SOFIE - Secure Open Federation for Internet Everywhere

779984

DELIVERABLE D6.5

Data Management Plan

Project title	SOFIE – Secure Open Federation for Internet Everywhere
Contract Number	H2020-IOT-2017-3 – 779984
Duration	1.1.2018 – 31.12.2020
Date of preparation	29.6.2018
Author(s)	Petri Laari (LMF), Giuseppe Raveduto (ENG), Yannis Oikonomidis (SYN), Priit Anton (GT), David Mason (ROV), Francesca Santori (ASM), Alessio Cavadenti (ASM), Tommaso Bragatto (ASM), Francesco Bellesini (EMOT), George Xylomenos (AUEB-RC), Dmitrij Lagutin (AALTO)
Responsible person	Petri Laari (LMF), petri.laari@ericsson.com
Target Dissemination Level	Public
Status of the Document	Completed
Version	1.00
Project web-site	https://www.sofie-iot.eu/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 779984.

Table of Contents

1. Introduction

The main goal of the SOFIE project is to enable diversified applications from various application areas to utilise heterogeneous IoT platforms and autonomous things across technological, organisational and administrative borders in an open and secure manner, making reuse of existing infrastructure and data easy. SOFIE work is guided by four pilots in three different areas: food-chain, mobile gaming, and energy (two different use cases). These pilots will provide feedback on the architectural work and their requirements will be used to identify potential synergies between these different areas.

The pilots will create instances of the SOFIE framework and utilize them in the specific use cases. The pilots will collect relevant data, which among other things will be used to analyze the functionality of the implementation. We surmise that the data is useful for other projects that are creating IoT systems with similar setups.

The purpose of this Data Management Plan is to provide guidelines on how to collect, maintain, and further distribute collected data for external usage. This document specifies the data sets that will be collected from the four pilots implementing instances of the defined SOFIE architecture and framework. In each specification, the content of the data is described, as well as the format and location where they are stored and from where they can be retrieved after the project has ended.

Data that can compromise commercialization prospects or has inadequate protection of, e.g., personal information, shall not be published. The rest of the data will be deposited in an open access repository such as Zenodo (https://www.zenodo.org). When the data is related to a publication, it will be linked to it via OpenAIRE (https://www.openaire.eu).

The rest of the document describes the collected data in more detail, and describes responsibilities related to the collection and securing the data.

This DMP is not a fixed document and it may evolve during the lifetime of the project.

2. Description of Collected Data

This chapter describes the collected data from the different pilots in more detail. These are the initial plans, which may change during the project's lifetime, as the pilots are being implemented and tested.

2.1 Food Chain Pilot

Dataset name	**Field Sensor Measurements**		
Dataset description	Data collected from the various field IoT sensors. Micro-climate data (e.g. air temperature, air humidity, wind direction, wind speed, rain volume, rain intensity), Soil and Crop related data (leaf wetness, soil type, soil temperature, soil humidity, soil conductivity) and Irrigation data (e.g. crop, irrigation frequency, irrigation time, irrigation water pipes pressure, volume of irrigation water consumed). Moreover, this data set will be used to calculate the crop growing degree days (ripening indicator). The data will be associated with time information and geospatial/location information provided by GPS.		
Security and privacy considerations			
Measurement data include product information which may be considered as sensitive from the ownership point of view. Therefore, this has to be taken into consideration if this data is to be considered as a potentially public dataset.			
Datatype name	Environmental conditions measurements		
Data description	Environmental conditions measurements		
Purpose of the data	In order to be able to properly monitor the state of a product in the field, data regarding the temperature, the wind speed and direction, the soil humidity and conductivity, as well as the environmental humidity and the solar radiation has to be collected.		
Relation to project objective	Field data will be a part of the product-related information that will be collected along its field-to-fork path.		
File types	.json		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
SynField IoT platform data	300KB per day per SynField node	Access will be provided, if needed for the project objectives.	Public access only to aggregated/anonymized data

