Multi-layered Security Technologies

for hyper-connected smart cities

D1.5: Data Management Plan

December 2018

Grant Agreement No. 814917

Multi-layered Security technologies to ensure hyper-connected smart cities with Blockchain, BigData, Cloud and IoT

Project acronym	M-Sec
Deliverable	D1.5 Data Management Plan
Work Package	WP1
Submission date	December 2018
Deliverable lead	WLI/NTTE
Authors	WLI/NTTE
Internal reviewer	TST/KEIO
Dissemination Level	Public
Type of deliverable	R
Version history	-V1.1, 22/November/2018, WLI, Content, Full Draft -V1.2, 29/November/2018, WLI, Contribution, First Draft -V1.3, 17/December/2018,NTTE, Comments Added -V1.4, 18/December/2018, WLI, Comments Checked -V1.5, 19/December/2018, TST, Internal Review -V1.6, 19/December/2018, WLI, Review Checked -V1.7, 20/December/2018, WLI, Changes on the logos order -V1.8, 21/December/2018, KEIO, Internal Review -V1.9,21/December/2018, NTTE, Input provided -V2.0, 21/December/2018, WLI, Final Version

The M-Sec project is jointly funded by the European Union's Horizon 2020 research and innovation programme (contract No 814917) and by the Commissioned Research of National Institute of Information and Communications Technology (NICT), JAPAN (contract No. 19501).

a.Table of Contents

b. List of Tables

c. Glossary

DMP	Data Management Plan
DPO	Data Protection Officer
EU	European Union
FAIR	Findable, Accessible, Interoperable and Reusable
GDPR	General Data Protection Regulation
IoT	Internet of Things
IPR	Intellectual Property Rights
NICT	National Institute of Information and Communications Technology
PIPA	Personal Information Protection Act

d. References

[EC] Guidelines on FAIR Data Management in Horizon 2020, http://ec.europa.eu/

[D22] Deliverable 2.2 "M-Sec pilots definition, setup and citizen involvement plan", M-Sec project, Month 8 (February 2019)

[OAIRE] OpenAire, www.openaire.eu

[RE3] Re3data, www.re3data.org

[DC] Datacite, www.datacite.org

[ODOAR] Opendoar, www.opendoar.org

[ZEN] Zenodo, www.zenodo.org

[D56] Deliverable 5.6 *"Market Analysis and Exploitation – first year",* M-Sec project, Month 12 (June 2019)

[D61] Deliverable 6.1 *"H-Requirement nº.1",* M-Sec project, Month 8 (February 2019)

[D62] Deliverable 6.2 *"H-Requirement nº.2",* M-Sec project, Month 8 (February 2019)

[D63] Deliverable 6.3 *"H-Requirement nº.3",* M-Sec project, Month 8 (February 2019)

[D64] Deliverable 6.4 *"POPD-Requirement nº.4",* M-Sec project, Month 8 (February 2019)

[D65] Deliverable 6.5 *"NEC-Requirement nº.6",* M-Sec project, Month 8 (February 2019)

[D66] Deliverable 6.6 *"NEC-Requirement nº.7",* M-Sec project, Month 8 (February 2019)

1. Introduction

This document is developed as part of the M-Sec (Multi-layered Security technologies to ensure hyper connected smart cities with Blockchain, BigData, Cloud and IoT) project, which has received funding from the European Union's (EU) Horizon 2020 Research and Innovation programme, under the Grant Agreement number 814917 and by the Commissioned Research of National Institute of Information and Communications Technology (NICT) under the Grant Agreement number 19501.

The purpose of the Data Management Plan (DMP) is to provide an overview of the available research data arising from the project, the data accessibility, management and terms of use. The DMP will follow the template that the European Commission suggests in the *"Guidelines on FAIR Data Management in Horizon 2020",* current version is 3.0, dated 26th July 2016 [EC], consisting of a set of questions that the project shall address and properly answer with a level of detail appropriate to the project. 'FAIR' data refers to data that is Findable, Accessible, Interoperable and Re-usable.

According to these guidelines, the DMP will include the following sections:

1. Data Summary
2. FAIR Data
    e. Making data findable, including provisions for metadata
    f. Making data openly accessible
    g. Making data interoperable
    h. Increase data reuse (through clarifying licenses)
3. Allocation of resources
4. Data security
5. Ethical aspects
6. Other issues

This deliverable presents an initial version of the DMP, and it does not intend to answer all these questions, but to present the information on how the actual DMP will be put together and its contents when data from the project will become available. This document will be updated over the course of the project and will be included within deliverable *"Project Progress Report"* updated consequently at the end of each year (M12, M24 and M36).

2. Data Summary

According to DMP guidelines, this section will address the following questions during the project lifetime:

1. What is the purpose of the data collection/generation and its relation to the objectives of the project?
2. What types and formats of data will the project generate/collect?
3. Will you re-use any existing data and how?
4. What is the origin of the data?
5. What is the expected size of the data?
6. To whom might it be useful ('data utility')?

