Grand Agreement No: 101016941

Full Title:	Open cooperative 5G experimentation platforms for the industrial sector NetApps
Start date:	01/01/2021
End date:	31/12/2023
Duration:	36 Months
Project URL:	https://www.5g-induce.eu/

Deliverable D7.2

Data Management Plan

Document type	Deliverable		
Title	D7.2 – Data Management Plan		
Contractual due date	31/03/2021 (M3)	**Actual submission date**	31/03/2021
Nature	ORDP	**Dissemination Level**	Public
Lead Beneficiary	8BELLS		
Responsible Author	Vasileios Samarinas (8BELLS)		
Contributions from	Theodora Kousta (8BELLS), Nikolaos Pitropakis (8BELLS)		

Revision history

Version	Issue Date	Changes	Contributor(s)
v0.1	17/03/2021	Initial version	Vasileios Samarinas (8BELLS), Theodora Kousta (8BELLS)
v0.2	23/03/2021	Minor changes	Theodora Kousta (8BELLS)
v0.3	29/03/2021	Minor changes	Theodora Kousta (8BELLS)
v1.0	31/03/2021	Version ready for submission	Theodora Kousta (8BELLS)

Disclaimer

The content of the publication herein is the sole responsibility of the publishers and it does not necessarily represent the views expressed by the European Commission or its services.

While the information contained in the documents is believed to be accurate, the authors(s) or any other participant in the 5G-INDUCE consortium make no warranty of any kind with regard to this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose.

Neither the 5G-INDUCE Consortium nor any of its members, their officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the 5G-INDUCE Consortium nor any of its members, their officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

Copyright message

Table of Contents

List of Figures

List of Tables

Glossary of terms and abbreviations used

Abbreviation / Term	Description
AGV	Automated Guided Vehicles
AI	Artificial Intelligence
API	Application Programming Interface
AR	Augmented Reality
DMP	Data Management Plan
EC	European Commission
ERP	Enterprise Resource Planning
ExFas	Experimentation Facilities
GDPR	General Data Protection Regulation
H2020	Horizon 2020
HD	High Definition
KPI	Key Performance Indicator
ML	Machine Learning
NetApps	Network Applications
PPC	Public Power Corporation
SAL	Server Admin Log
SDN	Software Defined Network
SIEM	Security Information and Event Management
SLAM	Simultaneous Localization and Mapping
UC	Use Case
VPN	Virtual Private Network
VR	Virtual Reality

Executive Summary

The content of this deliverable, namely 7.2 Data Management Plan, is targeting to equip the 5G-INDUCE project with a Data Management Plan (DMP). This will be feasible by first laying down the appropriate framework for governing issues which are related to data management. The DMP is meant to be used internally by the 5G-INDUCE consortium partners, thus achieving an effective management of the research data that will be generated within the context of the project. Additionally, it will permit a more efficient handling of the management of the upcoming publications. Within the context and the lifecycle of the 5G-INDUCE project, the DMP will describe the datasets that will be used along with the procedures that will make use of them. Specifically, these deliverable details, the data description, as well as the procedures responsible for documentation, organisation, and storage of the data. Additionally, the access control and sharing methodologies are going to be discussed under the scope of security, with the ethical considerations being also a concern of the DMP. The DMP is expected to mature during the project as more details regarding the use cases are going to become visible, thus constituting a living document which is going to be updated throughout the lifetime of the project. Its final version will be delivered at the end of the project.

1 Introduction

The 5G-INDUCE project is subject to the guidelines of the "Open Access to Scientific Publications and Research Data", as described in the Horizon 2020 (H2020) pilot program. The project aligns with the concept of open science since the novel solutions it introduces to the European and Global community allow the re-use of the data utilized in the scientific research. However, this act is susceptible to specific stipulations, which apply to certain datasets, as they are described in this document. Therefore, this project embraces the open access of the data and the findings occurring from the project's implementation, in a way that it does not contradict the pertinent conditions.

1.1 Scope and Objectives of the Deliverable

The aim of the Data Management Plan (DMP) is to describe the whole lifecycle of the data collected and generated under the scope of the implementation of the 5G-INDUCE project in a way that all major aspects of the process are covered. Therefore, the main objective of this deliverable is to provide an overview of:

- How and what data was either collected or generated and what its existing formats are,
- The lifecycle of the 5G-INDUCE datasets (see Fig. 1),
- The methodology and principles according to which the data was collected/generated,
- The conditions under which the data will be shared,
- How data will be preserved and secured.

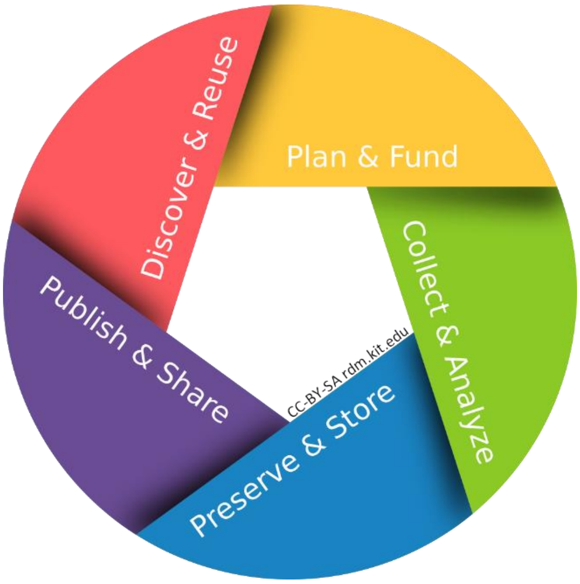

Figure 1: Research Data Lifecycle [1].

The legal protection of the project's data, and especially the personal ones, is of supreme priority in the 5G-INDUCE project and therefore the apt measures need to be taken into consideration. Therefore, any applicable restrictions and ethical aspects, with regard to the access to the data, have to be made valid.

Furthermore, since this project abides by the Open Research Data pilot by the ERC in Horizon 2020, the following aspects need to be outlined, in order for the datasets to be aligned with the "FAIR" [2] guidelines:

- Making data Findable,
- Making data Openly Accessible,

- Making data Interoperable,
- Increase data Re-use.

