



The Food Safety Market: An SME-powered industrial data platform to boost the competitiveness of European food certification

D8.5 – Data Management Handling Plan












DELIVERABLE NUMBER	D8.5
DELIVERABLE TITLE	Data Management Handling Plan
RESPONSIBLE AUTHOR	Nikolaus Forgó, Tima Anwana, Lukas Faymann (UNIVIE)



Co-funded by the Horizon 2020
Framework Programme of the European Union

GRANT AGREEMENT N.	871703
PROJECT ACRONYM	TheFSM
PROJECT FULL NAME	The Food Safety Market: An SME-powered industrial data platform to boost the competitiveness of European food certification
STARTING DATE (DUR.)	01/02/2020 (36 months)
ENDING DATE	31/01/2023
PROJECT WEBSITE	www.foodsafetymarket.eu
COORDINATOR	Nikos Manouselis
ADDRESS	110 Pentelis Str., Marousi, GR15126, Greece
REPLY TO	nikosm@agroknow.com
PHONE	+30 210 6897 905
EU PROJECT OFFICER	Stefano Bertolo
WORKPACKAGE N. TITLE	WP8 Management
WORKPACKAGE LEADER	Agroknow
DELIVERABLE N. TITLE	D8.5: Data Management Handling Plan for TheFSM
RESPONSIBLE AUTHOR	Nikolaus Forgó, Tima Anwana, Lukas Faymann(UNIVIE)
REPLY TO	tima.anwana@univie.ac.at , lukas.faymann@univie.ac.at
DOCUMENT URL	www.foodsafetymarket.eu
DATE OF DELIVERY (CONTRACTUAL)	31.01.2022
DATE OF DELIVERY (SUBMITTED)	31.01.2022
VERSION STATUS	2.0 Final
NATURE	Report (R)
DISSEMINATION LEVEL	PUBLIC
AUTHORS (PARTNER)	Nikolaus Forgó, Elisabeth Steindl, Tima Anwana, Ziga Skorjanc (UNIVIE)
CONTRIBUTORS	Nikos Manouselis, Francesca Tsaropoulou, Iliana Giannelou, Charalampos Thanopolous, Dimitris Fotakidis (AGROKNOW), Nikola Rusinov, Svetla Boytcheva (SAI), Danai Vergeti (UBITECH), Tanja Matosevic (AGRIVI), Ana Bevc, Tomaz Levak (PROSPEH), Yamine Bouzembrak, Hans Marvin (WFSR), George Gheorghiu (TUV AU ROMANIA), Anna Polloni (VALORITALIA), Stefan Hackel, Sabine Werb (TUV AU CERT), Stylianos Vaporidis (TAH)
REVIEWER	Zlatina Marinova (SAI)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
[0.1]	First Draft sent out for review by all contributors	03/07/2020	Nikolaus Forgó, Elisabeth Steindl, Tima Anwana, Žiga Škorjanc (UNIVIE)
[0.2]	Review by all contributors	10/07/2020	All contributors (see above)
[0.3]	Internal Review	23/07/2020	Nikola Rusinov (SAI)
[0.4]	Final Draft including all comments from the reviewing partners sent to Coordinator	27/07/2020	Nikolaus Forgó, Elisabeth Steindl, Tima Anwana, Žiga Škorjanc (UNIVIE)
[0.5]	Internal Review	28/07/2020	Dimitris Fotakidis (AGROKNOW)
[1.0]	Final Version	30/07/2020	Nikolaus Forgó, Elisabeth Steindl, Tima Anwana, Žiga Škorjanc (UNIVIE)
[1.1]	Updates made to reflect the current state of data management in the project	20/10/2021	Tima Anwana, Lukas Faymann (UNIVIE)
[1.2]	Review conducted by key contributors	01/11/2021	Dimitris Fotakidis (AGROKNOW), Danai Vergerti (UBITECH)
[1.3]	Final Updates made to new deliverable	10/01/2022	Tima Anwana, Lukas Faymann (UNIVIE)
[1.4]	Internal Review	29/01/2022	Simeon Petrov (SAI) and Nikolaus Forgo (UNIVIE)
[2.0]	Final Submission	31/01/2022	Tima Anwana, Lukas Faymann (UNIVIE)

PARTNERS		CONTACT
Agroknow IKE (Agroknow, Greece)		Nikos Manouselis (Agroknow) nikosm@agroknow.com
SIRMA AI EAD (SAI, Bulgaria)		Zlatina Marinova (SAI) zlatina.marinova@ontotext.com
GIOUMPITEK MELETI SCHEDIASMOS YLOPOIISI KAI POLISI ERGON PLIROFORIKIS ETAIREIA PERIORISMENIS EFTHYNIS (UBITECH, Greece)		Danai Vergeti (UBITECH) vergetid@ubitech.eu
AGRIVI DOO ZA PROIZVODNJU, TRGOVINU I USLUGE (Agrivi d.o.o., Croatia)		Filip Gerin (Agrivi d.o.o.) filip.gerin@agrivi.com
PROSPEH, POSLOVNE STORITVE IN DIGITALNE RESITVE DOO (PROSPEH DOO, Slovenia)		Ana Bevc (PROSPEH DOO) ana.bevc@tracelabs.io
UNIVERSITAT WIEN (UNIVIE, Austria)		Tima Otu Anwana (UNIVIE) tima.anwana@univie.ac.at
STICHTING WAGENINGEN RESEARCH (WFSR, Netherlands)		Yamine Bouzembrak (WFSR) yamine.bouzembrak@wur.nl
TUV- AUSTRIA ELLAS MONOPROSOPI ETAIREIA PERIORISMENIS EUTHYNIS (TUV AU HELLAS, Greece)		Kostas Mavropoulos (TUV AU HELLAS) konstantinos.mavropoulos@tuv.at
TUV AUSTRIA ROMANIA SRL (TUV AU ROMANIA, Romania)		George Gheorghiu (TUV AU Romania) george.gheorghiu@tuv.at
VALORITALIA SOCIETA PER LA CERTIFICAZIONE DELLE QUALITA'E DELLE PRODUZIONI VITIVINICOLE ITALIANE SRL (VALORITALIA, Italy)		Francesca Romero (Valoritalia) francesca.romero@valoritalia.it
TUV AUSTRIA CERT GMBH (TUV AU CERT, Austria)		Sousanna Charalambidou (TUV AU CYPRUS) sousanna.charalambidou@tuv.at

ACRONYMS LIST

TheFSM	The Food Safety Market
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable and Reuseable
QA	Quality Assurance
WP	Work Package
EC	European Commission
PO	Project Officer
GA	Grant Agreement
CA	Consortium agreement
PM	Project Manager
EU	European Union
GDPR	General Data Protection Regulation

EXECUTIVE SUMMARY

The purpose of the TheFSM project is to deliver an industrial data platform that has the potential to significantly modernize the procedure of how food certification takes place in Europe. By facilitating the exchange and connection of data between different food safety actors who are interested in sharing information critical to certification the project is going to accelerate the pace by which this group adopts digital innovation and offers data-driven services to its clients around the world.

Following the European Commission's recommendations, this deliverable describes the FAIR management of the research data sets processed for TheFSM: The Data Management Handling Plan (DMP) shall ensure that TheFSM follows the European Commission's guidelines on making our data FAIR (findable, accessible, interoperable, reusable). The DMP outlines the data management policy adopted in TheFSM Project. The document itself is a living document. The first version of the DMP was submitted in M6, this document represents the second iteration of the deliverable to be submitted in M24. A final version will be submitted in M36. Data management during pilots and validation phase of the project is addressed in a local DMP submitted in WP6, deliverable 6.1.

In addition, earlier or irregular updates will be made whenever significant changes arise (e.g. new data, changes in consortium policies or consortium composition).

The objective of a DMP is to define how data will be created, how it will be documented, who will be able to access it, where it will be stored, who will back it up and whether (how) it will be shared and preserved.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	6
1. INTRODUCTION.....	10
2. DATA SUMMARY.....	13
2.1 DATA SET 1: ADMINISTRATIVE DATA.....	13
2.2 DATA SET 2: PLATFORM DATA.....	14
2.3 DATA SET 3: TECHNICAL DATA	14
2.4 DATA SET 4: PILOT DATA	14
2.5 SYNOPSIS OF THEFSM DATA SETS	16
3. FAIR DATA.....	29
3.1 FAIR MANAGEMENT OF RESEARCH DATA.....	29
3.2 RESEARCH DATA LIFECYCLE	29
3.3 THE FAIR PRINCIPLE.....	31
3.3.1 Findable.....	32
3.3.2 Accessible	33
3.3.3 Interoperable.....	35
3.3.4 Re-usable	36
3.3.5 Allocation of resources	37
4. DATA SECURITY.....	38
4.1 ADMINISTRATIVE DATA	38
4.2 PLATFORM DATA	38
4.3 TECHNICAL DATA.....	39

5. PROTECTION OF PERSONAL DATA AND ETHICAL ASPECTS	40
5.1 PRINCIPLES OF PERSONAL DATA PROCESSING	40
5.2 LAWFULNESS OF PROCESSING	41
5.3 INTEGRITY AND CONFIDENTIALITY	42
5.4 ACCOUNTABILITY	43
5.5 RESPONSIBILITIES OF THE CONTROLLER AND THE PROCESSOR	43
5.6 JOINT CONTROLLERS	44
5.7 TRANSFER TO THIRD COUNTRIES	44
ANNEX	47
DATA MANAGEMENT PLAN <i>QUESTIONNAIRE</i>	47

List of Tables

Table 1: Data sets produced, generated or collected	14
Table 2: Data Production and Storage	16
Table 3: Organisation, documentation and metadata of data to be published.....	19
Table 4: Data Access.....	21
Table 5: Data sharing and re-use of data to be published	25
Table 6: Data preservation and archiving.....	26

List of Figures

Figure 1: Data Sets.....	13
Figure 2: Stages of data during a research process.....	29
Figure 3: The Fair Principle	31

1. INTRODUCTION

Scope of the Document

This document aims to outline the management of data in TheFSM project. For this purpose and in alignment with the European commission's open data policy and its objective to make research findable, accessible, interoperable and reusable (FAIR), this document represents the second iteration of the project-wide Data Management Plan (DMP). The document is based on the Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. The Guidelines detail how project data will be handled during & after the project; what data will be collected, processed or generated; what methodology & standards will be applied; and whether data will be shared /made open access/ how data will be curated and preserved.

Within the context of TheFSM, the DMP is used by the consortium partners for the effective management of the data that will be generated and collected within the context of the project and for the more efficient handling of the management of publications. After thoroughly evaluating the legal frameworks, the consortium will examine whether there is currently available data to which open access can be granted, always respecting the security and privacy requirements imposed. Concerning the dissemination of the scientific results, the consortium aims to establish and promote open access publications and partners will be encouraged to publish open access articles, to enable researchers to build upon previous research results, to foster collaboration, to avoid duplication of efforts, and to accelerate innovation.

