTHER	
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	dology applied	Data from sensors or industrial processes?	Variables (initial list of inputs and outputs or KPIs)	Expected data quality and coverage (high, medium, low)	Any special information to comment in relation to data	When the data will be generated (time range in project, eg. M6-M12)	When the data will be needed (time range in project, eg. M6-M18)	How long will the data remain available and findable?	How will the data be licensed to permit the widest re-use possible?	Are there any associated costs?
BfR	WP2.5	WP2	microbial cultivation; DNA extraction, PFGE, MALDI-TOF, PCR, realtime PCR; Southern Blotting; cloniing, MIC, conjugation, transformation, etc	data can result from products from industrial processes	N.A.	N.A.	N.A.	M6-M42	M6-M42	processed data in publications - unlimited	Gold open-access publication with CC BY-NC	yes, costs to publish data open access
BfR	WP2.5	WP2	Illumina NextSeq; MinIon sequencing,	data can result from products from industrial processes	N.A.	high quality	N.A.	M6-M42	M6-M42	after upload in public repositories - unlimited	Publicly avaiable	no
BfR	WP2.1	WP2	Agile methods for software development	N.A.	N.A.	high quality	N.A.	M1-M30	M1-M48	unlimited	GPL version 3	no

PRELIMINARY INFO			GENERAL INFO REGARDING DATA & DATASETS							
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Dataset name	Data Set Description & relation to project: purpose of the dataset, procedures for collection, storage, protection, preservation, destruction, etc)	Data source (generated, collected or reuse of existing data). Is the data generated from a task in the project or an external source ?	Type of data (numerical, categorical, textual, spectral, images, videos, combination of previous types)	Is the data real, synthetic or hybrid ?	Category type (e.g. experiment data, publication, personal data, environmental data, etc.)	Metadata & Standards (Which ones will be used? Will a new one be created?)	Does documentation exist for the data? (description of inputs/outputs, taxonomy for categories, etc.)
IRIS	WP4	All	FoodSafer_storage	Getting data by uploading at the website and transfer information thourgh end-points. All sensitive data is stored in a securitized database with protection from unauthorized access. From a technical perspective, the database is based on TypeORM and PostgreSQL. Preservation of data will be done in the databases of the project. Destruction of data will be performed on demand.	Data source comes from other WPs and Open Data sources.	Type of data are numerical, categorical, textual, spectral, images, videos, combination of previous types	Hybrid	Category type will be experiment data, publication, personal data and environmental data.	There will be metadata based on existing standards (known) - we will not create new standards	Data sources coming from the FoodSafeR project does not exist yet and data coming from open data sources does exist

PRELIMINARY INFO			DATA FEATURES			ACCESSIBILITY OF DATA & DATASETS			
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Formats that will be used/generated/collected? (e.g.csv, .xlsx, .txt, .xml, .json, .html, .jpeg, .doc...)	Requirements of Methods/Software for data access	Size estimation (Mb/Gb min/max) if a dataset has been collected	What is the access level? (Public / Restricted)	Will the data/metadata be stored in a trusted repository (e.g. Zenodo)? Or others (e.g. Google Drive, Teams folder, sharepoint, etc.)?	Provide a justification if some data/metadata cannot be shared.	Mention any Ethical/Legal issues that impact on sharing
IRIS	WP4	All	xlsx, Json., among others depending on the requirements	Via web browser or API connections	TBD	Both	Google Drive and Google Cloud	Any data that cannot be shared will be due to confidentiality and intellectual property constaints (to be defined)	Confidential data will not be disclosed without permission. Data Privacy and GDPR regulations will be fulfilled for all personal data handled on the platform.