This means that the manner in which the data owned by the consortium will be shared and become openly accessible to third parties will be determined by the owners themselves, in a way that facilitates their commercial benefits but at the same time complies with the FAIR guidelines of the Horizon 2020 pilot.

Finally, if any significant changes to the existing datasets or to the relevant guidelines occur, the DPM will be updated accordingly to reflect the actual state of data in every stage of the 5G-INDUCE project course. This DMP comprises deliverable D7.2 of the 5G-INDUCE project and is planned to be delivered on M03.

1.2 Structure of the Deliverable

The DMP deliverable is structured as follows:

- Section 1: The Introductory section of the deliverable aims to provide an overview of the DMP, its main objectives, as well as the structure of the deliverable.
- Section 2: This section provides a thorough description of the management plan of the data used in the 5G-INDUCE project, including the way in which data was collected/generated and what types of research data and datasets the project utilises.
- Section 3: The FAIR data describe how the data will be made findable, openly accessible, interoperable, and re-usable. Also, a description of how the allocation of the project's resources is going to occur, as well as what the collaborative responsibilities of the partners will be, is provided.
- Section 4: The Security section outlines the measures taken in order to protect the data as well as the conditions under which the data will be shared.
- Section 5: The Ethical Aspects section lists the ethical aspects affecting the data management in the 5G-INDUCE project.
- Section 6: In this section, an overview of the different use cases which make use of the research data is provided.
- Section 7: The final Section concludes the deliverable by summarizing the DMP.

Finally, the template of the questionnaire distributed to the consortium parties, in order for crucial information for the construction of the deliverable to be collected, is shown in the Appendix section.

2 Data Management in 5G-INDUCE

2.1 Data Summary

The main purpose of this section is to provide a brief summary of the data types, as well as of their origin and formats, that will be processed throughout the duration of this project.

A dedicated questionnaire has been carefully designed to accomplish this goal, where the partners of the consortium were asked to provide information regarding various data management categories. The main parts of the aforementioned questionnaire are related to areas such as data description, security and data management procedures. Moreover, this section includes a brief summary of the most significant results divided in the questionnaire's categories.

A template of the questionnaire can be found in the Appendix section of this report.

2.1.1 Types of 5G-INDUCE Datasets

This section includes the list of the most significant data categories which will be processed and generated in the scope of 5G-INDUCE. These categories have been identified based on the answers provided by the project partners in the distributed questionnaire.

The data processing rules will be determined alongside with the development of the project for both the project items and pilots where the involvement of partners is required. More specifically, clarifications will be provided regarding what data the project anticipates to be generated, per use case, and why the actual data would not be open.

Documents and reports will be made available to the project and uploaded to the project website.

2.1.2 Purpose of Data Collection/Generation

The following table describes the types of data that will be processed and generated alongside with their respective purpose in the scope of the 5G-INDUCE project.

Table 1: Purpose of data collection/generation.

Data Type	Purpose
Text	• Communication and Dissemination of project outcomes • Contributions to deliverables (contributors), managing of task (mailing list) • Partners' contact data will be published on international conferences and journals. • Project-Related Information Sharing • Name and surname, organisation/ workplace, country of origin, email address, online identifiers • Collection of geolocation information coming from the vehicles • login and access to web/mobile services
Sensor data, production data, failure probabilities, maintenance schedule	• Use-case demonstrations • Predictive maintenance scheduling

Data Type	Purpose
Audio/video streams (raw and processed)	• Use-case demonstrations • Remote inspection, surveillance and monitoring of industrial machinery
Average and peak values of consumption, historic data, optimized scheduling	• Use-case demonstrations
Images	• To present the workplace where tests are done • Dissemination of the project in web and social media • Demonstration of the work performed • to be processed in the 5G edge to provide specific information and orders to be executed in the AGVs, e.g., body and hand gestures
Videos	• Demos of applications for AGVs moving • Dissemination of the project in web and social media • Demonstration of the work performed • to be processed in the 5G edge to provide specific information and orders to be executed in the AGVs, e.g., body and hand gestures • Detection of people in the screen but at the same time without implication of identification to provide a 5G trial and validation in the context of Objective 4 • Troubleshooting
Production Line Sensor Data and Factory production data	• Analytics for predictive maintenance for the purposes of UC4
Research item artifacts (Models' outcome data, Platform monitoring data, Network measurement data, Questionnaires excluding personal data)	• Dissemination, • Reporting • WP2-WP6 objectives • Collaboration purposes with 3rd parties or other H2020 EU projects • Contributions to 5GPPP according to signed collaboration agreements
Software artifacts (Code, APIs, Containers, Dashboard)	• Platform operation, integration • Interworking • Contribution to WP3-WP5 objectives
Dataset artifacts (NetApp requirements (synthetic), Infrastructure topology and Interconnectivity (dynamic))	• Experimentation set-up and evaluation • Fulfilment of WP6 objectives
Numeric, currencies	• Calculation of costs • Employee IDs for organizational purposes • Collection of geolocation information coming from the vehicles

Data Type	Purpose
Network performance metrics (speed, delay, signal level RSRP/RSRQ)	• To support use cases and other project-related activities (UC 8 trial and validation in the context of Objective 4). • Communication, dissemination and field demonstration
Drone-based video stream of monitored industrial infrastructure will be recorded	• To support use case and other project-related activities (UC 8 trial and validation in the context of Objective 4). • Communication, dissemination and field demonstration
Geolocation information	• telematics processes • routing optimization • general geolocation information from vehicles
Network Numerical Data (bandwidth, traffic usage, packet latency, packet loss, etc.)	• To define the KPI's of the use cases • NetApps related to Patras 5G-VINNI facility.

2.1.3 Formats of Datasets

In the following section, a table which depicts the data types and their respective formats is demonstrated.

Table 2: Formats of data types.

Data Type	Format
Text	PDF, docx, xlcs, csv, pptx, xls
Sensor Data	text, numerical (integer, float, double, long), audio, image
Images	Jpg
Videos	MPEG2, MP4
Audio	MP3,MP4
Software, visualization and analytics	JSON

2.1.4 Origin of Data

The following table is showing the origins of various data types, excluding however the origin of data that are subjects on their own, such as industrial production equipment.

Table 3: Data origin.

