TheFSM | The Food Safety Market

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, Elisabeth Steindl, Tima Anwana, Žiga Škorjanc (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, Elisabeth Steindl, Tima Anwana, Žiga Škorjanc (UNIVIE)
REPLY TO	elisabeth.steindl@univie.ac.at, tima.anwana@univie.ac.at, ziga.skorjanc@univie.ac.at
DOCUMENT URL	
DATE OF DELIVERY (CONTRACTUAL)	31 July 2020 (M6)
DATE OF DELIVERY (SUBMITTED)	03 August 2020 (M7)
VERSION | STATUS	1.0 | Final
NATURE	Report (R)
DISSEMINATION LEVEL	PUBLIC
AUTHORS (PARTNER)	Nikolaus Forgó, Elisabeth Steindl, Tima Anwana, Žiga Škorjanc (UNIVIE)
CONTRIBUTORS	Nikos Manouselis, Francesca Tsaropoulou, Charalampos Thanopoulos, Eliana Giannelou, Dimitris Pitteros (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	Nikola Rusinov (SAI)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.5	First Draft sent out for review by all contributors	03/07/2020	Nikolaus Forgó, Elisabeth Steindl, Tima Anwana, Žiga Škorjanc (UNIVIE)
0.6	Review by all contributors	10/07/2020	All contributors (see above)
0.7	1. Internal Review	23/07/2020	Nikola Rusinov (SAI)
0.8	Final Draft including all comments from the reviewing partners sent to Coordinator	27/07/2020	Nikolaus Forgó, Elisabeth Steindl, Tima Anwana, Žiga Škorjanc (UNIVIE)
0.9	2. Internal Review	28/07/2020	Nikos Manouselis, Francesca Tsaropoulou, Charalampos Thanopoulos, Eliana Giannelou, Dimitris Pitteros (AGROKNOW)
1.0	Final Version	30/07/2020	Nikolaus Forgó, Elisabeth Steindl, Tima Anwana, Žiga Škorjanc (UNIVIE)

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 in around the world.

Following the European Commission's recommendations this deliverable describes the FAIR management of the research data sets processed for TheFSM.

Objective of the Data Management Handling Plan

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 document itself is a living document. This first version of the DMP due already in M6 reflects a relatively early stage of the project. Regular iterations will follow in M24 and in M36 where additional questions corresponding to the more advanced stage and knowledge about the project and its data will be tackled, the pilots will be addressed with a local DMP.

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 objectives 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.

PARTNERS		CONTACT
Agroknow IKE (Agroknow, Greece)	Agroknow	Nikos Manouselis (Agroknow) nikosm@agroknow.com
SIRMA AI EAD (SAI, Bulgaria)	ontotext	Nikola Rusinov (SAI) svetla.boytcheva@ontotext.com
GIOUMPITEK MELETI SCHEDIASMOS YLOPOIISI KAI POLISI ERGON PLIROFORIKIS ETAIREIA PERIORISMENIS EFTHYNIS (UBITECH, Greece)	UBITECH ubiquitous solutions	Danai Vergeti (UBITECH) vergetid@ubitech.eu
AGRIVI DOO ZA PROIZVODNJU, TRGOVINU I USLUGE (Agrivi d.o.o., Croatia)	agrivi	Tanja Matosevic (Agrivi d.o.o.) tanja.matosevic@agrivi.com
PROSPEH, POSLOVNE STORITVE IN DIGITALNE RESITVE DOO (PROSPEH DOO, Slovenia)	tracelabs origintrail Core Developers	Ana Bevc (PROSPEH DOO) ana.bevc@tracelabs.io
UNIVERSITAT WIEN (UNIVIE, Austria)	universität wien	Elisabeth Steindl (UNIVIE) elisabeth.steindl@univie.ac.at
STICHTING WAGENINGEN RESEARCH (WFSR, Netherlands)	WAGENINGEN UNIVERSITY & RESEARCH	Yamine Bouzembrak (WFSR) yamine.bouzembrak@wur.nl
TUV- AUSTRIA ELLAS MONOPROSOPI ETAIREIA PERIORISMENIS EUTHYNIS (TUV AU HELLAS, Greece)	TÜV AUSTRIA HELLAS	Kostas Mavropoulos (TUV AU HELLAS) konstantinos.mavropoulos@tuv.at
TUV AUSTRIA ROMANIA SRL (TUV AU ROMANIA, Romania)	TÜV AUSTRIA ROMANIA	George Gheorghiou (TUV AU Romania) george.gheorghiu@tuv.at
VALORITALIA SOCIETA PER LA CERTIFICAZIONE DELLE QUALITA'E DELLE PRODUZIONI VITIVINICOLE ITALIANE SRL (VALORITALIA, Italy)	VALORITALIA	Francesca Romero (Valoritalia) francesca.romero@valoritalia.it
TUV AUSTRIA CERT GMBH (TUV AU CERT, Austria)	TÜV AUSTRIA	Stefan Hackel (TUV AU CERT) stefan.hackel@tuv.at

