

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.)	01/01/2020 (36 months)
ENDING DATE	31/12/2023
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 Al
DELIVERABLE N. TITLE	D2.3 Report on Data Population
RESPONSIBLE AUTHOR	Branimir Rakić (Prospeh d.o.o.)
REPLY TO	branimir.rakic@tracelabs.io
DOCUMENT URL	
DATE OF DELIVERY (CONTRACTUAL)	31 January 2021 (M12)
DATE OF DELIVERY (SUBMITTED)	31 January 2021 (M12)
VERSION STATUS	V1.0 Final
NATURE	Report (R)
DISSEMINATION LEVEL	Public (P)
AUTHORS (PARTNER)	Branimir Rakić (Prospeh d.o.o.)
CONTRIBUTORS	Pavlin Gyurov (SAI)
REVIEWER	Tanja Matošević (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 and review	26/01/2021	Pavlin Gyurov (SAI)
0.9	Internal Review	29/01/2021	Tanja Matošević (Agrivi)
1.0	Final Version 1	31/01/2021	Branimir Rakić (Prospeh d.o.o.)

PARTNERS		CONTACT
Agroknow IKE (Agroknow, Greece)		Nikos Manouselis (Agroknow) nikosm@agroknow.com
SIRMA AI EAD (SAI, Bulgaria)		Svetla Boytcheva (SAI) svetla.boytcheva@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)		Tanja Matosevic (Agrivi d.o.o.) tanja.matosevic@agrivi.com
PROSPEH, POSLOVNE STORITVE IN DIGITALNE RESITVE DOO (PROSPEH DOO, Slovenia)		Ana Bevc (PROSPEH DOO) ana@origin-trail.com
UNIVERSITAT WIEN (UNIVIE, Austria)		Elizabeth Steindl (UNIVIE) elizabeth.steindl@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 current 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 presenting the first one which focuses on methodology, observed data models and the upcoming timeline for the data population task within the project.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	6
1 Introduction.....	8
2 Methodology and Process	9
3 Dataset Models	10
4 Data Population Timeline	12
5 Conclusion	13

LIST OF TABLES

Table 1: Data Population Gantt Chart.....	12
---	----

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.

The current document is the first out of three such reports to be submitted in the duration of the Project (M12, M24 and M36). Since this is the first report, it covers the activity of the past 12 months in which the platform has been in the design and definition phase, rather than being actively used. Due to that fact, the deliverable cannot yet report on the flow of the data towards the platform. It rather focuses on **methodology utilised to perform the reporting** in the subsequent deliverables in M24 and M36, **overview of the work** done to ensure that handling, locally storing and processing of dataset that have a continuous flow of information and **timeline of activities** for Task 2.4 in the future.

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 supporting all organisations that are sharing food safety data of various types and formats through the platform with their uploads of 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 (e.g. spreadsheets, FMS exports) defined in certifications scenarios.

Putting these existing IT systems for storing and managing data on a common denominator will create interoperability bridges between them. This can be achieved either by adapting the spreadsheets or connecting systems with APIs.

All of the above will ensure that the final goal of improving the certification process will be achieved. To follow our progress in the subsequent combination of lead metrics and bottom-line metrics will be analysed:

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.
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. **Total data size** - A bottom-line metric showing what is the total size of data that users have provided to the platform.

All of the above-mentioned metrics will be followed by analysing the performance of the platform once released.

3 DATASET MODELS

In the first 12 months, the partners have already put significant effort to address successful achievement of the Key Performance Indicators (KPIs) presented above. One of the activities relevant to the Report has been performed with **all partners** in the consortium by collecting relevant datasets from their existing systems. Inputs on which systems were relevant for such sample data collection were based on the business scenarios and technical requirements as developed by Work Package 1. In the course of the data collection following data formats which will be supported by the ingestion of the platform:

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. As a format, such documents are highly problematic and should be replaced by other formats or direct access to data sources creating the PDF in the first place.
2. **Editable text documents (.doc, .docx, .txt)** - Another popular format by majority of non-technical consortium partners, especially frequently used in the audit process of the certification. Such formats are often prone to errors and should have ambition to be replaced by more suitable formats. For highly standardized editable text documents, TheFSM platform can still offer ingestion.
3. **Spreadsheet formats (.xls, .xlsx, .csv and similar)** - A popular way of storing tabular data by all non-technical consortium partners and stakeholders, present as a manually created document and exported document from systems. Such formats, although prone to occasional errors, are highly suitable for ingestion in the platform.
4. **Highly machine-readable formats (.json, .ttl, .xml)** - Very suitable formats for ingestion and processing, predominantly used by technical partners.

In the follow up activities to the sample data collection activity is defining the best ontologies and data models that will encapsulate as many as possible sample data inputs in an interoperable way. Current progress has already identified certain data standards that are fit for this purpose:

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. **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

As described in the first paragraphs of this Report, data population will be analysed and actively reported in the period after the release of the platform - in Reports due M24 and M36. In those reports we will see reports on all 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 actions			
Iterations based on corrective actions			
Measuring performance of new iterations			

Table 1: Data Population Gantt Chart

5 CONCLUSION

The report concludes that current work has been done in a relevant way to support achieving the Key Performance Indicators which have been introduced as the key methodology in this Report:

- Number of supported formats by ingestion services,
- Relative coverage of formats in business scenarios,
- Number of organisations providing data to the platform,
- Relative coverage of organisations in business scenarios,
- Number of data sources providing data to the platform and
- Total data size.

Based on those indicators and metrics, it is evident that the future work in Task 2.4 will have to involve even more contact with the platform users directly. Only by providing hands-on guidance on how more relevant data can be shared, will we increase the data flow towards the platform and ensure all of the KPIs are achieved.