Research & Innovation Actions

5G PPP Research and Validation of critical technologies and systems: Enabling Smart Energy as a Service via 5G Mobile Network advances.

Project: H2020-ICT-762013

Enabling Smart Energy as a Service via 5G Mobile Network advances

Deliverable 4.1

Open Data Management Plan & Integration Guidelines V1

Author(s): RWTH, POPs

Status -Version: 1

Delivery Date (DOW): 30 November 2017

Actual Delivery Date: 29 December 2017

Distribution - Confidentiality: Public

Code: 10.5281/zenodo.1042548

Abstract:

This document is the first version of the NRG5 Data Management Plan (DMP) due to in project month 6. The document is a living document and will be updated in a second version in month 18.

This deliverable provides a description of the data management procedures to be followed in trial sites as well as across the overall project including procedures for data security and privacy. The expected data to be used and generated in the project is categorized by data sets, which are characterized by factsheets. The different aspects of making data FAIR are discussed.

In the second part of this document software integration infrastructure and techniques for software and hardware quality assurance are described. The focus is set on continuous integration and continuous delivery of software developed during the project.

Disclaimer

The NRG-5 Consortium is the following:

Participant number	Participant organisation name	Short name	Country
01	Engineering-Ingegneria Informatica SPA	ENG	Italy
02	THALES Communications & Security	TCS	France
03	SingularLogic S.A.	SiLO	Greece
04	Ineo Energy & Systems	ENGIE	France
05	Romgaz S.A	RGAZ	Romania
06	ASM Terni SpA	ASM	Italy
07	British Telecom	BT	UK
08	Wind Telecomunicazioni S.P.A.	WIND	Italy
09	Hispasat S.A.	HIS	Spain
10	Power Operations Limited	POPs	UK
11	Visiona Ingenieria De Proyectos SI	VIS	Spain
12	Optimum S.A	OPT	Greece
13	Emotion s.r.l	EMOT	Italy
14	Rheinisch-Westfälische Technische Hochschule Aachen	RWTH	Germany
15	Jožef Stefan Institute	JSI	Slovenia
16	TEI of Sterea Ellada/Electrical Engineering Dept.	TEISTE	Greece
17	University Pierre et Marie Curie	UPMC	France
18	Centro Romania Energy	CRE	Romania
19	Rutgers State University of New Jersey	Rutgers	USA

Document Revision History

Date	Issue	Author/Editor/Contributor	Summary of main changes
11/09/17	0.1	M. Pitz	Initial document creation
14/09/17	0.2	M. Pitz	Added DMP table of contents
09/10/17	0.3	M. Pitz	Added integration guidelines to structure
13/10/17	0.4	M. Pitz	Updated dataset structure and added datatypes
26/10/17	0.5	A. Voulkidis	Added content for the Integration Guidelines
06/11/17	0.6	M. Pitz	Added new contributions from partners
08/11/17	0.7	A. Voulkidis	Added content for the Integration Guidelines
10.11.17	0.8	M. Pitz	Pre final version of D4.1
10/11/17	0.10	J. Lalueza	Video, Image, positioning and planning data files information
28/11/17	1.0	M. Pitz	Minor changes for the final version

Table of Contents

List of Abbreviations

5G-PPP	5G Infrastructure Public Private Partnership
ATEX	ATmosphères EXplosibles EU directive 99/92/EC or 94/9/EC
CCTV	Closed Circuit Television
CD	Continuous Delivery
CERN	European Organization for Nuclear Research
CI	Continuous Integration
DC	Data Centre
DMP	Data Management Plan
DOI	Digital Object Identifier
DPO	Data Protection Officer
DR	Demand Response
EU	European Union
FIT	Future Internet Testing Facility
KPI	Key Performance Indicator
NFV	Network Function Virtualization
NORM	Next generation Open Real-time Meter
NS	Network Service
SDN	Software defined network
UC	Use Case
uMTC	Massive Machine Type Communication
uRRLC	Ultra Reliable and Low latency Communication
VNF	Virtual Network Function
VNFFP	Virtual Network Functions Forwarding Path
VNFFG	Virtual Network Functions Forwarding Graph
xMEC	extended Mobile Edge Cloud

1 Introduction

The handling of data by means of the different data types, restriction for privacy and openness is an important part of projects with multiple partners. If there is software, development involved in the project the project partners need to clarify how and according to which coding standards they intend to develop software. Both aspects are considered in this document. The first part will be the Data Management Plan (DMP). In the second part, the software aspect is discussed further.

The Data Management Plan is a formal description of the procedures of data handling during and after a project. By providing an assessment of data used in a project and a structured approach for aspects as naming conventions, metadata and versioning, the DMP should also support data quality, efficient processing and sharing of data. It is required for projects in the Horizon 2020 framework program that have not opted out of the Open Research Data Pilot [1].

The Integration Guidelines are a description on how the developers of the different partners are developing and handling the source code. In a modern development environment, continuous integration plays a major role to support the developer with automated code testing and deployment. To make the transfer and reusability easier for the partners but also for possible third-party users the partners have to agree on the integration guidelines on a formal level.

Naturally, not all questions regarding the handling and type of data and source code can be answered at the start of a project, as procedures have to be coordinated and implemented and the infrastructure has to be set up. To meet this lack of knowledge, this document is meant to be a living document, which should be adapted and completed during the project lifetime. The first version will be released in deliverable D4.1 “Open Data Management Plan & Integration Guidelines V1” in M06. An update will be provided in deliverable D4.2 “Open Data Management Plan & Integration Guidelines V2” in M18.

2 Data Management Plan

2.1 Introduction

Chapter 2 is structured according to Annex 1 of the “Guidelines on Fair data management in Horizon 2020” [1]. The data appearing in the NRG5 project is assessed in chapter 2.2 and structured in datasets. In chapter 2.3, the approach towards making data “FAIR” (Findable, Accessible, Interoperable, Re-usable) is described in regard to the tools, conventions and procedures to be used in NRG5. The allocation of resources for making the data “FAIR” is explained in chapter 2.4.

2.2 Data Summary

Within NRG5 there will be trials conducted by different project partners. One of the trials will test drone behaviour in a swam and offloading of video processing to the edge cloud. This trial will tag place at “Pilot#1: which will be a Natural Gas network/LNG Terminal provided by ENGIE” [2]. At the second trial site the integration of smarter energy and 5G will be tested. This trial will take place a “Pilot#2: Optimized Energy Network Management and Control (Italy)” [3] Besides the trials, there will be software development, testing and simulation done by different project partners. Due to the heterogeneity of the data, which is to be collected and generated in the NRG5 project, chapter 2.2 is structured by a differentiation by datasets. Structuring by datasets allows for a detailed assessment of the data collection and generation as well as issues of data privacy and security. To assess the data, each dataset is described in a fact sheet.

2.2.1 List of datasets

The following datasets have been identified to appear in the NRG5 project and will be described in more detail in the following subchapters. Because most of the tasks producing the data have not started yet, the information about the datasets it not necessarily complete. The structure of the table is divided into two parts. First, there is a generic description and an example for the dataset then the datasets are divided into datatypes. This should take into account the diversity of data generated during the project. In a datatype, there is a differentiation between the different partners generating the data. With that information, the DMP can identify the potential of data generated by a partner even if another partner is generating the same data from a different source. This terminology is flexible in adding datasets, datatypes or a new partner who is generating the same kind of data but with different content.

The generic information in the datasets is based on the Grant agreement. The content for the dataset is based either on the grant agreement or on information gathered from the project partners.

Dataset 1: Topology and asset description

- IT System topology data
- Terrain map files
- Charge point description
- FIT infrastructure

Dataset 2: Video recordings / CCTV

- Drone video material
- Drone recording still images

Dataset 3: Measurement data

- Voltmeter/Current meter/Custom recordings
- NORM data

Dataset 4: Log and access data

- Alarm and heartbeat Logs
- System Logs
- Network Traffic Logs

Dataset 5: Prediction, forecast and planning data

- Flight planning data
- Power exchange data
- Demand response data
- Energy or Power forecast of PV generation

Dataset 6: Positioning and location data

- GPS position data

Dataset 7: Models and code

- Source code
- Software Models

Dataset 8: Configuration and Orchestration data

- Configuration files
- Management data

2.2.2 Topology and asset description

Fact sheet Dataset 1			
Dataset name	Topology and asset description		
Dataset description	The topology and asset description includes plans and documentation about assets and equipment. The description of network topologies of electrical, gas and other energy distribution grids is included. In addition, the topologies of IT networks wired and non-wired are included. For the IT networks, detailed information about the hardware is part of this dataset.		
General security and privacy considerations			
Information about critical infrastructure may need to be handled confidentially so the security of that infrastructure cannot be compromised. Further assessments of the data is needed, before a decision about making it public and in which extent can be made.			
Datatype specific answers			
Datatype name	IT System topology data		
Data description	This data describes the IT infrastructure of a facility, including the hardware specifications.		
Purpose of the data	Information about the available resources and their performance is needed for dynamic NFV resource allocation.		
Relation to project objective	In the project, there will be an automated way of allocation computational resources.		
File types	.vsd		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Information about hardware infrastructure for CI/CD will be provided	Not yet known	All the partners have access to the data	The data can be made public because the datacentre infrastructure is not classified
Information about the used SDN infrastructure will be provided	Not yet known	Not yet known	Not yet known
Datatype name	Terrain map files		
Data description	This data describes facilities topology (e.g. substations)		
Purpose of the data	For automated drone flights, information about the facilities and the terrain is needed.		
Relation to project objective	Part of the project is to do maintenance drone flights at trial sites with autonomous drones.		
File types	.map, .jpg, .kml, .kmz		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
A ground floor map of Pilot#1 will be provided	MBs to 1 GB	Access will be provided, if needed for the project objectives.	The facility is critical infrastructure therefore the data cannot be made public

A map with information about industrial structures e.g. buildings, energy lines, pylons of Pilot#1.	MBs to 1 GB	Access will be provided, if needed for the project objectives.	The facility is critical infrastructure therefore the data cannot be made public
Information about forbidden zones, public or industrial restrictions (e.g. ATEX zones) at Pilot#1 will be provided.	MBs to 1 GB	Access will be provided, if needed for the project objectives.	The facility is critical infrastructure therefore the data cannot be made public
Datatype name	Charge point description		
Data description	To be specified		
Purpose of the data	Emotion goal is to optimize the electric power consumption of the EVs charging point using energy from renewable sources		
Relation to project objective	Test the 5G network through an electric mobility service		
File types	Not yet known		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Information about web service infrastructure at Pilot#2 will be provided.	Not yet known	Not yet known	Not yet known
Information about the charge point infrastructure at Pilot#2 will be provided.	Not yet known	Not yet known	Not yet known
A map of charging station at Pilot#2 will be provided.	Not yet known	Not yet known	Due to the sensitive nature of the data they will only be available on application and their use
Datatype name	FIT infrastructure		
Data description	This data describes the FIT infrastructure (Future Internet Testing Facility).		
Purpose of the data	For performance and correctness evaluation		
Relation to project objective	It will be used to validate the NRG-5 IoT requirements, namely M2M and MCM communications, xMEC efficiency, self* VNFs using constrained devices and moving IoT nodes and sensors.		
File types	.txt, .csv, .jpg		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
(UPMC) will provide information about their FIT infrastructure	Not yet known	All the partners have access to the data	UPMC promote the open access data policy.

2.2.3 Video recordings / CCTV

Fact sheet Dataset 2			
Dataset name	Video recordings / CCTV		
Dataset description	Video recordings of landscape in the visible spectrum.		
Security and privacy considerations			
Video and image data recorded by a drone can contain information, which has to be assessed from a privacy but also a security point of view. Further information can be found in chapter 2.5			
Datatype specific answers			
Datatype name	Drone video material		
Data description	Video material gathered during drone flights		
Purpose of the data	Video and still images are needed to test the implementation of drone path planning algorithms and to do the on fly path planning. To analyse the results of the flight trial the video material can be of benefit too.		
Relation to project objective	The offloading of image data from a drone, to do path planning on a cloud system is part of the NRG5 project. To do so video or image data needs to be acquired and then transferred.		
File types	.mpeg, .mkv, .avi		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
(VIS) Video recordings from a drone perspective will be made at Pilot#1	Minimum per 10min flight: ~600MB @ 720p ~1,5GB @ 1080p ~3,5GB @ 4K	All the partners have access to the data	Not yet known
(VIS) Video recordings of the drone from a ground perspective at Pilot#1	Minimum per 10min flight: ~600MB @ 720p ~1,5GB @ 1080p ~3,5GB @ 4K	All the partners have access to the data	Not yet known
Datatype name	Drone recording still images		
Data description	Still images extracted out of video recordings		
Purpose of the data	To analyse and document the flight behaviour, still images extracted out of video recordings are needed		
Relation to project objective	After flight, analysis will be part of the drone trial. To do so still images of the flight are useful.		
File types	.jpg, .png, .tif		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Still images from the	Minimum per	All the partners have	Not yet known

drone video recording	image: ~150KB @ 720p ~500KB @ 1080p ~2MB @ 4K	access to the data	
Still images of the drone from a ground perspective	Minimum per image: ~150KB @ 720p ~500KB @ 1080p ~2MB @ 4K	All the partners have access to the data	Not yet known

2.2.4 Measurement data

Fact sheet Dataset 3			
Dataset name	Measurement data		
Dataset description	Data gathered from sensors.		
Security and privacy considerations			
The combination of the collected information about driving and charging habits, but also the forecast information could have an impact on the privacy of the user. Before making this data public, this issue needs to be addressed.			
Datatype specific answers			
Datatype name	Voltmeter/Current meter/Custom recordings		
Data description	Voltmeter/Current meter/Custom recordings		
Purpose of the data	To control the charging behaviour of an electric vehicle (EV) it is important to know the current state of charge of the EV battery, but also the current state of the power grid. A forecast about the use of the EV and the needed energy, based on historical information can also use information about the driving behaviour.		
Relation to project objective	In Pilot#2, multiple EV's will be controlled in state of charge and the charging schedule. The schedule takes the current state of the power grid into account. Therefore, it is important to measure the grid state. To calculate a charging plan it is important to know a forecast of the user's behaviour. For that, information about the user's behaviour needs to be collected.		
File types	.csv, .xls, .raw, .json		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Measurements on EV's at Pilot#2 will be done.	Not yet known	All the partners have access to the data	Not yet known
Information about the energy consumption of the EV chargers will be measured.	Not yet known	Not yet known	Not yet known
The state of charge of EV's will be measured.	Not yet known	Not yet known	Not yet known
(UPMC) Will provide data gathered from use of the FIT platform.	Not yet known	All the partners have access to the data	UPMC promote the open access data policy.
Datatype name	NORM data		
Data description	Measurements from grid extracted from NORM		
Purpose of the data	To be specified		
Relation to project objective	To be specified		
File types	.csv, .xls		

(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Data gathered from NORM units will be used.	Not yet known	Not yet known	Due to the sensitive nature of the data they will only be available on application and their use

2.2.5 Log and access data

Fact sheet Dataset 4			
Dataset name	Log and access data		
Dataset description	Data which documents the current state or change in state of a system.		
Security and privacy considerations			
Alarm and logging data can reveal information about critical infrastructure or the behaviour and identity of user who are connected to the systems storing the alarm and logging data. Therefor this data needs to be assessed before a decision about making it public can be made.			
Datatype specific answers			
Datatype name	Alarm and heartbeat Logs		
Data description	To be specified		
Purpose of the data	Alarm and heartbeat data is needed for analysis of a systems behaviour		
Relation to project objective	Electric vehicles will be smartly connected together. To do so status information (e.g. alarms and heartbeats) about the EVs are needed.		
File types	.json		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Alarm data from EVs will be logged.	Not yet known	Not yet known	Not yet known
Heartbeat data from EVs will be logged.	Not yet known	Not yet known	Not yet known
Datatype name	System Logs		
Data description	Logs with a history of system states.		
Purpose of the data	Log data is needed for analysis of a systems behaviour		
Relation to project objective	The project will do multiple trials and lab tests. Part of the project is to create and further develop VNF's. When testing this VNF's log files will be created to analyse the behaviour of the tested function.		
File types	.csv, .json, .xml		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
System log data from EVs will be stored.	Not yet known	Not yet known	Not yet known
Log data from trials will be gathered	Very large GBs	Not yet known	Not available
(VIS) Log data from SDNs will be used	Some MBs	Not yet known	Not yet known
(VIS) Access data from SDNs will be used	Some MBs	Not yet known	Not yet known

Datatype name	Network Traffic Logs		
Data description	Logs of Network traffic including packet raw data		
Purpose of the data	To offline analyse of generated or simulated network traffic a traffic recording is needed		
Relation to project objective	To tune/validate the behaviour of orchestration strategies before actual deployment		
File types	.pcap		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Network traffic will be recorded in trials and lab tests.	In hundreds of MB per trace	Not yet known	Not yet known

2.2.6 Prediction, forecast and planning data

Fact sheet Dataset 5			
Dataset name	Prediction, forecast and planning data		
Dataset description	Data to plan the path of an object. Also included is data with forecast or schedule character.		
Security and privacy considerations			
To be evaluated			
Datatype specific answers			
Datatype name	Flight planning data		
Data description	To be specified		
Purpose of the data	After flight, analysis is an essential part of evaluating the flight path planning algorithm. To do so the actual flight data and path planning data needs to be recorded.		
Relation to project objective	Testing the behaviour of a drone swarm with automated flight path planning is one part of the project. To evaluate the path planning and the 5G performance information about the path planning is important.		
File types	.kmz, .kml, .gpx		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
(VIS) Planning data from drone flights will be stored.	Some MBs	Not yet known	Not yet known
Datatype name	Power exchange data		
Data description	Power exchange within the charge point depending on the electrical output		
Purpose of the data	To manage the power flog in an electrical grid information about the current state is needed		
Relation to project objective	Test the 5G network through an electric mobility service		
File types	Not yet known		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Information about the power exchange within the charge point will be recorded.	Not yet known	Not yet known	Not yet known
Datatype name	Demand response data		

Data description	Demand response (DR) signals and available resources for DR		
Purpose of the data	These data will be both generated and then collected to train the entire system		
Relation to project objective	These data are needed for the calculation of the flexibility needed for the DR		
File types	.csv		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
DR data from trials and lab tests will be stored.	Not yet known	Not yet known	From public data about DR is it possible to infer behaviour of prosumer. Only aggregated data in a certain level could be public
Datatype name	Energy or Power forecast of PV generation		
Data description	Energy or Power forecast of PV generation		
Purpose of the data	Electrical vehicle charging plans /vehicle charging demand profile		
Relation to project objective	These kind of information it is the trigger of UC2		
File types	.csv		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
EVs charging plans will be stored.	Not yet known	Not yet known	Due to the sensitive nature of the data they will only be available on application and their use
EVs charging demand profiles will be stored.	Not yet known	Not yet known	Not yet known

2.2.7 Positioning and location data

Fact sheet Dataset 6			
Dataset name	Positioning and location data		
Dataset description	To be specified		
Security and privacy considerations			
To be evaluated			
Datatype specific answers			
Datatype name	GPS position data		
Data description	To be specified		
Purpose of the data	Information about the position of facilities or recorded positions of object movement are needed to navigate		
Relation to project objective	To forecast the EV usage information historic information about the car movement is important. In the drone trials there is a need to know the positions of facilities to monitor with the drones, but also the information about the current drone position is essential for the path planning		
File types	.json, .csv, .xlsx, .kmz, .gpx		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
EVs geo localization data will be recorded.	Not yet known	Not yet known	Not yet known
(VIS) Positioning data for drones will be recorded.	Some MBs	Not yet known	Not yet known

2.2.8 Models and code

Fact sheet Dataset 7			
Dataset name	Models and code		
Dataset description	Behaviour or software models and software source code including build instruction		
Security and privacy considerations			
To be evaluated			
Datatype specific answers			
Datatype name	Source code		
Data description	Source code		
Purpose of the data	To further develop models but also software, developed during a project, it is important to have the source files, but also instructions on how to build those files to generate a model or a binary.		
Relation to project objective	In the NRG5 project, we will develop new Virtual Network Functions (VNF). Part of that process is to develop further existing source code or to implement new functions. The project also aims to make as much information public and open source as possible. To do so the developed source code needs to be stored.		
File types	.c, .cpp, .h, .py, .java, .html		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Source code will be stored. During development.	Less than 1 GB	Access will be provided, if needed for the project objectives.	Probably most of it could be made available as a platform
Datatype name	Software Models		
Data description	Behavioural models of systems		
Purpose of the data	To simulate the behaviour of an object (for example a power grid) a model of all the parts but at least the of the whole system is needed. The level of detail of the model depends on the level of detail of the simulation.		
Relation to project objective	In the NRG5 project, developed software will be tested in a lab environment. Part of that is testing the software in a simulated environment		
File types	.m		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
Software models will be created during the lab trials	Not yet known	Access will be provided, if needed for the project objectives.	Not yet known

2.2.9 Configuration and Orchestration data

Fact sheet Dataset 8			
Dataset name	Configuration data		
Dataset description	To be specified		
Security and privacy considerations			
To be evaluated			
Datatype specific answers			
Datatype name	Configuration files		
Data description	Configuration files and login credentials		
Purpose of the data	Configuration of software components or login data for automated data push services		
Relation to project objective	Building a CI/CD environment is part of the project. To automatically deliver software configuration files are needed. Further information about CI/CD can be found in chapter 3.2.2.		
File types	.sh, .json, .xml, .custom1		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
(VIS) Configuration files for running software will be created.	A few MBs	Not yet known	Not yet known
(VIS) Credentials for different services will be provided.	A few MBs	Not yet known	Not yet known
Datatype name	Management data		
Data description	Management of network and functions.		
Purpose of the data	To be specified		
Relation to project objective	To be specified		
File types	Not yet known		
(Data provider) Origin of the data	**Size (xByte)**	**Access for Partners**	**Access for the public**
(VIS) MANO messages will be recorded.	Some MBs	Not yet known	Not yet known
(VIS) SDN messages will be recorded.	Some MBs	Not yet known	Not yet known

2.2.10 File type description

In this chapter, the file types identified in the datasets are described. For every file type there will be a description which will answer the following questions:

- Describe the file type
- What software is needed to open the file?
- Where to get that software.
- Is there free and/or open source software available to open the file, if not what are the expected costs for the software?
- Is there a standard, which describes the file in detail?
- If it is a file type defined by a partner describe the content definition

2.2.10.1 .avi (Audio Video Interleave)

Description:

AVI, is a multimedia container format. AVI files can contain both audio and video data in a file container that allows synchronous audio-with-video playback. AVI files support multiple streaming audio and video.

Software:

VLC is a good application to reproduce almost every video format.

Standard:

AVI is a subformat of the Resource Interchange File Format (RIFF) [4]

2.2.10.2 .custom1

Description:

These files contain encrypted data

Software:

Because of the encrypted nature of the data the software for opening it, needs to be specified from case to case

Standard:

The standard for data interchange of encrypted data will be negotiated per individual case

2.2.10.3 .gpx (GPS Exchange Format)

Description:

GPX, or GPS Exchange Format, is an XML schema designed as a common GPS data format for software applications. It can be used to describe waypoints, tracks, and routes.

Software:

Almost every GPS software can read this file. For example, OpenStreetMap, a collaborative project to create free editable maps using, among others, GPX traces.

GPSBabel can be used to upload/download/convert GPX files.

Standard:

The complete specification for GPX files can be found at [5].

2.2.10.4 .jpg

Description:

These files contain image information

Software:

There are a lot of different example for software with which .jpg files can be opened. GIMP is one of them which is open-source and available for the different operating system platforms. More information about GIMP and a download source can be found on the GIMP webpage http://www.gimp.org.

Standard:

The file type is defined in IEC 10918.

2.2.10.5 .json

Description:

These files contain data in a textual form.

Software:

The file can be opened for human reading or for machine interpretation. For human reading, a standard text editor can be used. For example the free software notepad++ on Windows or the command line editor vim on other operating systems. For machine interpretation the file type can be read with any programming language which can open files.

Standard:

The file type is standardized in two competing standards RFC7159 and ECMA-404.

2.2.10.6 .kmz (compressed .kml)

Description:

KMZ is a file extension for a placemark file used by Google Earth. KMZ stands for Keyhole Markup language Zipped. It is a compressed version of a KML (Keyhole Markup Language) file. Keyhole was the founding company of the Earth Viewer software that Google Earth was built upon.

KMZ files can contain placemarks featuring a custom name; the latitudinal and longitudinal coordinates for the location, and 3D model data. KMZ files can be opened by Google Earth, or unzipped with a compression utility, such as WinZip on Windows, MacZip for Macintosh users, and Zip and UnZip for UNIX systems.

Software:

KMZ files can be opened by Google Earth, or unzipped with a compression utility, such as WinZip on Windows, MacZip for Macintosh users, and Zip and UnZip for UNIX systems.

Standard:

KML is an open standard officially named the OpenGIS® KML Encoding Standard (OGC KML). It is maintained by the Open Geospatial Consortium, Inc. (OGC). The complete specification for OGC KML can be found at [6].

2.2.10.7 .mkv (Matroska Multimedia Container)

Description:

MKV is an open standard, free container format, a file format that can hold an unlimited number of video, audio, picture, or subtitle tracks in one file.

Software:

VLC is a good application to reproduce almost every video format.

Standard:

MKV specifications can be found at [7].

2.2.10.8 .vsd

Description:

These files contain vector graphics data.

Software:

This file can be opened with Microsoft Office Visio for Windows. The software is not open. The costs for a single license are below 100 Euro.

Standard:

This file type has a proprietary binary-format maintained by Microsoft.

2.3 FAIR data

The FAIR data principle is required to be used in EU-Projects by the "Guidelines on FAIR Data Management in Horizon 2020" [1]. It should support the exchange of scientific data and lead to knowledge discovery and innovation. The FAIR data approach is described by the acronym:

Findable data	Clear naming and versioning of (meta-) data, use of search keywords and Digital Object Identifiers
Accessible data	Specification how data is made available, what tools are needed to access data
Interoperable data	Use of standards and vocabularies for (meta-) data and datatypes
Reusable data	Specification when and for which duration data is made available, licensing of data

Parts of the data gathered during the project could be used to identify subjects, both individuals and business entities. Before making data public, compliance with EU 2016/679 [8] regulation must be checked.

Besides that, it may be part of the project, to gather private, sensitive or security relevant data. Therefore, data needs to be assessed. After that a decision about making that particular Data public or not.

The effort to assess the data should be reasonable compared to the foreseen value of the FAIR data.

2.3.1 Findable data

To make the data generated by the NRG5 project findable, there first needs to be a fixed naming and versioning convention. For documents, this convention was defined during the project kick-off. Deliverables and working documents should be named as follows:

Deliverables:	NRG5_D**w**.**d**_**ACR**_V**x**.**y**-**YYYYMMDD**.**ext**
Working Documents:	NRG5_W**w**_T**tt**_**ACR**_V**x**.**y**-**YYYYMMDD**.**ext**

Where

w	is the workpackage number
d	is the deliverable number
tt	is the task number
ACR	is the partner acronym
x	is the version major number
y	is the version minor number
YYYY	is the year
MM	is the month
DD	is the day
ext	is the file extension

Besides the naming convention, the “Guidelines on FAIR Data Management in Horizon 2020” [1] also propose to have Digital Object Identifiers (DOI) for the data generated during the project. To do so the project will use the zenodo platform [9] to fulfil the DOI requirement. Zenodo is a platform for long term storage of data. The service is provided by CERN and financed by the EU. The platform can handle single datasets with up to 50 GB size. If more storage per dataset is needed, CERN encourages the user to contact them. The amount of space is not limited. The platform intends to help research project to share data all over the world. To do so the platform also helps by defining and storing some additional metadata provided by the uploader. It is possible to grant access to the data only to a specific group of users or the public. The platform also gives the user the possibility to restrict or open access to data for a fixed period of time.

2.3.2 Accessible data

One important part of the FAIR concept is to make the data accessible to the project partners and when possible to other researchers and the public. To do so the NRG5 project will use the zenodo platform [9] for the data the consortium decides to be public.

The set of data collected and intended to be made available on a public platform, must have a broadcasting authorization by the operator of the site of origin. This to guarantee any confidentiality and risk of fraudulent use of this data in a competitive context. This is applicable for data issued from industrial captors, and all pictures and video’s.

In particular, video taken during drone overflights must be validated by the local operator to be stored and shared on the public platform. The angles of view will be predefined to preserve confidential industrial processes.

These shots on the public area are regulated by the local authorities and require special agreements (to be confirmed, depend of the country)

2.3.3 Interoperable data

Making data interoperable mainly relies on using suitable standards for data and metadata creation along with appropriate vocabularies (e.g. for providing search keywords).

These standards are not specified to date and therefore will be provided in D4.2.

2.3.4 Re-usable data

Data will be published after the release of the respective deliverable or after the end of the project.

The availability of data after the end of the project depends highly on the type and content of the data. Therefore storing data on a public platform needs to be discussed with the contributor of the data. The requirements for making data public also have an ethics aspect, which is discussed in WP8. More information and final decisions on a dataset or sub dataset level will be provided in deliverable D4.2

The licensing of data and deliverables is not yet specified but should be as open as possible. This information will be provided in deliverable D4.2.

2.4 Allocation of resources

The costs of making data FAIR depend on the amount of data, which may be published due to the cost of long term storage solution and additional effort for publication. An estimation cannot be delivered to date, as too many influencing factors are unknown at the moment.

The Data Manager is in charge of making sure the project and the data handled and stored during project lifetime is compliant to EU 2016/679 [8] and EU 2016/680 [10].

The responsibilities related to data management are already specified partially:

Data Management Plan:	Task 4.1 partners, lead: RWTH
Data storage and backup:	ENG by administrating the internal repository
Data archiving and publication:	long-term and public repository not specified, to be defined

2.5 Data security

Each partner is responsible for recoverability of own generated data (backup plan in place according to institution or company practice). The assessment of privacy issues and security risks, which may arise, with the content of gathered data is done by the entity who is collecting the data.

2.6 Ethical aspects

Reference is made to the upcoming deliverables of work package 8, Ethics requirements.

3 Integration Guidelines

3.1 Introduction

Development of software and hardware can be a very complex task. To ensure code quality, test the developed product and to track the changes made by different developers, integration guidelines are needed. The integration guidelines are defining for example coding standards to follow, to ensure that the code is readable and understandable by different developers involved in the process. To track the changes in software, but also in the hardware development process repositories are used, where the different versions committed by the developers are stored. In the following chapters, guidelines for code and hardware development are discussed.

Due to the early stage in the project and because not all tasks involved in development have started yet, the following guidelines may not be complete at the moment. The guidelines will evolve during the project.

3.2 Software

3.2.1 Coding conventions

Not yet specified.

3.2.2 What is CI/CD

Continuous Integration (CI) refers to having a framework allowing developers to work on both differentiated branches of code and on a core (master) branch, effectively referring to the integration of code into a known or working code base [11]. All source-code related actions and control operations (e.g. branching requests, review requests, merge requests, unit testing procedures, software package building etc.) are related to CI, as well [12]. CI functionality are generally exposed by advanced source code management tools like git, subversion, mercurial etc, git being the most prominent and widely used of all. However, nowadays, there exist multiple platforms offering added value services on top of these source code management tools, such as github, gitlab, bitbucket etc.

Continuous Delivery (CD) refers to the automated process of delivering a complete software package or service to an environment, be it integrated testing or production. In other words, it refers to controlling the deployment of the software package or service to designated target application containers or servers for immediate consumption. Most of the aforementioned source code management platforms also provide the ability to integrate CD technologies either as third party services (e.g. use Github combined with Jenkins [13] or Travis [14]) or integrated (e.g. Gitlab with Gitlab-CI [15]).

Based on the above definitions, CI/CD refers to integrating the entire source code management and service delivery processes into one single pipeline that would cater for the deployment of a software package or service whose code has been mutually accepted by either automated routines (e.g. related to testing) or human-based activities (e.g. code review requests, branch merge requests etc.).

In the above context, testing plays a critical role as to the quality of the service being delivered in the end of the process. Three main types of testing are defined in general:

- Unit testing refers to testing the behaviour of the smallest part of software design, like functions. The goal is to identify code-related issues, pertaining to the syntactical correctness of the code and the rationality of the produced outcomes.
- Integration tests combine several software components (already unit-tested), such as libraries and modules, and test them as a whole. They ensure certain subsets of the system work as expected, from an interworking perspective.
- System tests focus on the user's or customer's perspective, checking whether the overall system meets the given specification.

Figure 1 provides an overview of an integrated CI/CD framework:

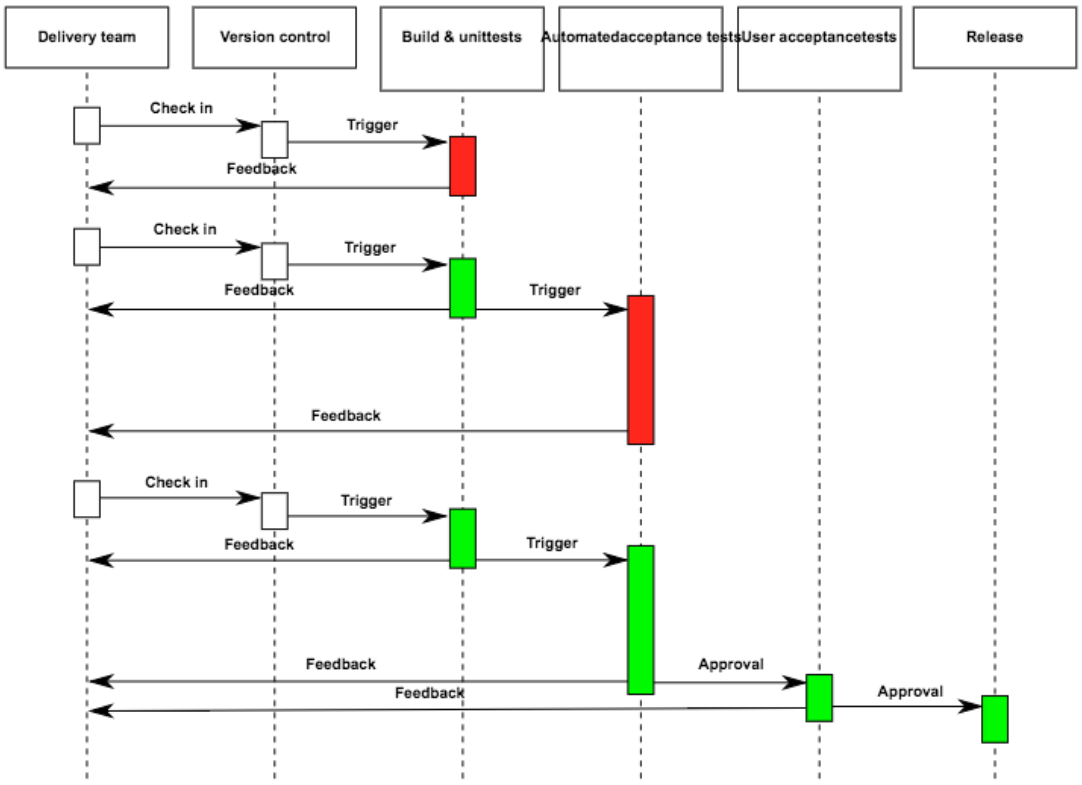

Figure 1: Principle CI/CD [16]

Note that the CI part practically ends on the Automated Acceptance tests; thereafter, the CD part takes place.

Completing the automated testing and quality enforcement puzzle of the CI/CD framework, one may apply source code inspection through automation frameworks such as SonarCube [17]. Although code quality is not vital for the successful testing of a system or even its deployment, it can largely affect its efficiency in terms of resource allocation and/or performance, hence it is highly recommended.

As per the proposal, the above notions have been already taken into consideration by Figure 2:

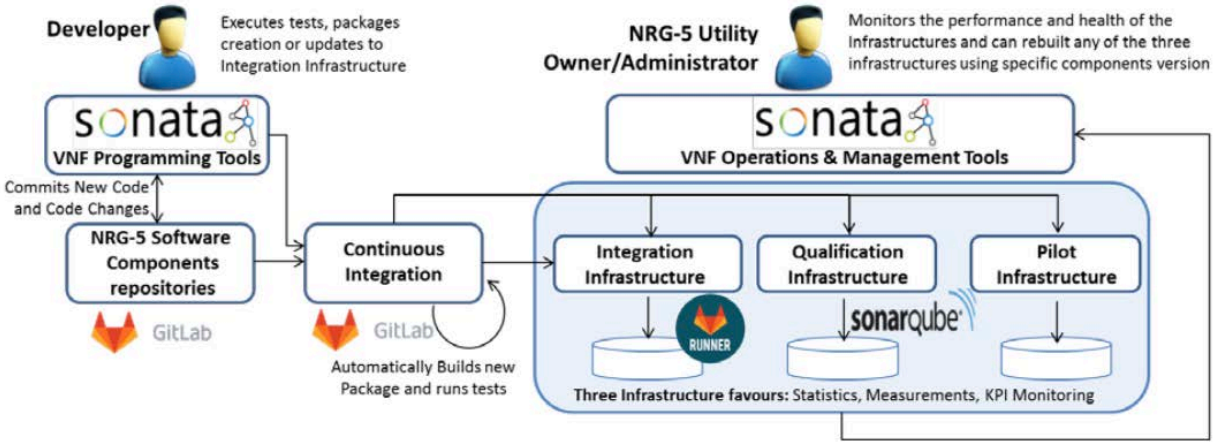

Figure 2: Development, validation, and integration lifecycle [18]

The relevant text extracted from the Grant Agreement of the NRG5 project, indicates that "*Gitlab will be used for software components and NFVs/VNFs versioning and as the core continuous integration driver (Gitlab-CI). In such a context, the latest version of the software will be automatically integrated upon successful source code updates and tests completion and a new version of it will be built and deployed on the Integration Infrastructure. As extensive testing is essential for achieving high-quality and hassle-free integration, unit and service-level tests will be given strong focus and will comprise both simple, scripts-based functionality testing and operational testing based on the VNFs testing and benchmarking framework provided by OPVNF, Yardstick. Note that NRG-5 will auto-feed the SONATA Catalogue so that stable releases/updates of certain VNFs will automatically update the SONATA catalogue.".*

Overall, such a strategy could highly benefit our work as automated testing, though cumbersome and very time consuming, can assure that development over heterogeneous infrastructure and of different scope remain working and inter-working in any time. Under a CI/CD approach and supposing that the base platform has been selected and instantiated (being either SONATA, OSM or OPNVF), the CI/CD approach would imply e.g.:

1. The developer implements a certain piece of software that wishes to integrate to the core system functionality, irrespectively of whether this piece of code corresponds to the development of a VNF or a platform component;
2. The developer introduces (or should have already introduced) a test case that unit tests the relevant piece of code at the level of rationality checking and outcomes consistency. The unit testing should be as thorough as possible.
3. The developer commits the code to a dedicated development branch defined in the CI infrastructure.
4. The developer initiates a request for code merging.
5. The CI platform performs the determined unit testing;
    a. If the unit testing is successful, the code is subject to further integrated system testing.
        i. If the integrated system testing is successful, then the merge request is accepted and the newly committed source code is part of the code master branch.
        ii. Otherwise, the merge request is rejected.
6. If either the unit or integrated system testing fails, the merge request is rejected. In this case, the developer should consider either a code rewrite to satisfy the unit testing procedures, or the unit testing rewrite (to avoid situations of erroneous unit testing. The workflow should be re-initiated at step 1.
7. The source code management platform move to the CD part, building the package and deploying it;

8. After successful CD the new code is running on the deployment infrastructure and is subject to user testing/production.

Although multiple CI/CD frameworks and platforms are available, Gitlab.com appears as a good candidate for hosting the developments of the NRG-5 project:

1. It is an open source management platform which suites the needs of the NRG5 project and can provide the following needs: Multiple projects are possible, grouped under groups and subgroups;
2. Multiple private projects are able to be created; no need for opting for a pro account to have private (non-public) projects. This is particularly important in the development phase; maybe we would prefer to proceed with the development up to a point, then making the projects public (open source).
3. It has native CD mechanisms

3.2.3 Integration of VNFs

Following the CI/CD integration guidelines documented in the previous section, the lifetime of the VNFs development, testing, qualification and deployment will be defined in a likewise manner. This approach is also in accordance to the overall CI/CD-oriented approach of SONATA (even if SONATA is not, in the end, selected as the MANO framework of NRG-5), which effectively defines the following steps as necessary for accepting new VNF- and platform-related code into the master source code branch:

1. Follow the steps 1 - 6 of the generic CI approach described in the previous paragraph
2. The successful merge of the code should be followed by a subsequent instruction to the MANO framework to integrate the updated component to an a-priori designated infrastructure which can vary as follows:
    - The xMEC if it is a VNF intended to be used at base station level (e.g. a VNF intended to fulfil a set of uRRLC or uMTC requirements);
    - At local DC level if it is a VFN intended to be used at non-real-time applications whose delay constraints are average;
    - At regional DC level if it is a VNF or VNF extract intended to be used for operations that imply large processing times, hence higher response times.

Note that in any case, the selected MANO framework will be responsible for creating or updating the respective VNF instance, together with the re-configuration of the VNFFPs, VNFFGs and NSs that such a creation/update might imply. At all times, the deployment process will be monitored in terms of time spent for the actual deployment, in an attempt to acquire metrics of performance of the particular deployment site. This continuous monitoring approach will assist the work of the determination of the optimal placement of the VNFs, avoiding to utilize deployment sites that exhibit lower performance profiles for delay intolerant VNFs and NSs.

3.3 Hardware

Hardware development is only a small part of the project. Mostly the project will use already existing hardware and combine it into a new system or add additional software functions.

3.3.1 Development

To ensure hardware development quality, there will be design reviews of the schematic plans. If possible, there will be a prototype build in house for testing. If the developed hardware is needed multiple times, the production may be outsourced.

3.3.2 Testing

To ensure that the developed hardware meets all the requirements there will be testing. Which tests are performed will be decided depending on the hardware. Typical tests with prototypes include timing/signal integrity, power supplies, energy efficiency, temperature and other tests.

4 References

[1] “Guidelines on FAIR Data Management in Horizon 2020,” 26 6 2016. [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. [Accessed 24 11 2017].

[2] *NRG5 Grant Agreement page 141.*

[3] *NRG5 Grant Agreement page 142.*

[4] “AVI RIFF File Reference,” [Online]. Available: https://msdn.microsoft.com/en-us/library/ms779636(VS.85).aspx.

[5] “GPX definition,” [Online]. Available: http://www.topografix.com/gpx/1/1/.

[6] “KML definition,” [Online]. Available: http://www.opengeospatial.org/standards/kml/.

[7] “MKV definition,” [Online]. Available: https://matroska.org/technical/specs/index.html.

[8] “Regulation (EU) 2016/679,” [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=DE.

[9] “https://zenodo.org/,” [Online].

[10] “Directive (EU) 2016/680,” [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016L0680&from=EN.

[11] “https://www.nwcadence.com/blog/continuousintegration-continuousdelivery,” [Online].

[12] “https://en.wikipedia.org/wiki/Continuous_integration,” [Online].

[13] “https://jenkins.io/,” [Online].

[14] “https://docs.travis-ci.com/user/deployment,” [Online].

[15] “https://about.gitlab.com/features/gitlab-ci-cd/,” [Online].

[16] “https://en.wikipedia.org/wiki/Continuous_delivery,” [Online].

[17] “https://www.sonarqube.org/,” [Online].

[18] *NRG-5 Grant Agreement page 139, Figure 10.*