This Data Management Plan (DMP) analyses the main elements of TheFSM data management policy. It is intended to cover the complete life cycle of the research data generated and processed. The DMP outlines the following:

- the types of research data that will be generated or collected during the project;
- how the research data will be processed and preserved;
- which parts of the datasets will be shared for verification or re-use;
- the standards that will be used; and
- the handling of research data after the end of the project.

In addition, this document is designed to monitor the privacy and confidentiality of the data sets in TheFSM and to set out the legal and ethical standards for data generation, use, storage and sharing in line with the overall-management of the project, as foreseen in grant agreement and consortium agreement, which will be applied throughout the project. Furthermore, the DMP seeks to ensure that TheFSM follows the H2020 Open Access policy wherever possible and that all consortium partners are compliant with the provisions of EU legislation such as the GDPR and national regulations.

All relevant information on the handling of data sets in TheFSM was gathered through a Questionnaire following the "Guidelines on FAIR Data Management in Horizon 2020" (see Annex 1), in M6 of the project. All partners contributed by submitting their answers to the Questionnaire. In M24 of the project, partners reviewed and updated their answers to reflect current developments in the project.

Intended Audience

This is a public document in accordance with TheFSM Grant Agreement; all persons interested in the topic will have access to this deliverable once it is published on the project website. However, it is assumed that this deliverable will be of primary interest to actors and stakeholders in the food industry and food safety supply chain.

Structure of the Document

The structure of this document aims at following the Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020.¹ This deliverable begins with introductory remarks regarding the scope of the TheFSM project. Thereafter, this report describes the principle of FAIR Data Management with reference to its implementation in TheFSM. This description is followed by a general analysis of the data generated and collected in course of the project and a detailed analysis of the life cycles of the various data sets as known in M24 of TheFSM.

Updates with respect to Previous Version (if any)

This is the second version of the Data Management Handling Plan for TheFSM; a final iteration is scheduled for M36. This section of the deliverable outlines the updates that have been made in M24 in each section.

Section 1: Introduction

- Section 1.4 has been updated to reflect the changes, which have been made with respect to the previous version of the deliverable.

Section 2: Data Summary

- The summary of TheFSM data sets has been updated to include Pilot Data, as reflected in Section 2.4.
- Section 2.5 focusing on the synopsis of TheFSM data sets has been updated to reflect the current approach to data management that has been adopted in the project.

Section 3: FAIR Data

- Section 3.3 has been updated to reflect the current measures employed in order to make data fair, accessible, interoperable and reusable.
- Section 3.3.5 focusing on the allocation of resources has been updated.

Section 4: Data Security

- This section has been updated to reflect the current data security measures implemented in the project.

¹ European Commission H2020 Programme, Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, Version 3.2 (2017). [Online] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

Section 5: Personal Data Protection and Ethical Aspects

- Section 5.2 focusing on lawful basis for personal data processing has been updated to include information on consent as a legal basis in TheFSM project and the measures implemented to obtain informed consent from external participants in project-related focus groups and pilots.

2. DATA SUMMARY

This section addresses the categorisation and mapping of data that will be produced, generated, and collected in TheFSM Project. The data generated, produced, or collected fall into one of the following four major categories as depicted in the diagram below (Figure 1).

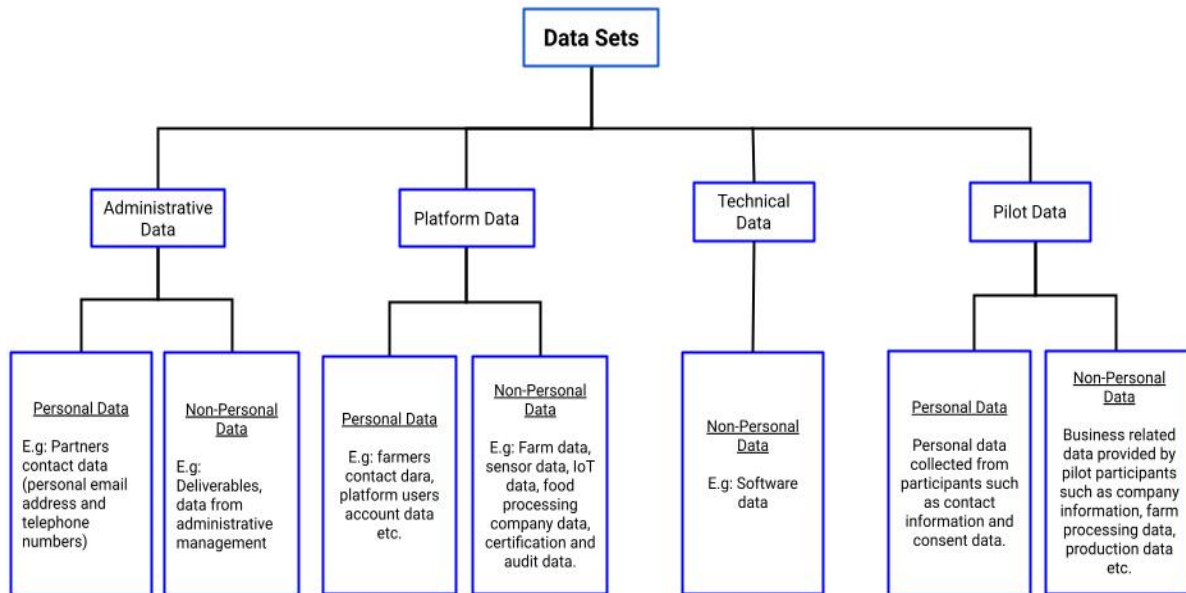


Figure 1: Data Sets

2.1 Data Set 1: Administrative data

Administrative data (personal and non-personal data), refers to data generated or collected for organisational, managerial, transactional and record-keeping purposes during the project. Some examples are:

- Partners contact
- Project Planning data including the coordination of partners' work, consortium meetings etc.,
- Templates of deliverables and reports,
- E-mails and minutes, documentation of communications among members of the projects,
- Data from administration and financial management,
- Data from marketing and commercialization process.
- Deliverables and reports

2.2 Data Set 2: Platform Data

Platform data refers to a variety of data that will be produced, collected and exchanged between actors in the supply chain on TheFSM Platform, in the various Business Scenarios. This data set will be developed in Task 1.4 Data Requirements of the project. Some examples of platform data are:

- Farm Data: farm management systems, sensory data, IoT data, sample data, pesticide database etc.
- Certification Data: audit reports, inspection results, certificates etc.
- Laboratory Data
- Producer data
- Supplier Profiles
- Consumer Complaints
- User account data
- Linked Open Data

2.3 Data Set 3: Technical Data

This category broadly refers to the main software development data generated during the project. This data set is developed in Task 1.2 Technical Requirements of the project.

2.4 Data Set 4: Pilot Data

This category broadly refers to the data collected and generated in the course of TheFSM Pilots. This data set is developed in WP 6 in the course of the pilots. Pilot data is not the focus of this document because deliverable 6.1 contains the local DMPs that are specific to the Pilots.

The following table (Table 1) presents additional details on the data sets generated, produced or collected during the project. The table indicates descriptions and data formats as provided by each partner. Relevant partners of TheFSM consortium provided the information in the table through questionnaires.

Table 1: Data sets produced, generated or collected

Partner	Data Set	Description of Data	Personal or Non-personal Data	Data Format
AGRIVI D.O. O	Platform Data and Technical data.	AGRIVI will generate or produce software code (technical data). AGRIVI will collect farmer contact data, crop production growing data, food processing company data and farmer contact data (platform data).	Personal and non-personal data.	Software code in the format of C# programming language.

Agrokno w	Platform Data	<p>Agrokno will generate or produce food safety records. Some examples include supplier profile, hazards and risk reports, hazards analytics, Certification schema information, certificate information.</p> <p>Agrokno will collect food safety records. Some examples include food recalls, border rejections, inspections/audits, lab data, supplier information, food data etc.</p>		Personal and non-personal data.	<p>MariaDB, PostgreSQL, elasticsearch, Html, PDF, Remote DBs e.g. GlobalGap, Audit Systems, ERPs, CRMs, Html, spreadsheets, pdfs, Remote systems e.g. Agrivi, sensor installations on farms.</p>
PROSPE H	Technical data, Administrative Data and Platform data	<p>PROSPEH will generate or produce Software code.</p> <p>PROSPEH will collect input from other partners for project development as well as platform data in the form of traceability, certification, auditing data etc.</p>		Personal and non-personal data.	<p>JavaScript and other programming language. Google Spreadsheets.</p>
SIRMA AI EAD (SAI)	Administrative data, Platform data and Technical data	<p>SAI will generate or produce Semantic Knowledge Graphs.</p> <p>SAI will collect food data, production data, laboratory data, certification data, inspection data, and data from public databases, linked open data.</p>		Personal and non-personal data	<p>RDF triples, JSON format, TTL format, CSV files, Excel files, Word documents, PDF, XML, SPARQL queries, Java Software and Python Software, Google Drive repository,</p>
TUV Austria CYPRUS	Platform Data and Administrative Data	<p>TUEV Austria CERT will generate or produce Standards and Food Certification requirements.</p> <p>TUV CYPRUS will collect general data of the partner organisations, product and raw material specification, supplier data, process description, operative checklists and records and Food Certificates.</p>		Personal and non-personal data.	<p>Partner and Organisational Contact data: pdf, vision, word, excel. Product and raw material specification: pdf, GSI database. Supplier data: pdf, visio, word, excel. Process description: pdf, visio, word, excel. Operative Checklists and records: word, excel, pdf. Certificates: pdf.</p>
TUV Austria Hellas	Platform data	<p>TUV Austria Hellas will generate or produce Certification and Audit data.</p> <p>TUV Austria Hellas will collect customer's application data.</p>		Personal and non-personal data	<p>Certification and Audit data: archives and docx Customer's application data: pdf format or docx.</p>
TUV Austria Romani a	Platform data	<p>TUV Austria Romania will generate or produce data related to audit services and conformity certificates.</p> <p>TUV Austria will collect audit and certification data. This includes information about company authorisations, management system documentation, product information, supplier's information and employee's information.</p>		Personal and Non-personal data.	<p>Docs, Excel, pdf.</p>

UBITEC H	Technical data, platform data.	UBITECH will generate or produce technical data generated within TheFSM Platform for its proper functionality and integration. User account data and access policies for each data set imported into TheFSM platform. UBITECH will collect data related to transactions, input output sources, IoT data and relational structured data.		Personal and non-personal data.	Java spring software code, python code, hyperledger fabric smart contracts.
UNIVIE	Administrative data	UNIVIE will generate or produce Deliverables, Reports and communications with partners. UNIVIE will collect project management related data (minutes, agenda etc.) and communications with partners.		Personal and non-personal	Word Docs, PDF files, Excel, Google drive repository, Outlook and Thunderbird.
VALORITALIA	Platform data	Valoritalia will generate or produce audit reports and lab data. Valoritalia will collect information from clients and institutions		Personal and non-personal data.	Audit reports: Excel CVS files and PDF files. Lab data: excel and CVS files. Information from clients and institutions: PDF files.
WFSR	Platform Data	Data generated is not specified. WFSR will collect data from actors in the supply chain (i.e. Farm data, industry data, slaughterhouse data and retail data).		Non-personal data.	Pdf files excel or CVS files and Google drive repository.