Table of Contents

List of Tables

List of Figures

1 INTRODUCTION

1.1 Scope of the Document

This document aims to tackle the handling and management of project data in TheFSM. For this purpose and in alignment with the European commission's open data policy and its objective to make research findable, accessible, interoperate and reusable (FAIR), a Data Management Plan (DMP) will be formulated and a first version will be delivered in M6 of the project. The document is based upon the Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, detailing (a) how project data will be handled during & after the project; (ii) what data will be collected, processed or generated; (iii) what methodology & standards will be applied; and (iv) whether data will be shared /made open access/ how data will be curated and preserved.

Within the context of TheFSM, the DMP will be used by the consortium partners for the effective management of the data that will be generated 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 are currently available data to which open access can be granted, always respecting the security and privacy requirements imposed. With regards to the dissemination of the scientific results, the consortium will establish and promote open access publications and partners will be encouraged to publish open access articles, so as 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) analyzes the main elements of TheFSM data management policy. It is intended to cover the complete life cycle of the research data generated and processed and will outline:

- 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.

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 share in line with the overall-management of the project, as foreseen in grant agreement and consortium agreement, which will be applied throughout the project. It reflects the early stage of M6 of TheFSM. Many developmental steps in the technological implementation are yet to be done before the industrial data platform is ready to run and accordingly there are many technological factors concerning data sets still that are still to be discussed and defined.

In addition, the DMP seeks to ensure that TheFSM follows the H2020 Open Access policy wherever possible and that all partners in TheFSM is compliant to all applications of relevant 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). All partners contributed by submitting their answers to the Questionnaire in the preparation of this document, which reflects a current picture of TheFSM in M6.

1.2 Intended Audience

The Data Management Handling Plan for TheFSM is a public document. Everybody interested in the topic will have access to this deliverable, however, it is to be assumed that it will be of primary interest for main actors and stakeholders in the food and food safety market.

1.3 Structure of the Document

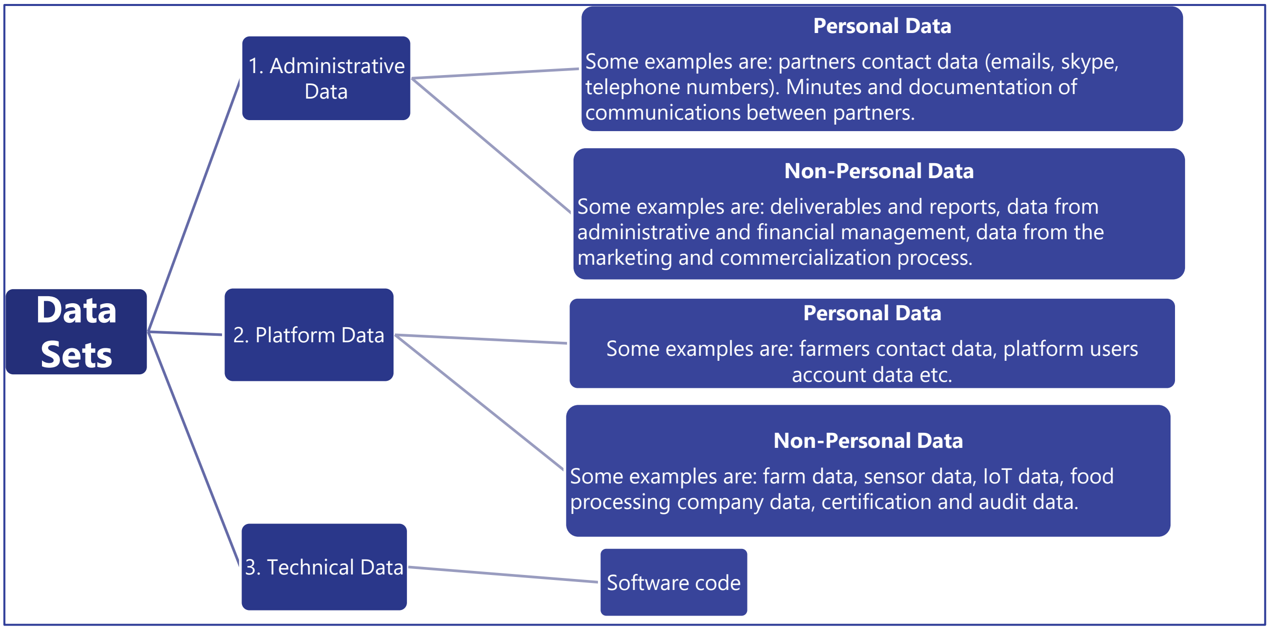
The structure of this document aims at following the Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. After first introductory remarks regarding the the scope of the TheFSM project, this report describes the principle of FAIR Data Management with reference to its implementation in TheFSM, 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 M6 of TheFSM.

1.4 Updates with respect to Previous Version (if any)

This is the first version of the Data Management Handling Plan for TheFSM, regular iterations are scheduled for M24 and M36.

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 three major categories as depicted in the diagram below (Figure 1).



Personal Data
Some examples are: partners contact data (emails, skype, telephone numbers). Minutes and documentation of communications between partners.

Non-Personal Data
Some examples are: deliverables and reports, data from administrative and financial management, data from the marketing and commercialization process.

Personal Data
Some examples are: farmers contact data, platform users account data etc.

Non-Personal Data
Some examples are: farm data, sensor data, IoT data, food processing company data, certification and audit data.

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.

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 will be developed in Task 1.2 Technical Requirements of the project.

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. The information in the table is derived from questionnaires filled out by relevant partners of TheFSM consortium.

Partner	Data Set	Data Set Number	Description of Data	Personal or Non-personal Data	Data Format
AGRIVI D.O. O	Platform Data and Technical data.	2 & 3	AGRIVI will **generate or produce** software code. AGRIVI will **collect** farmer contact data, crop production growing data, food processing company data and farmer contact data.	personal and non-personal data.	Software code in the format of C# programming language.

Agroknow	Platform Data	2	Agroknow will **generate or produce** food safety records. Some examples include, supplier profile, hazards and risk reports, hazards analytics, Certification schema information, certificate information. Agroknow will **collect** food safety records. Some examples include food recalls, border rejections, inspections/audits, lab data, supplier information, food data etc.	Personal and non-personal data.	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.
PROSPEH	Technical data, Administrative Data and Platform data	1,2,3	PROSPEH will **generate or produce** Software code. PROSPEH will **collect** input from other partners for project development as well as platform data in the form of traceability, certification, auditing data etc.	Personal and non-personal data.	Javascript and other programming language. Google Spreadsheets.
SIRMA AI EAD (SAI)	Administrative data, Platform data and Technical data	1, 2, 3	SAI will **generate or produce** Semantic Knowledge Graphs. SAI will **collect** partners contact information, food data, production data, laboratory data, certification data, inspection data, data from public databases, linked open data.	Personal and non-personal data	RDF triples, JSON format, TTL format, CSV files, Excel files, Word documents, PDF, XML, SPARQL queries, Java Software and Python Software, Google Drive repository,
TUEV Austria CERT	Platform Data and Administrative Data	1 & 2	TUEV Austria CERT will **generate or produce** Standards and Food Certification requirements. TUEV Austria CERT will **collect** partner contact data, general data of the partner organisations, product and raw material specification, supplier data, process description, operative checklists and records and Food Certificates.	Personal and non-personal data.	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.
TUV Austria Hellas	Platform data	2	TUV Austria Hellas will **generate or produce** Certification and Audit data. TUV Austria Hellas will **collect** customer's application data.	Personal and non-personal data	Certification and Audit data: archives and docx Customer's application data: pdf format or docx.
TUV Austria Romania	Platform data	2	TUV Austria Romania will **generate or produce** data related to audit services and conformity certificates. TUV Austria will **collect** audit and certification data. This includes information about company authorisations, management system documentation, product information, suppliers information and employees information.	Personal and Non-personal data.	Docs, Excel, pdf.

UBITECH	Technical data, platform data.	2 & 3	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	1	UNIVIE will **generate or produce** Deliverables, Reports and communications with partners. UNIVIE will **collect** partners contact data, 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	2	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	2	**Data generated** is not specified. WSFR 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.

Table 1: Data sets produced, generated or collected

2.4 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). 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.

DATA PRODUCTION AND STORAGE	
Data Generated /Collected	Administrative Data: partners contact details, project management data, and communications between partners. Minutes, Templates of Deliverables and Reports. Platform Data: food production data, certification data audit reports, customer's application data, data from public databases.