From all these questions, in this initial DMP we are starting to address the first two questions, while the other remaining four questions will be analyzed as soon as the progress of the project provides more concrete information on the datasets.

What is the purpose of the data collection/generation and its relation to the objectives of the project?

Mainly, data generated during the project's life will come from the specific needs of the M-Sec pilots, but also some data will be generated for measurement and assessment purposes of the M-Sec platform. This data generation is directly connected with M-Sec project objectives:

- *Objective 1: To design the future decentralized architecture of IoT that will unlock the capacity of smart objects, by allowing to instantly search, use, interact and pay for available assets and services in the IoT infrastructures.*

  Data and metadata will be generated by risk assessment study for threat and security threats, and mechanisms to establish seamless hyper-connectivity over heterogeneous communication channels.

- *Objective 2: To enable seamless and highly autonomous and secure interaction between humans and devices in the context of smart city, through the use of blockchain and for business contexts relevant to specific smart city use cases enabling innovative machine-human and machine-machine interactions.*

  The content will not be generated by the M-Sec platform; however, the management of security of some of this content will derive into blockchain transactions; some of them may contain associated M-Sec metadata. These metadata will be useful not only for pilots and for the M-Sec platform evaluation but also for third stakeholders with similar pilots and intending to adopt M-Sec solution.

- *Objective 3: To engineer new levels of security and trust in large scale autonomous and trust-less multipurpose smart city platforms and integrate privacy enhancing technologies at the design level of M-Sec architecture.*

  M-Sec platform will implement different mechanisms and security layers in order to facilitate end-to-end data security. Whether these datasets will be made publicly available or not will have to be decided case by case depending on several sharing criteria such as their nature, ownership or exploitability. Preference will always be given to openness, while private datasets shall be the exception, properly justified.

- *Objective 4: To create reference implementations of future decentralized IoT ecosystems and validate their viability and sustainability.*

  M-Sec will create demonstrators and ecosystems in two real IoT environments (Fujisawa, Japan and Santander, Spain) provided by smart cities through real-life use cases (six different use cases, 2 at a Europe level, 2 at a Japanese level and 2 Cross-borders) and from a sensor to business model. In addition, a novel marketplace will be implemented where smart objects will be able to exchange information and/or services through the use of virtual currencies.

  The availability of such datasets for public domain will be entirely dependent on each use case. If such datasets are already open they will continue being open, but those of private nature will not be disclosed unless the corresponding use case owner has the right to take such decision and decides to do so.

- *Objective 5: To maximize the impact of the project benefits.*

  This activity should not generate or manage any specific project dataset. However, in general, all data related to stakeholders involved in community building will be made open as long as it does not include any private data, which will be either anonymized if possible or completely removed prior to disclosure.

What types and formats of data will the project generate/collect?

Data generated by the M-Sec platform will mostly consist on open data sources, smart cities repositories, and blockchain transactions available on the public ledger. In addition to the data generated by M-Sec itself, the execution of the M-Sec pilots will also require accessing and collecting different types of data related to IoT devices, or data generated by mobile applications being managed by the M-Sec platform.

Once each specific dataset is identified, the consortium will decide on the precise format considering that, as explicitly mentioned in the DoA, the main goal is to, as much as possible, use not only open formats to store the data but also make the software open to provide the scripts and other metadata necessary to reuse it.

M-Sec's technical developments and results will be validated and demonstrated through six pilot use cases, as defined in deliverable "D2.2 M-Sec pilots definition, setup and citizen involvement report"[D22]. These pilot use cases will include several data related activities.

3. FAIR Data

3.1 Datasets identification and description

As specified in the guidelines of the European Commission on Data Management, the data to be made available for open access in Europe will have to be described using the following dataset description template (see Table 1). These descriptions will be stored in the project's internal repository and will be provided within the periodic Project Progress Report.

Table 1: Dataset description template

Dataset reference and name	**Identifier for the dataset to be produced**
Dataset description	Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.
Standards and metadata	Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.
Data sharing	Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.). In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).
Archiving and preservation	Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

3.2 Data Management Platforms

Regarding the EU partners, M-Sec will use OpenAIRE [OAIRE] in cooperation with re3data [RE3], to select the proper open access repository and/or deposit publications for its research results storage, allowing also for easy linking with the EU-funded project. This will increase the accessibility to the obtained results by a wider community, which can be further enhanced by including the repository in registries of scientific repositories, such as DataCite [DC] and OpenDOAR [ODOAR], or Zenodo [ZEN]. These are the most popular registries for digital repositories and, along with re3data, they are collaborating to provide open research data. For Japanese partners, as an approval form NICT is necessary for each case, will promptly confirm to obtain an approval.

3.3 FAIR Data Template

'FAIR' data (Findable, Accessible, Interoperable and Re-usable) aim to provide a framework to ensure that research data can be effectively reused. During the project lifetime, and according to DMP guidelines, the following questions shall be addressed:

Making data findable, including provisions for metadata

- Outline the discoverability of data (metadata provision).
- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?
- Outline naming conventions used.
- Outline the approach towards search keyword.
- Outline the approach for clear versioning.
- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how.

