# D10.7
# Data Management Plan. Initial version M6

| | | | |
|---|---|---|---|
| **WORKPACKAGE** | WP10 | **PROGRAMME IDENTIFIER** | H2020-ICT-2020-2 |
| **DOCUMENT** | D10.7 | **GRANT AGREEMENT ID** | 101016681 |
| **REVISION** | V1.0 | **START DATE OF THE PROJECT** | 01/01/2021 |
| **DELIVERY DATE** | 30/06/2021 | **DURATION** | 3 YEARS |

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No 101016681.

# DISCLAIMER



# ACKNOWLEDGEMENT

This document is a deliverable of the 5G-ERA project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 101016681.

The opinions expressed in this document reflect only the author's view and in no way reflect the European Commission's opinions. The European Commission is not responsible for any use that may be made of the information it contains.

| | |
|---|---|
| **PROJECT ACRONYM** | 5G-ERA |
| **PROJECT TITLE** | 5G Enhanced Robotic Autonomy |
| **CALL ID** | H2020-ICT-2020-2 |
| **CALL NAME** | Information and Communication Technologies |
| **TOPIC** | ICT-41-2020 5G PPP – 5G innovations for verticals with third party services |
| **TYPE OF ACTION** | Research and Innovation Action |
| **COORDINATOR** | ROBOTNIK AUTOMATION SSL (ROBOTNIK) |
| **PRINCIPAL CONTRACTORS** | BRINGAUTO S.R.O.; CAL-TEK SRL; COGNITECHNA SRO; EBOS TECHNOLOGIES LIMITED; HAL ROBOTICS LTD; IQUADRAT INFORMATICA SL; NEC LABORATORIES EUROPE GMBH; ORGANISMOS TILEPIKOINONION TIS ELLADOS OTE AE; TWI LIMITED; UNIVERSITY OF BEDFORDSHIRE BED; VYSOKE UCENI TECHNICKE V BRNE; WINGS ICT SOLUTIONS INFORMATION & COMMUNICATION TECHNOLOGIES IKE WINGS |
| **WORKPACKAGE** | WP10 |
| **DELIVERABLE TYPE** | REPORT |
| **DISSEMINATION LEVEL** | PUBLIC |
| **DELIVERABLE STATE** | Finished |
| **CONTRACTUAL DATE OF DELIVERY** | 30/06/2021 |
| **ACTUAL DATE OF DELIVERY** | 29/06/2021 |
| **DOCUMENT TITLE** | Data Management Plan |
| **AUTHOR(S)** | Rafael Lopez [ROB] |
| **REVIEWER(S)** | Renxi Qiu [BED] |
| **ABSTRACT** | SEE EXECUTIVE SUMMARY |
| **HISTORY** | SEE DOCUMENT HISTORY |
| **KEYWORDS** | Data Management, Open Research Data & Datasets |

# Document History

| Version | Date | Contributor(s) | Description |
|---|---|---|---|
| V0.1 | 01/06/2021 | Rafael Lopez | Content & Structure |
| V1.0 | 28/06/2021 | Renxi Qiu | Review |
| | | | |
| | | | |
| | | | |

# Table of Contents

# List of Tables

# List of Acronyms and Abbreviations

| | |
|---|---|
| DMP | Data Management Plan |
| EC | European Commission |
| ORD | Open Research Data |
| PPDR | Public Protection and Disaster Relief |
| QoE | Quality of Experience |
| | |
| | |
| | |
| | |
| | |

# Executive Summary

This deliverable is the first version of the 5G-ERA plan for handling data and knowledge issues. More specifically, it consists of two main parts. The former is the initial version (Month 6) of the Data Management Plan (DMP), while the latter will be the final version at month 24.

The 5G-ERA DMP has been drafted in accordance to the regulations of the Pilot action on Open Access to Research Data of the Horizon 2020 programme (H2020). The objective of the DMP is to support the data management life cycle for all data that will be collected processed or generated by the project. It contains preliminary information about the data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved.

5G-ERA declared its intention to participate in Open Research Data (ORD) Pilot, therefore this DMP has been prepared taking this fact into account. Actions taken in order to participate in ORD Pilot will be reported in subsequent versions of this deliverable.

To develop the present deliverable, an appropriate template was drafted based on the H2020 guidelines for the development of projects' DMP. This was circulated to all project partners so as to collect all relevant information concerning the datasets that are planned to be developed in the course of the project. On the basis of all partners' feedback, the preliminary data management plan of the project has been established, as described in the present deliverable.