2.5 Synopsis of TheFSM Data Sets

The following tables present a concise survey of the key features of the life cycle of this data. The information was derived from questionnaires filled out by relevant partners of TheFSM Consortium (AGRIVI, Agroknow, PROSPEH, SAI, TUEV Austria CERT, TUV Austria Hellas, TUV Austria Romania, UBITECH, UNIVIE, Valoritalia, WFSR) in M6 and updates provided by partners in M24. The subsequent sections will provide a more detailed analysis of the generation and collection processes for various data and the resulting characteristics of each group. The following tables provide additional details about each of the three data sets.

Table 2: Data Production and Storage

DATA PRODUCTION AND STORAGE	
Data Generated /Collected	<u>Administrative Data</u> Personal data: Partner's names, roles and contact details including email addresses, telephone numbers, Skype accounts and Google accounts.

	<p>Non-Personal Data: Project management Data, communications between partners, meeting minutes, Deliverables and Reports.</p> <p><u>Platform Data:</u> Food production data, certification data audit reports, customer’s application data, and data from public databases.</p> <p><u>Technical Data:</u> Software Code, Software Design.</p>
<p>Data Formats</p>	<p><u>Administrative Data:</u> PDF, Word processors (MS Office Word, Excel), Google drive repository (Google Docs and Google Spreadsheets).</p> <p><u>Platform Data:</u> PDF, GSI database, Excel, Word Docs, Visio, Archives, CVS files, DB records. XML.</p> <p><u>Technical Data:</u> Java spring software code, Python code, Hyperledger fabric, Smart contracts, SPARQL queries, RDF triples, JSON format, TTL format, C# programming language, CSV files, Excel files, Word Docs, Google Spreadsheets.</p>
<p>Reproducibility</p>	<p><u>Administrative Data:</u> Google Drive and local repositories</p> <p><u>Platform Data:</u> Local repositories, GitHub.</p> <p><u>Technical Data:</u> Data sources will be stored and back-up strategies applied to ensure recovery, reproduction and reuse of technical data.</p>
<p>Data Size</p>	<p><u>Administrative Data</u> The current size of the administrative data is 2.85 GB, consisting of 1509 different files stored in 381 folders.</p> <p><u>Platform Data</u> Currently (M24), the data sets that have been collected and exchanged through the platform are >100 datasets of foods, hazards, food incidents, food suppliers/companies and >100 data endpoints (accessed through REST APIs). They consist of approximately 50M data points. The current data volume is estimated to be approximately 70 GB.</p> <p><u>Technical Data</u> Currently (M24), the data sets that have been collected and exchanged through the platform are >100 datasets of foods, hazards, food incidents, food suppliers/companies and >100 data endpoints (accessed through REST APIs). They consist of approximately 50M data points. The current data volume is estimated to be approximately 70 GB.</p>

Software tools for creating/processing/visualising data	<p><u>Administrative Data</u> Google Drive</p> <p><u>Platform Data:</u> Excel tools, rest calls software, Customer database, Microsoft Tools, specific libraries.</p> <p><u>Technical Data:</u> Origin Trails Technologies, Agrivi – FSM (Farm Management Software), Chronograph, Grafana, Apache Strom, Apache Kafka.</p>
Use of pre-existing data	<p><u>Administrative Data</u> Publically available templates for deliverables and reports (e.g. Contractual Templates).</p> <p><u>Platform Data:</u> Pre-existing data will be integrated from Certification bodies’ databases, public databases.</p> <p><u>Technical Data:</u> Pre-existing data will be integrated from LOD (Linked Open data) Cloud, other EC funded research projects, partners systems and public/open data sources (Identified in data inventory list in WP2).</p>
Data Storage and Backup Strategies	<p><u>Administrative Data:</u> Google Drive, local repositories, partners’ IT departments and EC Portal. Project documents and deliverables are stored in the collaborative workspace of the project Google Drive repository (Google workspace, formerly GSuite, is an integrated service offered by Google to manage documents, purchased by the Coordinator for the TheFSM project).</p> <p>Final and official submitted versions of deliverables are stored by the Project Coordinators on the European Commission Portal. Public deliverables are uploaded on the project website (https://foodsafetymarket.eu/) and at the Community created in Zenodo repository (https://zenodo.org/communities/thefsm), after being officially submitted to the European Commission</p> <p><u>Platform Data:</u> Google drive repositories, GitHub, data retention points.</p> <p><u>Technical Data:</u> The technical data (data sets, source code and docker containers) are stored and hosted into a private git server (Gitlab) and the cloud infrastructure of UBITECH.</p> <p>UBITECH Cloud/Edge Experimental Infrastructure: UBITECH has deployed and maintains in its fire-protected computer room a highly available experimental Cloud infrastructure that includes more than 350 cores (about 1400 vCPUs) at 2.6GHz with virtualization capabilities as also 2.6TB</p>

	<p>of RAM. Currently, provisioning is made via Openstack (KVM-based hypervisors), CloudStack (KVB-based hypervisors) and ESXi (ESXi-based hypervisors) following the Infrastructure-as-a-Service (IaaS) model. It should be noted that every six months, the head version of Openstack and CloudStack are adopted. The computational infrastructure is complemented by multiple Network Attached Storages (NASs) with an effective capacity of 142 TB of data (in Synology Hybrid RAID). Furthermore, a secondary NAS backend with an effective capacity of 24.4 TB of data on a RAID 6 mode that can extend, on demand, the storage capacity of the aforementioned experimental Cloud Infrastructure to a total of 166.4 TB of effective data storage, with a Ceph-based API used to offer storage-as-a-service functionality. Such functionality is exposed in various modalities such as S3-based object storage, block storage and POSIX-based file system. Besides the Cloud infrastructure, UBITECH has deployed and maintains an Edge and Fog computational cluster composed of different devices. More specifically, the cluster is composed of 3 NVIDIA Jetson AGX Xavier with 32GB of RAM, 512-core NVIDIA Volta GPU and 64 Tensor Cores each, 3 NVIDIA Jetson Nano with 4GB of RAM and 128-core NVIDIA Maxwell GPU, 8 Raspberry Pi 4 with 8GB of RAM each and 6 Raspberry Pi 3+ with 1GB of RAM each. The Edge-Fog cluster is expandable to more devices and currently includes add-on devices like TPMs, Cameras, Microphones, 802.11s mesh WiFi cards, and LoRaWan cards. Additionally, UBITECH performs daily backups.</p>
--	--

Table 3: Organisation, documentation and metadata of data to be published

ORGANISATION, DOCUMENTATION AND METADATA OF DATA INTENDED TO BE PUBLISHED	
Standards for Documentation of Metadata	<p><u>Administrative Data:</u> All public deliverables are uploaded in Zenodo (https://zenodo.org/). Zenodo provides a unique DOI for each deliverable, ensuring its valid and safe disposal. A DOI (Digital Object Identifier) is a unique and never-changing string assigned to online (journal) articles, books, and other works. DOIs make it easier to retrieve works, which is why citation styles, like APA and MLA Style, recommend including them in citations.</p> <p><u>Platform Data:</u> DCAT metadata standards for publishing metadata of food records, Standard classifications for topics such as FOODEX2 and AGROVOC, Standards ontologies for the country information and GS1 standards for publishing traceability information.</p> <p><u>Technical Data:</u> W3C Verifiable Credentials, W3C Decentralized Identifiers, GS1 EPICS.</p>

<p style="text-align: center;">Best Practice/Guidelines for Data Management</p>	<p><u>Administrative Data:</u> All formal documents related to the project have to include on the front-page information about author(s), editor(s), work package, dissemination level and version in accordance with D8.1 Project Management Handbook v2_M18.</p> <p><u>Platform Data:</u> JSON</p> <p><u>Technical Data:</u> The technical data (data sets, source code and docker containers) are stored and hosted into a private git server (Gitlab) and the cloud infrastructure of UBITECH. The CI/CD paradigm and containerization is used for the platform integration and deployment.</p>
<p style="text-align: center;">Tools for Formatting Data</p>	<p><u>Administrative Data:</u> No automatic tools currently used.</p> <p><u>Platform Data:</u> JSON validation tools</p> <p><u>Technical Data:</u> GS1 Schema Validation modified JSON validators.</p>
<p style="text-align: center;">Directory and File Naming Convention.</p>	<p><u>Administrative Data:</u> Defined in Project Management Handbook. The project's Google drive has a predefined structure, as follows:</p> <ol style="list-style-type: none"> 1. Grant Agreement & Amendment 2. Consortium Agreement 3. Project Templates 4. Project Meetings <ol style="list-style-type: none"> 4.1. Bilateral Meetings 4.2. Monthly Online Meetings 4.3. Plenary Meetings 5. Review Meetings <ol style="list-style-type: none"> 5.1. Mid Term Review 5.2. Final Review 6. WPs <ol style="list-style-type: none"> 6.1. 1 Requirements 6.2. 2 Data 6.3. 3 Platform 6.4. 4 Application 6.5. 5 Legal 6.6. 6 Pilots 6.7. 7 Impact 6.8. 8 Management 6.9. 9 Ethics 6.10. Submission of Deliverables

	<p>All deliverables are stored inside of directory "6. WPs" under the relevant work package. After their submission to the EC portal, all deliverables are stored in the folder "Submission of Deliverables". In terms of file naming convention all deliverables should follow this structure:</p> <p>D[Deliverable Number] -[Deliverable Title]_[Version of Deliverable]_[Due date for submission of the deliverable]</p> <p>Example: D8.1 - Project Management Handbook_V2_M18</p> <p><u>Platform Data:</u> The git repositories follow the naming convention: [component-name]-[subcomponent name]</p> <p><u>Technical Data:</u> The git repositories follow the naming convention: [component-name]-[subcomponent name]</p>
--	---

