FUDGE-5G

FUlly DisinteGrated private nEtworks
for 5G verticals

Deliverable 6.2

Data Management Plan

Version 1.0
Work Package 6

Main authors	Luís Cordeiro, André Gomes (ONE)
Distribution	Public
Delivery date	26 February 2021 (M06)
Delivered date	26 February 2021 (M06)

Partners

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement № 957242.

Disclaimer

Project details

Project title: FUlly DisinteGrated private nEtworks for 5G verticals
Acronym: FUDGE-5G
Start date: September 2020
Duration: 30 months
Call: ICT-42-2020 Innovation Action

For more information

Project Coordinator
Prof. David Gomez-Barquero
Universitat Politecnica de Valencia
iTEAM Research Institute
Camino de Vera s/n
46022 Valencia
Spain

http://fudge-5g.eu
info@fudge-5g.eu

Acknowledgement

FUDGE-5G has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement № 957242. The European Union has no responsibility for the content of this document.

Abstract

The present document covers the underlying issues related to data management in FUDGE-5G, both for qualitative and quantitative data generated during the project. It encompasses the methodology and procedures that FUDGE-5G has put in place to generate, store and publish results, which include deliverables, open-source contributions, scientific publications and datasets. Furthermore, the document covers ethics and privacy issues, describing how the project follows GDPR principles and how personal data is handled during the project.

Versioning and contributions

Versioning

#	Description	Contributors
0.1	Table of contents	ONE
0.2	Initial contributions	ONE
0.3	ONE contribution to FUDGE-5G datasets	ONE
0.4	IDE contributions to datasets	IDE
0.5	5CMM contribution to datasets related to the Industry 4.0	5CMM
0.6	THA contribution to datasets related to the PPDR UC	THA
0.7	UPV contribution to datasets related to the Media UC	UPV
1.0	First D6.2 version.	ONE

Contributors

Partner	Authors
ONE	Luís Cordeiro, André Gomes, Marco Sequeira, André Ribeiro, António Borges
IDE	Sebastian Robitzsch
5CMM	Manuel Fuentes
CMC	José Costa-Requena
THA	Filippo Rebecchi
UPV	Carlos Barjau Esteven

Reviewers

Reviewer	Partner

Abbreviations

5GC – 5G Core
CA – Consortium Agreement
CC – Creative Commons
CID – Content Identifier
CO – Confidential (Deliverables)
DMP – Data Management Plan
DOI – Digital Object Identifier
DPO – Data Protection Officer
EC – European Commission
ePD – ePrivacy Directive
EU – European Union
FAIR – Findable, Accessible, Interoperable, Reusable
FQDN – Fully Qualified Domain Name
GDPR – General Data Protection Regulation
H2020 – Horizon 2020
HIPAA – Health Insurance Portability and Accountability Act
IPR – Intellectual Property Rights
KPI – Key Performance Indicator
NPN – Non-Public Network
PPDR – Public Protection and Disaster Relief
PU – Public (Deliverables)
RAN – Radio Access Network
RSRQ – Reference Signal Received Quality
SCP – Service Communication Proxy
SFVO – Service Function Virtualization Orchestrator
SINR – Signal to Noise Ratio
SMF – Session Management Function
UCX – Use Case X
UE – User Equipment
UPF – User Plane Function
WPX – Work Package X

Executive Summary

This deliverable covers the issues and measures concerning data management in FUDGE-5G, which are necessary to comply with the requirements for projects within the Horizon 2020 pilot action on open access to research data. Hence, all the relevant EC guidelines have been taken into consideration and were the key guidance for document preparation.

The Data Management Plan (DMP) that is presented in this document starts by the identification of what are the datasets that will result from the project, which include: project deliverables, scientific publications, promotional publications, dissemination tracking, open-source contributions, experimentation datasets and showcasing datasets.

For each type of result, the DMP establishes a methodology for collection, processing, sanitizing and public release towards open access data. Furthermore, it is with this methodology that FUDGE-5G plans to meet the objective of following the Findable, Accessible, Interoperable and Reusable (FAIR) principles.

By following FAIR principles, FUDGE-5G also considers how data is stored, which means that different repositories can be used depending on the data to be stored: the FUDGE-5G website, a FUDGE-5G repository and Zenodo (created by the EU-funded OpenAIRE project). All the repositories follow a data backup policy to make information available not only during the project but also for at least 5 years after its ending. For any data publicly available, preference is given to Zenodo, and every data set has a Digital Object Identifier (DOI) that enables its easy and unique identification.