Clearly, as the present deliverable has been drafted during the first project stages (M6), it can only reflect the intentions of the project partners toward developing the overall project's datasets. Thus, an updated version of the current deliverable is planned for M24 of the project.

# 1. Introduction

This document has main objective of drafting the 5G-ERA Data Management Plan (DMP), which aims to describe the data management life cycle for the data to be collected, processed or generated by the project.

The European Commission has recognised that, apart from scientific publications, research data of a project is also important. Thus, there is a need for a structured approach to the use of research data within a research project.

The 5G-ERA DMP will outline how research data will be handled throughout the project and after its completion. It will describe what data will be collected and processed and how this data will be shared and/or made open, and how it will be preserved.

The first version of the DMP is expected to be delivered within the first 6 months of the project. This deliverable should follow the template provided by the EU [1]. At the time of writing this deliverable, the information regarding the 5G-ERA datasets is incomplete, since most of the data are not available and will be collected/created at a later stage. More elaborated versions of the DMP shall be created at later stages of the project, or whenever important changes to the project occur due to inclusion of new data sets, changes in consortium policies or external factors.

This deliverable is expected to be a living document that will evolve throughout the project's lifespan.

# 2. Data Management Plan - General Principles

## 2.1. Purpose of data collection

The 5G-ERA is aimed towards a user-centric paradigm of integrating vertical knowledge into the existing standardised 5G testing framework to improve Quality of Experience (QoE) and focuses on building new NetApps for autonomous robots in PPDR, transport, logistics, and manufacturing processes. This NetApps will be evaluated and validated under the vertical domains using three existing testbeds from Greece, Spain, and U.K

A set of appropriate datasets needs to be identified early in the project, to evaluate the performance of the proposed technologies and to demonstrate the different Use Cases and scenarios. All these datasets have been defined after extensive discussions between the end-users and the technical partners of 5G-ERA. However, introduction of new dataset is still possible through the lifetime of the project, and an updated version of the current deliverable is planned for M24 of the project.

## 2.2. Participation in the Pilot on Open Research Data

5G-ERA participates in the Pilot on Open Research Data (ORD) launched by the European Commission along with the Horizon2020 programme. The consortium believes firmly in the concepts of open science, and the large potential benefits the European innovation and economy can draw from allowing reusing data at a larger scale. Therefore, an effort will be made to publish with open access a significant amount of data produced by the project. If certain datasets cannot be shared, a justification for opting out will be provided.

## 2.3. FAIR data

Within 5G-ERA, consortium members will respect the principles of FAIR data management, that is Findable, Accessible, Interoperable and Re-usable. To achieve this, a DMP should include information on:

- the handling of research data during and after the end of the project
- what data will be collected, processed and/or generated
- which methodology and standards will be applied

- whether data will be shared/made open access and
- how data will be curated and preserved (including after the end of the project).

## 2.4. Interoperability

Data interoperability is an important aspect within 5G-ERA to foster collaboration and efficiency between the partners. Whenever possible, existing, well-defined data-exchange standards will be used. To ensure interoperability in data exchange within 5G-ERA, the following requirements will be met:

- Use of vocabularies that follow FAIR principles.
- Include qualified references to other (meta)data.

## 2.5. Security

The datasets foreseen to be collected through 5G-ERA are of high value special care should be taken to prevent such datasets to leak or become hacked. This is another key aspect of 5G-ERA data management, and all data repositories used by the project will include effective protection.

A holistic security approach will be followed, in order to protect the pillars of information security (confidentiality, integrity, availability). The security approach will consist of a methodical assessment of security risks followed by their impact analysis. This analysis will be performed on the personal information and data processed by the proposed system, their flows and any risk associated to their processing.

# 3. Data summary

In this paragraph, we provide detailed information about the datasets that are planned to be captured by the partners of the 5G-ERA project. These are the foreseen datasets as of month 6 of the project. More datasets will be included in the course of the project and the DMP will be updated at month 24.

## 3.1. Datasets naming convention

Concerning the convention followed for naming the 5g-ERA datasets, it should be noted that the name of each dataset comprises: (a) a prefix 'DS' indicating a dataset, along with its unique identification number, e.g. 'DS1', (b) the name(s) of the partner(s) responsible to collect it, e.g. ROB, along with an identifier denoting the internal numbering of the dataset concerning the specific partner, e.g. -01 and (c) a short title of the dataset summarizing its content and purpose, e.g. Control system Dataset.