Data Type	Origin
Text	• Directly from partners, web sources (contact persons for organizations external to the project) • network and radio performance metrics
Sensor data, production data, failure probabilities, maintenance schedule	• Utilized data will be produced by deployed industrial IoT sensors

Data Type	Origin
Images	- Directly from partners - The data come from the NetApp itself, some of them collected by the user device App.
Videos	- Directly from partners, - Data provided from PPC, - The data come from the NetApp itself, some of them collected by the user device App.
Audio	- The data come from the NetApp itself, some of them collected by the user device App.
Production Line Sensor Data and Factory production data	- Sensors at the shop floor / databased at factory level.
Research item artifacts (Models' outcome data, Platform monitoring data, Network measurement data, Questionnaires)	- Generated within the project, directly from partners
Software artifacts (Code, APIs, Containers, Dashboard)	- Generated within the project, ExFas and NetApps
Dataset artifacts (NetApp requirements (synthetic), Infrastructure topology and Interconnectivity (dynamic))	- Generated within the project, ExFas and NetApps
Network performance metrics (speed, delay, signal level RSRP/RSRQ)	- Data will be generated by ININ's qMON solution and drone-based video streaming.
Drone-based video stream of monitored industrial infrastructure will be recorded	- Data will be generated by ININ's qMON solution and drone-based video streaming.
Geolocation information	- GPS devices installed at vehicles
Network Numerical Data (bandwidth, traffic usage, packet latency, packet loss, etc.)	- From network traffic when carrying out the use cases and running the NetApps. - Test results of the NetApps.

2.1.5 Expected Size of Data

The following table depicts the expected size of data and the respective ways of size measurement. Based on the responses received by the partners of the consortium, for various types of data, the WP leaders' contribution and strategy would be crucial to define the expected size.

Table 4: Projected size of data.

Data Type	Size / Measurement/ Consumption Types
Text	- Kilobytes of space, More than 1GB
Images	- 5GB post-processed
Videos	- 15GB (for processed ones), 5GB (for a 10 min file), 1GB depending on the AI/ML algorithm to be designed

Data Type	Size / Measurement/ Consumption Types
Performance Metrics	• Several GBs
Geolocation data (GPS)	• Up to 200GB
Network Numerical Data (bandwidth, traffic usage, packet latency, packet loss, etc.)	• Multiple Gigabytes of network data.

2.1.6 Data Utility

Project-generated and historical datasets are intended to contribute to the project actions and will be beneficial for various stakeholders and individuals such as the research community. In this direction, the data categories and types from use-cases 4,6 and 7 will be open-sourced. Moreover, any sharing of data outside the 5G-INDUCE consortium must be agreed upon by all partners and obey to the GA and CA rules. Furthermore, regarding the sharing agreement topic, those protected under the GA and related to the reporting of the results through the project deliverables, are an exception to the above. The issuing of agreements for the sharing of non-public data must be agreed upon by the project consortium. One such case is the sharing of data with the IAB members.

2.1.7 Data Management Procedures

One significant part of the 5G-INDUCE data management is constituted by the procedures that are in place from the partners of the consortium. More specifically, such policies do exist and are related to regulations such as the Binding Personal Data Protection Regulations, GDPR and internal partners' procedures and policies, local regulations/rules, as well as research data management procedures.

In the context of compliance with the requirements of the ISO 27001 standards, various partners have developed data management instructions concerning, among others, the physical data storage and the use of test data and background information. Furthermore, some partners had put in place staff training on IT security and data management topics.

Finally, one of the partners of the consortium (Infocom) is ready to set up a 5G-INDUCE project specific data management policy and to contribute to the definition of such policy at project level.

2.1.8 Documentation, Organisation and Storage

In this section, there are different aspects concerning the documentation, organisation and storage of the files. One of the ways that the data will be organized and labelled will be per publishing material such as conference or journal when it is related to the contact details. In most of the cases, the personal data generated in the project will be stored in time necessary for the project results to come out.

In terms of labelling and organizing data, this will be determined by the WP leaders' approaches, and some criteria could be the type and time of acquisition of data. Furthermore, some cases will include data analysis and processing but not data handling. The table below depicts the labelling and organization of files based on the responses received from the partners.

Table 5: Examples of data labelling.

20210X0X(date)_5GINDUCE(nameProject)_TX.X(task)_Shortdescription(e.g AGVremoteControlled)
File names will include information about the 'version' (in the form 'v<xx>') and 'release date' (in the form '<yyyy_mm_dd>'): <file name>_v<xx>_yyyy_mm_dd
5G-INDUCE_Dataset_XYZ_YYMMDD_Location.
Deliverables: {Type}=Dxy Milestones: {Type}=MSx_WPx Results: {Type}=Txy_{short name} Paper: {Type}={ConferenceName}-yy_{short-Title}, e.g. EUCNC-21_NAO-OSS-platform

Regarding the 5G-INDUCE data storage, it will be preserved for up to 5 years after the 5G-INDUCE completion, where appropriate protocols will be implemented to control the storage period. Furthermore, some of the common ways in which partners of the consortium will store their data, are the use of physical (non-network attached) local mass storage devices, personal computers, private servers (e.g., project's repository), data centres, and cloud-based storage (Microsoft Azure, Amazon Web Services, Google Cloud).

Regarding the mobile transmission and storage, transfers will take place by direct connection of mobile devices with partners' internal computers, where the data will be removed immediately, afterwards, from the mobile devices with the use of encrypted protocols. Also, mobile phone videos will be transferred and stored within an SD card, while other partners will set up internally a secured way of transmission of data between mobile terminals and backend components.

2.1.9 Accessibility of Data

The last section of this chapter contains a brief summary of the answers which have been provided by the partners related to the accessibility of the data that will be produced and processed during the lifetime of the project. This process is dependent on both the WP leaders' strategies and the objectives of the project.

In addition to that, one significant point regarding the data accessibility is related to the way in which the identity of the individual who accesses the data will be ascertained. In most cases, team members and use case owners, only, will access the data, and examples of the aforementioned ways include the SSH protocol as well as email exchange. Moreover, the most common software platforms that will be used to access the data from the partners are MS Office and Notepad++.

Finally, it should be mentioned that multiple partners retain private servers and access is allowed only from an internal network or Virtual Private Network (VPN).