Table 4: Data Access

DATA ACCESS	
Risks	<p><u>Administrative Data</u> Unauthorised access.</p> <p><u>Platform Data</u> Loss or destruction of data, loss of availability, loss of confidentiality.</p> <p><u>Technical Data</u> Unauthorised access.</p>
Risk Management	<p><u>Administrative Data:</u> User and Password controls implemented in the EC Portal while restricted access is given to the project's Google Drive repository. Documents stored on the Google Drive repository are protected from possible external damage and attacks.</p> <p><u>Platform Data & Technical Data</u> Technical partners adopt ISO 27001 and apply its procedures and policies for information security management. UBITECH has established and operates a certified (by TÜV AUSTRIA HELLAS) quality management framework and system. This framework is compatible with the EN ISO 9001:2008 quality standard, for the "design, development, integration, production, installation, deployment, hosting and technical support of software solutions and information technology systems". This framework is also compatible for the "design, development, execution, management</p>

	<p>and delivery of research and technological development information technology projects", determining the corporate policies and defining the quality responsibilities for all corporate processes.</p> <p>UBITECH has implemented and deployed internally a certified (by TÜV AUSTRIA HELLAS) management system for information security in accordance with the ISO 27001:2005 quality standard. The deployed information security management framework and system reassures the protection of information and information infrastructure assets of both UBITECH and the customers of UBITECH against the risks of loss, misuse, disclosure or damage.</p>
<p>Correct execution of the Access process</p>	<p><u>Administrative Data</u> Only people added to the repository can access the Google Drive. TheFSM project manager (Dimitris Fotiadis) is in charge of monitoring access to the project Google Drive Repository. Access is limited to specific persons indicated by the Partners Projects and those who are granted access by the project coordinator. Furthermore, the administrative manager (Charalambos Thanopoulous) is in charge of accessing the EC Portal.</p> <p><u>Platform Data & Technical Data</u> UBITECH manages access to technical and Platform data developed during the course of the project. UBITECH has established and operates a certified (by TÜV AUSTRIA HELLAS) quality management framework and system. This framework is compatible with the EN ISO 9001:2008 quality standard, for the "design, development, integration, production, installation, deployment, hosting and technical support of software solutions and information technology systems". Furthermore the framework is compatible for the "design, development, execution, management and delivery of research and technological development information technology projects", determining the corporate policies and defining the quality responsibilities for all corporate processes.</p> <p>UBITECH has implemented and deployed internally a certified (by TÜV AUSTRIA HELLAS) management system for information security in accordance with the ISO 27001:2005 quality standard, for the " design, development, integration, production, installation, deployment, hosting and technical support of software solutions and information technology systems ", the "implementation and management of the information technology projects", and the "delivery of Software-as-a-Service and Platform-as-a-Service services". The deployed information security management framework and system reassures the protection of information and information infrastructure assets of both UBITECH and the customers of UBITECH against the risks of loss, misuse, disclosure or damage.</p>
<p>Procedures to Follow a Data Breach</p>	<p><u>Administrative Data</u> The security of documents stored on TheFSM Google Drive Repository is based on Google policy for data security and management of data incidents. Google specifies</p>

	<p>that specific mitigation actions are put in place in a precise process to address any potential incidents affecting the confidentiality, integrity, or availability of customers' data.</p> <p>Should an incident of data breach occur, remedial action would be taken to mitigate the harm or damage, the following actions will be implemented:</p> <p>1. Immediate gathering of essential information relating to the breach by TheFSM project manager, assisted by the technical manager and the administrative manager. The dedicated team shall promptly gather the following essential information:</p> <ul style="list-style-type: none"> • When did the breach occur? • Where did the breach take place? • How was the breach detected and by whom? • What was the cause of the breach? • What kind and extent of personal data was involved? • How many data subjects were affected? • Who needs to be made aware of the breach? • Are there any methods to recover any losses and limit the damage the breach may cause? <p>2. Assessing the risk of harm Some data security breaches will not lead to risks beyond possible inconvenience, an example is where a laptop is irreparably damaged, but its files were backed up and can be recovered. While these types of incidents can still have significant consequences, the risks are very different from those posed by, for example, theft or identity fraud. Each data breach will follow the risk assessment process below:</p> <ul style="list-style-type: none"> • The kind of personal data being leaked; • The amount of personal data involved and the level of sensitivity; • The circumstances of the data breach i.e. online or traceable; • The likelihood of identity theft or fraud; • Whether the leaked data is adequately encrypted, anonymised or otherwise rendered inaccessible, e.g. if passwords are needed for access; • Whether the data breach is ongoing and whether there will be further exposure of the leaked data; • Whether the breach is an isolated incident or a systematic problem; • In the case of physical loss, whether the personal data has been retrieved before it can be accessed or copied; • Whether effective mitigation / remedial measures have been taken after the breach occurs • The ability of the data subjects to avoid or mitigate possible harm; • The reasonable expectation of personal data privacy of the data subject. <p>3. Contacting the interested parties, containment and recovery Once the risk has been assessed, the dedicated team will take actions to stop the breach and if necessary this may involve law enforcement agencies i.e. police. The following containment measures will be followed:</p> <ul style="list-style-type: none"> • Stopping the system if the data breach is caused by a system failure;
--	--

- Changing the users' passwords and system configurations to contract access and use;
- Considering whether internal or outside technical assistance is needed to remedy the system loopholes and/or stop the hacking;
- Ceasing or changing the access rights of individuals suspected to have committed or contributed to the data breach;
- Notifying the relevant law enforcement agencies if identity theft or other criminal activities are or will be likely to be committed;
- Keeping the evidence of the data breach that may be useful to facilitate investigation and the taking of corrective actions.

4. Notification of breaches

In case of a personal data breach, without undue delay and where feasible we aim to notify the data subject within 72 hours of becoming aware of the breach.

5. Review of the Incident

It is important not only to investigate the causes of the breach but also to evaluate procedures taken to mitigate possible future incidents. TheFSM project attempts to learn from the experience, review how data collected is being handled to identify the roots of the problem, allow constant review to take place and to devise a clear strategy to prevent future recurrence.

The review will take into consideration:

- Ongoing improvement of security in the personal data handling processes;
- The control of the access rights granted to individuals to use personal data. Are principals "need-to-know" and "need-to-access" being adopted;
- The adequacy of the IT security measures to protect personal data from hacking, unauthorised or accidental access, processing, erasure, loss or use;
- Ongoing revision of the relevant privacy policy and practice in the light of the data breach
- The effective detection of the data breach. The keeping of logs and trails of access enabling early warning signs to be identified;
- The strengthening of the monitoring and supervision mechanism of data users, controllers and processors;
- Review of the ongoing training to promote privacy awareness and to enhance the prudence, competence and integrity of the employees particularly those who act as controllers and processors;
- Review of this policy and procedures listed.

Platform Data & Technical Data

Step 1: Contain the data breach to prevent any further compromise of personal information.

Step 2: Assess the data breach by gathering the facts and evaluating the risks, including potential harm to affected individuals and, where possible, taking action to remediate any risk of harm.

Step 3: Notify individuals and the Commissioner if required. If the breach is an 'eligible data breach' under the NDB scheme, it may be mandatory for the entity to notify.

	<p>Step 4: Review the incident and consider what actions can be taken to prevent future breaches.</p> <p>The Data Protection Officer and Chief Information Security Officer of UBITECH manage this procedure.</p>
--	---

Table 5: Data sharing and re-use of data to be published

DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION	
Re-Use of Data	<p><u>Administrative Data</u> Only consortium partners use confidential project documents. Public deliverables are published on the project website (https://foodsafetymarket.eu/) and in Zenodo (https://zenodo.org/) for public use.</p> <p><u>Platform Data</u> Metadata of food safety data records.</p> <p><u>Technical Data</u> Open Source application code and Open Linked Data.</p>
Organization/ Labelling of Data for Easy Identification	<p><u>Administrative Data</u> All administrative data are organised in 6 folders in Google Drive. In terms of labelling all deliverables should follow this structure: D[Deliverable Number] -[Deliverable Title_[Version of Deliverable]_[Due date for submission of the deliverable] Example: D8.1 - Project Management Handbook_V2_M18</p> <p>Furthermore, all public deliverables are uploaded in Zenodo (https://zenodo.org/) after the submission in the EC portal. Zenodo provides a unique DOI for each deliverable, ensuring its valid and safe disposal. A DOI (Digital Object Identifier) is a unique and never-changing string assigned to online (journal) articles, books, and other works. DOIs make it easier to retrieve works, which is why citation styles, like APA and MLA Style, recommend including them in citations. For each deliverable, extra labels (keywords) are also inserted when uploaded.</p> <p><u>Platform Data</u> Catalogue of the Food Safety Marketplace.</p> <p><u>Technical Data</u> Open Linked Data structured according to W3C open standards and with relevant identifiers and tags for querying.</p>
Data Sharing Requirements	<p><u>Administrative Data</u> No requirements identified for publicly available data published on the project website. However, for documents uploaded in the Zenodo repository there is a standard curation policy: "All public documentation and/or dissemination material produced by the TheFSM project will be included under this community. Any material either not belonging to the above category or characterized as restricted</p>

	<p>will be declined". The responsible curator is the TheFSM project coordinator (Agroknow).</p> <p><u>Platform Data</u> Metadata of the datasets will be available in machine-readable format following standards such as DCAT.</p> <p><u>Technical Data</u> Conformity with open data models.</p>
Audience for Re-use	<p><u>Administrative Data</u> In the case of confidential project documents, the audience for re-use is limited to Consortium Partners. In the case of publicly disseminated deliverables, the audience for re-use is the public, specifically actors in the food supply chain and researcher.</p> <p><u>Platform Data</u> Researchers, Technology companies in the food industry, stakeholders in the food industry.</p> <p><u>Technical Data</u> Developers of food supply chain IT systems, auditors, customers.</p>
Restrictions on Re-use of Data	<p><u>Administrative Data</u> Personal data and confidential project documents will not be re-used.</p> <p><u>Platform Data</u> Restrictions defined by commercial subscription licenses, data from National Authorities will be subject to non-commercial use restrictions.</p> <p><u>Technical Data</u> Restrictions defined by the data owners.</p>
Publication	<p><u>Administrative Data</u> Accessible for consortium partners.</p> <p><u>Platform Data</u> Food Safety Marketplace.</p> <p><u>Technical Data</u> Open Source application code, available on GitHub for reuse and review. Open Linked Data published to the Origin Trial decentralized network.</p>

Table 6: Data preservation and archiving

DATA PRESERVATION AND ARCHIVING	
Archiving of Data for Preservation and Long-term Access	<p><u>Administrative Data</u> Google Drive, EC Portal and local repositories.</p>