Dataset name	Storage Sensor Measurements		
Dataset description	Data collected from the various storage and distribution center IoT sensors (Aberon WMS platform, http://www.optimum.gr/us/solutions/wrh-management/aberon-wms)		
Security and privacy considerations			
Measurement data include product information which may be considered as sensitive from the ownership point of view. Therefore, this has to be taken into consideration if this data is to be considered as a potentially public dataset.			
Datatype name	Environmental conditions and product tracking measurements		
Data description	Environmental conditions and product tracking measurements		
Purpose of the data	In order to be able to properly monitor the state of a product while being kept at the storage and distribution center, temperature and other measurements as well as location and proximity (e.g., RFID) data has to be collected.		
Relation to project objective	Storage center data will be a part of the product-related information that will be collected along its field-to-fork path.		
File types	.json		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Aberon IoT platform data	MBs per month	Access will be provided, if needed for the project objectives.	Public access only to aggregated/anonymized data

Dataset name	Transportation Sensor Measurements		
Dataset description	Data collected from IoT sensors mounted on a vehicle during transportation. Temperature data, RFID data. The data will be associated with time information and geospatial/location information provided by GPS.		
Security and privacy considerations			
Measurement data include product information as well as vehicle location and state, which may be considered as sensitive both from the ownership point of view and from a security point of view. Therefore, this has to be taken into consideration if this data is to be considered as a potentially public dataset.			
Datatype name	Environmental conditions and vehicle tracking measurements		
Data description	Environmental conditions and vehicle tracking measurements		
Purpose of the data	In order to be able to properly monitor the state of a product while being transported, temperature and other measurements, as well as vehicle location data, has to be collected.		
Relation to project objective	Transportation data will be a part of the product-related information that will be collected along its field-to-fork path.		
File types	.json		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Transportation IoT platform data	MBs per month	Access will be provided, if needed for the project objectives.	Public access only to aggregated/anonymized data

2.2 Energy Pilot (Italy)

Dataset name	**Topology and asset description**		
Dataset description	The topology and asset description includes plans and documentation about assets and equipment. The description of network topologies of electrical, gas and other energy distribution grids is included. In addition, the topologies of IT networks, wired and non-wired, are included. For the IT networks, detailed information about the hardware is part of this dataset.		
Security and privacy considerations			
Information about critical infrastructure may need to be handled confidentially so the security of that infrastructure will not be compromised. Further assessment of the data is needed, before a decision about making it public (and in which extent) can be made.			
Datatype name	Charge point description		
Data description	This data describes Electric Vehicle Supply Equipment (EVSE) status		
Purpose of the data	Optimize the electric power consumption of the electrical vehicles (EVs) charging point using energy from renewable sources		
Relation to project objective	Test the SOFIE platform and blockchain technology through an electric mobility service		
File types	.json		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Information about the charge point infrastructure at Terni pilot site will be provided.	Some KBs per day	Access will be provided, if needed for the project objectives.	The data will not be made public

Dataset name	Measurement data		
Dataset description	Data gathered from smart meters (energy meters) and data gathered from EVs.		
Security and privacy considerations			
The combination of the collected information about driving and charging habits, but also the forecast information could have an impact on the privacy of the user. Before making this data public, this issue needs to be addressed.			
Datatype name	Voltmeter/Current meter/Custom recordings/EV data		
Data description	Voltmeter/Current meter/Custom recordings/EV data (battery state of charge, residual autonomy, minutes to full charge, doors car state, engine car state)		
Purpose of the data	To control the charging behaviour of an electric vehicle (EV) it is important to know the current state of charge of the EV battery, but also the current state of the power grid. A forecast about the use of the EV and the needed energy, based on historical information, can also use information about the driver's behaviour.		
Relation to project objective	Multiple EV's will be controlled in terms of their state of charge and their charging schedule. The schedule takes the current state of the power grid into account. Therefore, it is important to measure the grid state. To calculate a charging plan it is important to make a forecast of the user's behaviour. For this reason, information about the user's behaviour needs to be collected.		
File types	.csv, .xls, .raw, .json		
(Data provider) Origin of the data	Size (xByte)	Access for Partners	Access for the public
	Some MBs per day	Access will be provided, if needed for the project objectives.	Public access only to aggregated/anonymized data

Dataset name	Log and access data
Dataset description	Data which documents the current state or change in state of a system.

Security and privacy considerations

Alarm and logging data can reveal information about critical infrastructure or the behaviour and identity of users who are connected to the systems storing the alarm and logging data. Therefor this data needs to be assessed before a decision about making it public can be made.

Datatype name	Alarm and heartbeat Logs
Data description	Log with a history of alarm and heartbeat states.
Purpose of the data	Alarm and heartbeat data is needed for analysis of a systems behaviour
Relation to project objective	Electric vehicles (EVs) and electric vehicle supply equipments (EVSEs) will be smartly connected. To do so, status information (e.g. alarms and heartbeats) about the EVs are needed
File types	.json

(Data provider) Origin of the data	Size (xByte)	Access for Partners	Access for the public
Alarm data from EVs/EVSEs will be logged.	Some kBs per day	Access will be provided, if needed for the project objectives.	The data will not be made public
Heartbeat data from EVs/EVSEs will be logged.	Some kBs per day	Access will be provided, if needed for the project objectives.	The data will not be made public

Dataset name	Prediction, forecast and planning data
Dataset description	Micro Grid and EV data to plan demand response (DR) campaigns. Also included is data with forecast or schedule character.

Security and privacy considerations
To be evaluated

Datatype name	Power exchange data
Data description	Power exchange within the charge point depending on the electrical output
Purpose of the data	To manage the power flow in an electrical grid, information about the current state is needed
Relation to project objective	Test the SOFIE platform and blockchain technology through an electric mobility service
File types	.json

(Data provider) Origin of the data	Size (xByte)	Access for Partners	Access for the public
Information about the power exchange within the charge point will be recorded.	Some kBs per day	Access will be provided, if needed for the project objectives.	The data will not be made public

Datatype name	Demand response data
Data description	Demand response (DR) signals and available resources for DR
Purpose of the data	This data will be generated and then collected to carry out electric vehicle charges when the PV plant energy surplus is present
Relation to project objective	This data are needed for the calculation of the flexibility needed for DR
File types	.csv, .json

(Data provider) Origin of the data	Size (xByte)	Access for Partners	Access for the public
DR data from trials and lab tests will be stored.	Some kBs per day	Access will be provided, if needed for the project objectives.	The data will not be made public

Datatype name	Energy or Power forecast of PV generation		
Data description	Energy or Power forecast of PV generation		
Purpose of the data	Electrical vehicle charging plans /vehicle charging demand profile		
Relation to project objective	To be defined		
File types	.csv		
(Data provider) Origin of the data	Size (xByte)	Access for Partners	Access for the public
Smart meter of PV plant	Some MBs per day	Access will be provided, if needed for the project objectives.	Public access only to aggregated/anonymized data

Dataset name	Positioning and location data		
Dataset description	Data which documents EVs current position		
Security and privacy considerations			
Positioning and location data can reveal information about the behaviour of EVs that are connected to the systems			
Datatype name	GPS position data		
Data description	Geolocation recordings		
Purpose of the data	Information about the position of EVs		
Relation to project objective	To forecast the EV usage, historic information about car movement is important		
File types	.json		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
EVs geo location data will be recorded	Some kBs per day	Access will be provided, if needed for the project objectives.	The data will not be made public

2.3 Energy Pilot (Estonia)

Dataset name	**Smart meter measurement data**		
Dataset description	Data from the smart meters, providing information about the energy consumption on a specific geographical location.		
Security and privacy considerations			
Positioning and location data can reveal information about the behaviour of customers that are connected to the system. Depending on the country, the smart meter ID and energy consumption can be subject to GDPR. During the SOFIE project we use anonymous smart meter devices that have no relation to customer behaviour and location.			
Datatype name	Energy consumption data		
Data description	Metering point ID, energy consumption KW/h, date and time		
Purpose of the data	Information about the energy consumption		
Relation to project objective	To provide basic energy consumption information about the participant as an input to future trading and agreeing a contract between parties.		
File types	.json		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Customers end-point physical smart meter on site / lab environment simulated data	300 kb per device / day	Access will be provided, if needed for the project objectives.	Simulated data can be made public / customer data will not be made public

Dataset name	System Logs and access data
Dataset description	Data about the Smart meter authentication process, storing the Physically Unclonable Function (PUF) attributes, Strong ID related access information, permissioned nodes “white list”. Monitoring information about access and operations of trusted nodes.

Security and privacy considerations
Dataset consists of private information and cannot be made public.

Datatype name	System log data
Data description	Smart meter ID, PUF ID, log files of system operation
Purpose of the data	Core element of controlling and managing the access to network, adding and removing the smart meters from the grid, monitoring abnormal behaviour of system, enabling access with Public key infrastructure.
Relation to project objective	Enabling to join the trusted network and providing secure access and validation of the input data to the SOFIE federated platform. Covering part of the security element between the IoT and adapter connected to SOFIE.
File types	.json

(Data provider) Origin of the data	Size (xByte)	Access for Partners	Access for the public
Smart meter owners premises (TSO, DSO etc.)	Up to 10 kb per smart meter a day	Access will be provided, if needed for the project objectives.	The data will not be made public

Dataset name	Customer data, positioning and location data
Dataset description	Dataset contains all relevant information about the person/entity who owns the smart meter. This includes the entities name, address, smart meter location and other information needed for the contract agreement between the energy provider and consumer.

Security and privacy considerations
Dataset cannot be made public.

Datatype name	Customer data		
Data description	Smart meter ID, customer related information.		
Purpose of the data	Confirmation of the entity and obligatory from the energy service contract side.		
Relation to project objective	Information that is handled by Smart meter provider customer request management side. Not directly linked to SOFIE platform, but in case of disputes it makes it possible to create a link between the ID and the legal entity.		
File types	.json		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Smart meter “data hub” premises (TSO, DSO etc.)	Up to 500 kb for one smart meter	Access will be provided only for the simulated data	Access will be provided only for the simulated data

Dataset name	Ownership, access and permission rights data
Dataset description	Data about who has a right to give access to smart meter data to third parties (already trusted by the network)

Security and privacy considerations
Anonymous data can be shared publicly, customer related data is not public.

Datatype name	Distributed Ledger Notary
Data description	Requests for data access, granting access based on Smart meter ID, agreement details between participants about the data transfer (parties involved, Smart Meter IDs, access protocol)
Purpose of the data	Share the requests between parties, agreeing access rights and enabling data transmission.
Relation to project objective	Core element of SOFIE federated platform, enabling data exchange between parties, not storing the data but making sure that requirements are met before data exchange.
File types	.json

(Data provider) Origin of the data	Size (xByte)	Access for Partners	Access for the public
Distributed ledger information, all nodes having consensus over access rights	2-3 kBs per transaction	Access will be provided, if needed for the project objectives.	The data will not be made public

2.4 Mobile Gaming Pilot

Dataset name	**Game content DNA**		
Dataset description	Data written to the blockchain for in-game content. This will enable swapping or buying with other players (e.g. characters, weapons, equipment, parts), leveraging DLTs to provide player ownership of the asset, an open market for trading transactions, transparency and consistency of asset attributes and transactions. This will also allow mini-games to be built on top of the game content.		
Security and privacy considerations			
This information will be transparent to consumers and potentially held on a public DLT. This data will contain no personal or commercially sensitive data.			
Datatype name	Game content DNA		
Data description	Game content DNA		
Purpose of the data	The unique attributes of the game content.		
Relation to project objective	Key content data for core game and mini-games.		
File types	.json		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Rovio Sofie	About 1kB per transcation	Access will be provided, if needed for the project objectives.	Dependant on DLT - public or private