	Technical Data: Software Code, Software Design.
Data Formats	Administrative Data: PDF, Google drive repository, Word Docs, Excel. Platform Data: PDF, GSI database, Excel, Word Docs, Visio, Archives, CVS files, DB records. XML. Technical Data: 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.
Reproducibility	Administrative Data: Google Drive and local repositories Platform Data: local repositories, GitHub. Technical Data: Data sources will be stored, and back-up strategies applied to ensure recovery, reproduction and reuse of technical data.
Data Size	Data volume and velocity for all data sets will be established at a later stage of the project. This information will be included in a future version of the DMP. It is currently estimated that total number of data points collected will be more than 200 million.
Software tools for creating/processing /visualising data	Administrative Data: Google Drive Platform Data: Excel tools, rest calls software, Customer database, Microsoft Tools, specific libraries. Technical Data: Origin Trails Technologies, Agrivi – FSM (Farm Management Software), Chronograph, Grafana, Apache Strom, Apache Kafka.
Use of pre-existing data	Administrative Data: Publically available templates for deliverables and reports. Platform Data: Pre-existing data will be integrated from Certification bodies databases, public databases Technical Data: Pre-existing data will be integrated from LOD (Linked Open data) Cloud, other EC funded research projects, partners systems and public/open data sources (to be identified in data inventory list in WP2).
Data Storage and Backup Strategies	Administrative Data: Google Drive, local repositories, partners IT departments and EC Portal. Platform Data: Google drive repositories, GitHub, data retention points. Technical Data: Origin Trials Decentralized Network, Origin Trails Node Clusters, cloud infrastructures, NOSQL databases, Mongo DB, Apache HBase, Apache Cassandra, Apache Hive, Hadoop Distributed File Systems (HDFS), Internal servers, ISO 27001, RAID5, internal storage and backup policies.

Table 2: Data production and storage

ORGANISATION, DOCUMENTATION AND METADATA OF DATA INTENDED TO BE PUBLISHED	
Standards for Documentation of Metadata	Administrative Data: No particular standardisation protocols observed. Platform Data: 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. Technical Data: W3C Verifiable Credentials, W3C Decentralized Identifiers, GS1 EPICS.
Best Practice/Guidelines for Data Management	Administrative Data: 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 D6.1 Project Management Handbook v1.00. Platform Data: JSON Technical Data: Swagger Framework for API Documentation,
Tools for Formatting Data	Administrative Data: No automatic tools currently used.

	Platform Data: JSON validation tools Technical Data: GS1 Schema Validation modified JSON validators.
Directory and File Naming Convention.	Administrative Data: Defined in Project Management Handbook. Platform Data: Not yet defined, this information will be included in future versions of the DMP. Technical Data: Not yet defined, this information will be included in future versions of the DMP.

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

DATA ACCESS	
Risks	Administrative Data: unauthorised access. Platform Data: Loss or destruction of data, loss of availability, loss of confidentiality. Technical Data: unauthorised access.
Risk Management	Administrative Data: User and Password controls and by other specific security. Platform Data: Formal risk management strategies are still to be decided; ISO 27001 and GDPR requirements will be applied. Technical Data: For most technical partners, risk assessment formalisation is in progress, this information will be included in a future version of the DMP. Technical partners adopt ISO 27001 and apply its procedures and policies for information security management.
Correct execution of the Access process	At this stage of the project, each partners' IT departments, quality assurance teams, systems administrators and data protection officers are responsible for correct execution of the access process. Access process involve the use of login or identity credentials. Updates or changes to this will be included in future versions of the DMP.
Procedures to Follow a Data Breach	At this stage of the project, partners follow internal privacy and data protection procedures. Procedures specific to TheFSM project will be defined at a later stage and documented in a future version of the DMP.

Table 4: Data access

DATA SHARING AND REUSE OF DATA INTENDED FOR PUBLICATION	
Re-Use of Data	Administrative Data: For consortium partners only. Platform Data: Metadata of food safety data records. Technical Data: Open Source application code and Open Linked Data.
Organization/ Labelling of Data for Easy Identification	Administrative Data: Access limited to consortium partners. Platform Data: catalogue of the Food Safety Marketplace. Technical Data: Open Linked Data structured according to open standards and with relevant identifiers and tags for querying.
Data Sharing Requirements	Administrative Data: No requirements identified for public disseminations. Platform Data: Metadata of the datasets will be available in machine-readable format following standards such as DCAT. Technical Data: Conformity with open data models.
Audience for Re-use	Administrative Data: Consortium partners.

	Platform Data: Researchers, Technology companies in the food industry, stakeholders in the food industry. Technical Data: Developers of food supply chain IT systems, auditors, customers.
Restrictions on Re-use of Data	Administrative Data: Personal data, privacy requirements, confidential data and IPR. Platform Data: Restrictions defined by commercial subscription licenses, data from National Authorities will be subject to non-commercial use restrictions. Technical Data: Restrictions defined by users.
Publication	Administrative Data: Accessible for consortium partners. Platform Data: Food Safety Marketplace. Technical Data: Open Source application code, available on GitHub for reuse and review. Open Linked Data published to the Origin Trial decentralized network.

