SecureIoT

Project Acronym:	SecureIoT
Grant Agreement number:	779899 (H2020-IoT03-2017 - RIA)
Project Full Title:	Predictive Security for IoT Platforms and Networks of Smart Objects

DELIVERABLE D1.3 - Data Management Plan

Deliverable Number	**D1.3**
Deliverable Name	**Data Management Plan**
Dissemination level	Public
Type of Document	ORDP
Contractual date of delivery	30/06/2018
Deliverable Leader	INTRASOFT International
Status & version	2.1 Final
WP / Task responsible	INTRASOFT International (INTRA)
Keywords:	Data Management Plan, data accessibility, data repositories, data interoperability, security.
Abstract (few lines):	The deliverable describes the data management life cycle for the data to be collected, processed and/or generated by SecureIoT.

Deliverable Leader:	George Koutalieris (INTRA)
Contributors:	Nikos Kefalakis (INTRA), Filippos Raditsas (INTRA), Sofoklis Kyriazakos (iSPRINT), Abdelkader Lahmadi (INRIA), John Soldatos (AIT)
Reviewers:	JérômeFrançois (INRIA) Sofianna Menesidou (UBI)
Approved by:	George Koutalieris (INTRA)

Executive Summary

This report includes a preliminary version of the SecureIoT Data Management Plan, where datasets that are likely to be opened and shared as part of the SecureIoT ecosystem have been identified. SecureIoT will release updates to the present DMP, in-line with the evolution of the specification and implementation of validating use cases.

Document History			
Version	**Date**	**Contributor(s)**	**Description**
0.01	18/05/2018	Nikos Kefalakis	Initial ToC,
0.02	13/06/2018	All partners	Added collected data related information from Partners.
0.03	18/06/2018	Sofoklis Kyriazakos	Additional contributions regarding GDPR
0.04	21/06/2018	All partners	Additional dataset information collection`
0.05	22/06/2018	Nikos Kefalakis, all partners	Added descriptions about the Dataset information fields/additional editing
0.09	25/06/2018	JérômeFrançois (INRIA) Sofianna Menesidou (UBI)	Peer review
1.0	29/06/2018	All partners	Final submission

Table of Contents

Definitions, Acronyms and Abbreviations

Acronym	Title
API	Application Programming Interface
CC2U	CloudCare2U
CMIP	Common Management Information Protocol
CoAP	Constrained Application Protocol
CPU	Central Processing Unit
DMP	Data Management Plan
DoA	Description of the Action
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
HDD	Hard Disk Drive
HTTPS	Hypertext Transfer Protocol Secure
IACS	Industrial Automation and Control Systems
JSON	JavaScript Object Notation
M2M	Machine to machine
RAM	Random Access Memory
ROS	Robot Operating System
SNMP	Simple Network Management Protocol
TBD	To be Determent
URL	Uniform Resource Locator
XML	Extensible Markup Language

Project Title: SecureIoT
Contract No. 779899
Project Coordinator: INTRASOFT International S.A.

Project Title: SecureIoT
Contract No. 779899
Project Coordinator: INTRASOFT International S.A.

1 Introduction

1.1 Overall Objective

As part of its exploitation and sustainability strategy, SecureIoT will be releasing part of its platform as open source software, which will be made available through the project's ecosystem portal that is developed in WP7 of the project. Along with software, the project plans to release datasets as well, as means of facilitating third-parties (i.e. members of the SecureIoT platform community) to test, validate and possibly extend SecureIoT developments. This intention is fully in-line with SecureIoT's strategy for data management, as the latter is reflected in the project's DoA (Description of the Action) document. In this context, this part of the deliverable is devoted to the presentation of the project's Data Management Plan (DMP).

In principle, the release of data in the scope of SecureIoT is aimed at the following objectives:

- **Validation of SecureIoT components**: SecureIoT needs to provide partners and third-parties with an easy way for using and validating its developments. In most cases, this requires the availability of some data that can be used to validate the operation of SecureIoT components.
- **Demonstration of SecureIoT components**: In addition to boosting the validation of SecureIoT components, datasets are also needed for running demonstrations of the various prototypes. Demonstrations is an essential element for ecosystem building, as third-parties are usually starving for one-click demonstrations that could easily help them understand the operation of certain software components.
- **Training and Education**: Open datasets can be an invaluable resource for developing training and education modules, such as the ones developed in the scope of the SecureIoT training services.
- **Follow the GDPR guidelines**: In May 2018, the new European Regulation on Privacy, the General Data Protection Regulation, (GDPR) came into effect. In this DMP we will describe the measures to protect the privacy of all data provided in the light of the GDPR.