Finally, the DMP discusses ethics and privacy issues, establishing rules for how personal data is handled and defining a process that involves both the project's Data Protection Officer (DPO) and a representative DPO from all the involved partners.

Table of Contents

1 Introduction

FUDGE-5G is a H2020 funded Innovation Action project which will enable highly customized cloud-native deployment of private 5G networks in five vertical trials (leveraging the 5G-VINNI testbed): Media Showroom; PPDR; 5G Virtual Office; Industry 4.0; and Interconnecting private NW. In this context, the FUDGE-5G Data Management Plan (DMP) gives an overview of the data and information collected throughout the project and specifies which data will be open access and which will be confidential within the consortium.

The FUDGE-5G DMP describes the data management life cycle for the data to be collected, processed and/or generated. This document addresses the critical aspect of making research data findable, accessible, interoperable and re-usable (FAIR), including details on how to handle research data during and after the end of the project; what data will be collected, processed and/or generated; which methodology and standards will be applied whether data will be shared/made open access; and how data will be curated and preserved (including past the end of the project). All tasks to be carried out by FUDGE-5G will consider ethical problems, privacy and data management as a fundamental part of its activities. There will be continuous monitoring to ensure compliance with H2020 standards and regulations, including GDPR data protection.

FUDGE-5G will participate in the Open Research Data Pilot and is committed to make research data accessible keeping data FAIR. Data gathered during the trials, related to the objectives of the project, will be made publicly accessible following the open data management guidelines of the H2020 programme.

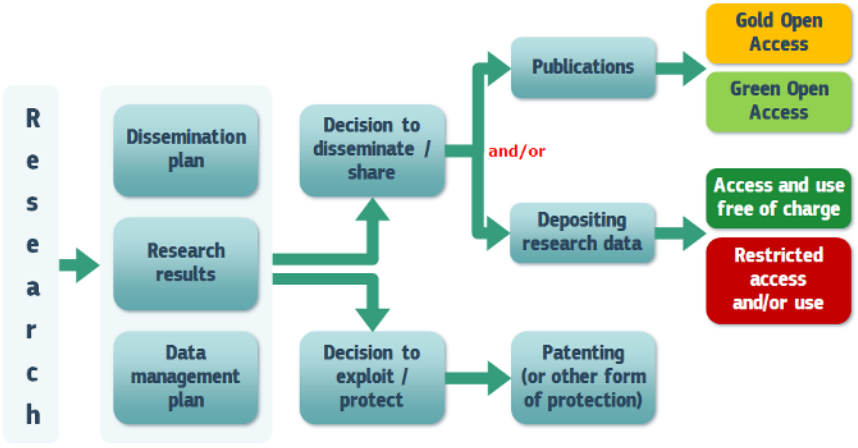

Figure 1 - Research data open access (from H2020 Open Access Online Manual)

The project will generate different types of data, requiring different ways of making the data available so external parties can validate the results. These results will also be presented in scientific publications through open access publishing. FUDGE-5G considers open research data as a keystone in advancing EU research and fostering innovation. In

order to ensure open access, for a maximum diffusion of project results, submitted version of IEEE papers and journals will be deposited in open access repositories.

The H2020 Open Access Online Manual decision flow, depicted in Figure 1, will be followed by the project. Whenever possible, the project will choose to disseminate results and publish them in Green Open Access repositories.

In this document, among others, details of procedures for data collection, anonymization, storage, protection, retention, destruction, and re-use. The DMP is a living document, which will evolve during the project, particularly whenever significant changes arise such as dataset updates, changes in the consortium or external sources. This document is the first version of the DMP, and the upcoming versions (depicted in Table 1) will provide more detail on datasets to be generated by the project, data interoperability and data management procedures.

Table 1 - DMP versions

	Date	Description
v1	M06	This document.
v2	M19	Document review based on DMP application in the project activities, including possible required adaptations and improvements to the definitions and methodologies. Refined version of the vertical trials and showcases datasets.
v3	M29	Final version reporting the implemented data management processes. Project assessment with regards to the data management.

This document made use of the Horizon 2020 Fair Data Management Plan Template[1], the Guidelines to FAIR data management in Horizon 2020[2] [3] and the GDPR Regulation (EU) 2016/679[4] [5] and Directive (EU) 2016/680[6].

