

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

D2.3 - Report on Data Population

DELIVERABLE NUMBER	D2.3
DELIVERABLE TITLE	Report on Data Population
RESPONSIBLE AUTHOR	Branimir Rakić (Prospeh d.o.o.)



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	WP2 Data
WORKPACKAGE LEADER	Sirma AI
DELIVERABLE N. TITLE	D2.3 Report on Data Population (update)
RESPONSIBLE AUTHOR	Branimir Rakić (Prospeh d.o.o.)
REPLY TO	branimir.rakic@tracelabs.io
DOCUMENT URL	
DATE OF DELIVERY (CONTRACTUAL)	31 January 2022 (M24)
DATE OF DELIVERY (SUBMITTED)	31 January 2022 (M24)
VERSION STATUS	V2.0 Final
NATURE	Report (R)
DISSEMINATION LEVEL	Public (P)
AUTHORS (PARTNER)	Branimir Rakić (Prospeh d.o.o.), Manos Karvounis (Agroknow), Filip Gerin (Agrivi), Margherita Vialetto (Valoritalia)
CONTRIBUTORS	Pavlin Gyurov (SAI)
REVIEWER	Maja Majdak (Agrivi)

VERSION	MODIFICATION(S)	DATE	AUTHOR(S)
0.1	Draft	27/01/2021	Branimir Rakić (Prospeh d.o.o.)
0.7	Added sections 1-5	27/01/2021	Branimir Rakić (Prospeh d.o.o.)
0.8	Partner Contribution	28/01/2021	Pavlin Gyurov (SAI)
0.9	Internal Review	29/01/2021	Tanja Matošević (Agrivi)
1.0	Final Version	31/01/2021	Branimir Rakić (Prospeh d.o.o.)
1.1	Draft	09/02/2022	Branimir Rakić (Prospeh d.o.o.)
1.7	Updated sections 1-5. Added sections 6-7	09/02/2022	Branimir Rakić (Prospeh d.o.o.)
1.8	Partner Contribution	09/02/2022	Manos Karvounis (Agroknow), Filip Gerin (Agrivi), Margherita Vialetto (Valoritalia)
1.9	Internal Review	30/01/2022	Maja Majdak (Agrivi)
2.0	Final Version	31/01/2022	Branimir Rakić (Prospeh d.o.o.)

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)		Martina Poberaj (PROSPEH DOO) martina.poberaj@tracelabs.io
UNIVERSITAT WIEN (UNIVIE, Austria)		Tima 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 CYPRUS (TUV AU CYPRUS, Cyprus)		Sousanna Charalambidou (TUV AU CYPRUS) sousanna.charalambidou@tuv.at

ACRONYMS LIST

TheFSM	The Food Safety Market
EPCIS	Electronic Product Code Information Services
W3C	World Wide Web Consortium

EXECUTIVE SUMMARY

In the D2.3 document we present the progress of the Data Population task of Work Package 2 of the development of TheFSM platform. A total of three data population reports will be presented during the course of the project. This document presents the second version of the report, which focuses on presenting the progress at the end of year two of the project, observed data sources integrated over TheFSM platform, corresponding data models and the overall progress of the data population task.

TABLE OF CONTENTS

1	INTRODUCTION	8
2	METHODOLOGY AND PROCESS	9
3	DATASET MODELS	10
4	DATA POPULATION TIMELINE.....	11
5	MEASURING PERFORMANCE	12
6	GAP ANALYSIS AND CORRECTIVE ACTIONS	14
7	CONCLUSION.....	15

LIST OF TABLES

Table 1: Data Population Gantt Chart.....	11
Table 2: KPI overview at M24.....	12
Table 3: Foodakai & Food Inspector Data records overview	13

1 INTRODUCTION

The Report on Data Population is a direct outcome of Task 2.4. Its purpose is to analyse and report on the data volumes, types and formats that have been shared and managed through the platform.

This document presents the second version of three such reports to be created in the duration of the Project (at M12, M24 and M36). This report covers the activity of the past 12 months (between M12 and M24) in which the platform development has moved from the design and definition phase towards implementation and initial integrations with data sources. As the platform software design, based on the business scenarios, identified user stories and functional requirements has yielded updates to the platform architecture, several metrics have been altered to accommodate the changes.

2 METHODOLOGY AND PROCESS

TheFSM platform's key value strives from the ability to ingest data from multiple data sources to provide as seamless an experience for its users as possible. In Task 2.4 our main focus is putting this overarching goal in practice.