3 5G-INDUCE FAIR Data Principles

3.1 FAIR Data

The FAIR data policy requires that all data generated during a H2020 project has to be Findable, openly Accessible, Interoperable and Reusable (FAIR). It is highlighted though, that this specific policy does not affect the ways and methodology in which data will be processed and analysed, and no specific kind of technology is recommended. According to the H2020 FAIR data guidelines, the data generated/collected and shared during the project should conform to the following:

- **Data is findable**: Data should assist the discoverability of metadata, by both humans and computer devices, for further use, as well as their identifiability with unique and persistent identifiers, according to the "GO FAIR" FAIR principles [3]. Naming conventions, the search keywords and the clear versioning methods also need to be determined. More specifically:
    - (Meta)data should be determined by a worldwide unique and persistent identifier,
    - Data has to be described by rich metadata,
    - The identifier of the data it describes is plainly and distinctly contained in the metadata, and
    - (Meta)data is recorded in a searchable repository.
- **Data is openly accessible:** Data should be accessible:
    - Through an open, free, and universally implementable identifier protocol, by a data repository. This protocol should permit an authentication and authorization procedure, when required. If access to certain data is restricted, proper justification should be provided. Additionally,
    - Access to metadata should be possible even when the data is no longer accessible.
- **Data is interoperable:** According to the FAIR principles, data has to be coherent with other data and work in conjunction with applications, as well as with analysis and storage procedures. More precisely:
    - The exchange and re-use of the (meta)data among the partners should be assisted by vocabularies and shared knowledge representation codes for semantic consistency among researchers,
    - These vocabularies should abide by the FAIR principles, and
    - (Meta)data should contain references to further (meta)data.
- **Data is re-usable:** The optimization of data reusability is the eventual objective of the FAIR principle. In order for this to be accomplished, (meta)data should be properly described so that it can be reproduced and integrated in various settings.
    - Clarifying licensing is needed in order for data to be as much re-usable as possible.
    - Origin of (meta)data must be precisely mentioned.
    - (Meta)data should conform to domain-relevant community standards.

  The time period for which data will be re-usable, and especially by third parties after the end of the project, has to be specified and accompanied by the relevant justifications and/or restrictions.

In order for the EC to facilitate the openness of data, under the FAIR data scope, several exemplars and standardized procedures are recommended. For instance, various vocabularies for metadata used in several areas of study are existent in the Metadata Standards Directory [4].

The FAIR policy has been created in order for data management and data curation to be as close to optimal as possible. Furthermore, the FAIR principles are framed in a way that they can be applicable to a wide range of research fields and data management objectives, from a simple data collection plan to large scale research projects. The enforcement of the FAIR principles by the H2020 guidelines aims to provide a guideline for efficient lifecycle data management, making sure that all of the principal aspects of the data's lifecycle are taken into account.

According to the EC policies and mandates on Open Access to publications and research data, presented on OpenAIRE [5], various economic, social and educational benefits can arise from making research data and publications open to access, after eliminating any financial, legal and technical barriers. More precisely, with open access:

- Research on a national level can be integrated with an interoperable network of global knowledge,
- The effect of the national research can be enhanced,
- New research partnerships emerge, and
- Professional isolation is eliminated.

As a result, the benefits of the open access also reflect on a society level, where research becomes more efficient and productive, finer and more rapid outcomes are conveyed, and economies are reinforced through the growth of a robust and substantive science base. Also, the greater return on investment that open access can bring, through the increased effect of the research in which countries have invested public resources, is proved to benefit the countries involved.

A project's DMP is the key to an adequate data management and the FAIR principles reflect well on this objective. The 5G-INDUCE's DMP is created during the first 3 months of the project's course, but will periodically be updated so that it keeps up with any possible significant changes to the key components of its structure (e.g., the data).

3.2 Data Sharing

Data Sharing describes the action of making private data available to authorized partners and will occur according to each specific dataset's form and the access limitations, or not, that characterize it. Access to scientific publications will be available through the project repositories. The set of research data which is not applicable to access restrictions and is free for open use will be pinpointed during the execution of the project.

In order to establish and reassure the privacy and integrity of the data shared, in a way that assists the tracking and control of all procedures, certain actions and measures need to be prompted among all authorized partners. In addition to this, partners should keep in mind that private data is usually subject to copyright restrictions to prevent the malevolent use of copies. In case that private data needs to be shared among the partners of the 5G-INDUCE consortium, the following conditions should be met:

- Only platforms and services approved by the 5G-INDUCE project must be used,
- If any private data is shared via e-mail, it is mandatory to be properly encrypted,
- Encryption keys also need to be transmitted as securely as possible among the partners prior to the data exchange,
- A process will be implemented to rotate cryptographic keys and access control models in case of compromise.

In the 5G-INDUCE project, the sharing of data mainly concerns research items and scientific results, specific examples that apply to certain use cases of the project, as well as information such as names and surnames, organizations, country of origin, email addresses, online identifiers, etc. It is important to pinpoint here that the disclosure of such data does not raise any privacy, ethical, or confidentiality concerns. The research items and results to be shared, according to the vast majority of the research partners, will be:

- Images of the workplaces where tests are taking place and equipment images,

- Images, videos and audio recordings,
- Demonstrations of the applications,
- Sensor data,
- Reports,
- Vehicle locations, for routing purposes, and smart logistics operations.

No personal data are supposed to be disclosed, according to the consortium members.

The main purposes of the disclosure of the data listed above are:

- Dissemination, exploitation and communication activities,
- Demonstration of the project's results (mainly applies to the specific use cases),
- Other collaboration activities according to pertinent agreements of the partners.

Any sharing of data will occur only when the relevant or any preliminary results are available, and only among the use cases partners and the 5G-INDUCE consortium members. For dissemination reasons, data might also become available for the scientific community, the Commission and the general public, but only after the conditions of such an action are decided by a relevant agreement. Nevertheless, the majority of partners do not demand the existence of a sharing agreement in order for their data and research results to be shared among the partners. The rest of them considers the existence of such an agreement as crucial enough, especially when it comes to personal data, claiming that no disclosure might be made without the consent of all the partners. Finally, a couple of partners suppose that they will not need to share any data at all.