	<p><u>Platform Data</u> Internal ERP – Archives, Cloud-based repositories, internal servers</p> <p><u>Technical Data</u> Partners (UBITECH, Agroknow, SIRMA) Cloud infrastructure, OriginTrail Decentralised Knowledge Graph,</p>
Data Retention	<p><u>Administrative Data</u> 5 years after the end of the project in accordance with the requirements of the Grant Agreement.</p> <p><u>Platform Data</u> 3- 5 years</p> <p><u>Technical Data</u> 1-5 years</p>
File Formats	<p><u>Administrative Data</u> Google Drive repository, PDF files, Word Docs, Excel</p> <p><u>Platform Data</u> Digital Archives, Database/backup, PDF files, Excel, CVS files, DB records</p> <p><u>Technical Data</u> NoSQL databases MongoDB, Apache Hbase, Apache Cassandra, JSON-LD, XML, images Apache Hive, Hadoop Distributed File System (HDFS)</p>
Data Archives	<p><u>Administrative Data</u> EC Portal</p> <p><u>Platform Data</u> Local institutional repositories</p> <p><u>Technical Data</u> Orgin Trail Decentralised Knowledge Graph</p>

<p>Long-term Maintenance of Data</p>	<p><u>Administrative Data</u> After the project ends, all the administrative data will still be online and accessible in the Google Drive repository and maintained, for at least five years. After that period will be archived and kept in a safe storage of the coordinator for at least an extra 5 years. Moreover, the long-term availability and maintenance of the data is guaranteed for all the public deliverables of the project, since every deliverable is uploaded in the EC portal and in Zenodo repository.</p> <p><u>Platform Data</u> FSM Coordinators (Agroknow), Certification bodies supervised by DPO, ISO 27001 policies</p> <p><u>Technical Data</u> Data is maintained by Technical Partners through the OriginTrail Decentralised Network and their internal systems</p>
---	--

A concise survey of the key features of the life cycle of Pilot data is contained in Deliverable 6.1.

3. FAIR DATA

3.1 FAIR Management of Research Data

TheFSM is committed to following the “Guidelines on FAIR Data Management in Horizon 2020”², recommended by the European Commission Directorate – General for Research & Innovation. These Guidelines foresee four principles that govern the management of research data in order to make them more easily understood, exchanged and open for re-use. Scientific data management thus has to ensure that research data is FAIR, meaning findable, accessible, interoperable and re-usable.

3.2 Research data lifecycle



Figure 2: Stages of data during a research process³

Planning research:⁴

- design research
- plan data management

² https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

³ See UK Data Service: Research data lifecycle <https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx>

⁴ See UK Data Service: Research data lifecycle <https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx>

- plan consent for sharing
- plan data collecting, processing protocols and templates
- explore existing data sources

Collecting data:

- collect data
- capture data with metadata
- acquire existing third party data

Processing and analysing data:

- Enter, digitize, transcribe and translate data
- check, validate, clean, anonymize
- derive data
- describe and document data
- manage and store data
- analyse and interpret data
- produce research outputs
- cite data sources

Publishing and sharing data:

- establish copyright
- create user documentation
- create discovery metadata
- select appropriate access to data
- publish/share data
- promote data

Preserving data:

- migrate data to best format/media
- store and backup data
- create preservation documentation
- preserve and curate data

Re-using data:

- conduct secondary analysis
- undertake follow-up research
- conduct research reviews
- scrutinize findings
- use data for teaching and learning

3.3 The FAIR principle⁵

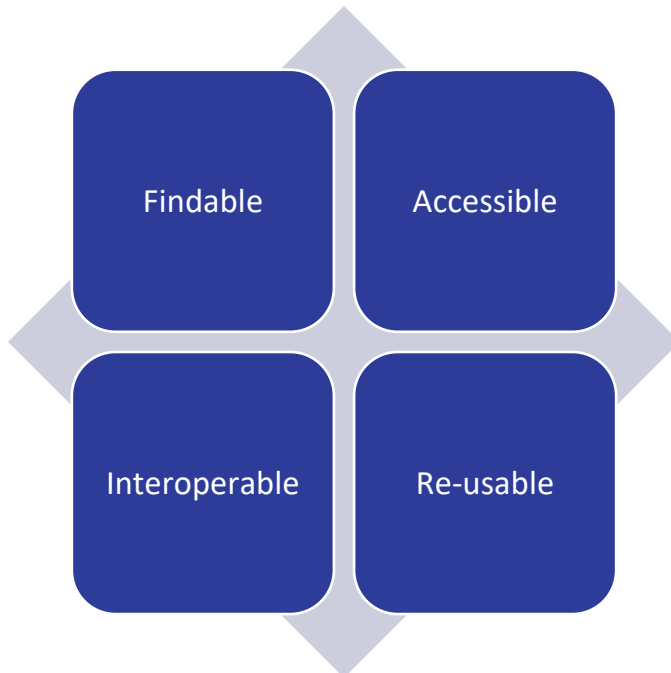


Figure 3: The Fair Principle

This FAIR principle defines how research outputs should be organised in order to adequately facilitate discovery, exchange and reuse by third parties. The elements of the FAIR Principles are related, but independent and separable. Major funding bodies, including the European Commission, promote FAIR data to maximize the integrity and impact of their research investment.⁶

TheFSM project aims to develop an industrial data platform for the exchange of business sensitive data in the food safety market to boost the competitiveness of the European food certification. The commercial objective of the project itself can sometimes collide with the FAIR principles. TheFSM is however committed to participate in the Pilot on Open Research Data in Horizon 2020, in a way that respects the security and privacy requirements of the pilots and the commercial interests of the industrial partners. As previously mentioned, a pilot specific Data Management Handling Plan has been developed in Work Package 6 to address the management of data that is processed in the course of TheFSM pilots. The Pilot DMHPs are documented in Work Package 6, Deliverable 6.1.

⁵ The Open Data Foundation <http://www.odaf.org/>

⁶ The Open Data Foundation <http://www.odaf.org/>

3.3.1 Findable

Data and metadata should be easily findable for humans as well as for computers. A crucial aspect for data to be findable is whether “the data produced or used in the project is discoverable with metadata, identifiable and locatable by means of a standard identification mechanism.”⁷ To be findable any Data Object should be uniquely and persistently identifiable. The same Data Object should be persistent and contain basic machine actionable metadata that allows it to be distinguishable from other Data Objects. Identifiers for any concept used in Data Objects should thus be Unique and Persistent. To make data findable the following procedure should be considered:

- **F1.** (Meta)data are assigned a globally unique and persistent identifier
- **F2.** Data are described with rich metadata (defined by R1 below)
- **F3.** Metadata clearly and explicitly include the identifier of the data they describe
- **F4.** (Meta)data are registered or indexed in a searchable resource⁸

In TheFSM, a variety of datasets are generated and collected. The data collection and processing workflow include discrete steps in which different versions of the data are stored. The original data collected from the data sources are stored and maintained in one-step to ensure that the provenance of the data can be traced back. Metadata for each data source is stored in a catalogue. After applying processing methods, a new version of the data is stored. Each set of data produced (dataset, deliverables, etc.) will be named in a uniform way and will include a table with a version control.

The approach to naming documents of the project are as follows:

- Choose easily readable identifier names (short and meaningful);
- Do not use acronyms that are not widely accepted;
- Do not use abbreviations or contractions;
- Avoid Language-specific or non-alphanumeric characters;
- Add a two-digit numeric suffix to identify new versions of one document.
- Dates should be included back to front and include the four-digit years: YYYYMMDD.

For deliverables: Project_ [Deliverable Code]-[Deliverable Title]_[Partner]-vA.BB i.e.: Project_D6.1-Project Management Handbook-v1.00 (for submission to the Commission)

For datasets: WP [Work Package number] P [Pilot number; pilot activity number] - [description of the activity] i.e.: WP4 P1.3 Results of demonstration performance.

With regard to project related contractual documents, copies are kept on the project Google Drive. Partners maintain their collected data and metadata about the collected data in different

⁷ See https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

⁸ See <https://www.go-fair.org/fair-principles/>

versions with backup versions e.g. in the Google Drive, GitHub or partner's cloud infrastructure (UBITECH Cloud/Edge Experimental Infrastructure). The Project repository provided through Google Drive serves as a collective working space where the partners can exchange documents and collaborate in the development of deliverables. Deliverable names/identifiers are agreed with the European Commission and documented in the Grant Agreement. Information about the source of the data sets will be tracked on a data inventory Google Sheet.

Metadata fields such as country, product, hazard and summary of a food safety record can be created automatically using text mining methods. In the context of technical and platform data, several standards for the documentation of metadata are adopted in the project. These standards include W3C Verifiable Credentials, W3C Decentralized Identifiers and GS1 EPICS.

The platform structure itself and the data that will be used for and transferred via the platform is still in the development process. TheFSM will make metadata available under a commercial subscription license. The standards used include DCAT metadata standard for publishing the metadata of the food safety records, Standard classifications for the topics such as FOODEX2 and AGROVOC, Standard ontologies for the country information and GS1 standard for publishing traceability information. TheFSM will follow the best practice of publishing metadata and data using a well-accepted format such as JSON.

3.3.2 Accessible

Accessibility requires that machines and humans upon appropriate authorisation through well-defined protocols can always obtain data. It should be determined what generated or used data will be made openly available. Where datasets cannot be shared or are restricted a clear explanation is required highlighting the legal or contractual reasons for the restriction of accessibility.

- **A1.** (Meta)data are retrievable by their identifier using a standardised communications protocol
 - **A1.1** The protocol is open, free, and universally implementable
 - **A1.2** The protocol allows for an authentication and authorisation procedure, where necessary
- **A2.** Metadata are accessible, even when the data are no longer available⁹

In the context of administrative data, deliverables and reports, which are marked as public in the Grant Agreement, will be published on the project website. In addition, some technical data will be made openly accessible. This includes open source application code available on GitHub for reuse and review as well as open linked data that is published to the Origin Trial decentralized network.

⁹ See <https://www.go-fair.org/fair-principles/>

Data access will vary depending on the storage location. Public deliverables and reports contained on the project website will be accessible to the public. Concerning the use case data, measures will be taken to enable third parties to access, re-use, analyse, exploit, and disseminate the data (bound by the license specifications). Different access procedures will be implemented as indicated in the **Table 4**, enabling the export of an entire dataset as well as the provision of a querying interface for the retrieval of relevant subsets. Access mechanisms will also be supported as much as possible by metadata enabling search engines and other automated processes to access the data using standard web mechanisms.

On the partner's side, IT departments, quality assurance teams, systems administrators and data protection officers will guarantee the correct execution and give access only to staff involved in the project. While the partners themselves follow the respective company's standard, a risk assessment formalization to guarantee safety of data for the platform is in progress.