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

DATA PRESERVATION AND ARCHIVING	
Archiving of Data for Preservation and Long-term Access	Administrative Data: Google Drive, EC Portal and local repositories. Platform Data: Internal ERP – Archives, Cloud-based repositories, internal servers Technical Data: Partners Cloud infrastructure, OriginTrail Decentralised Knowledge Graph,
Data Retention	Administrative Data: 5 years after the end of the project in accordance with the requirements of the Grant Agreement. Platform Data: 3- 5 years Technical Data: 1-5 years
File Formats	Administrative Data: Google Drive repository, PDF files, Word Docs, Excel Platform Data: Digital Archives, Database/backup, PDF files, Excel, CVS files, DB records Technical Data: NoSQL databases MongoDB, Apache Hbase, Apache Cassandra, JSON-LD, XML, images Apache Hive, Hadoop Distributed File System (HDFS)
Data Archives	Administrative Data: EC Portal Platform Data: Local institutional repositories Technical Data: Orgin Trail Decentrailised Knowledge Graph
Long-term Maintenance of Data	Administrative Data: Not yet defined, this information will be included in future versions of the DMP Platform Data: FSM Coordinators (Agroknow), Certification bodies supervised by DPO, ISO 27001 policies Technical Data: OrginTrail Decentralised Network, internal systems

Table 6: Data preservation and archiving

3 FAIR DATA

3.1 FAIR Management of Research Data

TheFSM is committed to following the "Guidelines on FAIR Data Management in Horizon 2020"[1], recommended by the European Commission Directorate – General for Research & Innovation. These Guidelines foresee four principles that govern the management of research date in order to make them more easily understood, exchanged and open for re-use: Scientific data management thus has to make sure research data is FAIR, meaning findable, accessible, interoperable and re-usable.

Data often have a longer lifespan than the research project itself. The European Commission has therefore provided for guidelines that allow future research to benefit from former projects, Scientists thus can make use of the data after the project has been completed or use it in follow-up projects.

3.2 Research data lifecycle

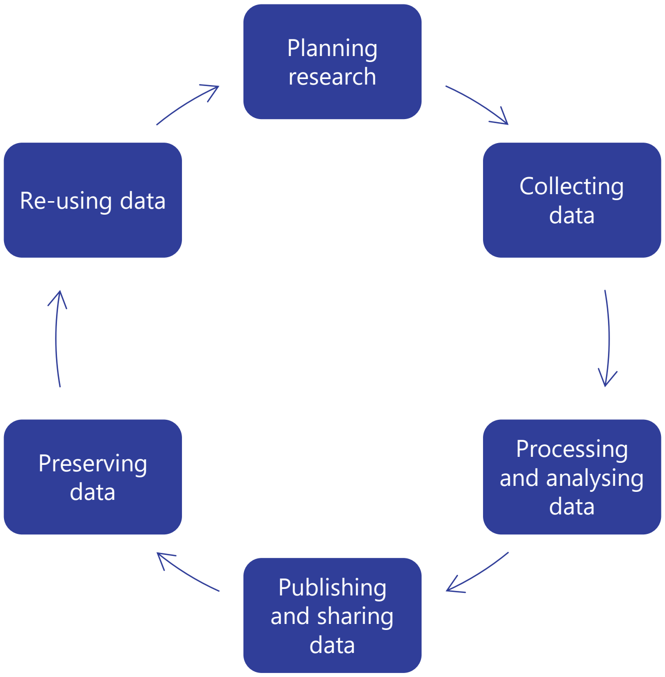


Figure 2: Stages of data during a research process[2]

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

[2] See UK Data Service: Research data lifecycle https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx

Planning research:[3]

- design research
- plan data management
- 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] See UK Data Service: Research data lifecycle https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx

3.3 The FAIR principle[4]

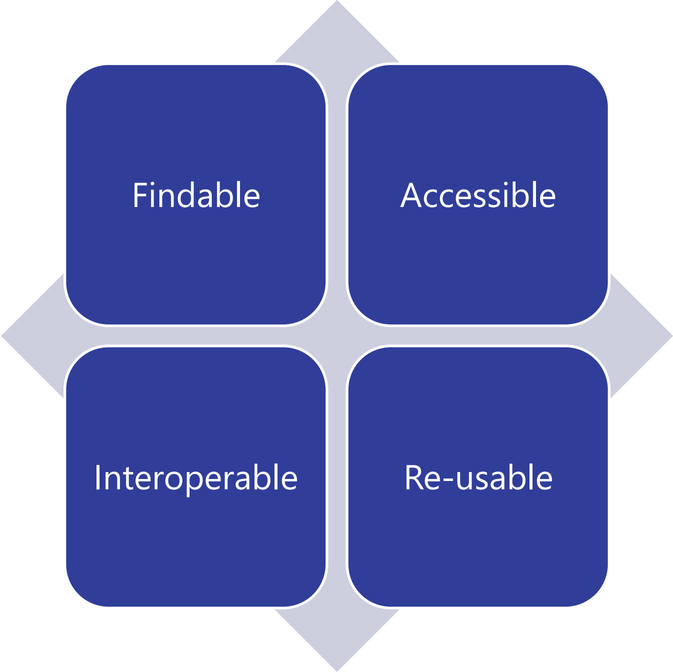


Figure 3: The Fair Principle

The elements of the FAIR Principles are related, but independent and separable. This principle defines how research outputs should be organised in order to adequately facilitate discovery, exchange and reuse by third-parties. Major funding bodies, including the European Commission, promote FAIR data to maximize the integrity and impact of their research investment.[5]

The TheFSM project was designed to develop an industrial data platform for the exchange of business sensitive data in the food and food safety market to boost the competitiveness of the European food certification. The commercial objective of the project itself can sometimes collide with complete FAIRness of data. TheFSM is however committed to participate in the Pilot on Open Research Data in Horizon 2020, though in a way that respects the security and privacy requirements of the pilots and the commercial interests of the industrial partners.

Data that will be processed for the pilots will be addressed in local iterations of the Data Management Handling Plans.

[4] The Open Data Foundation http://www.odaf.org/

[5] The Open Data Foundation http://www.odaf.org/

Findable

Data and metadata should be easily findable for humans as well as for computers. Crucial 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."[6] To be findable any Data Object should be uniquely and persistently identifiable. The same Data Object should be persistent, the Data Object should 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 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 resource[7]

In TheFSM a variety of data sets is 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 are 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 recommendations to name 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 contractual documents, copies are maintained on the Participant Project site. Partners maintain their collected data and metadata about the collected data in different versions

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

[7] See https://www.go-fair.org/fair-principles/

with backup versions e.g. in GitHub or partner's cloud infrastructure; for the platform this is still to be defined.

The Project repository provided through Google Drive serves as a collective working space where the partners exchange documents and collaborate in the development of deliverables. Deliverables names/identifiers are agreed with the EC and stated in the GA. 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. However, at this stage, there is no automatic tool yet for formatting data and no automatic creation of metadata.

The platform structure itself and the data that will be used for and transferred via the platform is still in the development process and not yet running at the time this document is being produced. TheFSM will make metadata available data 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.

Accessible

Accessibility requires that data can be always obtained by machines and humans upon appropriate authorisation through well-defined protocols. 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 available[8]

At the time of this first version of the Data Management Handling Plan information about which data can and will be published or made openly available is still imprecise. Data access will vary depending on the storage location. Starting with the use case data, measures will be taken to

[8] See https://www.go-fair.org/fair-principles/

enable third parties to access, re-use, analyze, exploit, and disseminate the data (bound by the license specifications). Different access procedures will be implemented, 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 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. Whereas the partners themselves follow the respective company's standard risk management to guarantee safety of data the procedure of a risk assessment formalization for the platform is in progress. It is being considered to implement a security ISO standard that includes a formal risk assessment.

Project related datasets (deliverables, reports, code, DBs) will be stored in Coordinator's premises or Consortium agreed tools with proper access rights. Project administrative data is not intended to be publically shared or otherwise made available to third parties, access to the repository is granted by the coordinating partner 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. Access will be provided upon positive confirmation of the required access criteria to the database, 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 are in conformity with open data models. Metadata of the datasets will be available in machine readable format following standards such as DCAT.

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)data[9]

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

[9] See https://www.go-fair.org/fair-principles/

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. However, many facts regarding the platform structure are yet to be defined and clarified.

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 easily 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.

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 will data be made available for re-use, the time-frame of the intended reuse and data quality assurance processes. In cases where an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

- **R1**. 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 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 commited 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 selfarchiving. The Consortium partners will therefore privilege 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 be in compliance with the requirements of the EC Guidelines on Open Access to Scientific Publications and Research Data in H2020. 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). 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 License (ASL) license and the LGPL software licenses. The ASL license 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 (i) generate business interest around future expansions and (ii) 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. Partners contact details, communications between partners and meeting minutes include personal data and privacy of correspondence and telecommunications have to be considered. What other platform and technical data will be made re-usable for research still has to be established. Given the very nature of the data exchanged as being business sensitive re-use will however have to be restricted for substantial parts to be in compliance with contractual commitments such as non-disclosure agreements necessary to get consent for the use of this data in the project, users themselves can and will restrict re-usability. 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.