3.3 Allocation of Resources

In order for data to be FAIR certain actions need to be taken, the cost of which is unknown at the present phase of the project. However, these costs can be perceived as eligible costs. The budget for dissemination activities will be allocated among the 5G-INDUCE members who aim to issue scientific publications. The allocation of these resources should be sufficient enough to accommodate, at least partially, the requirements of the project for making data FAIR. The costs for data preparation and management will also be part of the project's expenses.

3.4 Collaborative Responsibilities

In the Grant Agreement of the 5G-INDUCE project, the consortium members have agreed that any private information, research data, research results, as well as every relevant piece of information shared among the partners during the implementation of the project and its exploitation, should not comprise any personal data. This is also defined by the European General Data Protection Regulation (GDPR), Article 4, Section (1), with which all partners are obliged to comply. Hence, all partners have to ensure that every requisite measure is taken in order to prevent access to or, ideally, eliminate every personal data from the project's repositories.

4 Data Security

Under the scope of the 5G-INDUCE project, all of the researchers have to abide by the “Guide on Good Data Protection Practice in Research” [6], in order to guarantee the secured sharing of data. They have to make sure that only authorized users have access to private data repositories and establish a monitoring tool in order to reassure the optimum control of the process. What is more, the access prerequisites have to be made clear to all partners. Finally, if a partner has to authorize a third party to process their data, they have to make sure that all the appropriate security measures are taken by them.

In practical terms, the above could be achieved with:

- A user authentication tool, in order to properly identify every authorized user,
- An access control tool which allows/denies access to specific private datasets,
- A repository security mechanism, in order to prevent unauthorized users to get access to confidential data. This could be accomplished by operating system controls, usage of passwords, encrypting techniques, etc.,
- A secure channel for data transmission. Security solutions such as firewalls, antivirus programs and security information and event management (SIEM) systems need to be present in both sides of the channel to ensure protection against common threat vectors. Additionally, encrypted transmission is mandatory for online communication of data to avoid third parties’ eavesdropping, along with reinforced physical security in case of physical transmission to avoid cyber physical attacks.

Additionally, according to the “Open Access to Scientific Publications and Research Data” guidelines, all participants to the H2020 projects are obliged to take the following measures:

- Data and metadata necessary for the validation of the research outputs and research publications should be immediately stored in the relevant repository of the project. The same applies for any other data or metadata which researchers regard as important to be stored and preserved.
- Any pertinent measures that facilitate the access, inspection, usage, copying and distribution of the data by third parties should be taken.
- Validation techniques and relevant information, such as algorithms and analysis codes, should be made available in every repository necessary, to help with the validation process of the research results.

4.1 Data Security Policy in 5G-INDUCE

All the consortium members will have to conform to the General Data Protection Regulation (GDPR) regarding “the protection of natural persons with regard to the processing of personal data and on the free movement of such data” [7]. In the 5G-INDUCE project, the major risks regarding the data security, as they were outlined by the partners, are the loss of the data, data leak, data breach, and the access to it and its transmission by unauthorized parties. Server crashing or cyber-attacks are also issues addressed by the partners. The 5G-INDUCE members aim to tackle these potential threats, and ensure a safe data storage, sharing and management plan, by taking one or more of the actions listed below:

- Establish secure access protocols (e.g., multi factor authentication) to the data repositories,
- Set up a data backup mechanism and cloud storage services,
- Follow internal security policies and procedures, applicable to each partner and dataset,
- Perform frequent server maintenance,

- Use robust end-to-end channel encryption and pseudonymization techniques.

EU-Restricted datasets (of classified deliverables) will be treated in accordance with National and European regulations, in particular Commission Decision (EU, Euratom) 2015/444 of 13 March 2015 on the security rules for protecting EU classified information [8], and the SAL-related matters. The identification of classified information and attribution to the legal rules will be conducted in classified deliverables in due course according to the DoA (SDN).

Moreover, personal data will be treated in accordance with the partners' policy only for the lifetime of the project where encryption will be applied in order to store it. More specifically, passwords saved as Salted Hash (SHA-256) Databases and customer data in the file system are encrypted (AES-256). Backups are also encrypted (AES-256).

According to the "Principles relating to processing of personal data" of the GDPR, Chapter II, Article 5 [7], with which all partners have to comply, personal data has to be:

a) Processed lawfully, fairly and in a transparent manner,
b) Collected and processed only for specific and legitimate purposes, and further process should not deviate from these purposes,
c) Limited and relevant to the purposes of the data collection process and sufficient enough,
d) Accurate and updated, when required,
e) Stored in formats which prevent identification of data subjects after the end of the time period for which data are processed,
f) Processed in a way that ensures data integrity and confidentiality.

The data controller has to abide by and be accountable for the above.

Finally, it is claimed that all data (digital or not) are preserved in secure locations by each research partner. However, all partners are required to apply all necessary data assessment procedures and risk mitigation policies, to manage to mitigate the impact of common threats.

5 Ethical Aspects

In the research field, a common challenge is the sharing of data when at the same time keeping the privacy of their owners intact. This is where the need for the protection of personal data comes to the surface. Data protection aims to regulate the collection of data, the processes that grant access to it, the data transmission, the preservation of data, as well as other actions that are susceptible to malicious exploitation and privacy violation. Privacy violation issues might occur either when the personal data shared under a research project is in a digital form or not. Therefore, proper measures need to be taken into consideration in order to tackle potential threats.

After all the contingent ethical issues have been listed and analysed, it can be concluded that the 5G-INDUCE project does not deal with any aspects related to ethics. What is more, according to Article 2, Section (a) of the Data Protection Directive (95/46/EEC)[9], it is decided that any information, data or analysis outcomes, shared among the partners, should not contain any personal data. Therefore, all consortium partners should proceed accordingly by eliminating and restricting any access to personal data before disclosing any information to the rest of the partners. In case of any deviation of a partner from the above, they have to abide by the following conditions:

1. The partner disclosing any personal data must be authorized to do so,
2. They need to have the consent of all the individuals and institutions related to the information disclosed, and make sure that all the necessary legal actions have been taken, and
3. No restrictions can apply on the disclosed data for further usage of them by the partners for the project's objectives.

Personal data are also protected by the European General Data Protection Regulation (GDPR) of 2016 regarding the "Protection of natural persons with regard to the processing of personal data and on the free movement of such data" [7][7][7], with which all partners are obliged to comply.

In the 5G-INDUCE project, no ethical issues are addressed. The partner companies do not intent to make any use of data which belong to special categories, neither to have children or vulnerable groups as data subjects. Additionally, the vast majority of the partners are not going to collect personal/sensitive data from people who have not given their explicit consent to be part of the project, and if anyone does, it is going to be data coming from public sources and that will be used only for communication and dissemination purposes. In addition to that, a couple of partners have claimed to engage in large scale and/or big data processing, and some others have already received consent for data preservation and sharing from data subjects. Finally, no partner intends to disclose any personal or sensitive information to entities outside the EU.

6 5G-INDUCE Use Cases Overview and Data Sourcing

5G-INDUCE focuses on the establishment of an end-to-end orchestration platform, which facilitates the experimentation on high-level 5G NetApps that can deal with various 5G use case scenarios in the industry 4.0 sector, in order to validate the NetApps capability of achieving main targets for 5G KPIs on a technical and business level (Fig. 2). This is achieved through the collaboration of various sectors, such as manufacturing, logistics, maintenance, power management, security/surveillance, and more. The main objective is to evaluate the 5G-readines in both the telecom and the applications fields, in order to tackle the barriers related to interaction and porting issues of NetApps in the industry 4.0 sector.

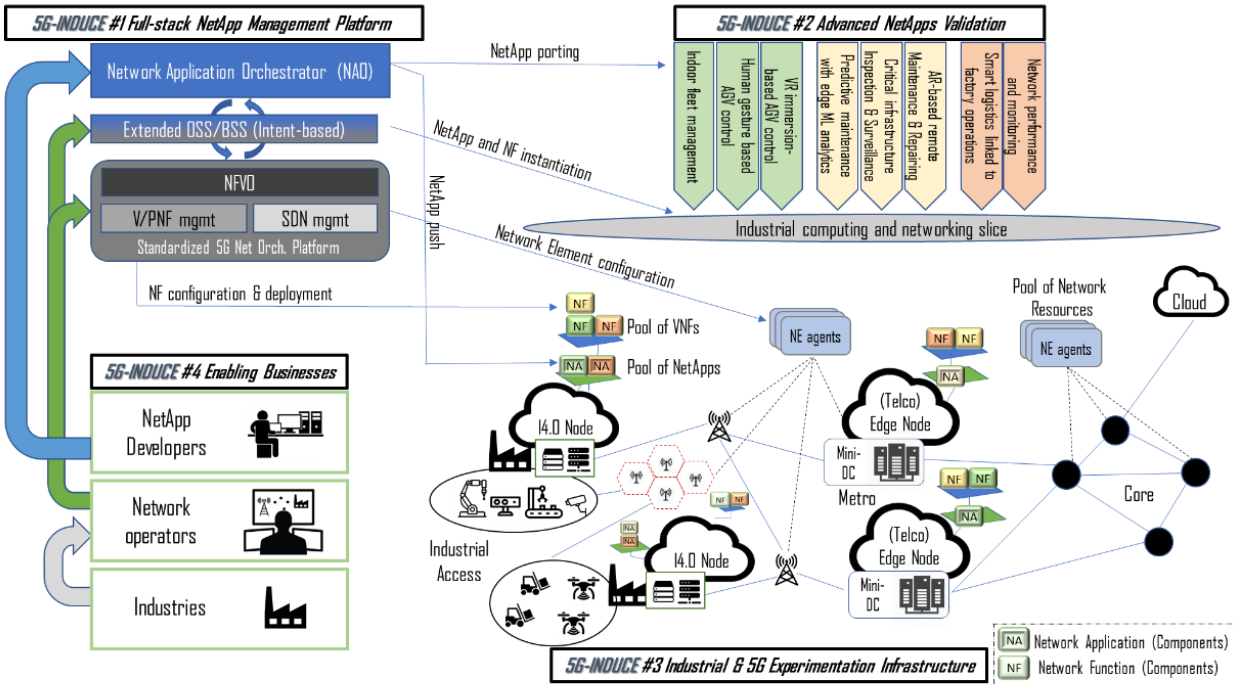
5G-INDUCE #1 Full-stack NetApp Management Platform
Network Application Orchestrator (NAO)
Extended OSS/BSS (Intent-based)
NFVO
V/PNF mgmt
SDN mgmt
Standardized 5G Net Orch. Platform
NF configuration & deployment
NetApp porting
NetApp and NF instantiation
NetApp push
Network Element configuration
5G-INDUCE #2 Advanced NetApps Validation
Indoor fleet management
Human gesture based AGV control
VR immersion-based AGV control
Predictive maintenance with edge ML analytics
Critical infrastructure Inspection & Surveillance
AR-based remote Maintenance & Repairing
Smart logistics linked to factory operations
Network performance and monitoring
Industrial computing and networking slice
Pool of VNFs
Pool of NetApps
NE agents
Pool of Network Resources
Cloud
I4.0 Node
(Telco) Edge Node
Mini-DC
Metro
Core
Industrial Access
5G-INDUCE #4 Enabling Businesses
NetApp Developers
Network operators
Industries
5G-INDUCE #3 Industrial & 5G Experimentation Infrastructure
NA Network Application (Components)
NF Network Function (Components)

Figure 2: The 5G-INDUCE vision and the adopted approach for easy (i) porting and/or (ii) development of industry 4.0 NetApps over advanced Experimentation Facilities combining real 5G and private industrial networks.

A short description of every use case of the 5G-INDUCE project is provided below. Each use case aims to accommodate the needs of different industry sectors' operation procedures, by implementing 5G network applications. In most of the use case scenarios listed below, more than one NetApp infrastructures are involved in order for the final service to be provided.

Use Case 1: Autonomous indoor fleet management

The main objective of use case 1 is focused on a small group of automated guided vehicles (AGV) that assist logistics operations both within and among Ford's warehouses and their suppliers' warehouses, using simultaneous localization and mapping (SLAM) navigation through 5G platforms.

Data Sourcing: The above will be achieved by establishing a set of AGV cameras and 3D laser sensors, which are going to be interconnected to smartphone apps and other NetApps, to assist the transmission of information to the AGV, by the controller, and thus determine its movements.

Use Case 2: Smart operation based on human gesture recognition

Use case 2 aims to guide the future of industrial experience and security, which lies in smart operation systems, through intelligent monitoring and natural gesture control. The main advantage of this attempt is the lack of extra, and probably costly, gear that would have to compose the controller's equipment.

Data Sourcing: The relevant information that will facilitate the above scenario will be obtained through High Definition (HD) video and audio and sensors. Therefore, the controller will be able to control a vehicle's movements physically, using oral commands and gestures, and this transmission of information will occur electronically, through smartphone apps and other NetApps connections.

Use Case 3: VR immersion and AGV control

Use case 3 also focuses on the industrial sector and AGV control. This time VR technology is used, in combination with a 5G network, to provide an immersive view from the AGV's perspective to a remote viewer. The main objective of this is to provide an inclusive update on all the vehicles' condition, but also to enhance services such as security and person recognition competence.

Data Sourcing: A camera will be installed on the vehicle, and, along with complementary information from sensors and NetApps, immersive visualizations will be sent to the viewer's VR equipment or mobile device. In this way, the viewer will be able to send and receive information from the AGV, to prevent a possible collision of the vehicle, and more.

Use Case 4: ML-Supported Edge Analytics for Predictive Maintenance

Use case 4 aspires to optimize industrial predictive maintenance, by utilizing machine learning techniques to perform federated learning using real-time data. In this way, the experimentation aims to increase prediction accuracy of distinct production lines operations, taking also into account the overall desired production and the industry's entire schedule. This process also aims to reinforce the decision maintenance process with enhanced security features.

Data Sourcing: Data for near real-time predictive maintenance will be collected from sensors and current production data bases. The resulting outcomes will be a maintenance schedule and failure probabilities. NetApps and mobile apps will be used for monitoring of the process and for receiving the relevant notifications.

Use Case 5: Inspection and surveillance services for critical infrastructures

The main objective of use case 5 is to build an enhanced surveillance system for industrial infrastructures, in order to identify on time any crucial corrosion of materials, rough handling and/or parlous operation levels. This AI-assisted monitoring of object status will mainly focus on identifying any potential malevolent presence of humans and/or animals.

Data Sourcing: A manually operated drone will provide all the relevant (encrypted) video information to the system, where it will be processed from a local 5G terminal or fog node. If any suspicious object is identified, the system will be alerted accordingly.

Use Case 6: AR-based remote maintenance, repairing and upgrade

Use case 6 focuses on the maintenance applications of complex industrial equipment and aims to attest that 5G platforms are able to enhance safety and security for remote assistance in maintenance applications.

Data Sourcing: For any workers' injury to be prevented and for any confidentiality breach to be avoided, encrypted audio and video streams from a live remote assistance videocall will be processed and analysed, possibly enabling camera steering and sensor management competence.

Use Case 7: Smart logistics over supply chain linked with factory operations

Use case 7, by utilizing a minimum number of vehicles-travels, aims to optimize supply chain operations such as the stock inventory, the supply route, and the damage detection. Continuous monitoring will be implemented in order for the appropriate external actions to be taken when required.

Data Sourcing: In order for any critical levels in the supply chain process to be detected, stock monitoring data, enterprise resource planning (ERP) data, operating data from critical equipment, as well as traffic, weather and route data will be collected. After the consumption and operational conditions are analysed, an optimized scheduling for the logistics operations will be produced. Finally, the live monitoring will enable immediate actions to be taken when critical values are detected, as well as external interactions when required.

Use Case 8: Drone assisted network performance and coverage monitoring for industrial infrastructures

Finally, use case 8 focuses on 5G smart factory environments, and aims to provide an end-to-end network performance and coverage monitoring of the critical communications infrastructure, using a drone-based monitoring and data collection system.

Data Sourcing: Monitoring can be both continuous and/or on-demand. In continuous mode, streaming data will be collected for real-time alerting and notifications purposes. In on-demand mode, the drone videos from the monitoring facility area will also be accounted in order for further analysis of the metrics to be conducted and the final optimization of the root-cause analysis to occur. In both cases, monitoring NetApps will be used in order to enable visual mapping of 5G radio, assisted by drones.

7 Conclusion

All in all, this deliverable is an explanatory description of the data collection and generation process that will take place under the scope of the 5G-INDUCE project and will serve the objectives of the use case scenarios. The DMP also provides a comprehensive outline of data access, sharing, and storage. This deliverable is dynamic and the DMP will be updated throughout the whole course of the 5G-INDUCE project, especially when significant changes to the crucial components of the project occur during the experimentation procedure.

The data collected and generated during the 5G-INDUCE project are subject to security measures and access limitation acts, especially when it comes to personal and confidential data. However, the project allows the sharing of research data among the partners within the project's repository, when applicable. These issues are explicatively addressed in this deliverable, as well as ethical aspects which might also have an impact on the confidentiality of the data. Additionally, the project abides by the FAIR data principles and, therefore, research data – free of privacy restrictions – and scientific publications should be made openly accessible via public repositories.

Finally, the datasets and the data analysis procedures, occurring from each different use case and pilot scenarios, will be specified during the course of the project and the experimentational procedures. The openness, or not, of each specific data information, will also be determined later on and the DMP will be updated accordingly. The progress of the 5G-INDUCE project will be reflected through the DMP, where all crucial updates will be incorporated, throughout the whole course of the project.

References

[1] [Online] R KIT – The Research University in the Helmholtz Association- Research data lifecycle https://www.rdm.kit.edu/english/researchdata_research_cycle.php, last accessed on 08/03/2021.

[2] [Online] Guidelines on FAIR Data Management in Horizon 2020, https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, last accessed on 08/03/2021.

[3] [Online] GO FAIR – FAIR Principles, https://www.go-fair.org/fair-principles/, last accessed on 19/03/2021.

[4] [Online] Metadata Standards Directory WG, https://www.rd-alliance.org/groups/metadata-standards-directory-working-group.html, last accessed on 08/03/2021.

[5] [Online] OpenAIR - EC policies and mandates, https://www.openaire.eu/ec-policies-and-mandates, last accessed on 19/03/2021.