Private project related administrative data is stored on the Google Drive Repository and the project coordinator monitors access. The project coordinator grants access to the repository to the partners actively involved in the development of the project.

The overall objective of TheFSM is to create an industrial data platform. This platform-to-be aims to facilitate the cross-border exchange of data between different food safety actors, and thus to digitalize and accelerate the food certification process. Due to the industrial nature of the project, large amounts of data exchanged are business sensitive and will therefore be protected by contractual agreements and license agreement, which are developed in WP5, Deliverable 5.1 and 5.2. Access to TheFSM platform will be provided upon positive confirmation of the required access criteria, verified accounts and login credentials.

In TheFSM, access to data assets will be regulated through Attribute-Based Access Control (ABAC) policies, based on the XACML OASIS standard that allows the data providers to protect and share their data assets, even when they do not have any prior knowledge of the potential individual data consumers in the food certification data value chain. XACML promotes common terminology and interoperability between access control implementations by multiple vendors.

As a general principle, the consortium is going to reuse conceptualizations and adopt broader standards where possible (dcterms, foaf, etc.). As the project supports a Linked Data approach, when applicable, the vast majority of resulting datasets are expected to comply with semantic standards (RDF/S), and additional standardisation activities done by the World Wide Web consortium (W3C), such as OAI-ORE's JSON-LD implementation.

TheFSM will generate its own valuable data assets in terms of metadata that will improve the description, interlinking, normalization, unification, and quality assessment of the collected datasets. The use of W3C standards such as PROV-O for provenance, and DCAT for data catalogue description will be encouraged.

Basic metadata will be used to facilitate the efficient recall and retrieval of information by project partners and external evaluators and contribute to easily find the information requested. To this end, all documents related to the project have to include in the front-page information about

author(s) & editor(s), WP, dissemination level and version. In case data and/or integrate services need to be exchanged with other partners as part of the technical implementation of s/w products, the swagger framework for API documentation is used.

Standards used for the documentation of metadata will include W3C Verifiable Credentials, W3C Decentralized Identifiers and GS1 EPCIS; tools for checking that the data are well formatted include GS1 Schema Validator, modified JSON validators. Regarding the project and data identifiers, already existing supply chain specific identifiers will be assigned (GS1 Identifiers); the assigning of identifiers will conform to the W3C Decentralized Identifiers recommendation. The community or industry standard for metadata sharing/integration used are GS1 standards related to product data sharing (master data) and W3C PROV for provenance of information.

Digital data will be released in machine-readable formats that supplement journal articles and presentations, sharing requirements that are in conformity with open data models. Metadata of the datasets will be available in machine-readable format following standards such as DCAT.

3.3.3 Interoperable

Interoperability refers to allowing data exchange and re-use between researchers, institutions, organisations, countries and other such parties. This element requires adherence to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins. The metadata vocabularies, standards and methodologies must ensure interoperability.

- **I1.** (Meta) data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- **I2.** (Meta)data use vocabularies that follow FAIR principles
- **I3.** (Meta)data include qualified references to other (meta)data¹⁰

The objective of the TheFSM project is to develop an industrial platform for the exchange of business sensitive data in the food certification process. One goal is to connect various different systems to achieve interoperability (system interoperability and data interoperability in particular). The technical partners in the consortium are therefore seeking to apply a generic approach for the infrastructural technical requirements needed for TheFSM, as well as to set fundamental technical boundaries in a way to provide the utmost value to end users and enable the stakeholders to take part in trusted data exchanges. Facts regarding the platform structure are further defined in WP2, WP3 and WP4.

Open source application code will be made re-usable or openly accessible, available on GitHub for re-use and review, and open linked data published. For researchers to be able to isolate their fields of interest in their study the open linked data is structured according to open standards and with relevant identifiers and tags for querying.

¹⁰ See <https://www.go-fair.org/fair-principles/>

3.3.4 Re-usable

Re-usability means that data can be easily utilised by third parties. The final goal of the European Commission's FAIR principles policy is to optimise the re-use of data. It is therefore important to determine how the data will be licensed to permit the widest re-use possible, when data will be made available for re-use, the period of the intended reuse and data quality assurance processes. In cases where an embargo is sought to give time to publish or seek patents, it will be specified why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

- **R1.1.** Meta(data) are richly described with a plurality of accurate and relevant attributes
- **R1.1.** (Meta)data are released with a clear and accessible data usage license
- **R1.2.** (Meta)data are associated with detailed provenance
- **R1.3.** (Meta)data meet domain-relevant community standards

During the TheFSM project, an industrial data platform will be developed to facilitate the data exchange between relevant actors in the certification process. Interested audience will include researchers (e.g. researchers developing risk estimation and prediction models), stakeholders in the food safety certification process, parties in the supply chain, technology companies in the food industry, stakeholders in the food industry, auditors or customers.

Public deliverables will be openly accessible on the project website and as such are open for re-use by all interested researchers and institutions. The website is available at the following URL: <https://foodsafetymarket.eu/>.

All partners in the consortium are committed to engage in dissemination activities such as webinars, papers and presentations to promote the project and its findings within the scientific and the food safety community. The output will be made openly accessible for re-use whenever it is in alignment with the organizing institution. Regarding scientific outcomes, the consortium is committed to the Horizon2020 Open Access mandates and is planning to embrace all possible Open Access options known today. These include Gold Open Access, Green Open Access and self-archiving. The Consortium partners will therefore give preference to Open Access journals or non-Open Access journals that support Green and Gold roads.

The data repositories used for storing, managing and disseminating research data sets will comply with the requirements of the EC Guidelines on Open Access to Scientific Publications and Research Data in H2020.¹¹ The intention of the consortium is to establish a multi-level approach towards managing the knowledge produced. Technology services will be copyright protected using a licensing scheme that is not violating the terms and conditions of the discrete components comprising it (e.g. the GraphDB software). Some components will be provided as open source implementations (e.g. the OriginTrail protocol) and delivered under such a license (e.g. CC-BY).

¹¹ See https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

The respective knowledge producer will keep the rights to the knowledge they have produced, whereas the software implementation will be shared with the rest of the consortium.

While the platform is still under development, possible licenses considered include the Apache Software License¹² (ASL) license and the LGPL software licenses.¹³ The ASL license¹⁴ seems to be suitable for the components and modules that can be delivered open source, as this license allows the redistribution of the program's source code in any form (compiled binary or plain text) without posing any limitations to the distributor. If released under the ASL license, the corresponding modules can generate business interest around future expansions and attract the attention of research communities. Components and modules that cannot be delivered open source will be copyright protected but of course will be freely available to consortium members to use for the production of foreground.

Administrative project data concerning management and finances is confidential within the consortium and not meant for re-use of third parties. All personal data collected in the course of the project will be kept confidential unless the informed consent of the relevant data subject is obtained for disclose. Given the business sensitive nature of the non-personal data collected in the project, re-use will be restricted to comply with contractual commitments such as non-disclosure agreements, license agreements and consent requirements. Platform Users can and will restrict the re-usability of business sensitive data which is exchanged during the course of the project. Plans for the re-use of data after the end of the project have still to be defined by the consortium.

One set of metadata that will be shared openly are the food safety data records collected from the National authorities (food recalls, border rejections, inspection results, country indexes). The metadata for these food safety records will be available at the catalogue of the Food Safety Marketplace. Standard classifications for origin, hazards and products will be used to facilitate the discovery of the field of interest.

3.3.5 Allocation of resources

The consortium will rely on its research funding in order to make scientific and research data FAIR. The effort necessary to make the data FAIR will be covered by the budget assigned to a partner or partners responsible for a specific task and/or for producing the relevant deliverable(s).

¹² See <https://www.apache.org/licenses/LICENSE-2.0>.

¹³ See <https://fossa.com/blog/open-source-software-licenses-101-lgpl-license/>.

¹⁴ See https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

4. DATA SECURITY

4.1 Administrative Data

Private administrative data will be shared between members of TheFSM consortium.

In addition to the external-facing project web site, a project-only collaboration space is set up using Google Drive, with a number of templates for reporting, deliverables, etc. This space serves as a private document repository that will be accessible only to partners, so that they can access and share all research and project documentation from final deliverables through to presentations and other relevant information. The project coordinator monitors access to the Google Drive. 256-bit Advanced Encryption Standards (AES) employed on all Google Drive servers protect the documents stored on the Google Drive. Furthermore, when data is in transit between users and Google Drive servers, Google uses the Transport Layer Security (TLS) protocol to prevent interception.¹⁵

In accordance with the Grant Agreement the data will be stored for a period of 5 years after the completion of the project, thereafter the data will be deleted.

4.2 Platform Data

Data security is of major importance in TheFSM project. Special attention will be given to the security of personal and business sensitive data. The protection of data will be ensured through procedures and appropriate technologies.

Security challenges associated with user authorization and access control to TheFSM platform will be addressed by designing an Authorization & Access Control Engine responsible for implementing the logical access control that prevents unauthorized access of any type of resource of the platform, including amongst others the data, the (cloud) services, and software applications (T3.2).

Security safeguards will complement these authorization and access control safeguards. They will guarantee security and integrity of information at-rest, namely of the data stored within the TheFSM platform storage, and of information in-transit, namely the security of the connection between the stakeholders, SSL connection, and the integrity of the data that will be transferred, e.g. by using asymmetric cryptography (T3.3).

Partners responsible for technical development and implementation of TheFSM platform also comply with and are certified under international quality standards, such as ISO 27001: 2005, providing requirements for an information security management system (ISMS). The deployed information security management framework and system reassures the protection of information and information infrastructure against the risks of loss, misuse, disclosure or damage. The data will be stored in a protected database with daily backups. If data will be kept in an external certified repository, then the security standards of that repository will apply.

¹⁵ Google Security Whitepaper, services.google.com/fh/files/misc/google_security_wp.pdf

4.3 Technical Data

As the lead technical partner, UBITECH has implemented and deployed internally a certified (by TÜV AUSTRIA HELLAS) management system for information security in accordance with the ISO 27001:2005. The deployed information security management framework and system reassures the protection of information and information infrastructure assets the risks of loss, misuse, disclosure or damage. Project related datasets such as software code would be stored in secure database where only authorized people from the consortium can access. The data will be backed up daily in a secure storage server. In instances where data is kept in an external certified repository, then the security standards of that repository will apply. Technical data will be retained for a period of 1-5 years.

5. PROTECTION OF PERSONAL DATA AND ETHICAL ASPECTS

The commitment to legal and ethical principles is a central concern of all research activities funded by the European Union. Thus, all activities carried out under the Horizon 2020 Framework Programme have to be legally and ethically compliant from beginning to end. Applying ethical principles and legislation to scientific research is a fundamental aspect of the data management policy adopted in TheFSM project.