## 3.2. Summary of foreseen 5G-ERA datasets

The following, Table 1 provides a list of the expected datasets, whereas the detailed description of each dataset, in accordance to the H2020 DMP template is provided in the following sections. At this stage (M6) there are 3 datasets foreseen in the project, covering a series of research dimensions on the skills the 5G-ERA project should develop. In the course of the project more Datasets will be added in the Data Management Plan if necessary.

| No | Dataset name |
|---|---|
| 1 | DS1_BED-01_3D_PointCloud |
| 2 | DS2_NEC-01_Radio Access Network |
| 3 | DS3_EBOS-01_Front-End-Dashboard |

Table 1 Summary of foreseen 5G-ERA datasets (as of month 6)

## 3.3. Datasets description

In order to meet the requirements of the DMP according to the Pilot Open Access of the Horizon 2020, each partner provided the description of their datasets using the template given in [1], which was formed by following the guidelines of the dataset aspects that should be reported in DMPs of the H2020 projects. Based on this information, partner ROB compiled the following tables:

| **DS1_BED-01_3D_PointCloud** | |
|---|---|
| GENERAL DESCRIPTION | |
| Labelled 3-D point cloud data collected by a moving robot platform in a simulated environment or collected from controlled testing environment. | |
| DATASET DESCRIPTION | |
| Data origin | Generated by Gazebo simulation or collected from robot sensors (Lidar & Laser scanner). |
| Nature and scale of data | Point cloud saved in RAW or saved inside ROSbag |
| To whom could the dataset be useful | Robot developers and 5G testbed providers |
| Related scientific publication(s) | N/A |
| Indicative existing similar data sets | LIDAR Point Cloud (https://data.gov.uk/dataset/977a4ca4-1759-4f26-baa7-b566bd7ca7bf/lidar-point-cloud)<br><br>Oakland 3-D Point Cloud Dataset (http://www.cs.cmu.edu/~vmr/datasets/oakland_3d/cvpr09/doc/) |
| METADATA AND ANNOTATION | |
| Labels for the simulated environment<br><br>Labels on testing environment | |

| | |
|---|---|
| OPEN ACCES TO DATA | |
| Access type | Consortium members |
| How will the data be made accessible | Private git repository |
| What methods or software tools are needed to access the data | GitHub (5G-ERA consortium GitHub repository) |
| Where will the data and associated metadata, documentation and code be deposited | Together with the raw data |
| DATA RE-USE | |
| Data preservation period | User authentication mechanism provided by GitHub |
| ALLOCATION OF RESOURCES | |
| Indicative associated costs for data archiving and preservation | GitHub monthly subscription plan. |

| | |
|---|---|
| Indicative plan for covering the above costs | Github Team (4$ per month) |
| PARTNERS ACTIVITIES AND RESPONSIBILITIES | |
| Partner Owner / Data Collector | 5G-ERA joint ownership |
| Partner in charge of the data analysis | ALL consortium members |
| Partner in charge of the data storage | BED |
| WPs and Tasks | WP2, Task 2.1, 2.2, 2.3 and 2.4 |

Table 2 DS1_BED-01_3D_PointCloud Experimental Data

| **DS2_NEC-01_Radio Access Network** | |
|---|---|
| GENERAL DESCRIPTION | |
| Radio Access Network metrics derived by an open source LTE/5G software emulator srsRAN (source-code available at https://www.srsran.com/). srsRAN is an Open-source 4G and 5G software radio suite developed by Software Radio Systems (SRS). The srsRAN suite includes:<br>• srsUE - a full-stack SDR 4G/5G-NSA UE application (5G-SA coming soon)<br>• srsENB - a full-stack SDR 4G eNodeB application (5G-NSA and 5G-SA coming soon)<br>• srsEPC - a light-weight 4G core network implementation with MME, HSS and S/P-GW<br>The above entities can run on dedicated hardware or as container instances. USRPs may be used to provide a real wireless channel between the UEs and the base station.<br>The data plane traffic can be generated by ROS instances and the corresponding topics exchange during runtime operations, or by means of existing rosbags, or exploiting generic traffic generators, e.g. iperf (https://iperf.fr/). | |
| DATASET DESCRIPTION | |
| Data origin | Experiments |
| Nature and scale of data | CSV format, MBs/GBs depending on simulation duration and complexity of the deployment (number of UEs, number of Base stations, etc) |
| To whom could the dataset be useful | 5GERA Project partners and 5G-related research. |
| Related scientific publication(s) | • L. Zanzi, V. Sciancalepore, A. Garcia-Saavedra, H. D. Schotten, and X. Costa-Pérez, "LACO: A Latency-Driven Network Slicing Orchestration in Beyond-5G Networks, "IEEE Transactions on Wireless Communications, vol. 20, no. 1, pp. 667–682, 2021. |