[6] [Online] Good Data Protection Practice in Research https://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf, last accessed on 08/03/2021.

[7] [Online] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679, last accessed on 08/03/2021.

[8] 5G-INDUCE Consortium Agreement

[9] [Online] DIRECTIVE 95/46/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=EN, last accessed on 19/03/2021.

Appendix A

5G-INDUCE Data Management Questionnaire

Data Description		
1.	What data will you process/generate? Please name all the data sets you are going to use.	*Ex. Name and surname, organisation/ workplace, country of origin, position/ rank, email address, age, sex, signature, online identifiers, images, videos, audio recordings*
2.	What is the purpose of the data collection for each data type?	*Ex. Communication, dissemination, information sharing, conducting training, use cases, field demonstration*
3.	Please explain the relation of the data processing/generation to the objectives of the project for each data type.	*Ex. Name and surname, email address, will be processed for the following project objectives:* *(i) creation and maintenance of mailing lists* *(ii) creation and maintenance of SharePoint*
4.	Please explain where you anticipate each data type will come from (i.e., what is the origin of the data?) Are you using data someone produced by someone else? If so, please state your source.	*Ex. The data come from the data subject itself, data from other sources such as web and social networks media profiles will be processed too.*
5.	How will you be processing the data? Provide separate answers for each data type/category you will be processing, collecting, or generating.	*Ex. Partner's name will process data according to GDPR, by collecting and recording. These files will be stored, according to their form, using encryption software.*
6.	What stages will the data go through?	*Ex. Raw, processed, analysed, published/made available.*
7.	Please state the format(s) for each data type, in which you expect that type of data to be collected/generated. (numerical data, image data, text sequences, audio data etc.)	*Ex. Most data will be collected in the form of text. Numerical data will be also collected in the form of age.*
8.	For each type of data what is the expected size?	*Ex. text sequences (name, surname, etc.) 1GB extendable* *Multimedia (images, audio, video) 100GB pre-processed, 20GB post-processing, extendable*

Data Management Procedures		
1.	Does your organisation have data management guidelines? If so, what are they?	*Ex. Have you developed a data management plan for the company or the project?*
2.	Does your organisation have a data protection or security policy that you will follow? If so, what is it?	*Ex. Define data protection policies, data protection breaches, lawful data processing*
3.	Does your organisation have a Research Data Management policy? What is it?	
4.	Are there any formal standards you will adopt when processing data for the project?	*Ex. Consult Ethics experts, ask for DPO's opinion, follow the organisations data protection policy and security policy.*
5.	Please provide the contact details of your organisation's DPO.	*Ex. Name, email*

Documentation, Organisation and Storage		
1.	Who will be responsible for documentation, organisation, and storage regarding your organisation?	*Name*
2.	How will you label and organise data, records, and files?	*Please provide an example of naming or labelling format for a dataset or record.*
3.	What is the length of time that you will be storing personal data generated in the project, after its completion in M36?	*Ex. Data will be preserved for up to 5 years after 5G-INDUCE completion.*
4.	Please specify how and where will the data be stored?	
5.	Are you using proprietary file formats to generate and store your research data? If so, will the documentation about the software needed to access the data be included?	
6.	What related information needs to be stored along with the data?	*Ex. Software, reports, research papers, metadata*
7.	If data are collected with mobile devices, how will you transfer and store the data?	
8.	If data are collected with mobile devices, do you wipe or keep the data in the mobile devices after their collection and transmission to your main storage space?	*Ex. Data are removed within one week after collection and transmission from our proprietary mobile devices; in the case of mobile devices owned by third parties (e.g. survey participants' smartphones), the device owner is given the option to remove the data immediately after transmission.)*
9.	If data are held in various places, how will you keep track of versions?	

Access		
1.	Who will manage access to this data within your organisation?	*Name(s)*
2.	Who will have access to the data processed for the 5G-INDUCE project within your organisation?	*Name(s)*
3.	How will the identity of the person accessing the data be verified?	
4.	Will there be conditions to gaining data access? If so, what will those conditions be?	
5.	What methods or software tools will be needed to access the data?	
6.	Will any other accompanying information be required to properly interpret the data?	
7.	What information, if any needs to be retained for the data to be read and interpreted in the future?	
8.	Will any data be made openly available for all project partners?	
9.	How will the data that is made openly available be maintained? In a repository?	

Sharing		
1.	What data will be shared?	*Ex. Name and surname, organisation/ workplace, etc*
2.	When will data be shared?	*Ex. Describe the dataset sharing timelines, especially if you have more than one datasets to share with different data sharing schedules.*
3.	For what purposes of the project will data be shared?	
4.	With whom will the data be shared?	
5.	Does sharing raise any privacy, ethical, or confidentiality concerns? If yes, what are they and why?	
6.	Will a data sharing agreement be required? If some data categories and types are kept closed provide a rationale for doing so.	

Security		
1.	What are the major risks to your data's security?	
2.	How are you planning to manage each of these risks?	
3.	Do you need to anonymise any of your data?	
4.	Have you prepared a formal risk assessment addressing each of the major risks to data security and how you will minimize each risk? If the answer is no, do you plan to perform a formal risk assessment, and if so, when?	
5.	What security measures do you anticipate being required for safe data storage, sharing and management?	
6.	Have you developed any data recovery protocols?	
7.	Have you developed or outlined any protocols for the safe transfer of personal data?	
8.	Have you developed or outlined any protocols for the safe transfer of special categories of data?	
9.	Have you implemented or outlined any procedures to follow in the case of a data breach? If so, what are they?	
10.	Are your digital and non-digital data, and any copies, held in a safe and secure location?	

Ethical Considerations		
1.	Do you intend to generate/process any personal data belonging to a special category?	
2.	Will any of the data subjects be children?	
3.	Will any of the data subjects be vulnerable people?	
4.	Will you be collecting personal or sensitive data from people who have not given their explicit consent to participate in the Project?	
5.	Have you already gained consent for data preservation and sharing from any data subject(s)?	
6.	Will you engage in large scale or big data processing?	
7.	Will any entity (including any service provider) outside of the E.U. have access to personal or sensitive data? If so, who? For what purpose? Where is each of these entities located?	