All personal data processed in the course of TheFSM project is subject to the protection of the General Data Protection Regulation (EU) 216/679 (GDPR).¹⁶ The GDPR is the most fundamental piece of European legislation where ethical considerations are enshrined. The GDPR is a binding legislative act; as such, it must be applied in its entirety across the EU. The GDPR applies whenever personal data is processed by organisations established in the EU whether they are functioning as processor or controller. Under certain conditions the GDPR applies even to companies that are not in Europe, e.g. when data processing activities are related to offering goods or services (even if for free) to data subjects situated in the EU (not restricted to EU citizens) and to monitoring of the behaviour of such data subjects. In TheFSM all Partners have an establishment in the EU and are thus subject to the rules set out in the GDPR.

The main purpose of the GDPR is the protection of natural persons with regard to the processing of their personal data.¹⁷ Personal data means any information relating to an identified or identifiable natural person ('data subject').¹⁸ Personal data processed wholly or partly by automated means or in a non-automated manner which forms part of, or is intended to form part of, a 'filing system' fall under the GDPR. However, only natural persons (including employees of businesses and public authorities) are protected by the GDPR. In the context of TheFSM project, personal data processed is limited to contact information such as names, email address and phone numbers. In this context, the relevant data subjects are the employees of Consortium Partners who are involved in the project and the individuals who participate in the Project focus groups and Pilots. Once TheFSM platform is fully developed and available for public use, the personal data of platform Users will be processed in accordance with the platform's Privacy Policy that is developed in WP5, deliverable 5.1.

5.1 Principles of Personal Data Processing

Article 5 of the GDPR lays down **seven principles** that have to be applied with regard to the processing of personal data. By respecting these principles, the requirement of legal compliance

¹⁶ Regulation (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

¹⁷ Art 1 (1) GDPR.

¹⁸ Art 4 (1) GDPR.

can be met and the necessary level of accountability and protection maintained. The rights of data subjects will thus be ensured.

1. **Lawfulness, fairness and transparency** — processing must be lawful (based upon a lawful basis), fair and transparent to the data subject.
2. **Purpose limitation** — data can only be processed for the legitimate purposes specified explicitly to the data subject when the data was collected.
3. **Data minimisation** — only as much data as necessary must be collected and processed for the purposes specified.
4. **Accuracy** — personal data must be accurate and up to date.
5. **Storage limitation** — personally identifying data must be stored only for as long as necessary for the specified purpose.
6. **Integrity and confidentiality** — processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g. by using encryption).
7. **Accountability** — the data controller is responsible for being able to demonstrate GDPR compliance with all of these principles.¹⁹

5.2 Lawfulness of Processing

For TheFSM as for every other actor a valid lawful basis is required for any processing of personal data. In principle, a lawful basis requires that processing is 'necessary' for a specific purpose. If that same purpose can be reasonably achieved without the processing, there is a lack of lawful basis. It is essential that this lawful basis be established before processing the data. The six lawful bases are outlined in Article 6/1 of the GDPR. According to the GDPR processing of personal data is only lawful if

- a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;
- b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;
- c) processing is necessary for compliance with a legal obligation to which the controller is subject;
- d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;
- e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
- f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests

¹⁹ Art 5 GDPR.

or fundamental rights and freedoms of the data subject, which require protection of personal data, in particular where the data subject is a child.²⁰

Whenever there is none of the above-mentioned legal bases, data processing, in particular data sharing of personal data is prohibited (Prohibition principle with permission reservation). To be in full compliance with the GDPR TheFSM must abstain from all data processing of personal data without a valid lawful basis.

Consent forms the legal basis for processing personal data, when external participants become involved in the project. This occurs through participation in project related focus groups or in the Pilots. Article 4 GDPR defines consent as a *“freely given, specific, informed and unambiguous indication of the data subject’s wishes by which he or she, by a statement of by a clear affirmative action, signifies agreement to the processing of personal data”*. In the course of the project, the informed consent of participants is sought through consent forms and corresponding information sheets. Participants in focus groups and pilots are required to sign the consent forms prior to participating and prior to the processing of their personal data. The consent forms provide the participant with detailed information about the envisaged data processing in an intelligible and easily accessible form, using clear and plain language (Article 7(2), GDPR). Deliverable 9.2 contains the templates of the informed consent forms that have been implemented in the project. These templates have been used in project focus groups and have been adapted for the pilots.

5.3 Integrity and Confidentiality

According to the GDPR all necessary precautions must be taken when processing personal data so that the **appropriate security of the data** is ensured, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures”. (Art 5(1) (f)). The protection of personal data intended by the GDPR goes from the strict application of the accountability principle to increased transparency and simpler information policies. Evidently, this principle is of utter importance for TheFSM.

Appropriate technical and organisational measures by the partners of TheFSM are required to provide **safeguards already in the stage of the establishment** of the platform.

[...] the controller should adopt internal policies and implement measures, which meet in particular the principles of data protection by design and data protection by default. Such measures could consist, inter alia, of minimising the processing of personal data, pseudonymising personal data as soon as possible, transparency with regard to the functions and processing of personal data, enabling the data subject to monitor the data processing, enabling the controller to create and improve security features. When developing, designing, selecting and using applications, services and products that are based on the processing of personal data or process personal data to fulfil their task, producers of the products, services and applications should be encouraged to take into account the right to data protection when developing and designing such products,

²⁰ Art 6 (1) GDPR.

services and applications and, with due regard to the state of the art, to make sure that controllers and processors are able to fulfil their data protection obligations.²¹

These measures include amongst others the principles of data protection by design and data protection by default.²² The principles of data protection by design and data protection by default have been implemented in all developmental stages of the project and in the establishment of the platform for TheFSM.

5.4 Accountability

Accountability in the context of the GDPR requires the implementation of **appropriate technical and organisational measures** and the ability **to demonstrate the effectiveness of the measures when requested**. As such, TheFSM should therefore be prepared to be able to prove its compliance with the GDPR and implement accordingly a complete personal data protection system (budget, compliance tools, procedures, staffing, technology & security) appropriately to the risk that TheFSM is generating.

Instruments implemented by the consortium to achieve accountability include:

- Informed consent forms and corresponding information sheets which contain information on what and how personal data are processed (which data is collected, how, to what purpose, how long, how it is used, where it is stored, who is responsible for it etc.).
- Informing participating Data Subjects about their rights contained in Art 12 – 23 of the GDPR. This information is provided in the informed consent forms.
- Access control procedures monitored by the project coordinator in order to protect personal data that is stored on the project Google Drive repository.
- Designated data protection responsibilities delegated to UNIVIE as the legal partners of the project.
- Joint Controllershship Agreement in place to govern the processing of personal data in the context of TheFSM pilots (further information reported in WP6, deliverable 6.1).
- Declaration of Compliance, signed by all consortium partners (reported in WP9, deliverable 9.1).
- Appointment of Data Protection officers by the partners who are responsible for the processing of personal data (reported in WP9, deliverable 9.4).

5.5 Responsibilities of the controller and the processor

Of utmost importance for the developmental activities of the partners in TheFSM are the responsibilities of the controller and processor outlined in the GDPR. **Both data controllers and data processors have obligations** under the GDPR. Fundamentally, data controllers have more accountability and liability, but processors do have new responsibilities and added layers of liability. The **data controller defines the purposes** for which **and the means** by which personal data is processed. The **data processor processes** personal data **only on behalf of the controller**.

²¹ Recital 78 GDPR.

²² Art 25 GDPR.

In addition, his duties towards the controller must be specified in an **agreement**. Typical activities of processors are to offer IT solutions, including cloud storage.

5.6 Joint Controllers

In the context of the Pilots, TheFSM pilot leaders jointly determine the “why” and the “how” of the processing of personal data. Therefore, they are classified as **Joint Controllers**. In this case, a **joint controllership agreement** has to be made between the partners wherein their respective responsibilities for complying with the GDPR rules are laid down. The joint controllers have the duty to disclose all necessary information to ensure fair and transparent processing. Data subjects whose data is being processed shall be informed about the relevant aspects of the agreement and can exercise their rights under the GDPR against each of the joint controllers.²³ A Joint Controllership Agreement has been developed to govern the processing of personal data in the course of the Pilots; this agreement is contained in the legal documents of WP6.

5.7 Transfer to third countries

Any transfer of personal data to recipients in a third country or international organisation is prohibited unless they fully comply with the conditions set out in Chapter V of the GDPR. (Art 44-50) The scope of the GDPR is to protect natural person’s data. Although the GDPR recognises the necessities of international trade and international cooperation and the increased flow of data, the core of the GDPR shall not be undermined by transferring data to and from countries outside the Union.²⁴ Thus the transfer of personal data to a third country or to an international organisation, known as ‘data export’,²⁵ is not allowed unless the jurisdiction in which the recipient is located is deemed to provide an adequate level of data protection, the data exporter puts in place appropriate safeguards or a derogation or exemption applies.

A **third country** is a country other than the EU member states and the three additional EEA countries (Norway, Iceland, and Liechtenstein) which have adopted a national law implementing the General Data Protection Regulation (GDPR). TheFSM is planning to run **pilots in Jordan and Egypt**. Both are considered third countries. Thus, any transfer of personal data to and from Jordan or Egypt are subject to the restrictions set out in the GDPR. Recognising these limitations, no personal data of EU participants will be transferred to Jordan and Egypt. The pilots that take place in third countries will involve non-EU participants and the transfer of personal data will be strictly prohibited. This information is communicated to participants in the informed consent forms.

Prospeh BDG located in Belgrade, Serbia, will have access to personal data on behalf of one of the partner’s in the consortium for software development and system maintenance purposes.

In addition, **Agroknow** will transfer some of the information obtained by user and the organisation or company with third parties who process them in order to optimize the Agroknow System and will provide to the user with custom-made services tailored to your needs. The services for which Agroknow may use third parties include Cloud and hosting services, Email sending

²³ Article 26 (3) GDPR

²⁴ Recital 101 GDPR.

²⁵ Gawronski (ed.) (2019): Guide to the GDPR, p. 99.

services, Online payment services, Analytics services, Online chat services, CRM services, Internet security. Some of these service providers may be outside E.U.