| | |
|---|---|
| | • Jose A. Ayala-Romero, Andres Garcia-Saavedra, Marco Gramaglia, Xavier Costa-Perez, Albert Banchs, and Juan J. Alcaraz, 2019, "VrAIn: A Deep Learning Approach Tailoring Computing and Radio Resources in Virtualized RANs", In The 25th Annual International Conference on Mobile Computing and Networking (MobiCom '19), Article 30, 1–16. DOI:https://doi.org/10.1145/3300061.3345431 |
| Indicative existing similar data sets | https://github.com/agsaaved/vrain |
| METADATA AND ANNOTATION | |
| In addition to standard RAN monitoring described in [https://docs.srsran.com/en/latest/usermanuals/source/srsenb/source/6_enb_commandref.html], base station transmission buffer sizes can also be collected through dedicated software changes. | |
| OPEN ACCES TO DATA | |
| Access type | Possibility to share directly within the consortium still under discussion and to be confirmed. |
| How will the data be made accessible | Online repositories (NEC hosted Git, NEC hosted Nextcloud, etc) |
| What methods or software tools are needed to access the data | Security credentials and web browser. Open access may also be possible after some internal procedures |
| Where will the data and associated metadata, | Within the same online repository |

| | |
|---|---|
| documentation and code be deposited | |
| DATA RE-USE | |
| Data preservation period | Till the end of the project, but extensions may be possible |
| ALLOCATION OF RESOURCES | |
| Indicative associated costs for data archiving and preservation | TBD |
| Indicative plan for covering the above costs | TBD |
| PARTNERS ACTIVITIES AND RESPONSIBILITIES | |
| Partner Owner / Data Collector | NEC |
| Partner in charge of the data analysis | NEC |
| Partner in charge of the data storage | NEC |
| WPs and Tasks | WP2, WP4, WP5 |

Table 3 DS2_NEC-01_Radio Access Network

| **DS3_EBOS-01_Front-End-Dashboard** | |
|---|---|
| GENERAL DESCRIPTION | |
| Data for the purposes of the front-end dashboard | |
| DATASET DESCRIPTION | |
| Data origin | Use case owners |
| Nature and scale of data | not personal data, large scale |
| To whom could the dataset be useful | Consortium |
| Related scientific publication(s) | not currently |
| Indicative existing similar data sets | N/A |
| METADATA AND ANNOTATION | |
| No metadata or annotation are needed to be stored for this dataset | |
| OPEN ACCES TO DATA | |
| Access type | Open to consortium members |
| How will the data be made accessible | the data will be accessible only to assigned authorized users |
| What methods or software tools are needed to access the data | N/A |

| | |
|---|---|
| Where will the data and associated metadata, documentation and code be deposited | Internal data repository of 5G-ERA |
| DATA RE-USE | |
| Data preservation period | For the lifetime of the project and up to 7 years after the project completion |
| ALLOCATION OF RESOURCES | |
| Indicative associated costs for data archiving and preservation | No costs are currently foreseen regarding its preservation. |
| Indicative plan for covering the above costs | N/A |
| PARTNERS ACTIVITIES AND RESPONSIBILITIES | |
| Partner Owner / Data Collector | Use case owners/ EBOS |
| Partner in charge of the data analysis | Use case owners |
| Partner in charge of the data storage | EBOS |
| WPs and Tasks | WP2, WP6, WP7, WP8, T2.5 |

Table 4 DS3_EBOS-01_Front-End-Dashboard

# 4. Conclusions

The current document provides preliminary, but detailed information about the datasets that are planned to be captured by the partners of the 5G-ERA project. These are the foreseen datasets as of month 6 of the project. A more complete list of datasets will be included in the future, as the project progresses. The DMP will be updated at month 24.

# 5. References

1. https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm