"5G for Drone-based Vertical Applications"

D6.1 – Data Management Plan and quality and risk management plan

Document ID:	**DMP**
Deliverable Title:	**Data Management Plan**
Responsible Beneficiary:	**UO**

Topic:	**H2020-ICT-2018-2020/H2020-ICT-2018-3**
Project Title:	**Unmanned Aerial Vehicle Vertical Applications' Trials Leveraging Advanced 5G Facilities**
Project Number:	**857031**
Project Acronym:	**5G!Drones**
Project Start Date:	**June 1st, 2019**
Project Duration:	**42 Months**
Contractual Delivery Date:	**When needed**
Actual Delivery Date:	**September 30th, 2019**
	DMP update November 16th, 2022
Dissemination Level:	**Open Research Data Pilot (ORDP)**
Contributing Beneficiaries:	**UO, THA, ALE, INV, HEP, AU, COS, AIR, UMS, NOK, CAF, EUR**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857031.

Document ID: DMP
Version: V8
Version Date: 16.11.2022
Authors: Jussi Haapola (UO)

Security: Open Research Data Pilot (ORDP)

Approvals

	Name	Organization	Date
Coordinator	Jussi Haapola	UO	16.11.2022
Technical Committee	Pascal Bisson	THA	30.9.2019
Management Committee	Project Management Team	UO, THA, AU, AIR, UMS, FRQ	30.9.2019

Document History

Version	Contribution	Authors	Date
V1	**Initial version of quality and risk management plan**	**Jussi Haapola (UO)**	**23.08.2019**
V2	**Initial version of Data Management Plan**	**UO, AIR, INV**	**28.08.2019**
V3	**Additions**	**Ilkka Kansala (NOK)**	**08.09.2019**
V4	**Review**	**THA, AIR, COS**	**11.09.2019**
V5	**Revision after PMT review**	**Jussi Haapola (UO)**	**16.09.2019**
V6	**Final version**	**Jussi Haapola (UO)**	**30.09.2019**
V7	**Data Management Plan update**	**Jussi Haapola (UO)**	**04.06.2021**
V8	**Data Management Plan update & isolation from D6.1**	**Jussi Haapola (UO), Mohamed Mekki (EUR)**	**16.11.2022**

Executive Summary

This report will provide a Data Management Plan for the 5G!Drones project. The Data Management Plan is a living document that will be updated during the 5G!Drones project lifecycle. The Version V8 of the document is a Data Management Plan update.

Table of Contents

List of Tables

List of Abbreviations

3GPP	3rd Generation Partnership Project
5G	5th Generation Cellular Technology
5G-PPP	5G Infrastructure Public Private Partnership
ADS-B	Automatic Dependent Surveillance – Broadcast
AI	Artificial Intelligence
AMF	Access Management Function
API	Application Interface
AR	Augmented Reality
BVLoS	Beyond Visual Line of Sight
CAA	Civil Aviation Authority
CC	Creative Commons
CoTS	Commercial Off-The-Shelf
DMP	Data Management Plan
EAB	External Advisory Board
eMBB	Enhanced Mobile Broadband
EPC	Evolved Packet Core
ETSI	European Telecommunications Standards Institute
FAIR	Findable, Accessible, Interoperable, Re-usable
FCT	Facility Coordination Team
FR	Financial Report
GA	General Assembly
GDPR	General Data Protection Regulation
GNSS	Global Navigation Satellite System
GUI	Graphical User Interface
ICAO	International Civil Aviation Organization
IMT	Innovation Management Team
IoT	Internet of Things
IPR	Intellectual Property Rights
IR	Internal Report

JSON	JavaScript Object Notation
KPI	Key Performance Indicator
KPIC	KPI Component
LiDAR	Light Detection and Ranging
LADN	Local Area Data Network
LTE	Long-Term Evolution
MANO	Management and Orchestration
MEC	Multi-access Edge Computing
ML	Machine Learning
mMTC	Massive Machine-Type Communications.
MoM	Minutes of Meeting
MS	Microsoft
PC	Project Coordinator
PCI	Physical Cell Id
PIA	Privacy Impact Assessment
PID	Persistent Identifier
PMT	Project Management Team
PSI	Public Service Identity
QMR	Quarterly Management Report
RAN	Radio Access Network
RRC	Radio Resource Control
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RSSI	Received Signal Strength Indicator
SNR	Signal to Noise Ratio
SORA	Specific Operations Risk Assessment
SSH	Secure Shell
TM	Technical Manager
ToC	Table of Contents
UAS	Unmanned Aerial Systems

UAV	Unmanned Aerial Vehicle
UCxScy	Use Case x, Scenario y
UE	User Equipment
uRLLC	Ultra-Reliable Low Latency Communications
UTM	UAS Traffic Management
VNF	Virtualised Network Function
VPN	Virtual Private Network
VR	Virtual Reality
WP	Work Package
WPL	Work Package Leader

1. DATA MANAGEMENT PLAN

1.1. Data summary

The 5G!Drones project will collect and generate data for mainly two purposes: 1) for trialling unmanned aerial systems (UAS) key performance indicators (KPIs) using 5G technology as a Use Case enabler; and 2) For trialling 5G technology components KPIs using UAS verticals as service users. For the former there exist three main use cases with a number of scenarios ranging from 3D mapping and data visualisation to operations in non-GPS locations to disaster recovery and unmanned aerial vehicle (UAV)-enhanced internet of things (IoT) data collection. A common Use Case then is UAS traffic management (UTM) command and control applications. For the latter effective UAV operations can significantly benefit from 5G infrastructure and the trials relate to how well 5G components, such as enhanced mobile broadband (eMBB), ultra-reliable low latency communications (uRLLC), massive machine type communications (mMTC), network slicing, multi-access edge computing (MEC), and management and orchestration (MANO) can service UAVs.

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used in the 5G!Drones project with regard to the project research and trials data. The DMP covers the complete data life cycle. It describes the types of research and trials data that will be generated or collected during the project, the standards that will be used, how the research data will be preserved and what parts of the datasets will be shared for verification and/or reuse.

The DMP is a living document, which will evolve during the project, particularly whenever significant changes arise such as data set updates, or changes in Consortium policies. The first version of the DMP was delivered in Month 4 of the project as a part of Deliverable D6.1. This version is an update to the DMP, generated at M24 of the project. It includes an updated overview of the datasets to be produced by the project, and the specific conditions that are attached to them. This updated version is made before the 1st set of trials and although it already covers a broad range of aspects related to the project data management, the future versions will provide more detail on particular issues such as data interoperability and practical data management procedures implemented by the 5G!Drones project.

The 5G!Drones project high-level objective 1 is: *Analysis of the performance requirements of UAV verticals' applications and business models in 5G.* The objective provides a deep analysis of the UAV use case requirements in terms of the needed network functionalities and the required application performance to validate. Business models will be also derived. The objective provides definition of the required 5G network functionalities to run the UAV use cases as well as the definition of the KPIs of the UAV applications involved in 5G!Drones use cases. In addition, the objective provides a techno-economic study of UAVs utilising cellular technology as a service enabler. This objective's relation to generated data relates to what actual data will be generated as analysis and evidence of the project trials. Furthermore, processing of the generated data and studies lead to scientific publications, white papers, standardisation input, etc.

The objective 2 is: *Design and implementation of the 5G!Drones software layer (or system) to execute UAV trials*. Designing and implementation of the 5G!Drones trial system, which will be in charge of running the UAV trials using the ICT-17 facility components and 5G!Drones enablers developed during the project. The envisioned 5G!Drones system abstracts the low-level details on the usage of 5G facilities resources, by providing a high-level application interface (API) to describe, run and obtain results on the specific KPIs. This is the trial enabler software that will not generate data itself but will be used for executing trials. The architecture and API descriptions will be public, whereas the software itself will be released by the Consortium Partners under varying openness and licensing terms. Some

software is sufficiently close to several of the Consortium Partners core business areas so that confidentiality is required.

The objective 3: *Design a high-level scenario descriptor language to run and analyse the results of the UAV trials* relates to designing a high-level (or Northbound) API to allow a UAV vertical to configure a trial and run the test. The research topics in this objective include devising a northbound API to describe the trial and to configure the UAV use case (e.g., type of 5G service it requires), using a high-level language, such as a JavaScript Object Notation (JSON)-based API. In addition, devising a graphical user interface (GUI) to display results and KPIs and providing data analytics tools to analyse the results are among the topics. Devising a management interface to allow a vertical to have access to the Network Slice(s) running its services, for high-level configuration and management is also required. Again, the specifications of the high-level API are public information, whereas the software itself will be released by the Consortium Partners under varying openness and licensing terms. The software is sufficiently close to several of the Consortium Partners core business area so that confidentiality is required in a number of cases.

The objective 5: *Validate 5G KPIs that demonstrate execution of UAV use cases*. According to the envisioned UAV use cases and scenarios, several 5G KPIs need to be demonstrated and tested to validate UAV application requirements. The research and development topics here are performance evaluation of the different 5G!Drones use cases focusing on 5G KPIs. This objective will generate the bulk of research data in the project in addition to the objective 6: *Validate* UAV KPIs using 5G. These objectives translate to a number of Use Cases in the project. The data will be divided into public and private data. Those data that are private relate to partners' own product line development and therefore cannot be disclosed. Public data, where they are perceived to bring additional value to the research and application developer communities will be made available using the FAIR (findable, accessible, interoperable, re-usable), for example via EUDAT [1] or IDA [2] research data storage at fairdata.fi [3].

The objective 8: *Dissemination, standardisation and exploitation of 5G!Drones* focuses on distribution of project outcomes. The generated and processed data will be openly distributed through scientific open access fora, standardization inputs, white papers, etc. Partners will exploit the generated knowledge in their own business lines.

1.2. Data Sets

The project Partners will generate data from multiple angles that may constitute as data sets. Some data sets are mainly for project partner, facility, or technology purposes with limited or no re-use value while others may have potentially the interest of a large audience. The data sets described as follows, mainly fall under categories: facility data, 5G data, UAS related data, Trial Controller related data, Use Case data, and Artificial Intelligence (AI) / Machine Learning (ML) data. The categories have significant overlap, but this serves as a rough categorisation on the data types of interest.

1.2.1. Facility data

1.2.1.1. EURECOM Microservices performance data

The dataset was generated from an experimental study to detect the misconfiguration of tenant resources. The datasets are collected for 3 types of applications: Web servers written in python and Golang, the RabbitMQ data broker, and OpenAirInterface's 5G core network Access and Mobility Management Function (AMF).

The dataset was collected using in-house developed tools combined with open source tools such as ApacheBench for the web servers, my5G-RANTester for the 5G core network, and RabbitMQ PerfTest for RabbitMQ benchmarking.

The dataset shows application performance in relation to load and allocated resources and resource utilization, and will be used to train auto-scaling models, understand application behavior, and provide appropriate configurations for edge applications.

1.2.1.1.1. Microservices performance

Table 1: Microservices performance in Cloud-native environments Data set

Function	Purpose	Objectives
Purpose of the data	Data used to validate the resources allocated to applications, in the edge facilities	
Type and format of Data	CSV files containing application's resources consumption and latency metrics	.csv
Reused-data (rd)	No	Nil
Data set is	Available	Nil
Quantity	~ 15 MB	Nil
Data Security & Storage	Nil	Nil
Data value (long term)	Nil	Nil
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Metadata using Zenodo: title, description, category, keywords, license	Nil
Identifiability of data	Data has a DOI: 10.5281/zenodo.6907619	Nil
Clear versioning approach	By timestamping.	Nil
Standards or procedures for metadata creation applied	No standards used	Nil
Making data openly accessible		
Data openly available	Data is openly available	Nil
Data kept closed	Nil	Nil

How data will be made available	Open research archive: Zenodo.	Nil
Repository for deposit of data, metadata, documentation and code	Publicly available data will use Zenodo repository with associated metadata.	Nil
Access restrictions	No access restrictions	Nil
Data interoperability assessment	Nil	Nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	Nil	Nil
Data licensing for wide reuse	Nil	Nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Data is made public.	Nil
Data usability by Third Parties (after the end of the project)	Publicly made data will be available for third parties.	Nil
Restrictions to data re-use	Nil	Nil
Quality assurance process	Nil	Nil
Length of time of data re-usability	Nil	Nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies.	Nil

1.2.1.1.2. Web servers performance

Data Set of web servers performance:
The dataset containing the web servers performance data is composed by:

Table 2: Web servers performance

Field	Description	Data type
time	Timestamp of collection of metrics	Long Integer

ram_limit	The memory allocated to the container in megabytes	String
cpu_limit	The CPU allocated to the container	Float
ram_usage	The amount of memory used by the container at the time of the metrics collection in byte	Integer
cpu_usage	The amount of CPU used by the container at the time of the metrics collection	Float
n	The number of requests sent to the container.	Integer
c	The concurrency level in the requests.	Integer
lat50	The least response time for the best 50% requests in microseconds	Integer
lat66	The least response time for the best 66% requests in microseconds	Integer
lat75	The least response time for the best 75% requests in microseconds	Integer
lat80	The least response time for the best 80% requests in microseconds	Integer
lat90	The least response time for the best 90% requests in microseconds	Integer
lat95	The least response time for the best 95% requests in microseconds	Integer
lat98	The least response time for the best 98% requests in microseconds	Integer
lat99	The least response time for the best 99% requests in microseconds	Integer
lat100	The least response time in microseconds	Integer

1.2.1.1.3. Data Set of AMF's performance

The dataset containing the AMF's performance data is composed by:

Table 3: AMF's measurements

Field	Description	Data type
time	Timestamp of collection of metrics	Long Integer
ram_limit	The memory allocated to the container in megabytes	String
cpu_limit	The CPU allocated to the container	Float
ram_usage	The amount of memory used by the container at the time of the metrics collection in byte	Integer
cpu_usage	The amount of CPU used by the container at the time of the metrics collection	Float

n	The number of parallel registration requests sent to the AMF	Integer
mean	The mean registration time for all the registration requests in microseconds	Integer
lat50	The median registration time for registration requests in microseconds	Integer
lat75	The least registration time for the best 75% registration requests in microseconds.	Integer
lat80	The least registration time for the best 80% registration requests in microseconds	Integer
lat90	The least registration time for the best 90% registration requests in microseconds	Integer
lat95	The least registration time for the best 95% registration requests in microseconds	Integer
lat98	The least registration time for the best 98% registration requests in microseconds	Integer
lat99	The least registration time for the best 99% registration requests in microseconds	Integer
lat100	The least registration time in microseconds	Integer

1.2.1.1.4. Data Set of RabbitMQ's performance

The dataset containing the RabbitMQ's performance data is composed by:

Table 4: RabbitMQ's measurements

Field	Description	Data type
time	Timestamp of collection of metrics	Long Integer
ram_limit	The memory allocated to the container in megabytes	String
cpu_limit	The CPU allocated to the container	Float
ram_usage	The amount of memory used by the container at the time of the metrics collection in byte	Integer
cpu_usage	The amount of CPU used by the container at the time of the metrics collection	Float
n	The number of producers sending messages to the RabbitMQ server	Integer
min	The minimum consumption time for the producer messages	Integer

lat50	The median consumption time for the producer messages	Integer
lat75	The least consumption time for the best 75% messages in microseconds	Integer
lat95	The least consumption time for the best 95% messages in microseconds	Integer
lat99	The least consumption time for the best 99% messages in microseconds	Integer

1.2.1.2. 5GTN Qosium data

1.2.1.2.1. Qosium Tool

Qosium a real-time passive software-based solution for monitoring and measurement Quality of Service (QoS), traffic load, and Quality of Experience (QoE) without generating additional traffic [4], [5]. Qosium is a distributed measurement system composed of three components:

1. Qosium Scope is the main controlling entity software responsible to implement the QoS Measurement Control Protocol (QMCP) and control measurements in a manual manner [5]. It is a full-scale analyzer software, which commands and controls Qosium Probes and their measurements, and depicts received results. Qosium Scope is key enabler for manual QoS, traffic load, and QoE measurements.
2. Qosium Probe is a lightweight measurement agent, which is installed to network devices from which traffic, QoS, and QoE are desired to be measured [5]. Network devices can be installed in Linux, Android or Windows based systems. The installed Probes can wait for the potential measurement points since it only listens for QMCP connection trials. Therefore, the Probes could be running all the time in the nodes where measurements can run eventually or constantly. The Precision Time Protocol (PTP) is a key enabler for better measurement accuracy and clock synchronization among the devices in the network. Qosium Probe is a multi-thread system that can be involved in several measurement instances [5].
3. Qosium Storage is a dedicated database system that store average results of measurements collected by the Qosium Probes [5]. Qosium Storage is especially useful when there are a lot of data results, and accessing the data is wanted to be centralized. Storage comes up with visualization and access to measurements via a web browser and a rest interface.

1.2.1.2.2. Measurements Results

The basic results of Qosium provides QoS metrics, which tell how well the network is performing from the perspective of the monitored application(s)/service(s) [5]. Traffic statistics, flow statistics, radio level statistics and events, and QoE measurements are KPIs that can be obtained by Qosium. The results are monitored as averaged values over the user-defined Averaging interval [5]. The Qosium results come with a single timestamp used to organize and read the average statistics results. The results can be associated to location to visualize in several types such as to plots, meters, numerical values, and heatmaps drawing for mostly Qosium metrics.

Qosium generates statistics that can be evaluated in real-time, e.g., with Qosium Scope. The results can also be printed into files for later analysis. The results files generated by Qosium Scope are in the TXT format during the measurement is ongoing. Furthermore, all average statistics results generated during the use case scenarios are saved in the Qosium Storage database.

Specifically, Qosium can generate four types of results:

- Average results - Average results are averaged over Averaging Interval, which is a parameter that can be defined in user interface of Qosium Scope [5]. The average results comprise the most important statistic set with over 60 specific statistics to be evaluated.
- Packet results - Packet results provide accurate QoS statistics for every single measured data packet in a two-point measurement (e.g., among two machines).
- Flow results - Flow results provide a simple view of who is communicating with who and how much traffic there is flowing [5]. It is able the flows detection of high bandwidth usage.
- Pcap results - These results provide optionally full packet captures of desired traffic in the Pcap format, allowing to analyze details with protocol analyzers e.g., Wireshark.

Currently, Qosium Storage saves mainly the most important metrics located at Average Results data set in the database. However, Qosium Scope can generate a TXT file for each type of result, as the User selected them. These files are saved in the machine file system where Qosium Scope is instantiated. Most of the Qosium's statistics are available for received (downlink) and sent (uplink) directions [5]. Hence, the average results are collected in both directions, meaning that sent and received traffic can be evaluated separately. The results can be averaged accurately per packet. Among the most important of average results statistics, there are:

1.2.1.2.3. Average Results of Qosium Dataset

General Statistics: This includes general measurement-level statistics that are informative rather than directly related to the networking performance.

Table 5: Qosium General Statistics

General/assisting statistics		
Field	Description	Unit
Control Packet Number	The number of the statistic control packet (an increasing counter)	Count
Measurement Duration	The duration of the measurement.	Seconds
True Averaging Period	The true (accurate) averaging period length.	Seconds
Timestamp	The timestamp at the time when the actual results were calculated. This value is taken at the Primary Probe.	Unix epoch

Single-Point General Statistics: This includes single-point general statistics that are Qosium Probe-specific.

Table 6: Qosium Single-Point General statistics

Single-point general statistics (specific for Primary or Secondary Probe)		
Field	Description	Unit

Malformed packets	The number of packets whose internal structure is not correct.	Packets
Malformed packets, total	The number of packets whose internal structure is not correct (total, measurement)	Packets
Drops, NIC [1]	The number of interface's dropped packets during the measurement. Many Pcap [2] versions (e.g., WinPcap) do not support this yet. This is calculated on the Primary Probe side only.	Packets
Drops, NIC, total	The same as above, but the total value over the measurement.	Packets
Drops, Pcap	The number of Pcap's dropped packets. This is calculated on the Primary Probe side only.	Packets
Drops, Pcap, total	The same as above, but the total value over the measurement.	Packets
Location [3]: Accuracy	Accuracy of the current location.	Meters
Location: Altitude	Altitude of the current location.	Meters
Location: Duration from last update	The age of the last location update. This parameter is handy in checking that the positioning is working correctly.	Seconds
Location: Heading	Heading at the current location if the node is moving.	Degrees
Location: Latitude	Latitude coordinate of the current location.	Degrees [-90°, 90°]
Location: Longitude	Longitude coordinate of the current location.	Degrees [-180°, 180°]
Location: Mode	Operation mode of the positioning (Qosium specific)	
Location: Speed	Speed of at the current location if the node is moving	Meters / s
Short packets	The number of packets that are too short for analysis	Packets
Short packets, total	The number of packets that are too short for analysis (total, measurement).	Packets

Physical Layer Statistics: This includes single-point Probe-specific statistics related to the physical layer properties, typically radio interface, including signal performance, cell information, etc. The availability of these statistics is dependent on the platform where Qosium Probe is running.

[1] Network Interface Controller (NIC) is a computer hardware component that connects a computer to the network.
[2] Pcap is an application programming interface (API) for capturing network traffic. While the name is an abbreviation of packet capture. Windows-based systems have a port called WinPcap while for Linux-based systems is Libpcap.
[3] The location measurements are dependent on measurement point location availability.

Table 7: Qosium Physical Layer Statistics

General single point statistics (specific for Primary or Secondary Probe)		
Field	Description	Unit
Base Station MAC address	The current Base Station MAC address	String
Cell ID	The cell ID of the current wireless network (e.g., in LTE).	Number
Network type	The type of the network (e.g., LTE, 5G).	Coded value
SINR	The signal to noise and interference ratio at the interface	dB
Signal [4] (RSRP)	The signal strength (Reference Signal Received Power) at the network interface.	dBm
Signal (RSRQ)	The signal quality (Reference Signal Received Quality) at the network interface.	dBm
Signal (RSSI)	The signal strength (Received Signal Strength Indication) at the network interface.	dB

Traffic Statistics: Traffic information that the single-point measurement can provide includes throughput, load, packet sizes, and more. The traffic statistics are available from both primary and secondary Probe.

Table 8: Qosium Traffic Statistics

Traffic statistics (specific for sent or received traffic of Primary or Secondary Probe)		
Field	Description	Unit
Duplicates	The observed duplicate packets (based on the packet identification method).	Packets
Duplicates, total	The observed duplicate packets as above but summed over the measurement.	Packets
Packet size	Average packet size during the last averaging interval. Note: The packet size means the total size at the link layer, so there is the link-layer header (Ethernet, etc.) also included.	Bytes
Protocol	Dominating protocol number in the measured stream.	Protocol number
Traffic, bits/s	The traffic load in bits/s (offered load and throughput).	Bits / s
Traffic, bytes	The amount of transmitted/received data.	Bytes

[4] Radio quality-related measurements are under development.

Traffic, bytes, total	The amount of transmitted/received data summed over the measurement.	Bytes
Traffic, packets	The amount of transmitted/received packets.	Packets
Traffic, packets, total	The amount of transmitted/received packets summed over the measurement.	Packets
Traffic, packets/s	The traffic load calculated in pkts/s (offered load and throughput).	Packets / s

QoS Statistics: QoS statistics are mostly two-point statistics that are calculated for a flow over a network path. They describe the flow's behaviour over the network path and include important statistics like delay, packet loss, etc. Some of the QoS statistics can also be calculated in a single-point measurement for a restricted set of protocols (e.g., RTP and MPEG-2 TS). All the QoS statistics Qosium produces are one-way statistics between the measurement points. These statistics are available in a two-point measurement only.

Table 9: Qosium Quality of Service Statistics

QoS statistics (specific for sent and received directions)		
Field	Description	Unit
Connection break count	The number of connection breaks during the last averaging interval.	Count
Connection break duration	Connection break duration (time-domain)	Seconds
Connection break length	Connection break length (packet domain), often as mean loss burst size.	Packets
Connection break maximum duration	The maximum connection break duration in the last averaging interval.	Seconds
Delay	Delay – one-way between two measurement points (Qosium Probes) – as all the QoS statistics.	Seconds
Delay threshold crossings	The number of packets whose delay exceeded the set threshold.	Packets
Delay threshold crossings, total	The number of packets whose delay exceeded the set threshold (total, measurement).	Packets
Delay, max	The maximum observed sample delay during the last averaging interval	Seconds
Delay, min	The minimum observed sample delay during the last averaging interval	Seconds
Delay samples	The number of Delay samples, i.e., how many values have been used to calculate the QoS statistics	Samples
Jitter	The pure delay variation	Seconds
Jitter threshold crossings	The number of packets whose jitter exceeded the set threshold	Packets

Jitter threshold crossings, total	The number of packets whose jitter exceeded the set threshold (total, measurement)	Packets
Jitter, max	The maximum observed sample jitter during the last averaging interval	Seconds
Jitter, MA	Moving average jitter (similar as in RTP)	Seconds
Lost Packets	The number of lost packets included in the QoS calculation	Packets
Lost Packets, total	The number of lost packets included in the QoS calculation (total, measurement)	Packets
Packet loss	Packet ratio of packets which were lost in the measured network path.	Percentage
Packet loss, total	Packet loss ratio (total, measurement)	Percentage
QoS samples	The number of packets that were used to calculate the QoS statistics.	Packets
Sent info not found	This special statistic is, in a way negative packet loss, meaning that a packet was received but was not indicated as sent in the other end. As understood, this is not a normal situation, and it is often a sign of false parameterization if is continuously nonzero values here. As a result, this statistic should be observed as an indicator of something weird that should be investigated.	Packets
Sent info not found, total	The summed value of the previous over the measurement.	Packets
Successful packets	The number of successfully transmitted packets included in the QoS calculation	Packets
Successful packets, total	The number of successfully transmitted packets included in the QoS calculation (total, measurement)	Packets

QoE Statistics: The Quality of Experience (QoE) is closer to the use's observations than the pure QoS. These statistics are available only for certain applications. While QoE refers to real user tests, the QoE statistics provided here are QoE approximations based on the measured QoS metrics and correlation with measurement data gathered with real user tests.

Table 10: Qosium Quality of Experience Statistics

QoE statistics (specific for sent and received directions)		
Field	Description	Unit
GQoSM [5]	QoE evaluation by using GQoSM algorithm	MOS (1-5)

[5] The Field Generic QoS Measure (GQoSM) algorithm is a parameter based QoS mapping algorithm allowing to map a single quality indicator from several parameters.

PSQA [6]	QoE evaluation by using PSQA algorithm	MOS (1-5)

Qosium use Mean Opinion Score (MOS) to measure QoE, from 1 to 5 scale. The value of 5 means excellent, non-impaired satisfaction, and the value of 1 that the application is not usable at all. Qosium use the MOS scale due to its widespread utilization.

1.2.1.2.4. Specific Data Set of Qosium Storage Database

It is important to consider the data set of Qosium Storage have similar fields of previous tables with the difference it contains the specific data fields stored in the database. Qosium Storage database contains specifically the following tables, fields, descriptions and data types:

Data Set of QM_IDs table: The table data set for measurement IDs at database level is composed by:

Table 11: Qosium Measurement IDs Table

Field	Description	Data type
caller_mapping_id_addr	Address part of the Caller mapping ID, generated automatically	Integer
caller_mapping_id_port	Port part of the Caller mapping ID, generated automatically	Integer
controller_Id	ID of the measurement controller, fixed	Integer
customer_Id	Unique ID of the "Customer", fixed	Integer
measurement_id	A unique measurement ID per given time in a single Qosium Probe.	Integer
meas_description	A voluntary and free style description of the measurement	Text
meas_end_time	The time when the flow was terminated. In results files, N/A is used if the flow had not ended at the end of the measurement.	Double
meas_start_time	Measurement stat time: if not set, the measurement has not yet been started	Double
qm_id	Measurement Identifier	Bigserial
secs_mapping_id_addr	Address part of the Secondary Service mapping ID, generated automatically	Integer
secs_mapping_id_port	Port part of the Secondary Service mapping ID, generated automatically	Integer
service_id	User given ID of the Qosium Probe (in QosiumProbe.ini)	Integer
service_mapping_id_addr	Address part of the Service mapping ID, generated automatically	Integer
service_mapping_id_port	Port part of the Service mapping ID, generated automatically	Integer
time	Epoch Time	Big Integer
user_id	User given ID of the user (given in subscribe(..) call)	Integer

[6] Pseudo-Subjective Quality Assessment (PSQA) is a neural network based model for estimating QoE.

user_quality_average	moving average quality 1.0-5.0 over the user_id measurements	Real
user_quality_symmetry	symmetry of the transmission (for VoIP for example)	Real

Data Set of averagestats table: The next data set contain all the average statistics composed by the traffic statistics, QoS statistics, single-point general statistics (e.g., location metrics), and physical-layer statistics (e.g., signal strength metrics), which are stored as-is in the Qosium Storage database:

Table 12: Qosium Average Statistics Table

Field	Description	Data Type
calc_period	The true averaging period [s]	Real
ctrl_pk_num	The number of the control packet (if sent over QMCP)	Integer
Id	index of results	Bigserial
meas_duration	The measurement duration in [s]	Real
measurement_id	Measurement case identification such as 5G	Integer
prim_accuracy	Accuracy of the position [7] in meter	Real
prim_age	Age of position information	Integer
prim_altitude	The altitude in meters	Real
prim_bs_mac_address	The MAC address of the Base Station	Text
prim_cell_id	Cell ID or similar of the network	Integer
prim_dropped_pk_if	The packets dropped by the interface	Integer
prim_dropped_pk_if_total	The total (since the beginning of measurement) packets dropped by the interface	Big Integer
prim_dropped_pk_pcap	The packets dropped by pcap	Integer
prim_dropped_pk_pcap_total	The total (since the beginning of measurement) packets dropped by pcap	Big Integer
prim_heading	Heading in degrees [8]	Small Integer
prim_latitude	The latitude in degrees	Decimal(9,6)
prim_longitude	The longitude in degrees	Decimal(9,6)
prim_malformed_pkts	The number of malformed packets	Integer
prim_malformed_pkts_total	The total number of malformed packets	Big Integer
prim_network_type	Network type (Qosium's own categories)	Small Integer
prim_position_mode		Small Integer
prim_rec_bits_s	Throughput (received traffic load by Primary Measurement Point) [bits/s]	Real
prim_rec_bytes	The amount of traffic received [bytes]	Big Integer
prim_rec_bytes_total	The total amount of traffic received [bytes]	Big Integer

[7] Position data set corresponds for primary measurement point (e.g., a mobile device with the Global Positioning System enabled).

[8] Location data set corresponds to for measurement points as mobile devices using the global positioning system.

prim_rec_dupl	The number of duplicates received	Integer
prim_rec_dupl_total	The total (summed) number of duplicates received	Big Integer
prim_rec_pkt_size	Packet size [B]	Integer
prim_rec_pkts	The number of received [packets]	Integer
prim_rec_pkts_s	Received traffic load in [packets/s]	Real
prim_rec_pkts_total	The total number of received packets [packets]	Big Integer
prim_rec_proto		Small Integer
prim_sent_bits_s	Throughput (sent traffic load by Primary Measurement Point) [bits/s]	Real
prim_sent_bytes	The amount of traffic sent [bytes]	Big Integer
prim_sent_bytes_total	The total amount of traffic sent [bytes]	Big Integer
prim_sent_dupl	The number of duplicates received	Integer
prim_sent_dupl_total	The total (summed) number of duplicates received	Big Integer
prim_sent_pkt_size	Packet size [B]	Integer
prim_sent_pkts	The number of sent [packets]	Integer
prim_sent_pkts_s	Received traffic load in [packets/s]	Real
prim_sent_pkts_total	The total number of sent packets [packets] as UL data[9]	Big Integer
prim_sent_proto		Small Integer
prim_short_pkts	The number of packets that are too short for analysis	Integer
prim_short_pkts_total	The total number of packets that are too short for analysis	Big Integer
prim_signal_strength	Signal strength [10] at the interface	Real
prim_snr	SNR strength at the interface	Real
prim_speed	Speed in m/s	Small Integer
qm_id	Index to QmId	Big Integer
rec_cbc	Connection break count (during last averaging interval)	Real
rec_cbl	Connection break length (pkts) (average of breaks during the averaging interval)	Integer
rec_cbl_tot	Connection break length (total average during measurement) [pkts]	Integer
rec_correct_qos_pk	The number of successfully transmitted packets that are included in the QoS calculation	Integer
rec_correct_qos_pk_total	The total number of successfully transmitted packets that are included in the QoS calculation	Big Integer
rec_d_samples	Samples	Integer
rec_delay_max	The maximum delay value during the averaging interval [s]	Real

[9] UL represents uplink data

[10] Signal strength-related measurements correspond to mobile devices.

rec_delay_min	The minimum delay value during the averaging interval [s]	Real
rec_delay_s	The delay [s] from the received traffic	Real
rec_jitter	The jitter [s] from the received traffic	Real
rec_jitter_ma	The moving average jitter [s]	Real
rec_jitter_max	The maximum jitter value during the averaging interval [s]	Real
rec_lost_qos_pk	The number of lost packets that are included in the QoS calculation	Integer
rec_lost_qos_pk_total	The total number of lost packets that are included in the QoS calculation	Big Integer
rec_max_cb	Maximum connection break duration during the last averaging interval [s]	Real
rec_notf_total	The total (summed) number of DL [11] (received) packets, whose sent info is not found [pkts]	Big Integer
rec_packet_loss	The absolute loss ratio value in range [0,1]	Real
rec_packet_loss_total	The total absolute loss ratio value in range [0,1]	Real
rec_qoe1	QoE-evaluation: GQoSM based Score (1 - 5)	Real
rec_qoe2	QoE-evaluation: PSQA (1 - 5)	Real
rec_sent_notf	The number of DL (received) packets, whose sent info is not found [pkts]. I.e., the packets have been received at the Primary Probe machine, but never sent from the Secondary Probe machine	Integer
rec_th_exceeded_delay_pkts	The number of packets whose delay exceeded the set threshold	Integer
rec_th_exceeded_delay_pkts_total	The total number of packets whose delay exceeded the set threshold	Integer
rec_th_exceeded_jitter_pkts	The number of packets whose jitter exceeded the set threshold	Integer
rec_th_exceeded_jitter_pkts_total	The total number of packets whose jitter exceeded the set threshold	Big Integer
results_type	The type of this average message Uint8_t	Small Integer
sec_accuracy	Accuracy of the position in meters	Real
sec_age	Age of position information	Integer
sec_altitude	The altitude in meters	Real
sec_bs_mac_address	The MAC address of the Base Station	Text
sec_cell_id	Cell ID or similar of the network	Integer
sec_dropped_pk_if	The packets dropped by the interface	Integer
sec_dropped_pk_if_total	The total (since the beginning of measurement) packets dropped by the interface	Big Integer

[11] DL: Down Link also represents the received packets in this context.

sec_dropped_pk_pcap	The packets dropped by pcap	Integer
sec_dropped_pk_pcap_total	The total (since the beginning of measurement) packets dropped by pcap	Big Integer
sec_heading	Heading in degrees	Small Integer
sec_latitude	The latitude in Degrees	Decimal(9,6)
sec_longitude	The longitude in Degrees	Decimal(9,6)
sec_malformed_pkts	The number of malformed packets	Integer
sec_malformed_pkts_total	The total number of malformed packets	Big Integer
sec_network_type	Network type (Qosium's own categories)	Small Integer
sec_position_mode		Small Integer
sec_rec_bits_s	Throughput (received traffic load by Secondary Measurement Point) [bits/s]	Real
sec_rec_bytes	The amount of traffic sent [bytes]	Big Integer
sec_rec_bytes_total	The total amount of traffic received [bytes]	Big Integer
sec_rec_dupl	The number of duplicates received	Integer
sec_rec_dupl_total	The total (summed) number of duplicates received	Big Integer
sec_rec_pkt_size	Packet size [B]	Integer
sec_rec_pkts	The number of received [packets]	Integer
sec_rec_pkts_s	Received traffic load in [packets/s]	Real
sec_rec_pkts_total	The total number of received packets [packets]	Big Integer
sec_rec_proto		Small Integer
sec_sent_bits_s	Throughput (received traffic load by Secondary Measurement Point) [bits/s]	Real
sec_sent_bytes	The amount of traffic sent [bytes]	Big Integer
sec_sent_bytes_total	The total amount of traffic sent [bytes]	Big Integer
sec_sent_dupl	The number of duplicates sent	Integer
sec_sent_dupl_total	The total (summed) number of duplicates received	Big Integer
sec_sent_pkt_size	Packet size [B]	Integer
sec_sent_pkts	The number of sent [packets]	Integer
sec_sent_pkts_s	Sent traffic load in [packets/s]	Real
sec_sent_pkts_total	The total number of sent packets [packets]	Big Integer
sec_sent_proto		Small Integer
sec_short_pkts	The number of packets that are too short for analysis	Integer
sec_short_pkts_total	The total number of packets that are too short for analysis	Big Integer
sec_signal_strength	Signal strength at the interface	Real
sec_snr	SNR strength at the interface	Real
sec_speed	Speed in m/s	Small Integer
sent_cbc	Connection break count (during last averaging interval)	Real

sent_cbl	Connection break length (pkts) (average of breaks during the averaging interval)	Integer
sent_cbl_tot	Connection break length (total average during measurement) [pkts]	Integer
sent_correct_qos_pk	The number of successfully transmitted packets that are included in the QoS calculation	Integer
sent_correct_qos_pk_total	The total number of successfully transmitted packets that are included in the QoS calculation	Big Integer
sent_d_samples	Samples	Integer
sent_delay_max	The maximum delay value during the averaging interval [s]	Real
sent_delay_min	The minimum delay value during the averaging interval [s]	Real
sent_delay_s	The delay [s] from the sent traffic	Real
sent_jitter	The jitter [s] from the sent traffic	Real
sent_jitter_ma	The moving average jitter [s]	Real
sent_jitter_max	The maximum jitter value during the averaging interval [s]	Real
sent_lost_qos_pk	The number of lost packets that are included in the QoS calculation	Integer
sent_lost_qos_pk_total	The total number of lost packets that are included in the QoS calculation	Big Integer
sent_max_cb	Maximum connection break duration during the last averaging interval [s]	Real
sent_notf_total	The total (summed) number of DL (received) packets, whose sent info is not found [pkts]	Big Integer
sent_packet_loss	The absolute loss ratio value in range [0,1]	Real
sent_packet_loss_total	The total absolute loss ratio value in range [0,1]	Real
sent_qoe1	QoE-evaluation: GQoSM based Score (1 - 5)	Real
sent_qoe2	QoE-evaluation: PSQA (1 - 5)	Real
sent_sent_notf	The number of DL (received) packets, whose sent info is not found [pkts]. I.e., the packets have been received at the Primary Probe machine, but never sent from the Secondary Probe machine	Integer
sent_th_exceeded_delay_pkts	The number of packets whose delay exceeded the set threshold	Integer
sent_th_exceeded_delay_pkts_total	The total number of packets whose delay exceeded the set threshold	Integer
sent_th_exceeded_jitter_pkts	The number of packets whose jitter exceeded the set threshold	Integer
sent_th_exceeded_jitter_pkts_total	The total number of packets whose jitter exceeded the set threshold	Big Integer

time	The timestamp when the message was created [Unix Epoch, s]	Big Integer

Qosium produces plethora of statistics that can be evaluated in real-time. The results can also be printed into files for later analysis. Specifically, the Qosium Scope can write average results to a file in TXT format. In addition, all average results are saved in the Qosium Storage database and such data can be accessed using the rest interface to read the results in JSON format using a web browser or a REST client (e.g., Postman software). In addition, there is an specific data set to be stored in the Elastic Search from the Trial Controller of the 5G!Drones system. This data set contains specific KPIs requested during the Trial in the 5GTN facility. The measurement results for the KPIs are associated to a Network Slice and measurement points selected by the User at the beginning of the Trial to fetch specific KPIs from the plethora of Qosium' data.

Table 13: Qosium Data Set - 5GTN

Function	Purpose	Objectives
Purpose of the data	KPI Measurement validation in the 5GTN network	Verify Trials
	Slicing, end-to-end KPI validation	Validate 5G KPIs
	Use Case KPI validation	Validate UAV KPIs
Type and format of Data	Log files for post processing	.txt, .csv, .JSON
Text	Readme files to understand data	.txt
Text	Log tables	.txt
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data captured in real-time during trials. Measurement data set is populated just when it is started the measurement from Qosium Scope after the Trial also starts. Qosium Scope is used manually to start and stop the measurements of KPIs. Qosium Probes collects all metrics and sent the results to the database and Qosium Scope. Qosium Storage contains all average statistics-related data set like the traffic statistics, QoS statistics, single-point general statistics, and physical-layer statistics.	Trial validation
Data set is:	Provide the status of 5GTN Network.	

	Grant a network traffic situation passing through Network Slice Instances.	
	Give the key performance indicators among the network components involved in the UAV-related trials over a 5G network.	
Quantity	For 15 minutes, in average 380 Kbytes of measurement data was generated for the Qosium Storage Database. Meanwhile the results file (.txt) of Qosium Scope has in average 490 Kbytes of measurement data. For both cases, an averaging time interval of 1 second was used to perform the measurements. A lower interval time would represent more measurement data per trial.	
Data Security & Storage	All average results of KPI measurements are mainly collected in the Qosium Storage database. These data sets can be later converted to an open data set. Institute network drive, IDA https://www.fairdata.fi/en/ida/, EUDAT https://eudat.eu/.	Accessible only with user credentials and strong authentication / VPN. IDA and EUDAT user credentials.
Data value (long term)	Data may be used for offline algorithm development regarding 5G network actual performance in UAV use-cases.	Data may have value for up to five years after the end of the project.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Metadata using Qvain. Descriptive, administrative and structural metadata.	Qvain fields: content description, actors, rights and licenses, temporal and spatial coverage, relations and history, and files.
Identifiability of data	Data will be made available using persistent identifiers (PIDs).	PID generated e.g. using **https://etsin.fairdata.fi/** https://etsin.fairdata.fi/.
Naming conventions used	Naming convention using 5G!Drones, Use Case, Trial Site, Qosium, version, and date.	
Search keywords approach	E.g., Qvain metadata search using etsin https://etsin.fairdata.fi/.	

Clear versioning approach	E.g., versioning using Qvain Relations and History fields.	Traceability, e.g. using DataCite https://schema.datacite.org/meta/kernel-4.1/doc/DataCite-MetadataKernel_v4.1.pdf. Last accessed: 27.8.2019. or Dublin Core metadata standard https://www.dublincore.org/resources/metadata-basics/.
Standards or procedures for metadata creation applied	E.g. DataCite metadata schema	
Making data openly accessible		
Data openly available	Data is the property of the 5G!Drones Partners participating in the trials. Data is shared among the 5G!Drones partners based on the Consortium Agreement.	Data that is not in Partners direct business interests can be made public, if they serve a purpose for the community at large.
Data kept closed	Industrial partners	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	Open research archives: EUDAT, IDA, Zenodo https://zenodo.org/. Use of persistent identifiers and metadata.	
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	
Repository for deposit of data, metadata, documentation and code	Private data will be stored in private Qosium Storage, Qosium Scope, GIT repositories, and Partners' internal storage drives. Publicly available data will use e.g. EUDAT, IDA, or Zenodo repositories with associated metadata using Qvain and persistent identifiers.	

Access restrictions	During trials data generation, processing, and analysis data is kept internal to 5G!Drones.	Data with re-use value will be made public after their respective 5G!Drones deliverables have been published.
Data interoperability assessment	Data interoperability will be ensured by use of e.g. DataCite metadata schema or Dublin Core metadata standard.	By default, no proprietary software is required access public data. Partners' proprietary software data will be exported and method of access to the data is defined in accompanied readme.txt file.
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	By use of e.g. ELSST https://elsst.ukdataservice.ac.uk/ thesaurus or Finto https://finto.fi/en/.	
Data licensing for wide reuse	Publicly made available data sets are by default under Creative Commons (CC) licensing. The Consortium may adopt specific license terms for specific data sets.	CC0 or CC by, sa, nc, nd, or any combination thereof. MIT is an option.
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Data can be made public after their reposting in respective deliverables.	A Partner may request an embargo period of up to one year.
Data usability by Third Parties (after the end of the project)	Publicly made data will be available for third parties for up to five years after the end of the project, after which the data is likely stale.	
Restrictions to data re-use	The 5G!Drones Consortium Agreement defines restrictions on making data public or with specific usage terms.	Certain data cannot be disclosed due to sensitivity of being close to Partner product line development or they would expose Partners product details.
Quality assurance process	Data to be made public is reviewed and accepted by the 5G!Drones consortium.	Data will be generated for trials. Analysis of the data will verify trial success. Trials are reported in deliverables, whose quality is assured by

		the quality assurance plan D6.1.
Length of time of data re-usability	It can be expected the data will be re-usable for up to five years after the end of the project.	
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. For example, UO policies can be found from https://www.oulu.fi/university/node/43683.	

1.2.1.3. X-Network data

1.2.1.3.1. Tools

Two levels are distinguished for the collected measurement at X-Network trial site, which are the orchestrator side and the UE side. The statics collected from the orchestrator level are based on a home-made solution deployed on the top Prometheus [20]. At the UE side, Aalto University makes use free tools such as netradar [21] and Ookla [22], in addition to Iperf. Furthermore, AU is developing home-made solutions to capture statistics at the UE side. The collected measurements are summarised in the two following tables.

Table 14: X-Network Statistics from the Orchestrator Level

Statistics from the orchestrator level		
Field	Description	Unit
Slice creation time	The time consumed for creating a slice	Second
Slice release time	The time consumed for releasing a slice	Second
VNF CPU usage	The percentage of CPU consumption of a specific VNF	Percentage
VNF RAM usage	The percentage of RAM consumption of specific VNF	Percentage
Node Disk usage	The percentage of disk consumption of a node	Percentage
Node CPU usage	The percentage of CPU consumption of a node	Percentage

Node RAM usage	The percentage of RAM consumption of a node	Percentage

Table 15: X-Network Statistics from the UE Level

Statistics from the UE level		
Field	Description	Unit
Throughput (UL/DL)	The throughput	Mbps
Latency (round trip)	Round trip latency from the UE to a given server	Second
RSSI	Received signal strength indication	

Table 16: X-Network Data Set

Function	Purpose	Objectives
Purpose of the data	KPI Measurement validation	Verify Trials
	Use Case KPI validation	Validate UAV KPIs
Type and format of Data	Log files for post processing	.txt, .csv, .JSON
Reused-data (rd)	No	Data will be generated in the trials.
Quantity	The quality will be defined according to the planned trials.	
Data Security & Storage	The measurements are stored in servers hosted at the university. The access to the measurement is restricted only to the consortium members.	Accessible only with user credentials.
Data value (long term)	Data may be used for offline algorithm development regarding 5G network actual performance in UAV use-cases.	
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Using standard vocabulary from well-known tools such as Prometheus	
Identifiability of data	Data id identified by PIDs	
Naming conventions used	Using standard vocabulary from well-known tools such as Prometheus	

Making data openly accessible		
Data kept closed	Industrial partners	Data close to own product line development. Data from partners own products
	Academic and research institute partners	Data exposing infrastructure based on background commitments.
Repository for deposit of data, metadata, documentation and code	Private storages are used to store the measured data.	
Access restrictions	During trials data generation, processing, and analysis data is kept internal to 5G!Drones.	
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies.	Using standard vocabulary from well known tools such as Prometheus.	
Increase data re-use (through clarifying licenses)		
Restrictions to data re-use	The 5G!Drones Consortium Agreement defines restrictions on making data public or with specific usage terms.	Certain data cannot be disclosed due to sensitivity of being close to Partner product line development or they would expose Partners product details.
Quality assurance process		Data will be generated for trials. Trials are reported in deliverables, whose quality is assured by the quality assurance plan D6.1.
Allocation of resources		
Costs estimates for making data FAIR	Costs estimates for making data FAIR	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies.	

1.2.1.4. Air interface data

Tools that will be used to gather 5G air interface data at the University of Oulu are Nemo Handy [6] and Nemo Outdoor [7]. The tools provide the following kinds of data.

- **Status**: System Time, Logging Status, Log File Size, Log File Path, Free Storage, Free Memory, Heap Size, Battery temp, Script status, Script file, GPS status, IMEI, IMSI, System Lock (LTE Only), Band lock (e.g. LTE FDD 2600 band 7), GSM BCCH lock, WCDMA Cell Lock, LTE Cell Lock, APN (e.g. 5GTN Oulu), PDP type (IP), Interface, Address (192.168...), Gateway (192.168...), DNS (185.38.2.36...), and NSAPI. These data are actual metadata that can be stored as metadata using Qvain [8] in order to understand the actual data in the research data storages.
- **Script**: Active transactions and Script commands. These data are not stored and they have no metadata value.
- **Signalling:** requests, master information, system information, radio resource control (RRC) connections, etc. These data can be stored as a part of the research data in case they bring added value to processing the data.
- **Summary:** Serving Carrier RSSI (dBm), Serving SNR (dBm), Serving RSRQ, Serving RSRP, Serving PCI, Detected RSRQ, Detected RSRP, Detected PSI. These data serve a purpose in validating the 5G KPIs and are core validation data generated by the tool.
- **Cell measurements:** Depicts various parameters of the mobile network cell. Also produces graphical plotting of the measured information. These data are important in the validations of the trials.
- **Application throughput:** Downlink and uplink. These data are essential in validation of trials and potential material for large data sets.

Specifically, the Nemo Handy and Outdoor provide log file exports in a compressed .zip file containing MS Excel comma delimited file as well as and a map file (.nmf). The above describes measurements for both LTE and 5G and data sets will be obtained for 5G radio interfaces.

Table 17: 5G Air-Interface Data Set

Function	Purpose	Objectives
Purpose of the data	5G air interface validation	Verify Trials, gather UAV link statistics
Type and format of Data	Log files for post processing	.nmf, .csv
Text	Readme files to understand data	.txt
Numeric	Log tables	.txt, .nmf, .csv
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data captured in real-time during trials	Trial validation
Data set is:	Revisable	
Quantity	1s to 10s of MB per trial.	

Data Security & Storage	Institute network drive, IDA [2], EUDAT [1].	Accessible only with user credentials and strong authentication / VPN. IDA and EUDAT user credentials.
Data value (long term)	Data may be used for offline algorithm development regarding radio environment of 5G outdoors.	Data may have value for up to five years after the end of the project.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Metadata using Qvain. Descriptive, administrative and structural metadata.	Qvain fields: content description, actors, rights and licenses, temporal and spatial coverage, relations and history, and files.
Identifiability of data	Data will be made available using persistent identifiers (PIDs).	PID generated e.g. using **https://etsin.fairdata.fi/** [9]
Naming conventions used	Naming convention using 5G!Drones, Use Case, Trial Site, version, and date.	
Search keywords approach	E.g., Qvain metadata search using etsin [9].	
Clear versioning approach	E.g., versioning using Qvain Relations and History fields.	Traceability, e.g. using DataCite [10] or Dublin Core metadata standard [12].
Standards or procedures for metadata creation applied	E.g. DataCite metadata schema	
Making data openly accessible		
Data openly available	Data is the property of the 5G!Drones Partners participating in the trials. Data is shared among the 5G!Drones partners based on the Consortium Agreement.	Data that is not in Partners direct business interests can be made public, if they serve a purpose for the community at large.
Data kept closed	Industrial partners	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
Data kept closed	Academic and research institute partners	Data exposing infrastructure based on background commitments.

How data will be made available	Open research archives: EUDAT, IDA, Zenodo [11]. Use of persistent identifiers and metadata.	
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	
Repository for deposit of data, metadata, documentation and code	Private data will be stored in private GIT repositories and Partners' internal storage drives. Publicly available data will use e.g. EUDAT, IDA, or Zenodo repositories with associated metadata using Qvain and persistent identifiers.	
Access restrictions	During trials data generation, processing, and analysis data is kept internal to 5G!Drones.	Data with re-use value will be made public after their respective 5G!Drones deliverables have been published.
Data interoperability assessment	Data interoperability will be ensured by use of e.g. DataCite metadata schema or Dublin Core metadata standard.	By default, no proprietary software is required access public data. Partners' proprietary software data will be exported and method of access to the data is defined in accompanied readme.txt file.
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	By use of e.g. ELSST [13] thesaurus or Finto [14].	
Data licensing for wide reuse	Publicly made available data sets are by default under Creative Commons (CC) licensing. The Consortium may adopt specific license terms for specific data sets.	CC0 or CC by, sa, nc, nd, or any combination thereof.
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Data can be made public after their reposting in respective deliverables.	A Partner may request an embargo period of up to one year.
Data usability by Third Parties (after the end of the project)	Publicly made data will be available for third parties for up to five years after the	

	end of the project, after which the data is likely stale.	
Restrictions to data re-use	The 5G!Drones Consortium Agreement defines restrictions on making data public or with specific usage terms	Certain data cannot be disclosed due to sensitivity of being close to Partner product line development or they would expose Partners product details.
Quality assurance process	Data to be made public is reviewed and accepted by the 5G!Drones consortium.	Data will be generated for trials. Analysis of the data will verify trial success. Trials are reported in deliverables, whose quality is assured by the quality assurance plan D6.1.
Length of time of data re-usability	It can be expected the data will be re-usable for up to five years after the end of the project.	
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. For example, UO policies can be found from https://www.oulu.fi/university/node/43683.	

1.2.1.5. User equipment data

User equipment (UE) within the context of 5G!Drones are attached to UAVs. The generated data significantly coincides with air interface data and the data set description and FAIR use of data essentially coincides with that of Table 17. In addition to the above-described air-interface data, the UE collect 5G standard based log files, including throughput, latency, jitter, etc. The UE will have an essential role in validating project 5G key performance indicators, such as enhanced mobile broadband, ultra-reliable low latency communications, and massive machine type communications.

1.2.1.6. Base station data

Regardless whether the base station is 4G, 5G non-stand alone, or 5G stand-alone the base station manages signalling, synchronisation and RRC. These data, derived from the 5G!Drones use case trials may be of re-use value and needs to be managed. The base station may be a macro base station servicing outdoors area or a small cell base station servicing from indoor or outdoor UAV use. The base station data set can be found from Table 18.

Table 18: Base Station Data Set

Function	Purpose	Objectives
Purpose of the data	4G, 5G non-stand alone and 5G stand-alone validation	Verify Trials
Type and format of Data	Log files for post processing	.txt, .xlsx .m, .csv
Text	Readme files to understand data	.txt
Numeric	Log tables	.txt, .xlsx, .m, .csv
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data captured in real-time during trials	Trial validation
Data set is:	Revisable	
Quantity	10s to 1000s of MB per trial.	
Data Security & Storage	Institute network drive, IDA [2], EUDAT [1].	Accessible only with user credentials and strong authentication / VPN. IDA and EUDAT user credentials.
Data value (long term)	Data may be used for offline algorithm development for optimization of radio resource scheduling both indoors and outdoors.	Data may have value for up to five years after the end of the project.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Metadata using Qvain. Descriptive, administrative and structural metadata.	Qvain fields: content description, actors, rights and licenses, temporal and spatial coverage, relations and history, and files.
Identifiability of data	Data will be made available using persistent identifiers.	PID generated e.g. using **https://etsin.fairdata.fi/** [9]
Naming conventions used	Naming convention using 5G!Drones, Use Case, Trial Site, version, and date.	
Search keywords approach	E.g., Qvain metadata search using etsin [9].	

Clear versioning approach	E.g., versioning using Qvain Relations and History fields.	Traceability, e.g. using DataCite [10] or Dublin Core metadata standard [12].
Standards or procedures for metadata creation applied	E.g. DataCite metadata schema	
Making data openly accessible		
Data openly available	Data is the property of the 5G!Drones Partners participating in the trials. Data is shared among the 5G!Drones partners based on the Consortium Agreement.	Data that is not in Partners direct business interests can be made public, if they serve a purpose for the community at large.
Data kept closed	Industrial partners	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
Data kept closed	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	Open research archives: EUDAT, IDA, Zenodo [11]. Use of persistent identifiers and metadata.	
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	
Repository for deposit of data, metadata, documentation and code	Private data will be stored in private Partners' internal storage drives. Publicly available data will use e.g. EUDAT, IDA, or Zenodo repositories with associated metadata using Qvain and persistent identifiers.	
Access restrictions	During trials data generation, processing, and analysis data is kept internal to 5G!Drones.	Data with re-use value will be made public after their respective 5G!Drones deliverables have been published.
Data interoperability assessment	Data interoperability will be ensured by use of e.g. DataCite metadata schema or Dublin Core metadata standard.	Partners' proprietary software data will be exported and method of access to the data is defined in accompanied readme.txt file.

Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	By use of e.g. ELSST [13] thesaurus or Finto [14].	
Data licensing for wide reuse	Publicly made available data sets are by default under Creative Commons licensing. The Consortium may adopt specific license terms for specific data sets.	CC0; or CC by, sa, nc, nd, or any combination thereof.
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Data can be made public after their reposting in respective deliverables.	A Partner may request an embargo period of up to one year.
Data usability by Third Parties (after the end of the project)	Publicly made data will be available for third parties for up to five years after the end of the project, after which the data is likely stale.	
Restrictions to data re-use	The 5G!Drones Consortium Agreement defines restrictions on making data public or with specific usage terms	Certain data cannot be disclosed due to sensitivity of being close to Partner product line development or they would expose Partners product details.
Quality assurance process	Data to be made public is reviewed and accepted by the 5G!Drones consortium.	Data will be generated for trials. Analysis of the data will verify trial success. Trials are reported in deliverables, whose quality is assured by the quality assurance plan D6.1.
Length of time of data re-usability	It can be expected the data will be re-usable for up to five years after the end of the project.	
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. For example, UO policies can be found from	

	https://www.oulu.fi/university/node/43683.	

1.2.1.7. Core network data

The core network has many functions. A schematic of 5G non-stand alone access with EPC flow can be found from [15]. Within the context of 5G!Drones slice management and end-to-end 5G KPI fulfilment are of significant interest. Significant amounts of data are generated for trialling purposes, but the re-use value of the data may be low. As such, 5G and 4G core network data will only be made public if during the project execution the consortium perceives re-use value in the data. The core network data set is described in Table 19.

Table 19: Core Network Data Set

Function	Purpose	Objectives
Purpose of the data	Slicing, end-to-end KPI (eMBB, uRLLC, mMTC) validation	Verify Trials
Type and format of Data	Log files for post processing	.txt, .xlsx .m, .csv
Text	Readme files to understand data	.txt
Numeric	Log tables	.txt, .xlsx, .m, .csv
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data captured in real-time during trials	Trial validation
Data set is:	Revisable	
Quantity	10s to 1000s of MB per trial.	
Data Security & Storage	Institute network drive, IDA [2], EUDAT [1].	Accessible only with user credentials and strong authentication / VPN. IDA and EUDAT user credentials.
Data value (long term)	Data may be used for slicing and management and orchestration algorithm development.	Data may have value for up to three years after the end of the project.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Metadata using Qvain. Descriptive, administrative and structural metadata.	Qvain fields: content description, actors, rights and licenses, temporal and spatial coverage, relations and history, and files.

Identifiability of data	Data will be made available using persistent identifiers.	PID generated e.g. using **https://etsin.fairdata.fi/** [9]
Naming conventions used	Naming convention using 5G!Drones, Use Case, Trial Site, version, and date.	
Search keywords approach	E.g., Qvain metadata search using etsin [9].	
Clear versioning approach	E.g., versioning using Qvain Relations and History fields.	Traceability, e.g. using DataCite [10] or Dublin Core metadata standard [12].
Standards or procedures for metadata creation applied	E.g. DataCite metadata schema	
Making data openly accessible		
Data openly available	Data is the property of the 5G!Drones facility owners participating in the trials. Data is shared among the 5G!Drones partners based on the Consortium Agreement.	Data that is not in Partners direct business interests can be made public, if they serve a purpose for the community at large.
Data kept closed	Industrial partners	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
Data kept closed	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	Open research archives: EUDAT, IDA, Zenodo [11]. Use of persistent identifiers and metadata.	
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	
Repository for deposit of data, metadata, documentation and code	Private data will be stored in private Partners' internal storages or repositories. Publicly available data will use e.g. EUDAT, IDA, or Zenodo repositories with associated metadata using Qvain and persistent identifiers.	

Access restrictions	During trials data generation, processing, and analysis data is kept internal to 5G!Drones.	Data with re-use value may be made public after their respective 5G!Drones deliverables have been published.
Data interoperability assessment	Data interoperability will be ensured by use of e.g. DataCite metadata schema or Dublin Core metadata standard.	Partners' proprietary software data will be exported and method of access to the data is defined in accompanied readme.txt file.
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	By use of e.g. ELSST [13] thesaurus or Finto [14].	
Data licensing for wide reuse	Publicly made available data sets are by default under Creative Commons licensing. The Consortium may adopt specific license terms for specific data sets.	CC0; or CC by, sa, nc, nd, or any combination thereof.
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Data can be made public after their reposting in respective deliverables.	A Partner may request an embargo period of up to one year.
Data usability by Third Parties (after the end of the project)	Publicly made data will be available for third parties for up to three years after the end of the project, after which the data is likely stale.	
Restrictions to data re-use	The 5G!Drones Consortium Agreement defines restrictions on making data public or with specific usage terms.	Certain data cannot be disclosed due to sensitivity of being close to Partner product line development or they would expose Partners product details.
Quality assurance process	Data to be made public is reviewed and accepted by the 5G!Drones consortium.	Data will be generated for trials. Analysis of the data will verify trial success. Trials are reported in deliverables, whose quality is assured by the quality assurance plan D6.1.

Length of time of data re-usability	It can be expected the data will be re-usable for up to three years after the end of the project.	
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. For example, UO policies can be found from https://www.oulu.fi/university/node/43683.	

1.2.1.8. Multi-access edge computing data

Edge computing as an evolution of cloud computing brings application hosting from centralised data centres down to the network edge, closer to consumers, verticals, and the data generated by applications. Multi-access edge computing is acknowledged as one of the key pillars for meeting the demanding KPIs of 5G, especially as far as low latency and bandwidth efficiency are concerned. [16]

As the MEC does not make any assumptions on the radio infrastructure, it is highly flexible within the scope of 5G!Drones and it can be utilised in trials on LTE, 5G non-stand alone, and 5G stand-alone trials. It can be used also in handovers between the technologies. Due to the virtualised characteristics of MEC, it has never been easier to monitor performance and resource needs of an application. MEC works as an enabler for multiple vertical applications among which the following appear appealing for 5G!Drones:

- Support of Local Area Data Network (LADN) by the 5G Core Network by providing support to connect to the LADN in a certain area where the applications are deployed. The access to a LADN is only available in a specific LADN service area.
- The Session and Service Continuity modes for different UE and application mobility scenarios.
- Local Routing and Traffic Steering: the 5G Core Network provides the means to select traffic to be routed to the applications in the local data network.

While the above have significant importance in the validation of both 5G and UAV verticals' KPIs, the data log files generated by MEC will most likely have little value as re-usable data sets. Still, the MEC data set is described in Table 20. Alternatively, the MEC data set can be populated with Qosium probes' data described in Table 13.

Table 20: Multi-access Edge Computing Data Set

Function	Purpose	Objectives

Purpose of the data	5G and UAV KPI validation	Validate Trials' performance
Type and format of Data	Log files for post processing	.txt, .xlsx, .csv
Text	Readme files to understand data	.txt
Numeric	Log tables	.txt, .xlsx, .m, .csv
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data captured in real-time during trials	Trial validation
Data set is:	Revisable	
Quantity	Few MBs to 1000s of MB per trial.	
Data Security & Storage	Institute network drive, IDA [2], EUDAT [1].	Accessible only with user credentials and strong authentication / VPN. IDA and EUDAT user credentials.
Data value (long term)	Data will most likely have no long term value.	Data may have value for up to three years after the end of the project.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Metadata using Qvain. Descriptive, administrative and structural metadata.	Qvain fields: content description, actors, rights and licenses, temporal and spatial coverage, relations and history, and files.
Identifiability of data	Data will be made available using persistent identifiers.	PID generated e.g. using **https://etsin.fairdata.fi/** [9]
Naming conventions used	Naming convention using 5G!Drones, Use Case, Trial Site, version, and date.	
Search keywords approach	E.g., Qvain metadata search using etsin [9].	
Clear versioning approach	E.g., versioning using Qvain Relations and History fields.	Traceability, e.g. using DataCite [10] or Dublin Core metadata standard [12].
Standards or procedures for metadata creation applied	E.g. DataCite metadata schema	
Making data openly accessible		

Data openly available	Data is the property of the 5G!Drones facility owners participating in the trials. Data is shared among the 5G!Drones partners based on the Consortium Agreement.	Data that is not in Partners direct business interests can be made public, if they serve a purpose for the community at large.
Data kept closed	Industrial partners	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
Data kept closed	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	Open research archives: EUDAT, IDA, Zenodo [11]. Use of persistent identifiers and metadata.	
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	
Repository for deposit of data, metadata, documentation and code	Private data will be stored in private Partners' internal storages or repositories. Publicly available data will use e.g. EUDAT, IDA, or Zenodo repositories with associated metadata using Qvain and persistent identifiers.	
Access restrictions	During trials data generation, processing, and analysis data is kept internal to 5G!Drones.	Data with re-use value may be made public after their respective 5G!Drones deliverables have been published.
Data interoperability assessment	Data interoperability will be ensured by use of e.g. DataCite metadata schema or Dublin Core metadata standard.	Partners' proprietary software data will be exported and method of access to the data is defined in accompanied readme.txt file.
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	By use of e.g. ELSST [13] thesaurus or Finto [14].	
Data licensing for wide reuse	Publicly made available data sets are by default under Creative Commons licensing. The Consortium may adopt	CC0; or CC by, sa, nc, nd, or any combination thereof.

	specific license terms for specific data sets.	
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Data can be made public after their reposting in respective deliverables.	A Partner may request an embargo period of up to one year.
Data usability by Third Parties (after the end of the project)	Publicly made data will be available for third parties for up to three years after the end of the project, after which the data is likely stale.	
Restrictions to data re-use	The 5G!Drones Consortium Agreement defines restrictions on making data public or with specific usage terms.	Certain data cannot be disclosed due to sensitivity of being close to Partner product line development or they would expose Partners product details.
Quality assurance process	Data to be made public is reviewed and accepted by the 5G!Drones consortium.	Data will be generated for trials. Analysis of the data will verify trial success. Trials are reported in deliverables, whose quality is assured by the quality assurance plan D6.1.
Length of time of data re-usability	Most likely the data will not be re-usable. In the case data can be re-used the time is expected to be up to three years after the end of the project.	
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. For example, UO policies can be found from **https://www.oulu.fi/university/node/43683**.	

1.2.2. Unmanned aerial vehicle data

UAV data are divided between traffic management and application data, and UAV sensor data. The unmanned aerial systems (UAS) traffic management (UTM) data is very close to several of the partners core business and therefore much of them will not be made publicly available.

1.2.2.1. Unmanned aerial systems traffic management data

Drones will be an important component of 5G communications, through their use within cellular architectures in which they facilitate both wireless broadcast and point-to-point transmissions. The command and control link provides safety critical information for the control of the drone to deal with flight and flight management.

Beyond visual line of sight (BVLoS) drone operations entails long-range commercial drone control, for applications such as drone delivery. The command and control application will demonstrate features such as automatic collision avoidance of drones, especially those flying in swarm, which requires sending large amounts of data in near real time to assess the potential risks in the sky and enable an enhanced flight awareness of all types of flying objects. Some BVLoS applications enable secure controlling of the drone (telepresence) using the VR/AR equipment. Already existing custom-made and/or openly available 3D data can be used here in such cases, hence existing data sets may be used. For example, the National Land Survey of Finland (**https://www.maanmittauslaitos.fi/en**) has made freely available spatial data from aerial images and laser scans.

In drone operations both pre-flight and after flight data must be generated. All use case trials must be performed according to regulations. The regulations for each trial site location are defined in 5G!Drones deliverable D1.2. As per agreement UTM will be part of all use cases. This includes strategic pre-flight steps:

- Registration of
  - Operator,
  - Drone, and
  - Crew.
- E-Identification of
  - Means to localize drones and
  - Means to identify drones.

The following important tactical pre-flight steps shall be considered as well:

- Risk Assessment – SORA,
- Mission Planning in coordination with the domain specific stakeholder,
- Flight Planning and Approval with the respective UTM counterpart, and
- Business Case related topics – E.g. financial aspects in getting approval for flights

After a flight tactical steps are part of all scenarios:

- Sufficient legal recording shall in be in place.
- Recorded monitoring and test data to be used for Study reports need to be collected
- Business case related topics need to be handled (closing transactions started in tactical pre-flight phase).

Coverage of these topics is ensured by the respective 5G!Drones's UTM Partners in the project. Even though significant amounts of data will be generated in the UTM use cases, it has little to none re-use value that should be shared. The UTM data set can be found from Table 21.

Table 21: Unmanned Aerial Systems Traffic Management Data Set

Function	Purpose	Objectives
Purpose of the data	Air traffic information aimed at unmanned traffic management	UTM Trials
Type and format of Data	Stream of live data	JSON
Text		JSON
Numeric		JSON
Discipline specific information	Aviation	ADS-B, FLARM, Mode S, Mode A/C
Reused-data (rd)	In 3D visualization – Yes, Otherwise – No.	e.g. **https://www.maanmittauslaitos.fi/en** Data will be generated from own hardware/software.
Observational	Real time information from standard aviation identification technology, coming from Partners' sensors' readings	UTM, Trial validation
Data set is:	Growing	
Quantity	About 5GB/sensor/day	
Data Security & Storage	Secure WebSocket	Private server protected by SSH authentication
Data value (long term)	Drone operators, Air Traffic Controllers, Sky Authorities. Data are confidential for business reasons.	
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data detailed in a specification document.	Metadata: GNSS coordinates, timestamps
Identifiability of data	Link with external aviation databases	Aircraft identifiable through their ICAO number / tail number
Search keywords approach	ICAO number / tail number	

Clear versioning approach	Based on timestamp	Based on timestamp
Standards or procedures for metadata creation applied	No reference, internal to Partners Metadata: timestamp, GNSS coordinates created by Partners' sensors'.	
Making data openly accessible		
Data openly available	Partner's ownership, not open data since it would disclose core information and parts of the production lines to direct competitors. Data essential to project trials openly available	
Data kept closed	Drone operators, Air Traffic Control, Sky Authorities.	It would disclose core information and part of Partners' production lines to direct competitors.
How data will be made available	WebSocket data stream	
Methods or software (SW) tools for data access	Application Interface, API	
SW documentation and other information needed	File specifying the stream protocol and architecture is available on demand and subject to confidentiality.	
Repository for deposit of data, metadata, documentation and code	Private data will be stored in private Partners' internal storages or GIT repositories.	
Access restrictions	Access to data is limited in time as per the Consortium Agreement.	
Data interoperability assessment	High level of interoperability (GNSS coordinates and timestamp)	no license / limitations
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	Specification document.	
Data licensing for wide reuse	N/A	N/A

Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Embargo on commercial use, especially for statistics	One year
Data usability by Third Parties (after the end of the project)	Restrictions to use for statistical analysis (e.g. use of the sky, traffic analysis, etc.) for commercial or non-commercial use.	
Restrictions to data re-use	Core business Confidentiality agreement	Data re-use could disclose large parts of Partners' core business to competitors and would thus be a threat to its survival. Existing confidentiality agreements also apply.
Quality assurance process	Cross validation, Multilateration, other algorithms	Partner internal control processes and quality guidelines
Length of time of data re-usability	N/A	
Allocation of resources		
Costs estimates for making data FAIR	20.000 €/year	Own resources
Data Management Responsibilities	Each Partner has its own data management policies. Appointed Data Management Leader.	

1.2.2.2. Alerion's PX4 data

The autopilot firmware PX4 logs various UAV metrics on the onboard SD card. The data is recorded at two messaging levels:

- uOrb messaging, which handles internal communication and results in uLog files
- MAVLink messaging, which acts as an interface between the autopilot and other components (*e.g.* sensors, groundstation) and results in tLog files (or csv files).

In both cases the data contains readings from the UAV sensors and actuators.

Table 22: Alerion's PX4 Data Set

Function	Purpose	Objectives
Purpose of the data	Debugging UAV sub-systems, real time monitoring of the UAV	Have feedback from the UAV and know the current state of

		its systems; understand the causes of possible failures.
Type and format of Data	Form	Format
Text	Status messages, field names	Human readable .csv
Numeric	Timestamps and various sensor and actuator readings	.ulog, .tlog, .csv
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data is captured in real-time during the trials	nil
Data set is:	Growing	nil
Quantity	~100 MB/h	nil
Data Security & Storage	Logs are recorded on the onboard SD card and on the computer hosting the GCS	Accessible only with user credentials and strong authentication.
Data value (long term)	The data can be used to debug issues with the UAV systems, replay the flights and log operational time. Data is valuable to Alerion.	Data may have value for up to five years.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With file naming and timestamps	nil
Clear versioning approach	Timestamping, file naming	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	The dataset comprises of mostly private data. Data is shared among the 5G!Drones partners based on the	This data has low research value for the larger

	Consortium Agreement. Data is owned by Alerion.	community and is specific to Alerion UAV setup.
Data kept closed	Data related to UAV systems that is close to product line development.	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	By sending the files upon request.	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	nil
Repository for deposit of data, metadata, documentation and code	The data will be stored in private storage.	nil
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is accessible with open-source software. Good interoperability.	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	nil	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Embargo on commercial use.	Five years
Data usability by Third Parties (after the end of the project)	Not available for third parties	nil

Restrictions to data re-use	Confidentiality agreement	The data cannot be disclosed due to sensitivity of being close to product line development.
Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. Appointed responsible postholder for data management.	nil

1.2.2.3. Unmanned aerial vehicle sensor data

UAV sensor data covers major amounts of data. The utilised sensors and their data collection rates depend on the particular use case. Without loss of generality at least the following sensors collecting data will be used:

- Light Detection and Ranging (LiDAR),
- 3D accelerometer,
- Gyroscope,
- Video camera,
- Still camera,
- Air Pressure,
- Sound power level,
- Various radio technology analysers,
- Bluetooth positioning (indoor applications),
- Magnetometer, and
- Drone telemetry.

The amount of data generated is particularly high for LiDAR (< 100 Mbps) making the data sets large. The collected sensor data have significant re-use value, both for public purposes and for Partners' product line development and testing purposes. With regards to privacy, all stored data shall not contain any information that would raise the EU General Data Protection Regulation (GDPR) Act (EU) 2016/679 [19] related privacy issues. The UAV sensor data set is described in Table 23.

Table 23: Unmanned Aerial Vehicle Sensor Data Set

Function	Purpose	Objectives
Purpose of the data	UAV operations, trial validation, collection of data for post-processing	Safe drone operations, UAV and 5G KPI testing, data sets for algorithm development.
Type and format of Data	Log files for post processing, video, still photos	.txt, .xlsx, .m, .mpeg-4 (.mp4), jpg, .raw
Text	Readme files to understand data	.txt
Numeric	Log tables	.txt, .xlsx, .m
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data captured in real-time during trials	Trial validation
Data set is:	Growing	
Quantity	100s MB to 10s of GB per trial.	
Data Security & Storage	Institute network drive, IDA [2], EUDAT [1].	Accessible only with user credentials and strong authentication / VPN. IDA and EUDAT user credentials.
Data value (long term)	Data will have mid-term value.	Data may have value for up to ten years after the end of the project.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Metadata using Qvain. Descriptive, administrative and structural metadata.	Qvain fields: content description, actors, rights and licenses, temporal and spatial coverage, relations and history, and files.
Identifiability of data	Data will be made available using persistent identifiers.	PID generated e.g. using **https://etsin.fairdata.fi/** [9]
Naming conventions used	Naming convention using 5G!Drones, Use Case, Trial Site, Sensor, version, and date.	
Search keywords approach	E.g., Qvain metadata search using etsin [9].	
Clear versioning approach	E.g., versioning using Qvain Relations and History fields.	Traceability, e.g. using DataCite [10] or Dublin Core metadata standard [12].

Standards or procedures for metadata creation applied	E.g. DataCite metadata schema	
Making data openly accessible		
Data openly available	Data is the property of the 5G!Drones Partners participating in the trials. Data is shared among the 5G!Drones partners based on the Consortium Agreement.	Data that is not in Partners direct business interests can be made public.
Data kept closed	Industrial partners	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
Data kept closed	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	Open research archives: EUDAT, IDA, Zenodo [11]. Use of persistent identifiers and metadata.	
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	
Repository for deposit of data, metadata, documentation and code	Private data will be stored in private Partners' internal storages or repositories. Publicly available data will use e.g. EUDAT, IDA, or Zenodo repositories with associated metadata using Qvain and persistent identifiers.	
Access restrictions	During trials data generation, processing, and analysis data is kept internal to 5G!Drones.	Data with re-use value will be made public after their respective 5G!Drones deliverables have been published.
Data interoperability assessment	Data interoperability will be ensured by use of e.g. DataCite metadata schema or Dublin Core metadata standard.	Partners' proprietary software data will be exported and method of access to the data is defined in accompanied readme.txt file.
Making data interoperable		
Standard vocabulary or mapping to	By use of e.g. ELSST [13] thesaurus or Finto [14].	

commonly used ontologies		
Data licensing for wide reuse	Publicly made available data sets are by default under Creative Commons licensing. The Consortium may adopt specific license terms for specific data sets.	CC0; or CC by, sa, nc, nd, or any combination thereof.
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Data can be made public after their reposting in respective deliverables.	A Partner may request an embargo period of up to one year.
Data usability by Third Parties (after the end of the project)	Publicly made data will be available for third parties for up to ten years after the end of the project, after which the data is likely stale.	
Restrictions to data re-use	The 5G!Drones Consortium Agreement defines restrictions on making data public or with specific usage terms.	Certain data cannot be disclosed due to sensitivity of being close to Partner product line development or they would expose Partners product details.
Quality assurance process	Data to be made public is reviewed and accepted by the 5G!Drones consortium.	Data will be generated for trials. Analysis of the data will verify trial success. Trials are reported in deliverables, whose quality is assured by the quality assurance plan D6.1.
Length of time of data re-usability	Data will in parts be re-usable. The time is expected to be up to ten years after the end of the project.	
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has its own data management policies. For example, UO policies can be found from **https://www.oulu.fi/university/node/43683**.	

1.2.2.4. Alerion's Bathymetry data

The multi-sensor used for bathymetry provides readings for

- Water temperature and
- Underwater depth.

These readings are stored locally on the UAV and remotely on the computer hosting the ground station. They are timestamped and logged in csv files.

Table 24: Alerion's Bathymetric Data Set

Function	Purpose	Objectives
Purpose of the data	Test data streaming over 5G network, provide 3D bathymetric reconstruction.	Real-time data streaming for smooth 3D bathymetric reconstruction.
Type and format of Data	Form	Format
Instrument specific	Timestamps and temperature and underwater depth readings	.txt
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data is captured in real-time during the trials	nil
Data set is:	Growing	nil
Quantity	~10 kB/h	nil
Data Security & Storage	Logs are recorded on the onboard SD card and on the computer hosting the GCS.	Accessible only with user credentials and strong authentication.
Data value (long term)	The data can be used to test further bathymetric data processing methods.	Data may have value for up to one year.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With file naming and timestamps	nil
Clear versioning approach	Timestamping, file naming	nil

Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	The dataset contains private data. Data is shared among the 5G!Drones partners based on the Consortium Agreement. Data is owned by Alerion.	This data has low research value for the larger community and is specific to Alerion UAV setup.
Data kept closed	The entire dataset	Data close to own product line development. Data from partners own products.
	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	By sending the files upon request.	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	nil
Repository for deposit of data, metadata, documentation and code	The data will be stored in private storage.	nil
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is accessible with open-source software. Good interoperability.	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	nil	nil
Data licensing for wide reuse	nil	nil

Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Embargo on commercial use.	One year
Data usability by Third Parties (after the end of the project)	Not available for third parties	nil
Restrictions to data re-use	Confidentiality agreement	The data cannot be disclosed due to sensitivity of being close to product line development.
Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. Appointed responsible postholder for data management.	nil

1.2.2.5. Alerion's video data

Two video camera providing First Person View are streamed from the UAV to the ground station. These streams are necessary to remotely pilot the UAV. These videos are then recorded on the computer hosting the ground station as mp4 files.

Table 25: Alerion's Video Data Set

Function	Purpose	Objectives
Purpose of the data	Used to remotely pilot the UAV and to test video streaming over 5G network.	Real-time video streaming, remote piloting in BVLOS.
Type and format of Data	Form	Format

Instrument specific	Video	.mp4
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data is captured in real-time during the trials	nil
Data set is:	Growing	nil
Quantity	~10 GB/h	nil
Data Security & Storage	Logs are recorded on the onboard SD card and on the computer hosting the GCS	Accessible only with user credentials and strong authentication.
Data value (long term)	The data can be used to test further video processing techniques and for communication purposes	Data may have value for up to five year.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With file naming and timestamps	nil
Clear versioning approach	Timestamping, file naming	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	The dataset contains private data. Data is shared among the 5G!Drones partners based on the Consortium Agreement. Data is owned by Alerion.	This data has low research value for the larger community and is specific to Alerion UAV setup.
Data kept closed	The entire dataset	Data close to own product line development. Data from partners own products.

	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	By sending the files upon request.	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	nil
Repository for deposit of data, metadata, documentation and code	The data will be stored in private storage.	nil
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is accessible with open-source software. Good interoperability.	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	nil	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Embargo on commercial use.	Five years
Data usability by Third Parties (after the end of the project)	Not available for third parties	nil
Restrictions to data re-use	Confidentiality agreement	The data cannot be disclosed due to sensitivity of being close to product line development.
Quality assurance process	nil	nil

Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. Appointed responsible postholder for data management.	nil

1.2.3. Trial Controller data

1.2.3.1. Web Portal 1 Data Set

1.2.3.1.1. General information

Web Portal 1 is an entry point to the Trial Controller. The experimenter interacts through it with the system and needs to login to Web Portal 1. For that reason, we need to collect some personal data regarding user/experimenter.
Collected data:

- User email address (used as a login)
- Password – temporary password is assigned for the first login, user is forced to change it to his own password during the first login
- First name and last name of the user – to identify person
- Time of user login to the portal

Collected data is stored in Keycloak Identity and Access Management (IAM) system, which is installed on the virtual machine provided by Aalto University (AU). Data is only collected for the project time being and planned to be removed when project ends.

1.2.3.1.2. Personnel managing the data

The IAM is managed by following personnel:

1) Cyril Dangerville (THA)
2) Yannick Schaeffer (INV)
3) Pawel Montowtt (INV)

1.2.3.1.3. Data collection for Web Portal 1 operation

It's planned to collect some data for the server resources usage and processes on which Web Portal 1 and IAM are located. This can be:

- CPU and RAM usage

- Throughput on interfaces
- Power consumption
- Operational logs
- Other

No statistical data on usage per person or organisation is planned.

1.2.3.1.4. After project ends

After the project is closed, the Web Portal 1 and IAM data are going to be removed from Aalto resources. Web Portal 1 installation files and installation procedure will be available in project GitLab repository. Keycloak up-to-date version is available on wwww.keycloak.org. The customised THA IAM Keycloak version will be accessible based on THA policies.

Data stored in Web Portal 1, representing the repositories status for defined UAVs, Operational Flight Plans or templates will be archived and stored in the common place with other data, which is identified as valuable after the end of the project.

Table 26: Web Portal 1 Data Set

Function	Purpose	Objectives
Purpose of the data	Accessing and using the Web Portal	To define, perform and analyse/download results.
Type and format of Data	Personal data, like email/name and password	Identify user. Give access to authorised persons.
	UAV details	Define UAV for the experiment.
	Operational Flight Plans	Describe the test plan for getting the UTM approval.
	Post-flight analysis	Analyse the results or download them for external analysis.
Text	Log records as strings	To analyse errors.
Numeric	Data in form of tables, charts, strings or raw measurements	To analyse the experiment.
Reused-data (rd)	No/Yes	No - data will be generated in the trials. Yes - trial description data can be re-used through stored templates.

Observational	Data is captured in real-time during the trials, which can be collected manually, as a notes.	Identify the conditions during the test, give remarks about encountered problems.
Data set is:	Revisable	User data is not changing, except password, which can be reset. UAV data can be modified. Operational Flight Plans can be modified before sending to validation by UTM.
Quantity	Depends on the trial - what is the payload characteristic	nil
Data Security & Storage	The user data is stored in Keycloak database. The UAV and Operational Flight Plan is stored in the repositories in Aalto infrastructure. The recorded logs are stored in storage managed by FRQ / THA.	Personal data accessible only by users and admins with strong authentication (OTP). Other data accessible by authenticated web portal users.
Data value (long term)	The data can be used to compare the results, can be useful for other projects for comparison or for reference in research.	Data may have value for up to three years, for project members, cellular and drone industry, research institutes.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With labels	nil
Clear versioning approach	Timestamping and test name	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		

Data openly available	The dataset with experiment results can be available publicly, excluding the data being private and belonging to 5G!Drones partners.	This data could have high research value for the larger community.
Data kept closed	Data from consortium members systems, which they decided to not disclose to public.	Data from partners own product line development. Data from partners own products.
	Academic and research institute partners.	Data exposing infrastructure based on background commitments.
How data will be made available	User credentials to the database.	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	nil
Repository for deposit of data, metadata, documentation and code	The data will be stored in a private storage.	nil
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is accessible with open-source software. Good interoperability.	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	Standard Ubuntu naming convention.	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	No embargo	nil
Data usability by Third Parties (after the end of the project)	Web Portal software and documentation will be available after the project will be finished.	nil

Restrictions to data re-use	No restrictions.	Accessible to everybody.
Quality assurance process	Internal project quality insurance and testing methods.	nil
Length of time of data re-usability	3-5 years after project is finished.	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. Appointed responsible postholder for data management.	nil

1.2.3.2. Delivery mode of Work Package 2 enablers

The Work Package 2 in responsible for developing the project's Trial Controller components. Those components have been defined and described extensively in the 5G!Drones Deliverables D2.1 - Initial definition of the trial controller architecture, mechanisms, and APIs and D2.2 - Initial implementation of the trial controller. The Table 27 describes various components of the Work Package 2 activities that have currently defined delivery more.

Table 27: Delivery Mode of Work Package 2 Enablers

Module	Owner	Delivery mode	Justification if closed source	Embargo period if open-source
Webportal1	INV	Open source License: MIT Link (to be provided)	/	No
Webportal2-EUR	EUR	Open source License: Apache v2 **Link**	/	Yes (till the end of the project)

Webportal2-NCSRD	NCSRD	Open source License: Apache v2 **Link**	/	No
Webportal2-UO	UO	Open source License: MIT Link (to be provided)	/	No
Webportal2-AU	AU	Closed source	The webportal is featured to run on the top of AU's facility, X-Network, at both backend and frontend levels. The university cannot open source this module. Therefore, the university will provide and maintain the access to the webportal.	/
Repositories	AU	Open source License: BSD **Link**	/	No
Trial validator	RXB/FRQ	Closed source	The framework used as basis for the backend as well as for the frontend components of the Trial Validator are baselined on the implementations used in the actual Frequentis UTM product. Therefore open-sourcing the Trial Validator is technically not feasible without impact on Frequentis intellectual property. Any API documentation needed by partners as well as support to use and integrate the Trial Validator will of course be provided by Frequentis.	/
Life-Cycle Manager	NOK	Open source License: Apache v2 **Link**[12]	/	No
Trial enforcement	NCSRD	Open source	/	No

[12] Currently open-sourced only to project partners via project own gitlab. Optional target is to deliver open-source via **https://github.com/nokia**

		License: Apache v2 **Link**		
KPI component	FRQ	Closed source	The internal messageing components and interfaces used as basis for implemenation of the KPI Component are baselined on the implementations used in the actual Frequentis UTM product. Therefore open-sourcing the KPI Component is technically not feasible without impact on Frequentis intellectual property. Any API documentation needed by partners as well as support to use and integrate the KPI Component will of course be provided by Frequentis.	/
U-space adapter	FRQ	Closed source	The components and interfaces which Frequentis contributes to the U-Space adapter within the 5G!Drones project are an integral part of the Frequentis UTM product suite and therefore cannot be open source. Of course any API documentation is available to all partners and Frequentis is providing support to integrate with these APIs.	/

1.2.4. Project Use Case data

1.2.4.1. UC1Sc1 Data

1.2.4.1.1. UC1Sc1 short description

The Use Case 1 Scenario 1 (UC1Sc1) represents a common use case where UAV is controlled by Ground Control Station (GCS) situated in the MEC server and Command and Control (C2) exchange between UAV and GCS is performed through 5G network. GCS is connected to traffic management (UTM) platform to send the real-time updates of UTM with drone telemetry and receive important messages on the air traffic in the vicinity of UAV operation.