In order to realize these objectives, SecureIoT is considering the release of certain datasets as open data. This DMP identifies candidate datasets, along with the preconditions for making them openly accessible as part of offerings to the project's ecosystem.

1.2 DMP Evolution

Project Title: SecureIoT
Contract No. 779899
Project Coordinator: INTRASOFT International S.A.

The DMP presented in this deliverable is characterized as preliminary, given that the project is still in the process of finalizing the specifications of validating scenarios and use cases, while actual data capturing has not commenced yet. SecureIoT will release updates to the present DMP, in-line with the evolution of the specification and implementation of validating use cases, including their deployment in the test environments.

As already outlined, this preliminary version of the DMP has a dual objective: First to identify available datasets that are likely to be opened and shared as part of the SecureIoT ecosystem. Second, to identify the conditions that should be met in order for these datasets to be opened. The identification of such conditions is particularly important, given that making data public is against the corporate policies of the manufacturers of the consortium. In certain case, this important barrier can be lowered following appropriate processing of the data (e.g., anonymization), as well as following reception of appropriate approvals.

1.3 GDPR

Since the 25th May 2018, GDPR is valid and obligatory and that applies also for SecreIoT project. Therefore, partners are following the same new rules and principles. In this section, we are describing how the founding principles of the GDPR will be followed in the SecureIoT project. More specifically, following points are taken into account:

- **Lawfulness, fairness and transparency**: Personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data user.
- **Purpose limitation**: Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.
- **Data minimization**: Personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.
- **Accuracy**: Personal data shall be accurate and, where necessary, kept up to date. All data collected will be checked for consistency.
- **Storage limitation**: Personal data shall be kept in a form, which permits identification of data for no longer than is necessary for the purposes for which the personal data are processed.
- **Integrity and confidentiality**: Personal data shall be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures.
- **Accountability**: The controller shall be responsible for, and be able to demonstrate compliance with the GDPR.

1.4 Datasets Description Template

In following paragraphs we provide an overview of the datasets that SecureIoT will be considering for release as part of its ecosystem. Note that the inclusion of a dataset in the list implies that it is considered to be offered in the project's portal, subject to the clearance of some precondition. Datasets information is divided in five different categories and in each category the information is described in a tabular form. The attributes of the information provided are:

General information

- **Ref. No**: Sequence Number
- **Title**: The title of the Dataset
- **Version**: The dataset info version
- **Description**: Briefly describe what data would represent
- **Type of data**: Data already existing OR date to be released
- **Dataset availability** : Date of the dataset availability
- **Future revisions anticipated**: Define if future revisions are anticipated
- **Owner**: Denotes the provider of the datasets.
- **Contact Person**: Person in charge of the release of the dataset and its inclusion in the SecureIoT portal.
- **Related Use Cases**: The set of SecureIoT use cases that the dataset related. The description of the use cases is performed with reference to deliverable D2.2.
- **Utility / Potential Use**: An illustration of why the particular dataset could be useful to the SecureIoT community. e.g.:
    - Research and experimentation
    - Service Development / Integration
    - Training & Education

Environment / Context

- **Directly observable device types**: i.e., Sensor, robot, vehicle board, monitor device, edge node, gateway
- **Directly observable software**: i.e., IoT application, gateway software, cloud service app…
- **Indirectly observable device**: i.e., Sensor, robot, vehicle board, monitor device, edge node, gateway (devices which are not directly monitored, be exhaustive to the extent possible)
- **Indirectly observable software**: List the software which is observed indirectly
- **Architecture/Topology description and communication protocols**: Figure showing where are the monitoring probes (some incertitude may occur)

Data access

Here there are three cases:

Project Title: SecureIoT
Contract No. 779899
Project Coordinator: INTRASOFT International S.A.

1. Data is already retrieved and stored as data files
2. Monitoring data can be retrieved through an interface
3. Data is present in sw/hw but no means exists yet to access them remotely, need for a probe to be developed

The first two may coincide. Data access has the following attributes:

- **Dataset provided as data file(s)**: Define if the dataset is provided as data file(s)
- **Remote accessibility**: Define if the data are remotely accessible and how.
- **If data is not yet accessible, how can they be retrieved?**: Define the method which the data can be accessed in the future.

Data description

- **Data format**: i.e., NetFlow, pcap, syslog, json (when an interface is used, the format of embedded data is needed to be described)
- **Encryption**: explain if and how the data are encrypted.
- **Data format description**: describe the syntax and semantics of data (very important for non-standard formats, e.g. describe the columns of a csv file, or the structure and semantics of what contains a JSON file)
- **For unusual format, tool to read it**: specify the required tool/library to read the data if their data type is not standard.
- **Dataset generation**: specify if the data was monitored in a system with real users? If no, how the data has been generated?
- **Attack**: specify if the dataset contain attacks? If yes, specify if the attacks are annotated? If yes, specify what is the granularity of the annotations?
- **Dataset statistics**: i.e., Duration, size(s) in appropriate format (MB, pkts), number of packets breakdown per IP address, protocols… (be exhaustive as possible)
- **Sample of data**: Provide a sample of data in this attribute or a link to them.

Data restrictions

- **Is the data open publicly?** : Specify if the data are public
- **If no, is there a plan to make data open?** : Specify if the data are not public a plan to make them public.
- **If no, will the data be accessible to the consortium, or to specific partner(s)?** : Specify if the data can be accessible to the consortium, or to specific partner(s) in case they cannot be public.
- **If yes, for how long?** : Specify the time period the data can be accessible to the consortium, or to specific partner(s) in case they cannot be public.
- **Can the data be used for public dissemination**: Specify if the data can be used for public dissemination (without revealing the full content of the data, aggregated view)
- **Who owns the data?** : Identify the data owner

- **Legal issues**: Specify the confidentiality level of the dataset and the license under which the dataset could be opened and offered publicly.

1.5 SecureIoT Datasets

This is the first version of the DMP deliverables and some of the Dataset information has not been determined yet. The missing fields will be completed in the coming versions of the deliverable. The information provided below has been collected in collaboration with WP3 and WP4.

1.5.1 Multi-Vendor Industrie 4.0 Usage Scenarios Data

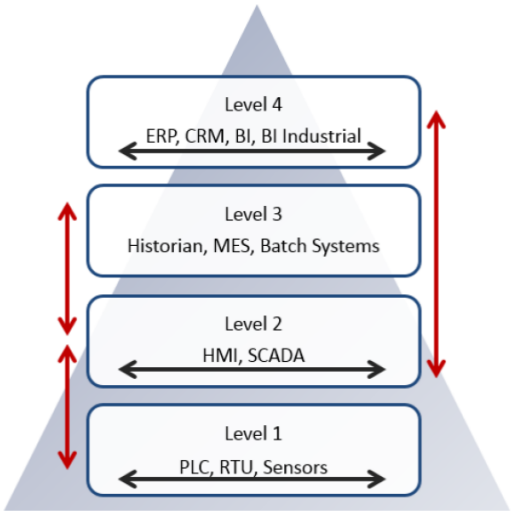
1.5.1.1 General information

Ref. No	*0001*
Title	Manufacturing data resulting from sensors, machines, IACS
Version	1.0
Description	This dataset contains or will contain different kind of data generated within manufacturing. These will be sensor data sometimes aggregated for a complete machine. Moreover, data generated by IACS shall be considered in the use case. The data may include application information, context information, status information, traffic data and much more. Details will be specified within the use case
Type of data	Application, context, performance, status, usage, alerts, etc.
Dataset availability	TBD

Future revisions anticipated	Yes
Owner	Weidmüller, Phoenix, it's OWL
Contact Person	David Schubert (d.schubert@its-owl.de)
Related Use Cases	TBD
Utility / Potential Use	TBD

1.5.1.2 Environment / Context

Directly observable device types	• IoT Gateways, e.g. FUJITSU Intelliedge • IACS systems
Directly observable software	• TBD in the use case. We consider several systems to be relevant: • P@SSPORT factory virtualization • SIEMENS Minsphere • FUJITSU IoT-Platform • FUJITSU Colmina Intelligent Dashboard
Indirectly observable device	*Sensor, robot, vehicle board, monitor device, edge node, gateway (devices which are not directly monitored, be exhaustive to the extent possible)* • Manufacturing devices connected to the IACS or gateways. These may be sensors, etc. The objective

	within the use case will be to use virtualized control systems and sensors.
Indirectly observable software	
Architecture/Topology description and communication protocols	*Figure showing where are the monitoring probes (some incertitude may occur)*  Probes may be placed on each level and within the vertical communications. Moreover, probes should be placed at the IoT-Platform level

1.5.1.3 Data access

Dataset provided as data file(s)	*Yes/No* TBD	
Remote accessibility	Yes/No	Usually: No
	Protocol	TBD

	Message format	TBD
	Pull/Push	TBD
	Provided interface	TBD
If data is not yet accessible, how can they be retrieved?	Describe the architecture and where the probe can deployed	TBD
	Probe development requirements	TBD
	Usable software API on device	TBD

1.5.1.4 Data description

Data format	• TBD inm the use case	
Encryption	*Is the data encrypted? (explain)* Yes, communication between all the components will rely on secure communication protocols, i.e., HTTPS.	
Data format description	• TBD in the use case	
For unusual format, tool to read it	TBD	
Dataset generation	Was the data monitored in a	We consider virtualized systems with scenarios and no real users

	system with real users?	
	If no, how the data has been generated?	*Actions triggered /performed/simulated, how many of them, methodology*
Attack	Does the dataset contain attacks?	In a first step we plan to provide normal operations data Later on the virtualized plant(s) shall be exposed to attacks and the data shall include attacks
	If yes, are the attack labeled?	No
	If yes, what is the granularity of the labels?	
Dataset statistics	TBD	
Sample of data	TBD	

1.5.1.5 Data restrictions

Is the data open publicly?	No
If no, is there a plan to make data open?	No
If no, will the data be accessible to the consortium, or to specific partner(s)?	Yes, whole consortium
If yes, for how long?	End of project
Can the data be used for public dissemination (without revealing the full content of the data, aggregated view)	TBD

Who owns the data?	The respective partners of use case scenario T6.2
Legal issues	There may be several issues regarding personal data of either customers or employees. Within the Industrie4.0 use case we shall consider anonymization of data for SecureIoT data collection. Moreover we will face M2M communications and thus telecommunication data will be a key part if the data collection. Finally, the data may contain business secrets, e.g. process parameters.

1.5.2 IoT-Enabled Socially Assistive Robots Usage Scenarios Data

1.5.2.1 QTrobot

1.5.2.1.1 General information

Ref. No	*0002*
Title	QTrobot
Version	1.0
Description	This dataset consists of traffic • In QTrobot • Between robot and its tablet GUI • Between robot and CC2U of iSprint • Between robot and internet • Between tablet and its cloud backup server

Type of data	• Raw sensory data (video stream, sound stream, robot's joint angles) • Perception data (recognized images, objects, faces, human gesture, speech, direction of voice) • Application and actuation data (video, sound and gesture outputs of QT, application events, recognized activity, proposed activities) • Robot and tablet config and performance (CPU, RAM, HDD and network bandwidth access and usage, network connection, running processes) • User data (user profile, application history, user performance and progress data, user-built applications) • Network traffic (packages)
Dataset availability	Mechanisms and Interfaces to capture and communicate the data to the destination device are to be developed.
Future revisions anticipated	Yes
Owner	LuxAI
Contact Person	Pouyan Ziafati (ziafati@luxai.eu)
Related Use Cases	Social Assistive Robots
Utility / Potential Use	Research and experimentation Training & Education

1.5.2.1.2 Environment / Context

Directly observable device types	*Sensor, robot, vehicle board, monitor device, edge node, gateway* • Robot Gateway • Tablet Gateway
Directly observable software	• Robot Operating System (ROS)
Indirectly observable device	• 3D camera • Microphone array • Motor sensors • Computer inside the robot • Android tablet • Router inside the robot • Wi-fi inside the robot
Indirectly observable software	Camera interface, microphone interface, motor interface, image recognition, face recognition, object and gesture recognition, sound play, video play, robot plan executor, gesture record and play
Architecture/Topology description and communication protocols	Robot --- ROS (JSON API through Websocket server can be developed)

1.5.2.1.3 Data access

Dataset provided as data file(s)	Yes	
Remote accessibility	Yes/No	Yes (but means have to be developed)
	Protocol	ROS (or its Websocket server interface)
	Message format	*ROS messages (or JSON equivalent of ROS messages)*
	Pull/Push	*Pull, push*
	Provided interface	*ROS service/pub-sub interface + message description (or Websocket URI to be developed)*
If data is not yet accessible, how can they be retrieved?	Describe the architecture and where the probe can deployed	*We use ROS to communicate between different pieces of software in the robot, and to communicate among the robot and tablet. ROS can be provided with a websocket JSON-based interface which we can use to develop a prob to access the robot. The other way around however would be to extend the SecureIoT data capturing interface to support direct communication with ROP.*
	Probe development requirements	*See previous answer.*
	Usable software API on device	*See previous answer.*

1.5.2.1.4 Data description

Data format	*NetFlow, pcap, syslog, json (when an interface is used, the format of embedded data is needed to be described)* Network traffic (could be pcap for instance) ROS Messages (proprietary format, or Jason equivalent)	
Encryption	Most of the data is not encrypted.	
Data format description	Full pcap file including payload Each type of data has its own format.	
For unusual format, tool to read it	*ROS messages are simple data structures similar to C structs.* *http://wiki.ros.org/Messages*	
Dataset generation	Was the data monitored in a system with real users?	*May be possible*
	If no, how the data has been generated?	*Data has not been generated*
Attack	Does the dataset contain attacks?	No
	If yes, are the attack labeled?	-
	If yes, what is the granularity of the labels?	*Per packet, per flow, timeline of anomalies*
Dataset statistics	TBD	
Sample of data	TBD	

1.5.2.1.5 Data restrictions

Is the data open publicly?	*No*
If no, is there a plan to make data open?	*Some parts can be made open*
If no, will the data be accessible to the consortium, or to specific partner(s)?	*Most part yes, Anonymization may be needed.*
If yes, for how long?	TBD
Can the data be used for public dissemination (without revealing the full content of the data, aggregated view)	*TBD*
Who owns the data?	TBD
Legal issues	***Flags:*** ☒ *data may be "personal data"* ☐ *we plan to combine/merge the data with this other data source:* _____ ☒ *data may be "telecommunication metadata"* ☒ *data may be "telecommunication content"* ☐ *data is encrypted* ☒ *data may contain business secrets*

1.5.2.2 CC2U

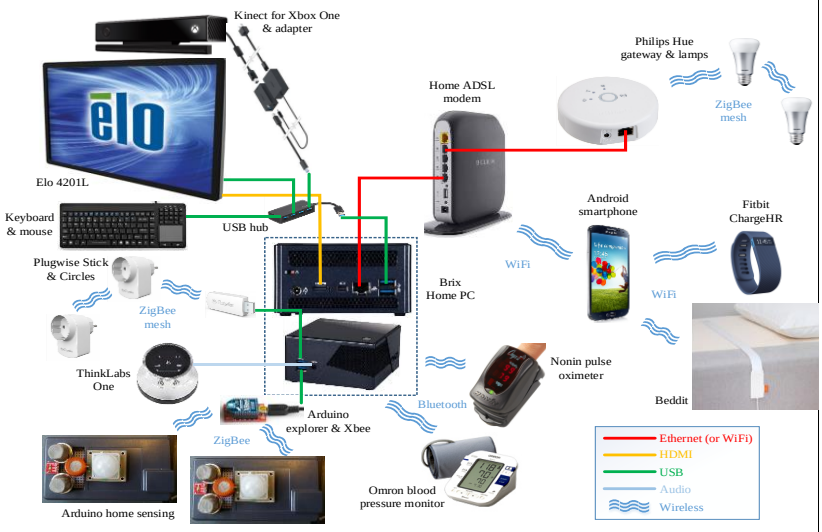
1.5.2.2.1 General information

Ref. No	*0003*
Title	CC2U
Version	1.0
Description	The Cloud Gateway if CC2U has different set of software interfaces used for exchange of commands and data primarily from and to Remote Proxy and that can be grouped to the following categories: • Control and configuration interfaces (I_1) • Sensing data interfaces (I_2) • Notifications interface (I_3) • Actuator interfaces (I_4)
Extended Description	Control and configuration Cloud Gateway interfaces (I1) are used for receiving registration and point of contact information (version, status, etc.) from remote proxy running in local home environments, synchronization of local and cloud data, obtaining device configuration data from cloud and remote configuration of local platform. Sensing data interfaces (I2) are used for receiving all sensing data from home environments and storing them using Data Manager. This includes user activity data, environmental sensing data (temperature, humidity, luminance, gas levels, movement, and presence), furniture sensing data, appliance sensing data, speaker sensing data, visual sensing and vitals data. Notification interface (I3) is used for exchanging notification messages from local reasoners in local environments and Notification Manager. Actuator interface (I4) is used for control and sensing actuator commands from cloud components towards local home environment.
Type of data	Application, context, performance, status, usage, alerts, etc.

Dataset availability	TBD
Future revisions anticipated	Yes
Owner	iSPRINT
Contact Person	Sofoklis Kyriazakos
Related Use Cases	Social Assistive Robots
Utility / Potential Use	TBD

1.5.2.2.2 Environment / Context

Directly observable device types	• Wearable devices (e.g. Fitbit tracker) • Medical devices (e.g. NONIN SpO2, OMRON blood pressure) • Environmental sensors (temperature, humidity) • A/V sensing (e.g. cameras, KINECT) • QT Robot
Directly observable software	• CloudCare2U – Cloud Gateway
Indirectly observable device	TBD
Indirectly observable software	TBD

Architecture/Topology description and communication protocols	*Figure showing where are the monitoring probes (some incertitude may occur)* 

1.5.2.2.3 Data access

Dataset provided as data file(s)	TBD	
Remote accessibility	Yes/No	Yes
	Protocol	TBD
	Message format	*JSON*
	Pull/Push	*Pull, push*
	Provided interface	TBD
If data is not yet accessible, how	Describe the architecture and where	TBD

can they be retrieved?	the probe can deployed	
	Probe development requirements	TBD
	Usable software API on device	TBD

1.5.2.2.4 Data description

Data format	TBD	
Encryption	Yes, communication between all the components will rely on secure communication protocols, i.e., HTTPS.	
Data format description	• TBD in the use case	
For unusual format, tool to read it	-	
Dataset generation	Was the data monitored in a system with real users?	Yes
	If no, how the data has been generated?	TBD
Attack	Does the dataset contain attacks?	No
	If yes, are the attack labeled?	*Yes/No*

		-
	If yes, what is the granularity of the labels?	-
Dataset statistics	TBD	
Sample of data	***Url*** *<HOST_URI>/Notif?user={user_name}&&date={date}&time={time}&type={type}&title={title}&content={content}&prior={priority}&source={source}* ***Parameters*** *user – user name (e.g. "Bob"), date – Date of the trigger (e.g. "08-05-2014"), time – Time of the trigger (e.g. "12.30.21"), type – The design type of the notification, in regards to its representation to the UI (e.g. "two-buttons"), title- The title of the notification shown on the UI (e.g. " Congratulations!"), content – The information content of the notification shown on the UI (a JSON formatted information), prior – the priority value for the given type of notification. This facilitates the possibility to sort the notifications based on priority (left for future use), source – the url address of the component that sends the notification.*	

1.5.2.2.5 Data restrictions

Is the data open publicly?	No
If no, is there a plan to make data open?	No
If no, will the data be accessible to the consortium, or to specific partner(s)?	TBD
If yes, for how long?	-
Can the data be used for public dissemination (without revealing the full content of the data, aggregated view)	Yes thru Anonymization

Who owns the data?	iSPRINT/LuxAI
Legal issues	TBD

1.5.3 Connected Car and Autonomous Driving Usage Scenarios Data

1.5.3.1 General information

Ref. No	*0004*
Title	Connected car and autonomous driving data
Version	1.0
Description	This dataset contains or will contain different kind of data related to the connected car and autonomous driving data, i.e., application information, context information, traffic data…
Type of data	Application, context, performance, usage, alert.
Dataset availability	Application data is available however this is not formally put into a log (e.g. JSON) so formally capturing this log and putting it into a dataset file is to be developed.
Future revisions anticipated	Yes
Owner	IDIADA
Contact Person	David Evans
Related Use Cases	Connected and Autonomous Vehicles
Utility / Potential Use	Research and experimentation

1.5.3.2 Environment / Context

Directly observable device types	• IDIADA IDAPT platform

Directly observable software	• IoT FIWARE related components:
	  ○ IoT Agent (JSON)
	  ○ Orion Context Broker

Indirectly observable device	• Vehicle components connected to the IDAPT platform. For instance:
	  ○ Vehicle Speed
	  ○ Braking information
	  ○ Steering Wheel Angle
	  ○ GPS Heading
	  ○ GPS Speed
	  ○ Yaw_Rate
	  ○ …

Indirectly observable software	
Architecture/Topology description and communication protocols	Vehicle components --- CAN bus --- IDAPT platform IDAPT platform --- MQTT / TCP + HTTPS + REST (monitoring probe)--- FIWARE IoT Agent FIWARE IoT Agent --- TCP + HTTPS + REST (monitoring probe) --- FIWARE Context Broker ---

1.5.3.3 Data access

Dataset provided as data file(s)	Yes	
Remote accessibility	Yes/No	No
	Protocol	-
	Message format	-
	Pull/Push	-
	Provided interface	-
If data is not yet accessible, how can they be retrieved?	Describe the architecture and where the probe can deployed	TBD
	Probe development requirements	TBD
	Usable software API on device	TBD

1.5.3.4 Data description

Data format	*NetFlow, pcap, syslog, json (when an interface is used, the format of embedded data is needed to be described)* • Application data / Context data ○ *ROS is available at the moment, but this is moreso for development purposes, JSON to be implemented,*

	however we are quite flexbile in relation to how the data is captured to a file. ○ At Cloud level: FIWARE NGSI model (http://fiware.github.io/context.Orion/api/v2/latest/) • Traffic data ○ Pcap files • Syslogs
Encryption	Yes, communication between all the components will rely on secure communication protocols, i.e., HTTPS.
Data format description	*Syntax and semantics of data (very important for non-standard formats, e.g. describe the columns of a csv file, or the structure and semantics of what contains a JSON file)* • Application data / Context data ○ *ROS is available at the moment, but this is moreso for development purposes, JSON to be implemented, however we are quite flexbile in relation to how the data is captured to a file.* ○ At Cloud level: FIWARE NGSI model (http://fiware.github.io/context.Orion/api/v2/latest/) • Traffic data ○ Pcap files • Syslogs
For unusual format, tool to read it	TBD

Project Title: SecureIoT
Contract No. 779899
Project Coordinator: INTRASOFT International S.A.

Dataset generation	Was the data monitored in a system with real users?	We are analysing several options: 1. *To use synthetic data generated by a simulator tool (http://www.dlr.de/ts/en/desktopdefault.aspx/tabid-9883/16931_read-41000/)* 2. *To use real data:* a. *Logs from vehicles and / or near-real-time streaming of data.*
	If no, how the data has been generated?	TBD
Attack	Does the dataset contain attacks?	No
	If yes, are the attack labeled?	-
	If yes, what is the granularity of the labels?	-
Dataset statistics	TBD	
Sample of data	TBD	

1.5.3.5 Data restrictions

Is the data open publicly?	No

If no, is there a plan to make data open?	No
If no, will the data be accessible to the consortium, or to specific partner(s)?	Yes, whole consortium
If yes, for how long?	End of project
Can the data be used for public dissemination (without revealing the full content of the data, aggregated view)	No
Who owns the data?	IDIADA / ATOS
Legal issues	We have identified that there may be some legal issues regarding the collected data. For instance, the GPS position of the vehicle (indirect identification), a vehicle identification number… Therefore, the data could lead to single out car driver.

2 Data Access and Sharing

Due to the nature of the data involved, some of the results that will be generated by each project phase will be restricted to authorized users, while other results will be publicly available. As is our commitment, data access and sharing activities will be rigorously implemented in compliance with the privacy and data collection rules and regulations, as they are applied nationally and in the EU, as well as with the H2020 rules. In the case end-user testing will be performed, SecureIoT users would be required to pre-register and consent using the system. Then they will need to authenticate themselves against a user database. If successful, the users will have roles associated with them. These roles will determine the level of access that a user will be given and what they will be permitted to do.

As the raw data included in the data sources will be gathered from the closed and controlled SecureIoT environment, collected measurements will be seen as highly commercially-sensitive. Therefore, access to raw data can only take place through the partners involved in the project. For the data analytic models to function correctly, the data will have to be included into the SecureIoT databases. The results of the IoT data collection and analysis will be secured and all privacy concerns will be catered during the design phase. In the cases of trend analytics, anonymization methods will be applied as part of the built-in cloud platform features.

Publications will be released and disseminated through the project dissemination and exploitation channels to make external research and market actors aware of the project as well as appropriate access to the data.

Within the project, our produced conference papers and journal publications will be Green Open Access and stored in an appropriate repository – such as OpenAIRE (European Comission, 2015), Registry of Research Data Repositories (German Research Foundation, 2015) or Zenodo (CERN Data Centre, 2015).

3 Data Management Plan Checklist

At the end of the project, we will be carrying out the following checklist to ensure that we are meeting the criteria to have successfully implemented an Open Access Data Management Plan. The required KPIs will be updated in subsequent versions of this document. By adhering to the items below, we are confident that the project will provide open access to the appropriate data and software, and thereby, enable researchers to utilize the findings of this project to further expand their knowledge capacity and personal gains as well as to provide the IoT industry with the necessary tools to advance their business and processes.

1. Discoverable:
    a. Are the relevant data that are to be made available, our project publications or any Open software that has been produced or used in the project, easily discoverable and readily located?
    b. Have we identified these by means of a standard identification mechanism?
2. Accessible:
    a. Are the data and associated software in the project accessible, where appropriate, and what are the modes of access, scope for usage of this data and what are the licensing frameworks, if any, associated with this access (e.g. licensing framework for research and education, embargo periods, commercial exploitation, etc.)?
3. Useable beyond the original purpose for which it was collected:
    a. Are the data and associated software, which are made available, useable by third parties even after the collection of the data?
    b. Are the data safely stored in certified repositories for long term preservation and curation?
    c. Are the data stored along with the minimum software, metadata and documentation to make them useful?
4. Interoperable to specific quality standards:
    a. Are the data and associated software interoperable, allowing data exchange between researchers, institutions, organizations, countries, etc. (e.g. adhering to standards for data annotation, data exchange, compliant with available software applications, and allowing re-combinations with different datasets from different origins)?

4 Conclusions

This deliverable has provided an initial framework on how to build the data collecting - and sharing plan during the course of the SecureIoT project and after the project will be finished. This plan will be updated as the project progresses, addressing issues such as dataset repository management and hosting of datasets after the end of the project, also considering public repositories. This deliverable is regarded as a live document which will be updated incrementally as the project progresses. This version sets the overall framework that will form the basis for two additional iterations on M18 and M36, towards the overall delivery of a comprehensive document at the end of the project.

In this version of the deliverable, we outlined the descriptions of the Use Case related Datasets, which are still in development and data access aspects have been addressed.

The upcoming revisions of this deliverable will focus -among other- to a fuller presentation of the datasets, description of the SecureIoT data models, update of data access and sharing and update of data interoperability priorities.

5 References

1 "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020", Version 3.2, 21 March 2017. Retrieved from: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

2 "Guidelines on FAIR Data Management in Horizon 2020", Version 3, July 2016. Retrieved from: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

3 European Comission. (2015, October 26). *OpenAIRE*. Retrieved from OpenAIRE: www.openaire.eu

4 German Research Foundation. (2015, October 26). *re3data.org*. Retrieved from re3data.org: www.re3data.org

5 Google Inc. (2015). *Static Transit*. Retrieved from Google Developers: https://developers.google.com/

6 Government Digital Service -UK. (2015). *Guidance, National public transport access nodes*. Retrieved from Gov.UK: http://www.naptan.org.uk/