The contributors are all organizations that are sharing food safety data of various types and formats through the platform with providing access to relevant data sets in various formats. In the process we are facilitating the use of data ingestion services that are being developed in accordance with inputs about the separate files defined in business scenarios.

The aim of TheFSM platform is to provide a seamless interface for data exchange between these existing IT systems for storing and managing data, providing several approaches of sharing data, with the goal of ensuring streamlined and more transparent certification processes.

To follow our progress of the data population we have identified the following lead metrics and bottom-line metrics to be analyzed over the course of the project:

1. **Number of supported formats by ingestion services** - A lead metric showing how many different data formats are supported by the platform's ingestion services. The higher the number of relevant formats that are supported, the greater inclusivity the platform is achieving and thus, greater usability.
2. **Relative coverage of formats in business scenarios** - A lead metric showing how much of the business scenarios have been covered by data formats and ingestion services of the platform. A sub-metric for such coverage is whether formats are only supported as inputs or also available as outputs for users to export data in the format best fitted for them.
3. **Number of organisations providing data to the platform** - A bottom-line metric showing how widely spread the usage of a platform is among those providing their data.
4. **Relative coverage of organisations in business scenarios** - A bottom-line metric showing how many of the organisations from the business scenarios are actively providing data to the platform. Within the WP2 we have decided that we will reformat this metric. Due to our organization that is not the most applicable and valuable metric and we are working on reformatting it which will be visible in the next version of the report.
5. **Number of data sources providing data to the platform** - A bottom-line metric showing how many different systems or devices have been successfully connected to the platform and can provide their data.
6. **Estimated number of records (formerly Total data size)** - A bottom-line metric showing what is the estimated number of records provided to the platform.

3 DATASET MODELS

After analyzing the initial data samples provided by the partners and designing the system to support the prioritized set of features, it has been decided to focus on several generic data models supported by the platform, potentially to be extended over the course of the third year of the project.

1. **Non-editable documents (.pdf)** - A common way of storing documents within food industry stakeholders, often also as an export document from an IT system.
2. **CSV as spreadsheet format**- A popular way of storing tabular data
3. **JSON format**- Very suitable formats for ingestion and processing, predominantly used by technical partners.

Over the course of the project several data exchange standard formats have been discussed and introduced limited support for or directly implemented as a core component of TheFSM platform. The following standards and recommendations have been in focus:

1. **GS1 EPCIS & CBV** - Electronic Product Code Information Services (EPCIS) is a GS1 standard that defines a common data model for visibility data and interfaces for capturing and sharing visibility data within an enterprise and across an open supply chain. The goal of EPCIS is to enable disparate applications to create and share visibility event data, both within and across enterprises.
2. **W3C WoT** - The Web of Things seeks to counter the fragmentation of the IoT, making it much easier to create applications without the need to master the disparate variety of IoT technologies and standards. Digital twins for sensors, actuators and information services are exposed to consuming applications as local software objects with properties, actions and events, independently of the physical location of devices or the protocols used to access them.
3. **W3C Verifiable Credentials together with W3C Decentralized Identifiers** - A verifiable credential can represent all of the same information that a physical credential represents, standardized in a JSON-LD serialization. The addition of technologies, such as digital signatures, makes verifiable credentials more tamper-evident and more trustworthy than their physical counterparts.
4. **TheFSM Data model** – The internal data model for TheFSM platform. It describes the inputs from identified data providers during the provisioned use case scenarios in WP1 as well as their interconnections and links to concepts in external ontologies. More detailed description of the data model is available in D2.1 Data Models & Representations.

4 DATA POPULATION TIMELINE

We present the updated data population timeline, originally outlined in the initial data population report (v1). With each successive report we are to present the progress on the key performance indicators defined in this document, highlighting any identified gaps as well as providing actionable suggestions for improvements.

Item	Phase 1 (M1-M12)	Phase 2 (M13-24)	Phase 3 (M25-M36)
Defining data sources and key indicators			
Measuring performance			
Gap analysis and corrective action identification			
Iterations based on corrective actions			
Measuring performance of new iterations			

Table 1: Data Population Gantt Chart

5 MEASURING PERFORMANCE

With the release of TheFSM and the first version of the platform the following metrics have been observed. It is important to note however that performance measurements have been slightly altered due to the nature of the platform and its API integrations allowing access to data outside of TheFSM platform, making it difficult to measure the total data size available through the platform. Therefore, we have decided to propose a different approach with the substitute metric being “estimated number of records” available through the platform, compiled by analysing data record estimates provided from each data source.