4 ALLOCATION OF RESOURCES

4.1 Administrative Data

Administrative data does not include any scientific or research data (see page 10 above). Therefore, it is – at least at this stage – not planned to make administrative data available under FAIR principles for third parties (see pages 23 and 26 above). No cost will incur.

4.2 Platform Data

The partners will rely on dedicated funding from their research projects and/or institutions In order to make scientific and research data FAIR.

If open access can be granted, research data and originals or pre-prints of the publications will be stored into their organization's repository (e.g. the PHAIDRA institutional repository of UNIVIE[10]) or, in absence of such repositories into OpenAIRE's Zenodo[11] repository for publications. Free of cost storage infrastructure that could be provided by partners, will be preferred (see section 1 above).

Task leaders are responsible for making the data FAIR. UNIVIE will support them with regard to legal aspects. Partners will be encouraged to make data and scientific results accessible under FAIR principles, to enable researchers to build upon previous research results, to foster collaboration, to avoid duplication of efforts, and to accelerate innovation.

4.3 Technical Data

The work 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). UNIVIE will support them with regard to legal aspects.

The technical outcomes of TheFSM will be made available in order to foster collaboration and sharing. Regarding source code we will use repositories that are free of charge (e.g. GitHub repository[12]).

[10] https://phaidra.univie.ac.at/.

[11] https://zenodo.org/.

[12] https://github.com/.

5 DATA SECURITY

5.1 Administrative Data

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 by partners, so that they can access and share all research and project documentation from final deliverables through to presentations and other relevant information (see proposal section 3.2.1.1).

In accordance with the grant agreement the data will be stored for a period of 5 years after the completion of the project. Afterwards the data will be deleted.

5.2 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/IEC 27001, 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.

5.3 Technical Data

Project related datasets such as software code will be stored in secure database where only authorized people from the consortium can access to. The data will be backed up daily in a secure storage server. If data will be kept in an external certified repository, then the security standards of that repository will apply.

6 PROTECTION OF PERSONAL DATA AND ETHICAL ASPECTS

The commitment to 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 ethically compliant from beginning to end and a thorough ethical evaluation is required from the conceptual stage of the proposal. Applying ethical principles and legislation to scientific research is fundamental for all projects and in all possible domains of research. Main objectives of the ethical considerations in this document are whether there are any ethical or legal issues, which might have an impact on data sharing.

Every personal datum in TheFSM is subject to the protection of the General Data Protection Regulation (EU) 216/679 (**GDPR**).[13] The GDPR is the most fundamental piece of European legislation where ethical considerations are enshrined. It is a **binding legislative act** that as such **must be applied in its entirety across the EU**. This means that the GDPR has binding effects in all Member States, without any further implementation being necessary.

The main purpose of the GDPR is the protection of natural persons with regard to the processing of their personal data.[14] **Personal data means any information relating to an identified or identifiable natural person ('data subject')**.[15] 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.

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.[16] **In TheFSM all Partners have an establishment in the EU and are thus subject to the rules set out in the GDPR.**

[13] 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).

[14] Art 1 (1) GDPR.

[15] Art 4 (1) GDPR.

[16] Art 3 GDPR.

6.1 Principles

In Article 5 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 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 absolutely 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.[17]

In the context of the TheFSM project some of these principles need more clarification than others. With regard to the specific needs of TheFSM and the role of the partners therein this document aims to explicate some of the essential terminology in the context of the principles of the GDPR.

6.2 Lawful basis

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;

[17] Art 5 GDPR.

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 or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.[18]

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.**

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 5/1f). 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,

[18] 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.[19]

Technical measures can for example mean instruments such as requiring your employees to use two-factor authentication on accounts where personal data are stored to contracting with cloud providers that use end-to-end encryption. Organisational measures include initiatives such as staff training, adding a data privacy policy to the employee handbook, or limiting access to personal data to only those employees in the organization who need it.

These measures include amongst others the principles of data protection by design and data protection by default.[20] The principle of **data protection by design and data protection by default should make an integral part in all developmental stages** of the project, but it is of course of particular importance during the establishment of the platform for TheFSM.

Accountability

According to Article 24 (1) GDPR controllers and processors shall "implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation**.**"

Accountability in the context of the GDPR requires the data controller to **put in place appropriate technical and organisational measures** and to **be able to demonstrate the effectiveness of the measures when requested**. The data controller must be able to demonstrate his compliance with the GDPR.

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.

Possible instruments to achieve accountability are to:

- accurately document 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 ecc)
- implement documented processes and procedures aiming at tackling data protection issues at an early state when building information systems or responding to a data breach
- designate data protection responsibilities to the team
- have Data Processing Agreement contracts in place with third parties that are contracted to process data

[19] Recital 78 GDPR.

[20] Art 25 GDPR.