[1] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm
[2] https://www.force11.org/group/fairgroup/fairprinciples
[3] https://www.nature.com/articles/sdata201618
[4] https://eur-lex.europa.eu/eli/reg/2016/679/oj
[5] https://ec.europa.eu/info/law/law-topic/data-protection_en
[6] http://data.europa.eu/eli/dir/2016/680/oj

2 Data summary

In order to provide an overview of the different datasets that will be produced in the FUDGE-5G project, we start by identifying in Table 2 the FUDGE-5G project expected datasets.

Table 2 - FUDGE-5G datasets

Dataset	Description
Project deliverables	The project will produce several deliverables according to the Grant Agreement and presented in the project website. These deliverables are either public (PU) or confidential (CO).
Scientific publications	Consortium partners throughout the project activities, in the scope of WP and tasks, will produce scientific publications, such as, journals papers, conference papers, white papers, books, workshops, tutorial sessions, etc., that will be made publicly available for the wider audience.
Other dissemination and communication publications	In the scope of the project Exploitation, Standardization and Dissemination activities (WP5), other publications will be produced like website pages, promotional materials (e.g., newsletter, flyers, posters, videos), press releases, website news, posts (e.g., Twitter, LinkedIn, blogs).
Data tracking	The FUDGE-5G website and social media platforms collect user's data. This is used to monitor the project dissemination and communication activities and audience interest.
Open-source contributions	FUDGE-5G partners are committed to contribute to open-source projects and release software as new open-source projects. It is expected that these initiatives are focused to one partner or a limited group of partners, and with specific licence requirements.
Tests and trials	The project outcomes will be tested and trialled in five verticals leveraging 5G-VINNI infrastructure and several other project and vertical industries partners infrastructures. These will produce and process datasets such as infrastructure (radio, 5GC, cloud, network, etc.) usage data, applications data, vertical specific data, KPI related data, etc.
Demos and showcases	Similarly to the tests and trials, the project will organize several vertical demos and showcases events where datasets will be produced and processed.

FUDGE-5G datasets processing will comply with the different data protection regulations that apply on FUDGE-5G and will be led by each project partner Data Protection Officer (DPO) in collaboration with the FUDGE-5G DPO. The DPO main role is to ensure that personal data are processed in compliance with the data protection rules. Table 3 identifies the FUDGE-5G project DPO and the Consortium partners DPOs.

Table 3 - FUDGE-5G DPOs

#	Partner	DPO	Email
-	FUDGE-5G	Luís Cordeiro, ONE	cordeiro@onesource.pt
1	UPV	Carlos Barjau	carbare1@iteam.upv.es
2	TNOR	Kashif Mahmood	kashif.mahmood@telenor.com
3	ATH	Alan Dahi	dpo@athonet.com
4	CMC	Mika Skarp	mika.skarp@cumucore.com
5	FHG	Pousali Chakraborty	pousali.chakraborty@fokus.fraunhofer.de
6	O2M	Peter Sanders	peter.sanders@one2many.eu
7	UBI	Eleonora Papatsoutsou	epapatsoutsou@ubitech.eu
8	ONE	Luís Cordeiro	cordeiro@onesource.pt
9	5CMM	Manuel Fuentes	manuel.fuentes@fivecomm.eu
10	IDE	Sebastian Robitzsch	sebastian.robitzsch@interdigital.com
11	HWDU	Joerg Thomas	joerg.thomas@huawei.com
12	THA	Candice Zimmermann	candice.zimmermann@thalesgroup.com

Depending on the dataset different responsibilities may apply. Table 4 identifies the default responsibilities for the datasets.

Table 4 – Datasets management responsibilities

Dataset	Responsible	Key Partner
Project deliverables	Editor, PMT	Editor
Scientific publications	Editor, PMT	Editor
Other dissemination and communication publications	Communication Manager	ONE
Data tracking	Communication Manager	ONE
Open-source contributions	Contributor, PMT	Contributor
Tests and trials	Trial owner, WP4 leader	Trial owner

Demos and showcases	Trial owner, WP4 leader	Trial owner

To ensure the compliance with the DMP each dataset lifecycle will follow the methodology depicted in Figure 2. This methodology is composed of the following main steps:

- Step 1: The data will be collected from the experiment.
- Step 2: Each partner collecting/generating datasets is responsible for following this document rules. The partner respective DPO must ensure that all required procedures are properly implemented.
- Step 3: The FUDGE-5G DPO reviews the datasets collected/generated and ensures they are in accordance.
- Step 4: The generated and processed data will be store following the DMP procedures.
- Step 5: The FUDGE-5G DPO reviews the store procedure and ensure that the DMP procedures were implemented.
- Step 6: The dataset responsible partner follows the DMP procedures for dataset distribution to make the respective dataset available.

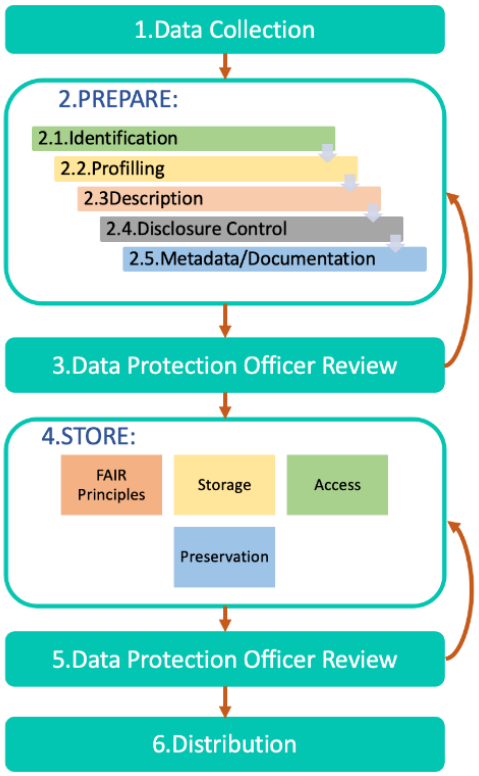

Figure 2 - Data management methodology

Next are presented the key processes, requirements and methodologies that will be applied to all FUDGE-5G datasets throughout the project and beyond the project activities.

2.1 Collection and generation

As a research and innovation project data will be collected from existing sources like scientific publications, open access datasets and standards, and generated by project partners or throughout project activities. The motivation for this is to facilitate the technological evaluation proposed by FUDGE-5G and support its trials evaluation and showcase in the project five verticals. Only data required to achieve the projects objectives will be collected and generated, and, if possible, no participants personal data will be asked or recorded.

2.2 Format

The choice of data format should consider open, well-documented and non-proprietary formats whenever possible. Long and short-term formats, dissemination and preservation formats must be assess depending on the purpose, whether to analyse, store and share. Table 5 presents the most relevant formats that are considered for FUDGE-5G.

Table 5 - Types of data

Type	Recommended formats	Acceptable formats
Textual Documents	• Plain text, ASCII (.txt) • XML (.xml) • JSON (.json) • Adobe PDF (.pdf)	• Hypertext Mark-up Language (.html) • MS Word (.doc/.docx) • Software-specific formats
Databases	• Comma-separated values (.csv) • MS Excel (.xls, .xlsx) • Clear text files (.txt) • Machine formats (.json)	• SQL (.sql)
Image	• JPEG (.jpeg, .jpg) • GIF (.gif) • PNG (.png)	• TIFF (.tif, .tiff) • Photoshop files (.psd) • BMP (.bmp)
Audio	• FLAC (.flac) • MPEG-1 Audio Layer 3 (.mp3)	• Audio Interchange File (.aif) • WAV (.wav)
Video	• MPEG-4 (.mp4) • OGG video (.ogv, .ogg) • Motion JPEG 2000 (.mj2)	• AVCHD video (.avchd)

2.3 Identification

FUDGE-5G identifies all generated and collected data with a FUDGE-5G unique identifier that provides an internal system for persistent and actionable datasets identification. When making publicly available the curated FUDGE-5G datasets, the project will rely on the platform identification mechanism (usually based on a Digital Object Identifier, DOI).

The FUDGE-5G internal identifier follows the project documentation identification and complements it with additional dataset specific identifiers:

FUDGE-5G_[type]_[UC]_[date]_[name]_[version]_[policy]

Where we identify the following fields:

- “type” describes the type of data (e.g., deliverable, database, video, code, publication, measured data) [REQUIRED].
- "UC" the use case whose trials generated the data (UC1, UC2, etc.) [OPTIONAL].
- “date” is the date in the format “YYYYMMDD” [REQUIRED].
- “name” is a short name for the data [REQUIRED].
- “version” is the version of the dataset in the format v(MAJOR).(MINOR) [REQUIRED]
- “policy” is the polity to apply to the dataset (e.g., restricted, embargoed, open)
- “_” (underscore) is used as the separator between the fields.

2.4 Data Processing

All deliverables, or in another way of reporting, that imply data possessing require a written note by the FUDGE-5G DPO and all consortium partners DPOs (see Table 3). The note must explicitly confirm and explain the compliance with the consortium privacy policy and EU regulations (that include GDPR).

2.5 Profiling

FUDGE-5G will perform data profiling on the collected or generated data. Data profiling aims to discover and investigate data quality issues, such as duplication, lack of consistency, and lack of accuracy and completeness. This is accomplished by analysing one or multiple data sources and collecting metadata that shows the condition of the data and enables the data steward to investigate the origin of data errors. Data profiling results metadata will be represented as data rules bound to the datasets made available.

The main data rules to be assessed will be:

- **Distinct count and per cent**: Analysing the number of distinct values within each column will help identify possible unique keys within the source data.
- **Zero, blank, and NULL per cent**: Analysing each column for missing or unknown data helps you identify potential data issues.
- **Minimum, maximum, and average string length**: Analysing string lengths of the source data is a valuable step in selecting the most appropriate data types and sizes in the target database.
- **Numerical and date range analysis**: Gathering information on minimum and maximum numerical and date values is helpful to identify appropriate data types to balance storage and performance requirements.
- **Key integrity**: After all-natural keys have been identified, the overall integrity by applying the zero, blank, and NULL per cent analysis to the data set. In addition, checking the related data sets for any orphan keys is extremely important to reduce downstream issues.

- **Cardinality**: Identification of the cardinality (e.g., one-to-one, one-to-many, many-to-many, etc.) between the related data sets is important for database modelling and business intelligence (BI) tool set-up.
- **Pattern, frequency distributions, and domain analysis**: Examination of patterns is useful to check if data fields are formatted correctly. This type of analysis can be applied to most columns but is especially practical for fields that are used for outbound communication channels.

To achieve this, a data profiling tool is recommended. Table 6 presents a list of possible data profiling tools that can be used in FUGDE-5G.

Table 6 - Profiling tools

Tools	Characteristics	Features
Atlan's data profiling software[7]	• Commercial application	• Auto-Generated Data Profiles • Quick Visual Analysis • Automated Anomaly Detection
Open Studio for Data Quality[8]	• Free open-source Apache license • Eclipse-based tooling • Windows and Mac OS versions	• Advanced statistics with indicator thresholds • Column set analysis • Advanced matching analysis • Time column correlation analysis
OSDQ[9]	• Open Source • Restful API • Apache Spark-based data quality • Windows and Unix versions • Detailed installation guides	• Export and import from XML, XLS or CSV format, PDF export • File Analysis, Regex search, Standardization, DB search • Statistical Analysis, Reporting (dimension and measure based), Ad Hoc reports and Analytics • Metadata Information, Reverse engineering of Data Model

The tool OSDQ will be considered as the project reference tools for data profiling because it is a free opensource tool that provides all the features required, and it can deal with all FUDGE-5G recommended formats. Nevertheless, each partner will be able to select a different tool.

[7] https://atlan.com
[8] https://www.talend.com/
[9] https://github.com/arrahtech

2.6 Description

High-quality data descriptions are required to understand the nature and provenance of data, including what the data is, the format the data is represented with, where the data can be retrieved from, what license associated with the dataset, how it was generated, when it was generated, and by whom it was generated. Such dataset descriptions should provide globally unique identifiers for specific versions and formats of datasets so that they may be used and referenced by others in downstream analyses. Table 7 presents the description template for the different dataset's types, aiming to improve and maximize access to and re-usage of the project generated data.

Table 7 - Dataset description template

Field	Details
Identifier	Each dataset will have a unique identifier
Name	Name of the dataset
Description	The dataset description will have a detailed report regarding where it was collected, what is the source and data will and can be used. If there is data that is being re-used, a reference is made.
Policy	One of the following sharing policies: • Restricted: The data is only available for project internal use. • Embargoed: The end of the embargo period is disclosed. After that date, the dataset will become public and widely available. • Open: The dataset is public and can be used by other persons or entities outside the project.
Licence	The dataset licence (e.g.: APACHE GPL, CC).
Origin	Identification of the dataset origin.
Responsibility	FUDGE-5G WP, task and partners responsible for the dataset.
Scale	Estimation of the dataset size.
Target	Description of the dataset target audience.
Format	Description of the dataset format.
Metadata	FUDGE-5G datasets are required to have detailed metadata files, describing completely the information contained. It will have descriptive metadata, structural metadata, technical metadata, administrative metadata, preservation metadata and profiling

	metadata. The metadata format will follow the dissemination platform requirements and specifications.
Security & storage	The location of storage and time period in which the dataset will be preserved.

2.7 Disclosure control

FUDGE-5G aims to apply a disclosure control regime that maximises data utility whilst minimizing disclosure risk. The datasets collected/generated by FUDGE-5G will be subject to the application of different disclosure methods, which are described in Table 8. Disclosure control methods usually harm data utility, comprising the data's analytical completeness and its analytical validity.

Table 8 - Disclosure methods

Method	Description	Application
Anonymization	The anonymization of data records, which is also referred to in the literature as de-identification, consists of removing from each data vector the formal identifiers before the data is disseminated.	Datasets with personal data that can lead to individual identification.
Suppression	Suppression will be applied when an extreme value or an extreme combination is contained in a data record. This data record will be unique in the sample and in the population, thus providing the means to identify a respondent and hence to disclose sensitive variables.	When very singular data records exist on the dataset that can lead to individual identification.
Reduction in detail	When the data is represented in categories defined by a range and there are one or more categories with a low frequency (allowing the individual identification). The number of categories is cut down - increases the range and absorbs the low-frequency categories.	When categorized data presents categories with low frequency.
Top Coding and Bottom Coding	When data is presented in intervals. If the lowest intervals and the highest have very low frequency, singularities in the sample and/or in the population occurs, which can lead to the identification of the respondent and then to the disclosure of sensitive variables. Therefore, the range of the interval for the lowest and the highest groups is modified.	Datasets with data in ranges that have the lower or highest ranges with low numbers of occurrences.

2.8 Re-usage

The FUDGE-5G project will make publicly available datasets to freely support other to access, mine, exploit, reproduce and disseminate. It is recommended to use copyright licences that protect the authors of specific rights but at the same time allowing free and open re-usage. All partners collection/generating datasets will license their data to allow the widest possible reuse and will make their data to third parties in public repositories. FUDGE-5G partners are recommended to use licences such as Creative Commons[10] (e.g., CC-BY and CC0). When datasets are published in a data repository, a license agreement will be applied to the data. Datasets accessibility will be extended for a certain period after the completion of the project. By default, the data will be made available for reuse. If there are any restrictions, an embargo period can be identified to keep the data private for a certain period.

Thus, FUDGE-5G will increase data reuse through:

- All the project partners collection/generating datasets will license the datasets to allow the widest reuse possible.
- By default, the data will be made available for reuse.
- Data will be made available for third parties through public repositories.

2.9 Reproducibility of Results

The reproducibility of data is a measure of whether results in an experiment can be attained by a different research team, using the same methods. This shows that the results obtained are not artefacts of the unique setup in one research lab. Reproducibility is desirable, as it reinforces findings and protects against rare cases of fraud, or less rare cases of human error, in the production of significant results. If an observation is reproducible, it should be able to be made by a different team repeating the experiment using the same experimental data and methods, under the same operating conditions, in the same or a different location, on multiple trials.

FUDGE-5G aims to make all the results obtained reproducible by producing reports containing:

- A description of all methods, instruments, materials, procedures, measurements, and other variables involved in the study.
- A description of the analysis of data and decisions for the exclusion of some data and inclusion of others.

[10] https://creativecommons.org/

- For results that depend on statistical inference, a description of the analytical decisions, including when these decisions were made and whether the study is exploratory or confirmatory.
- A discussion of the expected constraints on generality, such as which methodological features we think could be varied without affecting the result and which must remain constant.
- Reporting of precision or statistical power.

2.10 IPR

The Consortium Agreement (CA) includes IPR related rules about partners rights to intellectual property. IPR will be managed according to the CA.