PRELIMINARY INFO			HOW/WHICH DATA IS GENERATED					TIMINGS			OTHER	
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	dology applied	Data from sensors or industrial processes?	Variables (initial list of inputs and outputs or KPIs)	Expected data quality and coverage (high, medium, low)	Any special information to comment in relation to data	When the data will be generated (time range in project, eg. M6-M12)	When the data will be needed (time range in project, eg. M6-M18)	How long will the data remain available and findable?	How will the data be licensed to permit the widest re-use possible?	Are there any associated costs?
IRIS	WP4	All		N.A.	N.A.	N.A.	N.A.	M6-M48	From M9	At a minimum, that which is established in the Grant Agreement, however, in line with a sustainable Business Model the platform will be serviced and will continue to grow in the future	Depending on the data ownership	Storage costs dependent on the amount of data

PRELIMINARY INFO			GENERAL INFO REGARDING DATA & DATASETS							
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Dataset name	Data Set Description & relation to project: purpose of the dataset, procedures for collection, storage, protection, preservation, destruction, etc)	Data source (generated, collected or reuse of existing data). Is the data generated from a task in the project or an external source ?	Type of data (numerical, categorical, textual, spectral, images, videos, combination of previous types)	Is the data real, synthetic or hybrid ?	Category type (e.g. experiment data, publication, personal data, environmental data, etc.)	Metadata & Standards (Which ones will be used? Will a new one be created?)	Does documentation exist for the data? (description of inputs/outputs, taxonomy for categories, etc.)
BSI	Open-source data	WP3/T3.1	Satellite data	Satellite images consisting of surface reflectance and surface temperature at high spatial resoulution will be collected using scripts and used for development of prediction tools	External source, reuse of existing data	spectral and thermal images	real	environmental data	Exists from provider	Exists from provider

PRELIMINARY INFO			DATA FEATURES			ACCESSIBILITY OF DATA & DATASETS			
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Formats that will be used/generated/collected? (e.g.csv, .xlsx, .txt, .xml, .json, .html, .jpeg, .doc...)	Requirements of Methods/Software for data access	Size estimation (Mb/Gb min/max) if a dataset has been collected	What is the access level? (Public / Restricted)	Will the data/metadata be stored in a trusted repository (e.g. Zenodo)? Or others (e.g. Google Drive, Teams folder, sharepoint, etc.)?	Provide a justification if some data/metadata cannot be shared.	Mention any Ethical/Legal issues that impact on sharing
BSI	Open-source data	WP3/T3.1		Python scripts	Gb	Public	Open data	No	No

PRELIMINARY INFO			HOW/WHICH DATA IS GENERATED					TIMINGS			OTHER	
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	dology applied	Data from sensors or industrial processes?	Variables (initial list of inputs and outputs or KPIs)	Expected data quality and coverage (high, medium, low)	Any special information to comment in relation to data	When the data will be generated (time range in project, eg. M6-M12)	When the data will be needed (time range in project, eg. M6-M18)	How long will the data remain available and findable?	How will the data be licensed to permit the widest re-use possible?	Are there any associated costs?
BSI	Open-source data	WP3/T3.1	earning	No	Various vegetation indices and land surface temperature	high	No	Existing/M1-M36	M1-M36	Open data	Open data	No

PRELIMINARY INFO			GENERAL INFO REGARDING DATA & DATASETS							
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Dataset name	Data Set Description & relation to project: purpose of the dataset, procedures for collection, storage, protection, preservation, destruction, etc)	Data source (generated, collected or reuse of existing data). Is the data generated from a task in the project or an external source ?	Type of data (numerical, categorical, textual, spectral, images, videos, combination of previous types)	Is the data real, synthetic or hybrid ?	Category type (e.g. experiment data, publication, personal data, environmental data, etc.)	Metadata & Standards (Which ones will be used? Will a new one be created?)	Does documentation exist for the data? (description of inputs/outputs, taxonomy for categories, etc.)
BIO	n.a. (historic data)	WP 3 Task 1	DSM MTX Survey	Historic data on mycotoxin prevalence in cereal grain samples, originated in European countries, starting 2012.	Collected data. BIO/DSM collected data already before and during the project. Collection of data is not part of the project.	Numerical, categorical and textual	Real	Experiment data	tbd	Yes, on Databricks

PRELIMINARY INFO			DATA FEATURES			ACCESSIBILITY OF DATA & DATASETS			
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Formats that will be used/generated/collected? (e.g.csv, .xlsx, .txt, .xml, .json, .html, .jpeg, .doc...)	Requirements of Methods/Software for data access	Size estimation (Mb/Gb min/max) if a dataset has been collected	What is the access level? (Public / Restricted)	Will the data/metadata be stored in a trusted repository (e.g. Zenodo)? Or others (e.g. Google Drive, Teams folder, sharepoint, etc.)?	Provide a justification if some data/metadata cannot be shared.	Mention any Ethical/Legal issues that impact on sharing
BIO	n.a. (historic data)	WP 3 Task 1	.csv	Access only from Erber MS Azure Cloud.	75Mb	Restricted, data is confidential.	No. Data is only made available on Erber MS Azure Cloud. Access will be granted on a strict need-to-know basis combined with personally signed NDAs.	Data is of high value for BIO/DSM and sharing of it would conflict with commercial interests of BIO/DSM.	BIO/DSM has an absolute veto right of publication of this data.

PRELIMINARY INFO			HOW/WHICH DATA IS GENERATED					TIMINGS			OTHER	
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	ɔology applied	Data from sensors or industrial processes?	Variables (initial list of inputs and outputs or KPIs)	Expected data quality and coverage (high, medium, low)	Any special information to comment in relation to data	When the data will be generated (time range in project, eg. M6-M12)	When the data will be needed (time range in project, eg. M6-M18)	How long will the data remain available and findable?	How will the data be licensed to permit the widest re-use possible?	Are there any associated costs?
BIO	n.a. (historic data)	WP 3 Task 1						n.a.	M1-M35	M48	n.a.	no

PRELIMINARY INFO			GENERAL INFO REGARDING DATA & DATASETS							
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Dataset name	Data Set Description & relation to project: purpose of the dataset, procedures for collection, storage, protection, preservation, destruction, etc)	Data source (generated, collected or reuse of existing data). Is the data generated from a task in the project or an external source ?	Type of data (numerical, categorical, textual, spectral, images, videos, combination of previous types)	Is the data real, synthetic or hybrid ?	Category type (e.g. experiment data, publication, personal data, environmental data, etc.)	Metadata & Standards (Which ones will be used? Will a new one be created?)	Does documentation exist for the data? (description of inputs/outputs, taxonomy for categories, etc.)
Barilla	WP3	WP4	Furans-Mitigation	Operative parameters, guidelines, monitoring strategies devoted to mitigate furans content in industrial food sauces	Generated in the project task	Numerical, categorical, textual, spectral, images, combination of previous types	real	Experimental Data	Recipe, ingredients, etc.	Yes
Barilla	WP3	WP4	Toxins Fate - Gluten Free	Operative parameters, guidelines, monitoring strategies devoted to mitigate mycotoxins content in industrial gluten free products	Generated in the project task	Numerical, categorical, textual, spectral, images, combination of previous types	real	Experimental Data	Recipe, ingredients, etc.	Yes

PRELIMINARY INFO			DATA FEATURES			ACCESSIBILITY OF DATA & DATASETS			
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Formats that will be used/generated/collected? (e.g.csv, .xlsx, .txt, .xml, .json, .html, .jpeg, .doc…)	Requirements of Methods/Software for data access	Size estimation (Mb/Gb min/max) if a dataset has been collected	What is the access level? (Public / Restricted)	Will the data/metadata be stored in a trusted repository (e.g. Zenodo)? Or others (e.g. Google Drive, Teams folder, sharepoint, etc.)?	Provide a justification if some data/metadata cannot be shared.	Mention any Ethical/Legal issues that impact on sharing
Barilla	WP3	WP4	.csv, .xlsx, .jpeg, .doc	Microsoft Office	Mb	Public / Restricted (to be defined later on)	It can be considered this option	Confidentiality of sensible information IP	
Barilla	WP3	WP4	.csv, .xlsx, .jpeg, .doc	Microsoft Office	Mb	Public / Restricted (to be defined later on)	It can be considered this option	Confidentiality of sensible information IP	

PRELIMINARY INFO			HOW/WHICH DATA IS GENERATED					TIMINGS			OTHER	
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	dology applied	Data from sensors or industrial processes?	Variables (initial list of inputs and outputs or KPIs)	Expected data quality and coverage (high, medium, low)	Any special information to comment in relation to data	When the data will be generated (time range in project, eg. M6-M12)	When the data will be needed (time range in project, eg. M6-M18)	How long will the data remain available and findable?	How will the data be licensed to permit the widest re-use possible?	Are there any associated costs?
Barilla	WP3	WP4		It can be possible		medium		M12-M36	M36-M48	permanently	No specific restrictions, apart the data that will be classified as confidential, if any	No within the perimeter of the already agreed tasks/resources in the project
Barilla	WP3	WP4		It can be possible		medium		M12-M36	M36-M48	permanently	No specific restrictions, apart the data that will be classified as confidential, if any	No within the perimeter of the already agreed tasks/resources in the project

PRELIMINARY INFO			GENERAL INFO REGARDING DATA & DATASETS							
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Dataset name	Data Set Description & relation to project: purpose of the dataset, procedures for collection, storage, protection, preservation, destruction, etc)	Data source (generated, collected or reuse of existing data). Is the data generated from a task in the project or an external source ?	Type of data (numerical, categorical, textual, spectral, images, videos, combination of previous types)	Is the data real, synthetic or hybrid ?	Category type (e.g. experiment data, publication, personal data, environmental data, etc.)	Metadata & Standards (Which ones will be used? Will a new one be created?)	Does documentation exist for the data? (description of inputs/outputs, taxonomy for categories, etc.)
PIWet	WP2.4	WP2	Results on HEV detection in RTE food	Data on occurrence of hepatitis E virus in RTE products. Datasets will be used for safety assessment of wild boar meat and pork products. Analysis of the risks associated with illegaly imported food of animal origin to Europe. Digital storage of the data.	Data will be genereated in the project	Categorical and textual data as well as combination of diffrent types of data	Real data	Experimental data	Metadata will be attached	No
PIWet	WP2.4	WP2	HEV nucleotide sequences	Nucleotide sequences of a genome fragment of hepatitis E virus (HEV). Datasets will be used for identification of HEV subtypes. Digital storage of the data.	Data will be genereated in the project	Textual, graphical	Real data	Experimental data	Metadata will be attached	Yes

PRELIMINARY INFO			DATA FEATURES			ACCESSIBILITY OF DATA & DATASETS			
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Formats that will be used/generated/collected? (e.g.csv, .xlsx, .txt, .xml, .json, .html, .jpeg, .doc...)	Requirements of Methods/Software for data access	Size estimation (Mb/Gb min/max) if a dataset has been collected	What is the access level? (Public / Restricted)	Will the data/metadata be stored in a trusted repository (e.g. Zenodo)? Or others (e.g. Google Drive, Teams folder, sharepoint, etc.)?	Provide a justification if some data/metadata cannot be shared.	Mention any Ethical/Legal issues that impact on sharing
PIWet	WP2.4	WP2	xlsx, doc	Basic computer software (Microsoft office)	2Mb	Restricted	Teams folder, sharepoint	An embargo on data sharing will be applied to give time to publish of the results	Intelectual property rights
PIWet	WP2.4	WP2	AB1, SEQ	BioEdit, Basic computer software (Microsoft office)	15 Mb	Restricted	Teams folder, sharepoint	An embargo on data sharing will be applied to give time to publish of the results	Intelectual property rights