1.2.4.1.2. Data sources

For further investigations, data is collected in different locations. Following sources were identified:

- Facilities IT infrastructure, including MEC server – KPIs like RAM usage, CPU/GPU load, throughput on some interfaces, etc.), RNIS data,
- 5G KPIs collected in the gNodeB and core network,
- 5G KPIs collected in the UE or modem carried by drone,
- Data collected by IoT sensors carried by drone,
- Data collected in the 5G!Drones Trial Controller,

- Telemetry data from UAV and other data shared by UTM service, and
- Video signals or pictures taken by drone and transmitted to the ground.

Additionally, the logs from UAV, GCS, UTM can be extracted and stored after the trial is concluded. Some of them will be a standard log bookkeeping required by law, some will be specific to the research and scientific nature of 5G!Drones project.

1.2.4.1.3. Measurements Types

The basic results of UC1Sc1 provides the metrics:

- the network (5G RAN and core) and supporting infrastructure (servers) performance from the perspective of the application or service.
- Traffic statistics, flow statistics, radio level statistics, events and QoS measurements are KPIs that can be obtained from Eurecom facility.
- The results are averaged values over the user-defined interval. The different source results should come with a synchronised timestamps allowing data correlation between different streams.
- UAV measurements will be associated to 3D location to allow visualization of certain types of data on the 3D map.

Data accessibility and analysis

The main aim of the experiments is to gather the data, which could be analysed and used for drawing the conclusions and experience, how 5G network with its new characteristics and features can be useful and efficient for controlling the UAV over the distances longer than the range of RC links used today. All this not compromising people and object's safety in the air and on the ground.

It's expected that the gathered data will be analysed in many different ways, by many partners working in the project and also external to the project, as majority of this data will be available publicly. It is assumed that data can be also used in other scientific research activities. Hence, at the end of the project data are going to be accessible for anyone who wants to use it in his research, under later defined licences.

Exceptions: Please specify any.

Location in which data will be accessible: common project repositories – hotsted by FRQ and THA.

Licensing: Developed software for this specific test will be accessible under the conditions specified by the partner responsible for corresponding module.

Table 28: UC1Sc1 Data Set

Function	Purpose	Objectives
Purpose of the data	Experimental control of the drone flight utilizing 5G network.	To collect data on different interfaces, which can indicate the utility of such solution, reveal the risks and define minimum operational performance requirements

Type and format of Data	The network (5G RAN and core) and supporting infrastructure (servers) performance.	To evaluate 5G network from perspective of application and service.
	Traffic statistics, different counters, radio level statistics, events and QoS measurements.	Assess the KPIs that can be obtained from Eurecom facility.
	UAV measurements associated to 3D location, telemetry data provided by drone and supplementary tracker.	Visualization of certain types of data on the 3D map, assuring discoverability and safety.
	Supplementary UTM data from INVOLI central server.	Provision of airspace traffic awareness to UTM and drone pilot, increasing the safety.
Text	Log records as strings	To analyse errors
Numeric	Data in form of tables, charts, strings or raw measurements.	To analyse the experiment
Reused-data (rd)	No	No - data will be generated in the trials.
Observational	Data is captured in real-time during the trials, which can be collected manually, as a notes.	Identify the conditions during the test, give remarks about encountered problems.
Data set is:	Static	Once data is collected, it can be analysed and manipulated, but doesn't change.
Quantity	Depends on the trial conditions, periodicity of the measurements, time granurality – it's not known for the moment.	nil
Data Security & Storage	The recorded logs are stored in storage managed by FRQ / THA. Access security assured by Keycloak IAM.	Data accessible by authenticated web portal users.
Data value (long term)	The data can be used to compare the results, can be useful for other projects for comparison or for reference in research.	Data may have value for up to 3-5 years, for project members, cellular and drone industry, research institutes.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Primarily data will be made available to consortium partners based on the consortium agreement. At the end of the	nil

	project should be made public, where no restrictions given by involved partners.	
Identifiability of data	With labels, short test description.	nil
Clear versioning approach	Timestamping and test name	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	The dataset with experiment results can be available publicly, excluding the data being private and belonging to 5G!Drones partners.	This data could have high research value for the larger community.
Data kept closed	Data from consortium members systems, which they decided to not disclose to public.	Data from partners own product line development. Data from partners own products.
	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	User credentials to the database.	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	nil
Repository for deposit of data, metadata, documentation and code	The data will be stored in a private storage.	nil
Access restrictions	The data is kept internal to 5G!Drones before final report and closure of the project.	nil
Data interoperability assessment	When only possible, the data is stored in a form that is accessible with open-source software. Good interoperability.	nil
Making data interoperable		
Standard vocabulary or mapping to	Standard Ubuntu naming convention.	nil

commonly used ontologies		
Data licensing for wide reuse	MIT license applied.	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	No embargo	nil
Data usability by Third Parties (after the end of the project)	Data will be public and accessible to everybody.	nil
Restrictions to data re-use	No restrictions	Accessible to everybody
Quality assurance process	Internal project quality insurance	nil
Length of time of data re-usability	3-5 years after project is finished	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner involved in the tests has their own data management policies. Appointed responsible postholder for data management.	nil

1.2.4.2. UC1Sc3, UC2Sc3, and UC3Sc1 Sub-Scenario 1 Data

This section describes the handling of data both for developments by CAF during the 5G!Drones project and during preparations and execution trials in which CAF participates (UC1Sc3; UC2Sc3; UC3Sc1 Sub-Scenario 1).

During the project trials and preparations, data is collected by CAF. This data is intended to describe the preparations, course, and results of the trials and will be shared with other partners (e.g. videos, photos, results of 5G coverage measurements; drone flight routes, etc.). Data collection tools and mechanisms are based on the applications used, e.g. if using a drone, the video of the drone is recorded on the memory card on board the drone.

CAFA Tech handles the following data:

1. **Telemetry and flight paths datasets from CAFA drone Ground Control Software**
2. **Videos and photos from CAFA drones**
3. **5G Network coverage quality of service data**
4. **Computer Vision training datasets (c.f. Section 1.2.5.2)**

Telemetry and flight paths datasets from CAFA drone Ground Control Software. CAF uses UGCS (Universal Ground Control Software) software, which runs both on the CAF server and on the Edge server of the respective facility during the tests. Flight logs (e.g. positions of drone, velocity, direction, altitude etc.) and telemetry data are stored until the end of the project. Telemetry data export is described in document:

https://wiki.ugcs.com/index.php?title=Category:Telemetry_player&action=pdfbook The data is used to compile reports and analyses.

Videos and photos from CAFA drones. CAF takes both photos and videos during drone flights, which are retained by CAF until the end of the project. Some photos and videos are used to compile trial reports.

5G Network coverage quality of service data. CAF uses an on-board drone application to measure 5G network quality. It is planned to use UO Qosium tool, COS, NCSRD, and publicly available (e.g. Iperf, nPerf) network coverage measurement applications.

Table 29: Telemetry and Flight Paths Data Sets from CAFA Drone Ground Control Software

Function	Purpose	Objectives
Purpose of the data	Logging GCS and drone all actions	To log all actions by drone and GCS.
Type and format of Data	KML, XML, CSV	Machine readable format
Text	Logging GCS and drone all actions	To log all actions by drone and GCS.
Numeric	KML, XML, CSV	Machine readable format
Reused-data (rd)	No	No
Observational	nil	nil
Data set is:	Revisable	nil
Quantity	In a period of 60 minutes, about 1 MB to be stored.	nil
Data Security & Storage	The recorded logs are stored in a CAF server or 5G edge server	Accessible only with user credentials and strong authentication.

Data value (long term)	The data can be used to compare the previous results with subsequent ones. The valuable only to CAF.	Data may have value until end of project.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	File names	nil
Clear versioning approach	Timestamping	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	The entire dataset is private data belonging to CAF. Data is shared among the 5G!Drones partners based on the Consortium Agreement.	This data has low research value for the larger community.
Data kept closed	Data access is restricted to consortium members.	nil
	Academic and research institute partners	nil
How data will be made available	To upload to 5G!Drones MS Teams dedicated folder.	nil
SW documentation and other information needed	nil	nil
Repository for deposit of data, metadata, documentation and code	The data will be stored in a private storage.	nil
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is accessible with open-source software. Good interoperability.	nil
Making data interoperable		

Standard vocabulary or mapping to commonly used ontologies	Standard CSV ontology	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Data is available until end of the project.	Data is available until end of the project.
Data usability by Third Parties (after the end of the project)	Not available for third parties. The data will be useless after a certain period for practical purposes.	nil
Restrictions to data re-use	nil	nil
Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered by CAF budget.
Data Management Responsibilities	CAF follows internal data management rules.	nil

Table 30: Videos and Photos from CAFA Drones

Function	Purpose	Objectives
Purpose of the data	a. To produce 3D map b. For communication and dissemination activities	a. To produce 3D map b. For communication and dissemination activities
Type and format of Data	.jpg, .png., mp4 and other video formats	Standard formats
Text	Nil	nil
Numeric	Timestamps	nil

Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data is captured in real-time during the trials.	nil
Data set is:	Revisable	nil
Quantity	In a flight period of 10 minutes, about 1 GB of data to be stored.	nil
Data Security & Storage	The recorded photos and videos are stored in CAF computers and 5G!Drones MS Teams.	Accessible only with user credentials and authentication.
Data value (long term)	The data has mid-term value.	Data may have value until end of the project.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With tags, timestamps and field names	nil
Clear versioning approach	Timestamping	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	The entire dataset is private data. Data is shared among the 5G!Drones partners based on the Consortium Agreement. Data is owned by CAF.	This data has low research value for the larger community.
Data kept closed	The entire dataset described in this table.	Nil
How data will be made available	User credentials to the MS Teams or CAF computers.	Nil
SW documentation and other information needed	nil	nil
Repository for deposit of data, metadata,	The data will be stored in a private storage.	nil

documentation and code		
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is accessible with open-source software. Good interoperability.	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	nil	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Until end of the project.	One year
Data usability by Third Parties (after the end of the project)	Not available for third parties. Due to the data having low value.	nil
Restrictions to data re-use	No reuse foreseen.	nil
Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds of Euros per annum.	During project lifetime, costs can be covered from CAF.
Data Management Responsibilities	CAF follows internal data management rules.	nil

Table 31: 5G Network Coverage Quality of Dervice Data

Function	Purpose	Objectives
Purpose of the data	Analysing 5G network QoS	Analysing 5G network QoS

Type and format of Data	CSV	Standard format
Text	Explanatory text	Human readable .csv files
Numeric	CSV	Human readable .csv files
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data is captured in real-time during the trials	nil
Data set is:	analytics	nil
Quantity	In a test period of 10 minutes, about 0.1 MB of data was observed to be stored.	nil
Data Security & Storage	The data are stored in CAF computers and 5G!Drones MS Teams.	Accessible only with user credentials and strong authentication.
Data value (long term)	Short value only for compiling trial report.	Short value only for compiling trial report.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With file naming and timestamps	nil
Clear versioning approach	Timestamping, file naming	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	Data is shared among the 5G!Drones partners based on the Consortium Agreement.	This raw data has low research value for the larger community. Analytical reports will be published by 5G!Drones project.
Data kept closed	For consortium partners.	nil

How data will be made available	By publishing reports after trials.	nil
SW documentation and other information needed	SW documentation is published by SW provider (UO, nPerf, IPerf developers etc.)	nil
Repository for deposit of data, metadata, documentation and code	The data will be stored in CAF computers.	nil
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is accessible with open-source software.	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	nil	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Until end of the project.	No
Data usability by Third Parties (after the end of the project)	Not available for third parties	nil
Restrictions to data re-use	Confidentiality agreement	nil
Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds of Euros per annum.	During project lifetime, costs can be covered from CAF.
Data Management Responsibilities	CAF follows own data management rules.	nil

1.2.4.3. UC2Sc2 Data – Disaster Recovery

1.2.4.3.1. Description

This Use Case scenario replicates a “disaster recovery” situation in which autonomous UAVs will scan and detect affected people on the ground and subsequently provide on-demand network connectivity to these affected people. These autonomous UAVs will be deployed using Unmanned Systems’ proprietary software platform i.e. Unmanned Life Central Command Platform (UL-CCP) that manages the command and control (C2) of each individual UAV as well as the management of the entire fleet. The UAVs being deployed will be installed with different payloads depending on the functionality required like a video camera for detection and a 5G modem to provide hotspot connectivity.

1.2.4.3.2. Data sources

Based on an analysis of the Use Case scenario, data will be collected at multiple sources such as:

1. 5G!Drones trial facilities infrastructure KPIs like RAM usage, CPU/GPU load, throughput on some interfaces, etc., RNIS data.
2. 5G KPIs collected at the gNodeB and core network.
3. 5G KPIs collected in the modem carried by UAV.
4. Data collected in the 5G!Drones Trial Controller.
5. Telemetry data from UAV and other data shared by UTM service.
6. Video footage taken by the UAV and transmitted to the MEC for analysis.
7. Logs from UAV, UTM, and UL-CCP post trial conclusion

1.2.4.3.3. Measurements Types

As part of measurement, UC2Sc2 aims to collect the following metrics:

1. Performance of the network and 5G infrastructure to support the use case scenario deployment.
2. KPIs like traffic statistics, flow statistics, radio level statistics, events and QoS measurements obtained from Eurecom’s facility.
3. Telemetry data like GPS coordinates, flight height, and battery status.

1.2.4.3.4. Data accessibility and analysis

The goal of the project generally and Use Case scenario specifically is to collect data for analysis allowing the consortium to understand the need for a 5G network to deploy complex UAV specific cases. In the case of this scenario, a live video is planned to be streamed from the UAV to Unmanned Systems modules on the MEC infrastructure testing the bandwidth capabilities of the 5G network as well as the low-latency required to control and manage a swarm of autonomous UAVs. This scenario is being led by Unmanned Systems (UMS) with the support of partners like Eurecom (EUR), Hepta Airborne (HEP) and Frequentis (FRQ). Analysis of the results of this experiment can be used by entities within and outside the project to further research and develop the use of 5G networks in combination with UAVs.

Depending on the type of data requested from this use case scenario, it will be made available outside the project based on appropriate agreements in place.

1.2.4.3.5. Tools used by UMS for Data Collection and Analysis

1. YOLOv3

General information: UMS uses YOLOv3 (You Only Look Once, Version 3), an open-source real-time object detection tool that identifies specific objects in videos, live feeds, or images as part of this scenario. The tool is run on the video footage captured by the UAV to detect humans in real-time, the analysis of which triggers the second UAV to fly to the location and provide network connectivity.

In its normal operations, UMS does not store the video footage streamed from the UAVs. It simply uses it to detect humans following which the video is discarded. However, if required the video footage can be stored using an open-source tool like VLC.

Table 32: YOLOv3 Data Set

Function	Purpose	Objectives
Purpose of the data	Video footage streamed from the UAV is used for detection of humans on the ground	Detect human beings
Type and format of Data	.avi	
Data set (if stored) is:	Growing	
Quantity (if stored)	100s MB to 10s of GB per trial.	
Data Security & Storage		Accessible only with user credentials and strong authentication / VPN.
Data value (long term)	Infinite	Infinte
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With labels	nil
Clear versioning approach	Timestamping and test name	nil
Standards or procedures for	No standards used.	nil

metadata creation applied		
Making data openly accessible		
Data kept closed	The video footage data will be kept closed.	Following GDPR laws
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	Standard Ubuntu naming convention	nil
Data licensing for wide reuse	N.A.	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Data will not be available for re-use	nil
Data usability by Third Parties (after the end of the project)	N.A.	nil
Restrictions to data re-use	Data will not be available for re-use	nil
Quality assurance process	Internal project quality insurance and testing methods	nil
Length of time of data re-usability	N.A.	nil
Allocation of resources		
Costs estimates for making data FAIR	N.A.	nil
Data Management Responsibilities	If video footage is stored then a person will be assigned responsible for data management.	nil

2. Ros2Bag

Description of Tool: Ros2 bag is a command line tool for recording data published on topics in the system. It accumulates the data passed on any number of topics and saves it in a database. The data can be replayed to reproduce the results of the tests and experiments.

Table 33: Ros2Bag Data Set

Function	Purpose	Objectives
Purpose of the data	This data provides telemetry of UAV in-flight like GPS coordinates, flight height and battery status.	Real-time data of in-flight UAVs for UTM services which can also be collected for post-flight processing
Type and format of Data	.bag	
Data set is:	Growing	
Quantity	1 Mb to 10s of Gbs	
Data Security & Storage		Accessible only with user credentials and strong authentication / VPN.
Data value (long term)	Infinite	Infinte
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With labels	nil
Clear versioning approach	Timestamping and test name	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	The dataset containing UAV telemetry can be available publicly upon request	This data could have high research value for the larger community.
How data will be made available	User credentials to the database.	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a	nil

	readme.txt file describing the details required accessing the data.	
Repository for deposit of data, metadata, and documentation	The data will be stored in a private storage.	nil
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is accessible with open-source software. Good interoperability.	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	Standard Ubuntu naming convention	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	No embargo	nil
Data usability by Third Parties (after the end of the project)	UAV telemetry data and documentation will be available after the project will be finished	nil
Restrictions to data re-use	Following an agreement with UMS	nil
Quality assurance process	Internal project quality insurance and testing methods	nil
Length of time of data re-usability	1 year after project is finished	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs.
Data Management Responsibilities	If video footage is stored then a person will be assigned responsible for data management.	nil

1.2.4.4. UC3Sc1 Sub-Scenario 2 Data

1.2.4.4.1. Description of data collection tools

Promtail [24] **+ Loki** [25] **–** Will be installed on the on-board computer (OBC) of the UAV. Promtail will grab and send generic OS logs from Ubuntu 18.04 OS's systemd journal to Loki installed on an AWS EC2 instance controlled by Hepta. The logs are stored in Loki as chunks. This data can be used to diagnose possible problems with software running on the OBC. The journal is a component of systemd. It's a centralized location for all messages logged by different components in a Linux system. This includes kernel and boot messages, messages coming from syslog and different services. The logging classification corresponds to Syslog protocol RFC 5424. [23]

Telegraf [26] **+ InfluxDB** [27] **–** Will be installed on the OBC. Telegraf will read and send OBC metrics to InfluxDB installed on an AWS EC2 instance controlled by Hepta. Telefraf will send metrics outputs in InfluxDB Line protocol format to InfluxDB. This data will be used to assess OBC workloads in different stages of flight to determine any issues with resource utilisation and availability.

Ardupilot [28] **–** The autopilot firmware Ardupilot will store various UAV metrics to the SD card of the autopilot board as .bin files. This data contains readings from a multitude of sensors used for diagnosing issues and controlling the flight of the UAV. A subset of this data is forwarded to the ground control station computer with telemetry stream. There, it is stored as .tlog files, .rlog files, .csv files and .kml files.

DJI M600Pro UAV – DJI M600Pro will store telemetry into a binary .dat file on its internal memory during its flight. This data includes various metrics from its on-board sensors, similar to the .bin files created by Ardupilot.

Photos – A photo camera will be taking images of the mock-up power line test track during the use case 3 scenario 1 sub-scenario 2 trial. These images will be saved as jpg files onto the SD card of the camera and to the OBC's hard drive. The images will contain geographic location data, a timestamp and gimbal angle information in Exif.

Point cloud – The VLP-16 LiDAR sends data to the on-board computer through the Ethernet port as User Datagram Protocol (UDP) packets. The data itself consists of firing sequence, Laser ID, measured distances, azimuth angle, reflectivity of an object, model of sensor, laser return mode and timestamp. Further on, the measurement data is being converted from Spherical to Cartesian coordinates, by utilising ROS. The resulting point cloud is stored as a ROS message of type sensor_msgs/PointCloud2.

Table 34: Promtail - Loki Data Set

Function	Purpose	Objectives
Purpose of the data	Debugging UAV on-board computer performance.	Find and diagnose OS and application-level issues
Type and format of Data	Form	Format
Text	Log records as strings	Loki chunks
Numeric	indexing	indexes

Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data is captured in real-time during the trials. Populating the dataset starts with launching the Promtail service on startup after the network connection is established on the on-board computer. Data capture ends after the Promtail service has been stopped by the user or the OBC is shut down.	nil
Data set is:	Revisable	nil
Quantity	In a test period of 15 minutes, about 7 MB of data was observed to be stored.	nil
Data Security & Storage	The recorded logs are stored in a Loki log aggregation system that is running on an AWS EC2 instance.	Accessible only with user credentials and strong authentication.
Data value (long term)	The data can be used to compare the previous results with subsequent ones. The valuable only to Hepta.	Data may have value for up to one year.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With labels	nil
Clear versioning approach	Timestamping	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	The entire dataset is private data belonging to Hepta. Data is shared among the 5G!Drones partners based on the Consortium Agreement.	This data has low research value for the larger community and is specific to Hepta's OBC-s setup.

Data kept closed	Data access is restricted to consortium members.	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	User credentials to the database.	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	nil
Repository for deposit of data, metadata, documentation and code	The data will be stored in a private storage.	nil
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is accessible with open-source software. Good interoperability.	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	Standard Ubuntu naming convention.	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Embargo on commercial use.	One year
Data usability by Third Parties (after the end of the project)	Not available for third parties. The data will be useless after a certain period for practical purposes and a retention policy will be set to save storage space.	nil
Restrictions to data re-use	Reuse restricted	The data cannot be disclosed due to sensitivity of being close to product line development.

Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. Appointed responsible postholder for data management.	nil

Table 35: Telegraf + InfluxDB Data Set

Function	Purpose	Objectives
Purpose of the data	Debugging UAV on-board computer performance.	Find and diagnose OS and application-level issues
Type and format of Data	Form	Format
Text	Tags and fields	string
Numeric	Timestamps and measured metrics	numeric
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data is captured in real-time during the trials. Populating the dataset starts with launching the Telegraf service on startup after the network connection is established on the on-board computer. Data capture ends after the Telegraf service has been stopped by the user or the OBC is shut down.	nil
Data set is:	Revisable	nil
Quantity	In a test period of 15 minutes, about 4 MB of data was observed to be stored.	nil

Data Security & Storage	The recorded logs are stored in an InfluxDB database that is running on an AWS EC2 instance.	Accessible only with user credentials and strong authentication.
Data value (long term)	The data can be used to compare the previous results with subsequent ones. Data is valuable to Hepta only.	Data may have value for up to one year.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With tags, timestamps and field names	nil
Clear versioning approach	Timestamping	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	The entire dataset is private data. Data is shared among the 5G!Drones partners based on the Consortium Agreement. Data is owned by Hepta.	This data has low research value for the larger community and is specific to Hepta's OBC-s setup.
Data kept closed	The entire dataset described in this table.	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	User credentials to the database.	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	nil
Repository for deposit of data, metadata,	The data will be stored in a private storage.	nil

documentation and code		
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is accessible with open-source software. Good interoperability.	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	nil	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Embargo on commercial use.	One year
Data usability by Third Parties (after the end of the project)	Not available for third parties. Due to the data having low value already a short time after the trial, data retention policy will be set and data deleted after this period to save storage space.	nil
Restrictions to data re-use	No reuse foreseen.	The data cannot be disclosed due to sensitivity of being close to product line development.
Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. Appointed responsible postholder for data management.	nil

Table 36: Ardupilot Data Set

Function	Purpose	Objectives
Purpose of the data	Debugging UAV sub-systems, real time monitoring of the UAV	Have feedback from the UAV and know the current state of its systems; understand the causes of possible failures.
Type and format of Data	Form	Format
Text	Status messages, field names	Human readable .csv, .kml files
Numeric	Timestamps and various sensor and actuator readings	.bin files, .tlog files, .rlog files
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data is captured in real-time during the trials	nil
Data set is:	Growing	nil
Quantity	In a test period of 15 minutes, about 7 MB of data was observed to be stored.	nil
Data Security & Storage	The recorded logs are stored on the SD card attached to the flight controller and to the computer the GCS is installed on.	Accessible only with user credentials and strong authentication.
Data value (long term)	The data can be used to debug issues with the UAV systems, replay the flights and log operational time. Data is valuable to Hepta.	Data may have value for up to five years.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With file naming and timestamps	nil
Clear versioning approach	Timestamping, file naming	nil

Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	The dataset comprises of mostly private data. Data is shared among the 5G!Drones partners based on the Consortium Agreement. Data is owned by Hepta.	This data has low research value for the larger community and is specific to Hepta's UAV setup.
Data kept closed	Data related to UAV systems that is close to product line development.	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	By sending the files upon request.	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	nil
Repository for deposit of data, metadata, documentation and code	The data will be stored in private storage.	nil
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is accessible with open-source software. Good interoperability.	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	nil	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		

Timing of data availability for re-use (incl. indications on embargo).	Embargo on commercial use.	Five years
Data usability by Third Parties (after the end of the project)	Not available for third parties	nil
Restrictions to data re-use	Confidentiality agreement	The data cannot be disclosed due to sensitivity of being close to product line development.
Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. Appointed responsible postholder for data management.	nil

Table 37: DJI M600 Pro UAV Data Set

Function	Purpose	Objectives
Purpose of the data	Debugging UAV sub-systems, real time monitoring of the UAV	Have feedback from the UAV and know the current state of its systems; understand the causes of possible failures.
Type and format of Data	Form	Format
Text	Status messages, field names.	inside .dat files
Numeric	Timestamps and various sensor and actuator readings	inside .dat files
Reused-data (rd)	No	Data will be generated in the trials.

Observational	Data is captured in real-time during the trials.	nil
Data set is:	Growing	nil
Quantity	In a test period of 15 minutes, about 7 MB of data was observed to be stored.	nil
Data Security & Storage	The recorded logs are stored on an SD card on-board the UAV. After the flight, the .dat file is uploaded to a cloud based flight logging system and distributed with 5G!Drones partners if requested.	Accessible only with user credentials and strong authentication.
Data value (long term)	The data can be used to debug issues with the UAV systems, replay the flights and log operational time.	Data may have value for up to five years.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With file naming and timestamps	nil
Clear versioning approach	Timestamping, file naming	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	This dataset is shared among the 5G!Drones partners based on the Consortium Agreement. Data is owned by Hepta.	This data has low research value for the larger community as it is a generic 3rd party UAV log.
Data kept closed	Data related to UAV systems that is close to product line development.	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
	Academic and research institute partners	Data exposing infrastructure based on background commitments.

How data will be made available	By sending the files upon request.	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	nil
Repository for deposit of data, metadata, documentation and code	The data will be stored in private storage.	nil
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is accessible with open-source software. Good interoperability.	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	nil	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Embargo on commercial use.	One year
Data usability by Third Parties (after the end of the project)	Not available for third parties	nil
Restrictions to data re-use	Confidentiality agreement	nil
Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the

		project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. Appointed responsible postholder for data management.	nil

Table 38: Photos Data Set

Function	Purpose	Objectives
Purpose of the data	Testing streaming the photos over 5G connection, validating integration with Hepta's cloud-based infrastructure inspection platform uBird and validating the performance of automatic fault detection with trained ML models.	Be able to stream photos quicker than to take them on average, have seamless integration with uBird and auto-detect faults.
Type and format of Data	Form	Format
Audiovisual	Images	Geotagged jpg image files with additional gimbal angle information and timestamps attached.
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data is captured in real-time during the trials. A mock-up course will be set up to be photographed.	nil
Data set is:	Growing	nil
Quantity	In a test period of 15 minutes, about 9 GB of data was observed to be generated.	nil
Data Security & Storage	The images are stored on the camera's SD card and streamed to Hepta's cloud storage.	Accessible only with user credentials and strong authentication.
Data value (long term)	For comparing the detector model's performance compared to subsequent images taken from the same test track and to newer iterations of the detector model.	The dataset may have value for up to 1 year.
FAIR data		

Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With file names, timestamps and geotags	nil
Clear versioning approach	File naming, timestamping	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	The entire dataset is private data. Data is shared among the 5G!Drones partners based on the Consortium Agreement. Data is owned by Hepta.	This data has low research value for the larger community and is specific to Hepta's product setup.
Data kept closed	The entire dataset described in this table is restricted to access except for Consortium partners.	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	Sent upon request by partners.	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	nil
Repository for deposit of data, metadata, documentation and code	The data will be stored in a private storage.	nil
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is interoperable with open-source software. Good interoperability.	nil
Making data interoperable		

Standard vocabulary or mapping to commonly used ontologies	nil	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Embargo on commercial use.	One year
Data usability by Third Parties (after the end of the project)	Not available for third parties	nil
Restrictions to data re-use	Confidentiality agreement	The data cannot be disclosed due to sensitivity of being close to product line development.
Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. Appointed responsible postholder for data management.	nil

Table 39: Point Cloud Data Set

Function	Purpose	Objectives
Purpose of the data	Testing streaming the point cloud over 5G connection, validating further processing of the point cloud, visualising the scanned area in 3D.	Be able to stream the point cloud, visualise it smoothly and further process it.

Type and format of Data	Form	Format
Instrument specific	Velodyne VLP-16 measurement points imported to ROS. Contained data: Distance, azimuth angle, reflectivity, firing sequence, laser ID, laser return mode, timestamp, coordinates	Reflection points measured by a laser scanner with absolute coordinates attached. Streamed in ROS sensor_msgs/PointCloud2 message format.
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data is captured in real-time during the trials. A mock-up power line course will be set up to be scanned.	nil
Data set is:	Growing	nil
Quantity	In a test period of 15 minutes, about 1.9 GB of data was observed to be generated.	nil
Data Security & Storage	The point cloud is stored on the drone's OBC, then streamed and stored on the operator's computer.	Accessible only with user credentials and strong authentication.
Data value (long term)	For diagnosing issues with data capture, for testing further point cloud processing methods.	The dataset may have value for up to 1 year.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With file names, timestamps and geotags	nil
Clear versioning approach	File naming, timestamping	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		

Data openly available	The entire dataset is private data. Data is shared among the 5G!Drones partners based on the Consortium Agreement. Data is owned by Hepta.	This data has low research value for the larger community and is specific to Hepta's product setup.
Data kept closed	The entire dataset described in this table.	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
	Academic and research institute partners.	Data exposing infrastructure based on background commitments.
How data will be made available	Sent upon request by partners.	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	nil
Repository for deposit of data, metadata, documentation and code	The data will be stored in a private storage.	nil
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is interoperable with open-source software. Good interoperability.	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	nil	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Embargo on commercial use.	One year
Data usability by Third Parties (after the end of the project)	Not available for third parties	nil

Restrictions to data re-use	Confidentiality agreement	The data cannot be disclosed due to sensitivity of being close to product line development.
Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. Appointed responsible postholder for data management.	nil

1.2.4.5. UC3Sc3 Data

Nokia Drone
Nokia Drone uses PX4 autopilot firmware [33], which stores multiple different UAV metrics to the SD card as .bin files. The data includes readings from various sensors which are critical for the control of the drone. The data is forwarded to the ground control station with telemetry stream over LTE. After the flight is completed, this data can be used to diagnose drone flight or replay the flight in 3D visualisation.

Nemo Outdoor
Nemo Outdoor is software from Keysight [34], which is used to log 5G parameters using Qualcomm and Samsung UEs. Software runs on computer stick, which is installed in a payload that is connected to the drone. Payload is designed for the Nokia drone and if another drone is used then adapter is needed. Drone provides power and ethernet to the payload and thus remote connection is to the payload is possible. Measurement can be started stopped during the measurement flight. UE is connected to the payload computer via USB cable. Nemo Outdoor is used to collect 5G NR field measurements and quality-of-experience (QoE) metrics. Data is encrypted and Nemo analyser [35] is needed to decrypt the data. Output format can be for example .csv file which is easy to use later for different applications. Nemo analyser also supports different customizable data visualizations for human readable presentation of the data.

1.2.4.5.1. Data Management Plan for UC3SC3 Data Sets

UC3SC3 short description: The purpose of this scenario is to demonstrate how UAVs could obtain position information in situation where Global Navigation Satellite System (GNSS) is compromised. Example of this kind of situation could be flying indoors or nearby to high buildings.

Data from Nemo Outdoor is collected to train a neural network, as described in Section 1.2.5.4. Main parameters used as the input for neural network are SSB index, RSRP and RTK GPS position (longitude, latitude). SSB index and RSRP values are obtained from the UE and GPS position is obtained from Mavlink which means from the drone.

Table 40: Nokia Drone Data Set

Function	Purpose	Objectives
Purpose of the data	Real time monitoring of the UAV. Debugging UAV sub-systems.	Monitor current state of the drone and prevent possible failures. If failure happens, understand the causes that led to the failure
Text	Status messages, field names	Human readable .csv, .kml files
Numeric	Timestamps and various sensor and actuator readings	.bin files, .tlog files, .rlog files
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data is captured in real-time during the trials	nil
Data set is:	Growing	nil
Quantity	In a test period of 25 minutes, about 15 MB of data was observed to be stored.	nil
Data Security & Storage	The recorded logs are stored on the SD card attached to the flight controller and to the computer the GCS is installed on.	Accessible only with user credentials and strong authentication.
Data value (long term)	The data can be used to debug issues with the UAV systems, replay the flights and log operational time. Data is valuable to drone owner.	Data may have value for up to 2 years.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With file naming and timestamps	nil

Clear versioning approach	Timestamping, file naming	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	The dataset comprises of mostly private data. Data is shared among the 5G!Drones partners based on the Consortium Agreement. Data is owned by NOK	This data has low research value for the larger community and is specific to NOK's UAV setup.
Data kept closed	Data related to UAV systems that is close to product line development.	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	By sending the files upon request.	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	nil
Repository for deposit of data, metadata, documentation and code	The data will be stored in private storage.	nil
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is accessible with open-source software. Good interoperability.	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	nil	nil

Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	nil	nil
Data usability by Third Parties (after the end of the project)	Not available for third parties	nil
Restrictions to data re-use	Confidentiality agreement	The data cannot be disclosed due to sensitivity of being close to product line development.
Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. Appointed responsible postholder for data management.	nil

1.2.5. Artificial Intelligence / Machine Learning Data

1.2.5.1. OpenMobile data set used by THA

The openmobiledata_public is a public data set, result of data collected by an application: MobiPerf. This Measurement Lab's (or M-Lab) open-source application aims to provide network performance measurements on mobile platforms. Measurements can be made directly by users but MobiPerf also performs measurements regularly in the background.

Measurements are recorded since the fourth of June 2012 and according to the latest M-Lab communication data is collected over 612 active users of the MobiPerf application with 6GB transmitted measurement data per day (within 800 requests per day (maximum3000)).

How to access it? Anonymised data is stored in Google Storage. The most convenient way of downloading is to use the gsutil tool and use the command: ./gsutil cp -r gs://openmobiledata_public .

About the measurements: Their type may be "ping", "traceroute", "dns", or "http". A description of the measurements - and most globally the features available in the data set - with respect to the official documentation can be found here: **https://github.com/Mobiperf/MobiPerf**

Ping requests:

- ping_timeout_sec - the number of seconds we wait for a ping response.
- target_ip - the IP address of the target we ping against
- mean_rtt_ms - Mean RTT in milliseconds.
- min_rtt_ms - Min RTT in milliseconds
- max_rtt_ms - Max RTT in milliseconds
- stddev_rtt_ms - Standard deviation of RTT in milliseconds
- filtered_mean_rrt_ms - Mean RRT with outlier values filtered (milliseconds).
- packet_loss - Fraction of lost packets

Traceroute requests:

- the hostname or IP address to use as the target of the traceroute.
- packet_size_byte - the packet per ICMP ping in the unit of bytes.
- max_hop_count - the total number of hops we ping before we declare the traceroute fails. Default to 10.
- num_hops - Number of observed hops in the traceroute
- hop_N_addr_i - The ith IP address of the Nth hop along the observed route, where N ranges from 0 to num_hops-1.
- hop_N_rtt_ms - Observed RTT in milliseconds to this hop.

DNS lookup requests:

- target - Hostname of the target to resolve
- server - IP address of a DNS server to use as the resolver. If not present, the device's default resolver is used.
- address - IPv4 address of the target as returned by an A record
- real_hostname - True FQDN of the host that has been resolved
- time_ms - Time taken to perform the DNS lookup

HTTP requests:

- headers - String (possibly containing newlines) with additional headers to send with the request. Each header and value pair is in the form of "headerParam:value", with different pairs separated by "\r\n".
- body - String with the request body to send (if method is "POST") time_ms - Time in milliseconds to perform the complete request
- headers_len - Size in bytes of the original response headers
- body_len - Size in bytes of the original response body
- headers - Response headers - may be compressed, truncated, or elided in the case of a large response

- body - Response body - may be compressed, truncated, or elided in the case of a large response. It is a JSON encoded bytearray.

TCP Throughput requests:

- dir_up - Uplink or Downlink measurement (boolean)
- data_limit_mb_up - Uplink cellular network data limit (double)
- data_limit_mb_down - Downlink cellular network data limit (double)
- duration_period_sec - Downlink maximum experiment duration period (double)
- pkt_size_up_bytes - The size each packet in the uplink (int)
- sample_period_sec - The small interval to calculate current throughput result (double)
- slow_start_period_sec - Waiting period to avoid TCP slow start (double)
- tcp_timeout_sec - TCP connection timeout (double)
- tcp_speed_results - A list of throughput sampling results in kbps (list of double)
- data_limit_exceeded - A flag indicating transmitted data exceeding limit (boolean)
- duration

1.2.5.2. Computer vision training data sets used by CAF

Computer Vision training datasets. CAF collects data for Computer Vision video analysis application development. In order to train a Computer Vision application to find a specific object, it is necessary to provide a sufficient number of photos of some objects to develop artificial neural networks. Usually, freely available dedicated data environments are used for training by companies developing computer vision application. This data can be used in these environments and the ownership of the data does not pass to CAF. Some publicly available data sets for training Computer Vision applications:

- 15 Drone Datasets and Satellite Image Databases for Machine Learning **https://lionbridge.ai/datasets/15-best-aerial-image-datasets-for-machine-learning/**
- Stanford Drone Dataset **https://cvgl.stanford.edu/projects/uav_data/**
- Okutama-Action: An Aerial View Video Dataset for Concurrent Human Action Detection **http://okutama-action.org/** , also available in GitHub

Table 41: Computer Vision Training Data Sets

Function	Purpose	Objectives
Purpose of the data	Training Computer Vision application	To train with real data
Type and format of Data	.jpeg, .png photos and videos	To use standard formats
Text	Nil	Nil
Numeric	Timestamps	nil
Reused-data (rd)	Yes. Publicly available data sets for training Computer Vision applications:	nil

Observational	Data is collected or used from publicly available portals.	nil
Data set is:	Growing	nil
Quantity	Ca 200 GB.	nil
Data Security & Storage	Most of publicly available data sets for training Computer Vision applications stay inside training environments. Some datasets are downloadable.	nil
Data value (long term)	The training data sets are changing and growing.	nil
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data sets are available for all via Internet. Some data sets are chargeable.	nil
Identifiability of data	With file naming and timestamps	nil
Clear versioning approach	Timestamping, file naming	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	These data sets are not owned by nor by the project consortium. These data sets publication depends on dataset owner decision.	nil
Data kept closed	Some datasets are available by fee.	nil
	Some datasets are publicly accessible.	nil
How data will be made available	It is decision from data owners.	nil
SW documentation and other information needed	Mostly publicly available in the Internet.	nil
Repository for deposit of data, metadata,	Dedicated platforms by different service providers.	nil

documentation and code		
Access restrictions	nil	nil
Data interoperability assessment	nil	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	CAF cannot change data interoperability or access rules.	nil
Data licensing for wide reuse	CAF cannot change data interoperability or access rules.	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	CAF cannot change data interoperability or access rules.	One year
Data usability by Third Parties (after the end of the project)	Third parties can ask access from data sets owners.	nil
Restrictions to data re-use	nil	nil
Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of hundreds to a few thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs.
Data Management Responsibilities	CAFA follows own data management rules.	nil

1.2.5.3. Machine learning data sets used by HEP for UC3Sc1 Sub-Scenario 2

Table 42: HEP Machine Learning Data Set

Function	Purpose	Objectives
Purpose of the data	Training machine learning models	To enable automatic detection of infrastructure objects and faults from photos
Type and format of Data	Form	Format
Text	Annotation data. Text files containing image reference, bounding box coordinates and classes	.csv
Audiovisual	Image files of power line elements	.jpg
Reused-data (rd)	Yes	Images have been captured by HEP before the trials and outside the scope of the project.
Reused-data (rd)	No	Annotations
Observational	Images	nil
Derived/Compiled	Annotations	nil
Data set is:	Revisable	nil
Quantity	The size of the data set is ca 72 GB	nil
Data Security & Storage	Data is stored in a private cloud storage.	Accessible only with user credentials and strong authentication.
Data value (long term)	The data is useful for HEP in long term for training better ML models.	The dataset may have value for up to 20 years.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	The data is strictly related to HEP led use case scenario. No discoverability measures applied.	nil
Identifiability of data	With file names	nil
Clear versioning approach	nil	nil
Standards or procedures for	No standards used.	nil

metadata creation applied		
Making data openly accessible		
Data openly available	The entire dataset is private data. Data will be shared among the 5G!Drones partners if requested and dictated by the Consortium Agreement. Data is owned by HEP.	This data is specific to HEP product development.
Data kept closed	The entire dataset described in this table.	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
	Academic and research institute partners.	Data exposing infrastructure based on background commitments.
How data will be made available	nil	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	nil
Repository for deposit of data, metadata, documentation and code	The data will be stored in a private storage.	nil
Access restrictions	The data is kept internal to 5G!Drones.	nil
Data interoperability assessment	The data is stored in a form that is interoperable with open-source software. Good interoperability.	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	nil	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	nil	nil

Data usability by Third Parties (after the end of the project)	Not available for third parties	nil
Restrictions to data re-use	Confidentiality agreement	The data cannot be disclosed due to sensitivity of being close to product line development.
Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	Costs are expected to be in the range of thousands of Euros per annum.	During project lifetime, costs can be covered from Equipment costs. After the project, Partners individually bear the costs.
Data Management Responsibilities	Each Partner has their own data management policies. Appointed responsible postholder for data management.	nil

1.2.5.4. Machine learning data set used by NOK for UC3Sc3

Table 43: Nemo Outdoor Data Set used by NOK for UC3Sc3

Function	Purpose	Objectives
Purpose of the data	Provide alternative position information in situation when GNSS is compromised	Train neural network which will give GPS position as an output. Alternative objective is to collect as much alternative data as possible for future researches as a dataset
Text	Status messages	.csv files
Numeric	Longitude and latitude	inside .csv files
Reused-data (rd)	No	Data will be generated in the trials.
Observational	Data is captured in real-time during the trials.	nil
Data set is:	Growing	nil

Quantity	nil	nil
Data Security & Storage	The recorded logs are stored on computer stick on-board the UAV. After the flight, the .csv file exported for processing	Accessible only with user credentials and strong authentication.
Data value (long term)	The data can be used to train and test different neural network algorithms	Data may have value for up to five years.
FAIR data		
Making data findable		
Discoverability of data (metadata provision)	Data will be made available to consortium partners based on the consortium agreement.	nil
Identifiability of data	With file naming and timestamps	nil
Clear versioning approach	Timestamping, file naming	nil
Standards or procedures for metadata creation applied	No standards used.	nil
Making data openly accessible		
Data openly available	This dataset is shared among the 5G!Drones partners based on the Consortium Agreement. Data is owned by NOK.	This data has research value for the larger community as it can involve potentially many different interesting parameters for different use cases
Data kept closed	Data related to UAV systems that is close to product line development.	Data close to own product line development. Data from partners own products (e.g. log and telemetry files)
	Academic and research institute partners	Data exposing infrastructure based on background commitments.
How data will be made available	By sending the files upon request.	nil
SW documentation and other information needed	Any publicly made available data shall be accompanied by a readme.txt file describing the details required accessing the data.	nil

Repository for deposit of data, metadata, documentation and code	Nil	nil
Access restrictions	nil	nil
Data interoperability assessment	nil	nil
Making data interoperable		
Standard vocabulary or mapping to commonly used ontologies	nil	nil
Data licensing for wide reuse	nil	nil
Increase data re-use (through clarifying licenses)		
Timing of data availability for re-use (incl. indications on embargo).	Embargo on commercial use.	One year
Data usability by Third Parties (after the end of the project)	Could be available for third parties	nil
Restrictions to data re-use	Confidentiality agreement	nil
Quality assurance process	nil	nil
Length of time of data re-usability	nil	nil
Allocation of resources		
Costs estimates for making data FAIR	nil	nil
Data Management Responsibilities	nil	nil

1.3. Allocation of resources, data security, and data policy

The consortium will use the free-of-charge repositories, EUDAT [1], IDA [2], and Zenodo [11] for making the datasets that will be public accessible. On the other hands costs will incur, especially in personnel resources making the data available and maintaining the data that are live. An issue may arise as, for example, the live data in EUDAT are limited to 20 GB in size and six months usage restrictions. As several data sets are revisable, they need to be live data and therefore, IDA and Zenodo may be better alternatives. There are no such restrictions for frozen data. The project deliverables, which are public, will be published on 5G!Drones own web pages. The project deliverables, which are not public, will be stored in the project internal data repository platform, MS Teams, and submitted to the Commission through the EC Participant Portal. These actions are to ensure that data are safely stored and accessible to entities with access rights to utilize the data. EUDAT, IDA, and Zenodo function as repositories for long term preservation and curation if the project generates any data having more than mid-term value.

The handling of the EUDAT, IDA, and Zenodo repositories on behalf of 5G!Drones is the responsibility of the project Partners that generated the data. The coordinator's responsibility is to maintain track of the data management issues related to the project. The project Partners responsibility is to make the coordinator and the project aware that such FAIR data will be made available and to seek from the consortium.

As for the publications, where the analyses of the empirical research data will be presented, the consortium will publish them in scientific journals that allow various open access mechanisms. At minimum the authors shall use their institutions' data policy defined mechanisms for self-archiving of the final draft version of the publication. The costs related to open access will be claimed as part of the Horizon 2020 grant. For example, UO uses self-archiving in the university's open publication repository Jultika [17].

Members of the 5G!Drones consortium involved have their own data management policies. As examples, the UO responsible research data management policy can be found from [18]. In addition the UO Data Protection Officer is Niilo Vähäsarja, who can be contacted through dpo (at) oulu.fi. For AIR, they have Privacy Impact Assessment (PIA) conducted and data protection officer appointed for their Tactilon AGNET; a hybrid solution in Professional Mobile Radio industry that expands the critical communication perimeter with extended teams and new technical capabilities. It will be used in 5G!Drones.

1.4. Ethical aspects

As stated in the project plan: "No ethics issues have been entered in the ethical issues table in the administrative proposal forms." Therefore, in 5G!Drones information collected can be released without privacy restrictions because it does not constitute private information about identified human subjects.

The project will monitor if ethical or privacy implications arise and use a general strategy for the monitoring of the implications. Should ethical implications arise, the project shall follow the GDPR act [19] on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. The steps taken to monitor ethical and privacy implications are:

- Knowledge acquisition: This step includes the study of the needs of the trials of the project.
- Privacy impact assessment: a PIA will be conducted to study of all the use case trials in which, during the project lifecycle, personal data rights are identified to be at stake. Special attention will be paid to activities involving data collection from project external sources.

As evolution to this Data Management Plan, procedures shall be defined in order to collect and process data that have not been identified at this step. Such data includes items such as:

- legal basis,
- consent process,
- anonymization,
- storage,
- protection,
- retention,
- destruction,
- reuse,
- etc.

1.4.1. Informed Consent

For managing the Consortium and being able to disseminate the project information 5G!Drones Consortium has requested Partner and personnel consent on

- Taking photos of individuals during meetings, events, social dinners, etc.
- Publishing photos of individuals at 5G!Drones website, social media, newsletters, leaflets, posters, etc.
- Uploading photos at the Microsoft Teams repository.
- Mentioning individuals' names and organisations at 5G!Drones website, social media, newsletters, articles, etc.
- The project will not use personnel names unless agreed on beforehand.

When dealing with personnel external to the Consortium, similar request for informed consent shall be made prior to storage or publication of any personal identifiable information.

References

[1] https://eudat.eu/

[2] https://www.fairdata.fi/en/ida/

[3] https://www.fairdata.fi/en/

[4] Qosium Web Page. [Online] Available at: https://www.kaitotek.com/qosium. Last accessed: 15.04.2021.

[5] Qosium Documentation. Available at: https://www.kaitotek.com/resources/documentation. Last accessed: 15.04.2021.

[6] https://www.keysight.com/en/pd-2767485-pn-NTH00000A/nemo-handy?cc=FI&lc=fin

[7] https://www.keysight.com/fi/en/product/NTA00002B/nemo-outdoor-5g-nr-drive-test-solution.html

[8] https://www.fairdata.fi/en/qvain/qvain-user-guide/

[9] https://etsin.fairdata.fi/

[10] https://schema.datacite.org/meta/kernel-4.1/doc/DataCite-MetadataKernel_v4.1.pdf. Last accessed: 27.8.2019.

[11] https://zenodo.org/

[12] https://www.dublincore.org/resources/metadata-basics/

[13] https://elsst.ukdataservice.ac.uk/

[14] https://finto.fi/en/

[15] https://www.eventhelix.com/5G/non-standalone-access-en-dc/en-dc-secondary-node-addition-core-network-details.pdf . Last accessed: 27.8.2019.

[16] Sami Kekki, Walter Featherstone, Yonggang Fang, Pekka Kuure, Alice Li, Anurag Ranjan, Debashish Purkayastha, Feng Jiangping, Danny Frydman, Gianluca Verin, Kuo-Wei Wen, Kwihoon Kim, Rohit Arora, Andy Odgers, Luis M. Contreras, Salvatore Scarpina, “MEC in 5G networks”, ETSI White Paper No. 28, ISBN No. 979-10-92620-22-1, June 2018. [Online] Available at: https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp28_mec_in_5G_FINAL.pdf.

[17] http://jultika.oulu.fi/

[18] https://www.oulu.fi/university/node/43683

[19] European Union, “REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL,” 27 April 2016. [Online] Available at: https://publications.europa.eu/en/publication-detail/-/publication/3e485e15-11bd-11e6-ba9a-01aa75ed71a1/language-en. Last Accessed: 16.9.2019.

[20] prometheus, "prometheus web site," [Online]. Available: https://prometheus.io/. [Accessed 2021].

[21] Netradar, "Netradar web site," [Online]. Available: https://www.netradar.com/. [Accessed 2021].

[22] Ookla, "Ookla test speed webpage," [Online]. Available: https://www.speedtest.net/apps. [Accessed 2021].

[23] https://wiki.archlinux.org/title/Systemd/Journal

[24] https://grafana.com/docs/loki/latest/clients/promtail/

[25] https://grafana.com/oss/loki/

[26] https://www.influxdata.com/time-series-platform/telegraf/

[27] https://www.influxdata.com/

[28] https://ardupilot.org/

[29] UGCS telemetry data export manual https://wiki.ugcs.com/index.php?title=Category:Telemetry_player&action=pdfbook

[30] 15 Drone Datasets and Satellite Image Databases for Machine Learning https://lionbridge.ai/datasets/15-best-aerial-image-datasets-for-machine-learning/

[31] Stanford Drone Dataset https://cvgl.stanford.edu/projects/uav_data/

[32] Okutama-Action: An Aerial View Video Dataset for Concurrent Human Action Detection http://okutama-action.org/ , also available in GitHub.

[33] PX4 Open Source Autopilot. (2019). PX4 Open Source Autopilot. [online] Available at: https://px4.io/

[34] Keysight (2021). Nemo Outdoor 5G NR Drive Test Solution. [online] Keysight. Available at: https://www.keysight.com/fi/en/product/NTA00002B/nemo-outdoor-5g-nr-drive-test-solution.html [Accessed 2 Jun. 2021].

[35] Keysight (2021). Nemo Analyze Drive Test Post Processing Solution. [online] Keysight. Available at: https://www.keysight.com/fi/en/product/NTN00000B/nemo-analyze-drive-test-post-processing-solution.html [Accessed 2 Jun. 2021].