Making data openly accessible

- Specify which data will be made openly available. If some data is kept closed provide rationale for doing so.
- Specify how the data will be made available.
- Specify what methods or software tools are needed to access the data. Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited.
- Specify how access will be provided in case there are any restrictions.

Making data interoperable

- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.
- Specify whether you will be using standard vocabulary for all data types present in your dataset, to allow inter-disciplinary interoperability. If not, will you provide mapping to more commonly used ontologies?

Increase data re-use (through clarifying licenses)

- Specify how the data will be licensed to permit the widest reuse possible.
- Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.
- Specify whether the data produced and/or used in the project is usable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.
- Describe data quality assurance processes.
- Specify the length of time for which the data will remain re-usable.

This questions template will be filled as soon as datasets get defined as described in previous section 3.1, and will be provided within the periodic deliverable "Project Progress Report".

3.4 Source Code

M-Sec will make available the generated software and its source code to the Open Source Community. M-Sec consortium has not still identified which kind of Open Source License of source code will be applied. However, it may be possible that a dual license scheme could be considered in order to protect the business exploitation perspectives of the partners. Duality means that both the free software distribution mechanism and traditional software product business are combined. There is technically only one core product but two licenses: one for free distribution and free use, and another one for commercial use (proprietary). The business model will be explained in detail at deliverable *"D5.6 Market Analysis and Exploitation"[D56]*.

4. Other Data Management Aspects

The DMP guidelines also refer to the following aspects related to data management.

4.1 Allocation of Resources

All use case partners and technical partners, with their related role, are involved in data management activities, either collecting, processing, or creating datasets and the corresponding effort is embedded into the tasks in which they are undertaking these activities. Hence, all related costs for data management are already covered by the M-Sec project and no additional resources will be needed.

4.2 Data Security

Any issue regarding the Protection of Personal Data will be included in deliverable *"D6.4 POPD-Requirement No.4"[D64]* and hence is not repeated here. Given that almost all use cases require collection of data from the field of operation, in addition to personal data protection, M-Sec will use state-of-the-art technologies for secure storage, delivery and access of personal information, as well as managing the rights of the users. In this way, there is complete guarantee that the accessed, delivered, stored and transmitted content will be managed by the right persons, with well-defined rights, at the right time.

State-of-the-art firewalls, network security, encryption and authentication will be used to protect collected data. Firewalls prevent the connection to open network ports, and exchange of data will be through consortium known ports, protected via IP filtering and password. Where possible (depending on the facilities of each partner) the data will be stored in a locked server, and all identification data will be stored separately.

A metadata framework will be used to identify the data types, owners and allowable use. This will be combined with a controlled access mechanism and in the case of wireless data transmission with efficient encoding and encryption mechanisms, for example WPA2 (Wireless Protected Access II), a security method for wireless networks that provides stronger data protection and network access control.

4.3 Ethical Aspects

M-Sec partners are to comply with the ethical principles which states that all activities must be carried out in compliance with:

a) Ethical principles
b) Applicable international, EU and national law.

All information related with Ethical Aspects will be handled within WP6 "Ethics requirements" which include the submission of different deliverables on M8. WP6 aims to follow-up the ethical issues applicable to the M-Sec project implementations. It includes:

- The procedures and criteria that will be used to identify/recruit participants (Deliverable 6.1 [D61]),

- the informed consent procedures that will be implemented for the participation of humans (deliverable 6.2 [D62]),
- procedures for processing personal data, compilation, pseudonymisation, protection and deletion (deliverable 6.3 [D63]),
- in case of processing personal data, information about the appointment of a Data Protection Officer (DPO) (deliverable 6.4 [D64]),
- and finally, in case of personal data is transferred from EU to Japan or to another non-EU country , or the opposite from non-EU countries or international organization to an EU country (deliverable 6.5 [D65] and 6.6 [D66]), confirmation that such transfers are in accordance with GDPR, Japanese personal data protection law (PIPA, Personal information Protection Act)and manual concerning the handling of personal data stipulated by NICT, and also the laws of the country in which the data was collected.

GDPR and IPR (Intellectual Property Rights) protection issues also have a dedicated WP (WP5 "*GDPR, dissemination, exploitation and sustainability"*). This WP will provide a guide for compliance of the M-Sec project results with GDPR law and the intellectual property rights of the project results.

5. Conclusions

This deliverable gives an insight of the initial Data Management Plan of M-Sec. It is actually a guideline of the different aspects that need to be covered and tackled as soon as datasets gets identified. The document defines how those datasets have to be properly described. While the project progresses, it will be identified what kind of data or metadata can be publically accessible to other parties, considering both data generated by the own pilots and data generated by the M-Sec platform itself.

This deliverable also provides an insight about in which conditions source code will be made available and the respective platforms that will host data.

In this context, M-Sec will provide an updated and concrete Data Management Plan, including the description of the identified datasets, on the deliverable “Project Progress Report” which will be submitted at the end of each year.