If a third country has not yet been approved by the Commission by means of an adequacy decision, like in the case of Serbia, Egypt and Jordan, this does not necessarily foreclose any data transfer to this country. The transfer of personal data to that third country from the EU is still allowed, if there are certain **safeguards** in place. These safeguards include:

- a) a legally binding and enforceable instrument between public authorities or bodies;
- b) binding corporate rules in accordance with Article 47;
- c) standard data protection clauses adopted by the Commission in accordance with the examination procedure referred to in Article 93(2);
- d) standard data protection clauses adopted by a supervisory authority and approved by the Commission pursuant to the examination procedure referred to in Article 93(2);
- e) an approved code of conduct pursuant to Article 40 together with binding and enforceable commitments of the controller or processor in the third country to apply the appropriate safeguards, including as regards data subjects' rights; or
- f) an approved certification mechanism pursuant to Article 42 together with binding and enforceable commitments of the controller or processor in the third country to apply the appropriate safeguards, including as regards data subjects' rights. (Art 46 (2));
- g) contractual clauses between the controller or processor and the controller, processor or the recipient of the personal data in the third country or international organisation (Art 46(3)(a));
- h) provisions to be inserted into administrative arrangements between public authorities or bodies, which include enforceable and effective data subject rights (Art 46(3) (b)).

The crucial element in all data transfer to third countries is that **the controller must ensure** by other means **that the personal data of the data subject will be sufficiently protected** by the recipient.

The most common grounds for data export are **standard data protection clauses** (in case of established business-to-business relations) or a **contract** (in case of business to employee relations, business to customer or similar).²⁶ Contractual clauses or provisions need permission from the respective supervisory authority. With legally **binding corporate rules**,²⁷ applying to every party involved in the data processing companies are allowed to transfer data to third countries without adequacy decision. The binding corporate rules have to specify at the least who will be affected by the transfer, what types of personal data are being transferred and how the information about the data will be communicated. (Art 47/2) The rules have to also confer rights on the data subjects whose data is being transferred to a third country. The respective supervisory authority shall approve a company's rules if they meet the criteria outlined in the GDPR (Art 47/1).

The partners that use service providers from a third country have to make sure that they have data sharing agreements in place with these providers and all the providers are compatible with GDPR.

²⁶ Gawronski (ed.) (2019): Guide to the GDPR, p. 102.

²⁷ Recital 110 GDPR.

At this stage of the project, no third country service providers are being used. However, independent pilots will take place in Egypt and Jordan. Pilots that take place in third countries are facilitated by affiliated entities of Consortium Members, TUV AU Hellas, TUV AU Romania and TUV AU Cyprus (TUV AU Group). The pilots that take place in third countries will involve non-EU participants and the transfer of personal data will be strictly prohibited.

5.8 Ethical Aspects

All ethical requirements in the project are covered under WP9. To ensure ethical compliance, in particular the following measures have been implemented:

- All Partners have checked and verified if a declaration on compliance and/or authorisation is required under national law for collecting and processing personal data. All declarations of compliance and/or authorisation are documented in Deliverable 9.1.
- Templates of informed consent forms/assent forms and information sheets are documented in Deliverable 9.2.
- Each Partner, which is legally required to appoint a Data Protection Officer (DPO), has confirmed said appointment; this information is documented in Deliverable 9.4. All Partners in the consortium have an appointed DPO.
- The data minimisation policy implemented in the project is documented in Deliverable 9.5. This policy ensures that the collection of personal information is limited to what is directly necessary for the research purposes on the project.

ANNEX

Data Management Plan *Questionnaire*

For each data category/data type you plan to generate, collect and/or process, please provide a **separate answer** to the following questions. For example, if you are going to process food safety data, and to generate software, you are managing two different data categories and you need to answer the below questions for both data categories separately!

Possible data categories/types are following:

- Traceability information
- Farm data
- Sensor data
- Supplier information
- Food recalls & border rejections
- Certification scheme parameters
- Consumer complaints
- Retailers' list of certificates required
- Transactions, input output sources
- Lab data
- Data from actors in the supply chain
- Supplier information
- Food recalls and border rejections
- Software code (including the IT systems)

If you are collecting/processing or generating additional categories of data, we encourage you to add them and answer the below questions also with respect to those additional categories.

The questions concern data collected and processed **during the lifetime of the Project only**. This questionnaire does not address the management of data once the envisioned FSM platform is on the market. However, please note the questions address **all data types or categories and not just personal data** collected or generated during the Project.

If you are uncertain how to answer a question, please refer to anyone from the team of UNIVIE: tima.anwana@univie.ac.at

Your answers to this Questionnaire will be seen in an annex to the Data Management Plan deliverable.

Please provide your answers in a different **colour** or in "revision mode" so that they are easily legible.

Definitions and Reference Material

Personal data: any information relating to an identified or identifiable natural person ('**data subject**'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person

Sensitive data:

- personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs;
- trade-union membership;
- genetic data, biometric data processed solely to identify a human being;
- health-related data;
- data concerning a person's sex life or sexual orientation.

Processing means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction. (Art 4/2 GDPR)

Pseudonymisation means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person. (Art 4/5 GDPR) Pseudonymised data is considered as personal data under the GDPR.

In contrast to pseudonymisation there is the concept of **anonymisation**. Whereas Pseudonymisation substitutes the identity of the subject so that additional information is required for re-identification anonymisation irreversibly destroys any means of identifying the data subject.

Controller means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law. (Art 4/7 GDPR)

Joint controller means that two or more parties determine the purpose and means of processing. (Art 26/1 GDPR)

Processor means a natural or legal person, public authority, agency or other body, which processes personal data on behalf of the controller. (Art 4/8 GDPR)

Legal basis for processing: (Art 6/1 GDPR)

Processing shall be lawful only if and to the extent that at least one of the following applies:

- freely given informed consent
- processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract
- processing is necessary for compliance with a legal obligation
- processing is necessary in order to protect the vital interests
- processing is necessary for the performance of a task carried out in the public interest
- processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party

Data sharing agreement means a formal contract that identifies inter alia the parameters which govern the collection, transmission, storage, security, analysis, re-use, archiving, and destruction of data.

Guidelines on FAIR Data Management in Horizon 2020:

https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Data Management: https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

QUESTIONS

The following answers are on behalf of _____

1. Data Types and Storage

The following questions are intended to understand what types of data will be collected, processed and generated during the project.

Please, provide a brief description of each data set. In the following sections kindly, specify always to which data set your answer applies.

- (a) What type of data will you **produce or generate** during the Project? Moreover, in what formats? (e.g. *software code in the format of Java language*)
- (b) What type of data will you **collect** during the Project and in what formats? (e.g. *partner contact information in the format of Google drive repository*)
- (c) In which WP/Task will the data type be relevant?
- (d) How will you trace the collected data? How do you trace the provenance of the data collected or metadata you maintain about the collected data? (e.g. *maintain data in different versions*)
- (e) Will the process of data generation or production be reproducible? What would happen if collected data gets lost or becomes unusable later?
- (f) How much data will be collected, and at what growth rate? How often will it change?
- (g) Are there specific tools or software needed to create/process/visualize the data?
- (h) Will you use pre-existing data? If this data is personal data, what is the legal basis (e.g. *consent*)? From where?
- (i) What are the storage and backup strategies?

2. Data Organization, Documentation and Metadata:

The following questions are intended to understand the plan for organizing, documenting, and using descriptive metadata to assure quality control and reproducibility of these data. *Answer to the following questions only WRT the portion of data that you will publish (i.e. make available to people external to the project).*

- (a) What standards will be used for the documentation of metadata? (e.g. *Digital Object Identifiers*)
- (b) Do you use any best practices/guidelines for managing the data to be publish (i.e., made available to third parties)?
- (c) Do you use any tool for checking that the data are well formatted? (e.g. *Standard Oracle Java formatting*)
- (d) What directory and file naming convention will be used? (e.g. *Standard Java naming conventions*)
- (e) What project and data identifiers will be assigned?
- (f) Is there a community or industry standard for metadata sharing/integration?

- (g) Can any metadata be created automatically?

3. Data Access and Intellectual Property

The following questions aim to identify any data access and ownership concern.

- (a) What are the major risks to
- Loss or destruction of data?
 - Loss of availability
 - Loss of integrity
 - Loss of confidentiality?
 - Data breach.
 - Unauthorized alteration, transmission and storage of data?
- (b) Have you prepared a formal risk assessment addressing each of the major risks to data security and potential solutions?
- (c) Does your data have any access concerns? Describe the process someone would take to access your data.
- (d) Who checks the correct execution of the access process? (e.g. PI, lab, University, funder, developer)
- (e) What procedures have you developed or plan to develop for the safe transfer of data including personal or sensitive data?
- (f) Have you implemented or outlined any procedures to follow in the case of a data breach (e.g. Data Privacy Impact Assessment, Data Protection Officer in place, contact with Data Protection Authority, Chief Information Security Officer)?

4. Data Sharing and Reuse

The following questions are intended to clarify how the collected data will be released for sharing. *Answer to the following questions only WRT the portion of data that you will publish (i.e. make available to people external to the project)*

- (a) If you allow others to reuse your data, how will the data be discovered and shared? List the categories of data that will be made re-usable or openly accessible.
- (b) If so, how will you organize/label the data so that researchers may easily isolate fields of interest in their study?
- (c) Any sharing requirements? (e.g., funder data sharing policies often require that the digital data be released in machine-readable formats that supplement journal articles and presentations)
- (d) Audience for reuse? Who will use it now? Who will use it later?
- (e) Any restrictions on who can re-use the data and for what purpose?
- (f) When will you publish it and where? If it is personal data will, you anonymised or pseudonymised it.

5. Data Preservation and Archiving

The following questions are intended to clarify how the collected data will be preserved and archived.

- (a) How will the data be archived for preservation and long-term access? (*e.g. cloud-based repositories*)
- (b) How long the data should be retained? (*e.g. 3-5 years, 10-20 years, permanently*)
- (c) What file formats?
- (d) Are there data archives that are appropriate for your data (subject-based or institutional)?
- (e) Who will maintain the data for the long-term?
- (f) Who decides what data or what categories of data will be kept and for how long?
- (g) The GDPR requires personal data not be kept longer than necessary for the purpose for which it was stored. What protocol(s) will you put in place to ensure you delete personal data that is no longer required to be stored?

6. Data protection and Ethical Aspects

- (a) What types of personal data (*e.g. partner's contact details*) do you intend to collect, generate or process?
- (b) What types of sensitive data (if any) do you intend to collect, generate or process?
- (c) Will you be controller or processor of personal data? If you act as a controller, will you do so jointly with another party (*Joint controllership*)?
- (d) What is the legal basis?
- (e) If your legal basis is consent: Have you already gained consent for processing from data subject(s) (*e.g. data preservation and sharing*)?
- (f) How will you protect the identity of Project participants (*e.g. pseudonymisation*)?
- (g) Will you engage in large scale or big data processing?
- (h) Will any entity (including any service provider) outside of the E.U. have access to personal data?
 - If yes, who?
 - For what purpose?
 - Where are each of these entities located?
 - Is there a data sharing agreement in place?

7. Pilots

This question is directed at pilot leaders only.

- (a) Who (or which entity or entities) will be responsible for determining what data is produced/generated/collected for your Pilot