- appointment of a Data Protection Officer by the partners

6.3 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. Whoever therefore decides "why" and "how" personal data shall be processed is the data controller. The **data processor** himself **processes** personal data **only on behalf of the controller**. 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.

6.4 Joint Controllers

Should partners in TheFSM determine together the "why" and the "how" of the processing of personal data they will be so-called **Joint Controllers.** In this case, a **joint controllership agreement** has to be made between these 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 thus be informed about the relevant aspects of the agreement and can exercise their rights under the GDPR against each of the joint controllers.[21]

6.5 Data Protection Impact Assessment (DPIA)

Whenever the controller instructs a processor to carry out an activity that is considered to be of high-risk a **Data Protection Impact Assessment (DPIA)** must be made. A DPIA is "a process designed to describe the processing of personal data, access its necessity and proportionality."[22] By making a DPIA the potential risks to the rights and freedoms of natural persons shall be minimised. Examples for high-risk activities are e.g. trying out new technologies, carrying out large scale profiling, large scale processing of special category data, mixing or matching data from multiple sources.

For TheFSM a possible profiling of farmers and the mixing or matching of data from multiple sources might qualify as relevant high risk activities. Also, the use of big data technologies may result in an increased risk to data subjects.

[21] Article 26 (3) GDPR

[22] Article 29, Data Protection Working Party, 'Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679. Adopted on 4 April 2017. Page 6.

The assessment needs to contain at least:

- a description of the processing operations and the purposes of the processing,
- an assessment of the necessity and proportionality of the processing operations
- an assessment of the risks to the rights and freedoms of data subjects
- measures to address and minimise the risks. (Art 35/7)

6.6 Records

Data controllers are required to keep accurate **records of processing activities** when processing sensitive information or when the data controller is an organisation with more than 250 employees. These records shall include

- details of the controller
- Processing purposes
- Description of types of data collected
- Categories of data recipients
- Data transfers including data transferred to third countries
- Erasure details
- Overview of data security measures[23]

According to Article 7, the controller has to keep a record of consent given or withdrawn by the data subjects. In accordance with Article 33, the controller is obliged to record and document any personal data breaches, comprising the facts, effects and actions taken in regards to the data breach.[24]

Data processors are equally required to keep records that are related to the processes they were asked to carry out on behalf of the controller:

- Name and details of processors, controllers and Data Protection Officer (if applicable)
- Categories of processing
- Data transfers to third countries or international organizations
- General description of security measures according to Article 32

All above mentioned records must be both in writing and electronic form and ready to be presented upon the request of the Supervisory Authority.

[23] Article 30 (1) GDPR.

[24] Article 33 (5) GDPR.

6.7 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 hereby increased flow of data, the core of the GDPR shall not be undermined by transferring data to and from countries outside the Union.[25] Thus the transfer of personal data to a third country or to an international organisation, known as 'data export',[26] 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.

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 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);

[25] Recital 101 GDPR.

[26] Gawronski (ed.) (2019): Guide to the GDPR, p. 99.

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)

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).[27] Contractual clauses or provisions need permission from the respective supervisory authority. With legally **binding corporate rules**[28] 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. A company's rules shall be approved by the respective supervisory authority 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.

With regard to the pilots in Jordan and Egypt and the access from partner's partners in Serbia partners in TheFSM involved in processing operations should therefore review the planned business operations accordingly, identify all circumstances in which personal data is being transferred to recipients located in one of those three countries and ensure that, for each such transfer, there is a data transfer mechanism in place that complies with the requirements of the GDPR.

6.8 Remedies, Liabilities and Penalties

In Chapter 8 the GDPR outlines the remedies, liabilities and penalties that data subjects are entitled to. Compared to previous regulations the GDPR sets the potential penalties significantly

[27] Gawronski (ed.) (2019): Guide to the GDPR, p. 102.

[28] Recital 110 GDPR.

high. The GDPR allows DPAs to issue fines for serious infringements up to a maximum of the greater of €20 million or four percent of worldwide turnover. Therefore, it is crucial that partners in TheFSM in every stage of the project are aware of their data protection compliance whenever they process data.

All partners in TheFSM must be aware that both controller and processor can be made liable for material and immaterial damage and that in cases of joint controllership both can be held liable for the entire damage.

By its very nature, the GDPR will play a crucial role when drafting a **Privacy Policy** for TheFSM later in the course of the project.

7 ANNEX

7.1 Data Management Plan *Questionnaire*

Deadline: 8 June 2020

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:
elisabeth.steindl@univie.ac.at
tima.anwana@univie.ac.at
ziga.skorjanc@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 color 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.

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

(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 anonymize or pseudonomynize 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 should the data 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. *pseudonomysation*)?
(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