PRELIMINARY INFO			HOW/WHICH DATA IS GENERATED					TIMINGS			OTHER	
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	dology applied	Data from sensors or industrial processes?	Variables (initial list of inputs and outputs or KPIs)	Expected data quality and coverage (high, medium, low)	Any special information to comment in relation to data	When the data will be generated (time range in project, eg. M6-M12)	When the data will be needed (time range in project, eg. M6-M18)	How long will the data remain available and findable?	How will the data be licensed to permit the widest re-use possible?	Are there any associated costs?
PIWet	WP2.4	WP2	Data will be obtained on experimental conditions and collection from food processing companies	Yes, some data will be obtained from industrial processes		High	No	M6-M42	M6-M42	4 years (project duration time)	Creative commons licence realted to scientific publications	Publication costs
PIWet	WP2.4	WP2	Data will be obtained on experimental conditions	No		High	No	M6-M42	M6-M42	Data will be available in NCBI GenBank data base	NCBI license	No

PRELIMINARY INFO			GENERAL INFO REGARDING DATA & DATASETS							
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Dataset name	Data Set Description & relation to project: purpose of the dataset, procedures for collection, storage, protection, preservation, destruction, etc)	Data source (generated, collected or reuse of existing data). Is the data generated from a task in the project or an external source ?	Type of data (numerical, categorical, textual, spectral, images, videos, combination of previous types)	Is the data real, synthetic or hybrid ?	Category type (e.g. experiment data, publication, personal data, environmental data, etc.)	Metadata & Standards (Which ones will be used? Will a new one be created?)	Does documentation exist for the data? (description of inputs/outputs, taxonomy for categories, etc.)
FSAI	WP 4.3 - Living Lab (LL) 2	WP 4	LL2 output	This will be a final report of the LL2 rather than a defined dataset	Primary source of the data will be generated from the LL	Textual, numerical and images	Real	LL output - Report	We won't be generating a defined dataset. It will be a report of findings, learnings and observations from the LL2 exercise	no

PRELIMINARY INFO			DATA FEATURES			ACCESSIBILITY OF DATA & DATASETS			
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	Formats that will be used/generated/collected? (e.g.csv, .xlsx, .txt, .xml, .json, .html, .jpeg, .doc...)	Requirements of Methods/Software for data access	Size estimation (Mb/Gb min/max) if a dataset has been collected	What is the access level? (Public / Restricted)	Will the data/metadata be stored in a trusted repository (e.g. Zenodo)? Or others (e.g. Google Drive, Teams folder, sharepoint, etc.)?	Provide a justification if some data/metadata cannot be shared.	Mention any Ethical/Legal issues that impact on sharing
FSAI	WP 4.3 - Living Lab (LL) 2	WP 4	pdf	MS office	less than 1GB	Restricted	SharePoint and FSAI internal secure IT system	All relevant data will be shared	Not for public sharing - these outputs are from Regulatory agencies and should be treated as confidential within the FoodSafeR community

PRELIMINARY INFO			HOW/WHICH DATA IS GENERATED					TIMINGS			OTHER	
Partner responsible for the data	WPs or Tasks in which the data is generated	WPs or Tasks in which the data will be used	ɔology applied	Data from sensors or industrial processes?	Variables (initial list of inputs and outputs or KPIs)	Expected data quality and coverage (high, medium, low)	Any special information to comment in relation to data	When the data will be generated (time range in project, eg. M6-M12)	When the data will be needed (time range in project, eg. M6-M18)	How long will the data remain available and findable?	How will the data be licensed to permit the widest re-use possible?	Are there any associated costs?
FSAI	WP 4.3 - Living Lab (LL) 2	WP 4		Generated during a LL	Not yet defined	Not generating a defined dataset. It will be a report of findings, learnings & observations from the LL2 exercise	no	Oct-Dec 2023 (M10-12)	M12 +	Until the end of the FoodSafeR project	This data will not be re-used outside of FoodSafeR	No, not expected