Metric	M24	Comment
Number of supported formats by ingestion services	3	Generic formats: PDF, JSON and CSV all are supported as all of the connected data sources at this stage are compatible with such formats. This includes API integration response formats.
Relative coverage of formats in business scenarios	100%	The supported formats have shown viable support for all business scenarios.
Number of organisations providing data to the platform	3	Data is currently directly provided by Agrivi, Agroknow and Valoritalia. It is important to note however that Agroknow’s applications (Foodakai and Food Inspector) are providing datasets aggregated from a large number of varied data sources, as seen by the large number of data sources available through the platform.
Number of data sources providing data to the platform	9889	The large number of data sources comes from source data aggregation performed by Agroknow and provided through the Foodakai and Food Inspector applications, together with Agrivi and Valoritalia.
Estimated number of records	1.15B	Details of the number of records and sources are presented in the table below.

Table 2: KPI overview at M24

Together with the presented metrics above, we include a specific table for the Foodakai and Food Inspector record estimates provided by Agroknow as a significant number of records comes from those systems. All data records reported are accessible inside TheFSM ecosystem and Platform in the following ways:

- For end-users, through the FOODAKAI and/or FoodInspector applications or through (a subset of all available) datasets uploaded and available directly in TheFSM Platform (in the future also accessible through the Data Marketplace)
- For internal TheFSM Platform services, as well as for other apps and services inside TheFSM ecosystem, through the Data API (as reported in D4.2)

Data Type	Sources	Records (as of 1/22)
Global Food Product Recalls	50	78,466
Global Food Product Border Rejections & Import Refusals	12	333,842
Global Lab Tests	34	102,190,099
Inspection Results	3	274,562
Global Fraud Reports	51	13,739
News Reports of Food Product Recalls	423	72,638
Foodborne Disease Outbreaks	2	67,395
Food Safety Legislation	1	279
Food Supply Chain Companies	50	702,915
Greek Consumer Complaints & Reports in Social Media	9,000	1,000,000,000
International Publications on Food & Agricultural Sciences	250	10,000,000
Country Risk & Corruption Indicators	2	3,594
Price Data on Key Commodities & Products	6	415,670
Production Data of Key Commodities & Products	1	503,329
Trade Data about Key Commodities & Products	1	41,424,465
Animal Health Data from OIE	1	426,661

Table 3: Foodakai & Food Inspector Data records overview

6 GAP ANALYSIS AND CORRECTIVE ACTIONS

We assess the overall progress of the data population task to be at satisfactory level and that initial results provide the necessary capabilities needed in supporting the business scenarios and requirements envisioned for the platform. The above presented KPIs demonstrate a significant level of coverage of a variety of heterogeneous data available for access and consumption through TheFSM platform.

The corrective action recommendations have been identified in the following three key areas:

- **Action 1 - facilitating growth of available records from existing sources:** to ensure a wider applicability of the platform in business scenarios, for which input is expected to be determined through piloting efforts and feedback by end users.
- **Action 2 - growth through enablement of data marketplace capabilities:** the data marketplace is expected to yield growth in the estimated available number of records and data sources, which will be enabled through the open nature of the platform and its interoperability capabilities (such as semantic enrichment)
- **Action 3 - increased interoperability through standard capabilities:** with the growth of available data records, implementing standardized, semantic data sharing approaches should be implemented in order to enable higher utility to the data exchanged. Several proposed standards and recommendations (presented above in Section 3) are suggested to provide improvement in this category, specifically when it comes to interoperability and usability of novel decentralized solutions such as the blockchain and decentralized knowledge graph

Having the platform facilitating data exchanges on various levels, the value provided by TheFSM platform is in direct proportion to the number of available records and their interoperability. Therefore, the corrective actions are expected to be implemented within the final year of the project, to the extent necessary and useful in relation to the end users and business scenarios.

7 CONCLUSION

The report indicates the current progress of the data population task to support achieving the Key Performance Indicators which have been introduced as the key methodology in this Report. TheFSM platform's current estimated available records and interoperability capabilities sufficiently support the relevant business scenarios and are enabled by current functionalities of the platform, however focusing on growth and extended interoperability as indicated in the corrective actions are expected to yield further growth in the value of the platform. We expect to iterate with measurements and corrective suggestions over the course of the final year of the project, focusing on increasing the value of the platform, as a proxy of the number of available data records and their overall interoperability and usability in TheFSM platform.