Dataset name	Mobile Game Analytics		
Dataset description	Session data collected from the consumer playing the game. Used to understand player behaviour so as to improve and tune the game. Collected from game's smartphone client, game server and game services (payment etc.) Will be collected and stored by using Rovio's data analytics pipeline.		
Security and privacy considerations			
This data may be considered commercially sensitive as likely to reveal how the game runs and operates - could be misused to cheat. This data will pseudonymised to ensure compliance with GDPR and is unlikely to be provided as a public dataset, even if anonymised.			
Datatype name	Game events		
Data description	Analytic game events		
Purpose of the data	Data used to improve and tune the game including game design, economy balancing, game play optimisation, cheat detection etc.		
Relation to project objective	Key data for game development.		
File types	.json		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Mobile Game Client, Game Server, Game Services	MBs per user per day	No access due to personal data and commercial sensitivity	No access due to personal data and commercial sensitivity

Dataset name	IoT device data		
Dataset description	Data collected from IoT devices used to as part of IoT mini-game(s). These game(s) would leverage IoT devices via the Sofie platform to provide game experiences, for example using Beacons for a scavenger hunt completing 'collection' missions built on the assets from the core game, potentially providing a reward in the game of from the locations (e.g. retailer).		
Security and privacy considerations			
Parts of this data should not commercially sensitive for the pilot unless it includes partner data or sensitive location data. This data should include no personal data in order to be shareable.			
Datatype name	IoT device events		
Data description	IoT device event data used to interact with the game		
Purpose of the data	Event data used to inform the game of interaction with the IoT device and environmental information required for the game.		
Relation to project objective	IoT data required to enable gaming pilot		
File types	.json		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Mobile Game Client, Game Server, Game Services	KBs per user per day	Open access if data contains no commercial or personal data	Potentially open access if sharing the data does not undermine the game play

3. Resources Required for Storing Data

Collecting the data and storing it requires both working hours and data storage that create cost for the partners. Also other things e.g. licensing may create costs. In this section, the potential cost targets are described for each of the pilot projects.

Food Chain pilot: Decisions regarding data access and licensing costs and strategies will be made in due time, once the datasets are available.

Energy pilots (both use cases): The cost of making data accessible also depends on the amount of data, the cost of long term storage solution and the effort required for publication. An estimation cannot be delivered at this time, as too many influencing factors are unknown at the moment.

The responsibility for the long term data archiving and publication is not specified yet.

Mobile Gaming pilot: The cost of making data accessible also depends on the amount of data, the cost of long term storage solution and the effort required for publication. An estimation cannot be delivered at this time, as too many influencing factors are unknown at the moment.

4. Data Handling and Security

Each of the pilots have different security requirements for their data. Although all pilots follow the generic rules, they still have some specific issues to be considered from their perspective. In the following, data security for each pilot project is described with their own requirements.

Food Chain pilot: The exact access policy has not been defined at this stage, as issues related to data privacy, confidentiality and anonymity have to be taken into consideration first. For data which cannot be shared, the reasons will be mentioned and these data will be preserved in repositories with limited access.

Each pilot partner will be responsible for its own generated data, including storage, data recovery, and transfer.

To facilitate a good level of collaboration between the consortium's partners, pilot test data repositories will be available at Synelixis SynField cloud platform.

Energy pilots (both use cases): Each partner is responsible for recoverability of their own generated data. The assessment of security risks, which may arise, with the content of gathered data will be done by the entity who is collecting the data.

Mobile Gaming pilot: The data security will comply with Rovio's privacy notice http://www.rovio.com/privacy and terms of service http://www.rovio.com/terms-of-service.

5. Ethical Aspects

The SOFIE partners will comply with the GDPR legislation. Ethical principles are described in more detail in Section 5.1 of the Annex 1 to the SOFIE Grant Agreement (Description of the Action, Part B).