2.11 Security

FUDGE-5G will apply at all stages of dataset lifecycle the appropriate security principles (e.g., confidentiality, integrity, availability, non-repudiation) in order to ensure data protection. At collection/generation and disclosure control phases partners will have to ensure data security following their internal data management policies (with close monitoring by the FUDGE-5G DPO). The storage, access and presentation phases of the project datasets will be managed by the project and will rely on the selected repositories security mechanisms and policies (c.f. Section 2.12).

2.12 Repositories

FUDGE-5G project uses several tools for communication, organization/management and data storage. For datasets storage the project has identified three repositories: the project website, the project repository and the Zenodo[11] platform. Other tools like the project Wiki (supported by the Confluence platform[12]), source code management (supported by the GitLab platform[13]), Twitter, Linked, etc., are not considered as dataset repositories by the project. All Consortium partners include their own internal tools and methodologies that are independent from the project ones.

2.12.1 FUDGE-5G website

The FUDGE-5G website (https://fudge-5g.eu) will be used as a project communication tool, and where several of the project datasets will be made available to the public. The website will mainly contain public project details, news and events, the public project deliverables

[11] https://about.zenodo.org/
[12] https://www.atlassian.com/software/confluence
[13] https://about.gitlab.com/

and other dissemination and communication materials (e.g., white papers, flyers, posters, papers). In many cases the website will refer the user to the Zenodo repository, but many of the public project datasets will also be available on the website. The website has a private area restricted to the communication team. The FUDGE-5G website is hosted in ONE datacentre in Coimbra, Portugal, and has a backup policy that implements a thirty-day history daily backup and is 24/7 monitor by a maintenance and support team. The website development, deployment and maintenance follow the most modern and strict security requirements and guidelines. FUDGE-5G website will be made available during the project and for at least five years after the project end.

2.12.2 FUDGE-5G repository

FUDGE-5G repository (https://cloud.fudge-5g.eu) is the reference repository for all project datasets, both confidential, embargoed and open. All partners must storage their datasets in this repository. This repository is not intended to make available any dataset publicly and aims to be a project internal only repository. This repository is a deployment of the NextCloud platform[14], hosted in OneSource datacentre in Coimbra, Portugal, and has a backup policy that implements a thirty-day history daily backup and is 24/7 monitor by a maintenance and support team. This platform is HIPAA[15] and GDPR[16] compliant and includes auditing and file access control policies. Regarding access, the FUDGE-5G repository implements the following security policies: authentication, each project partner has its own authentication credentials to access the repository, and access is possible only to authenticated user (two-factor authentication is available); authorization, access control policies and mechanisms are implemented (including restricted physical access) and enforced, being the responsibility of each dataset owner the management of such datasets access by other partners users; accounting: a secure log policy is implemented to register any access and modification to a resource by any user. FUDGE-5G repository will be made available during the project and for at least five years after the project end.

2.12.3 Zenodo

FUDGE-5G selected the Zenodo[11] online platform as the project reference public dataset repository. Zenodo is a general-purpose open-access repository operated by CERN[17] OpenAIRE program. Zenodo stores files and metadata. Zenodo provides version control and assigns DOIs to all uploaded elements. Zenodo is an open and accessible repository that enables access to data without restrictions. The files are stored in CERN EOS[18] service disk

[14] https://nextcloud.com/
[15] https://www.hhs.gov/hipaa
[16] http://data.europa.eu/eli/reg/2016/679/oj
[17] https://about.Zenodo.org/infrastructure/
[18] https://eos-web.web.cern.ch/eos-web/

cluster. Metadata and persistent identifiers are stored in a PostgreSQL DB[19] with a 12-hour back-up operation with one backup sent to storage once a week. Zenodo will retain datasets for the lifetime of the repository, at least 20 years[20]. Zenodo uses a JSON schema as the internal representation of metadata and offers export to other formats such as Dublin Core[21], MARCXML[22], BibTeX[23], CSL[24], DataCite[25] and export to Mendeley[26]. The data record metadata will make usage of the vocabularies applied by Zenodo. For certain terms, these refer to open, external vocabularies, e.g.: license (Open Definition[27]), funders (FundRef[28]) and grants (OpenAIRE[29]). Reference to any external metadata is done with a resolvable URL.

[19] https://www.postgresql.org/
[20] https://about.Zenodo.org/policies/
[21] https://dublincore.org/
[22] https://www.loc.gov/standards/marcxml/
[23] http://www.bibtex.org/
[24] https://citationstyles.org/
[25] https://datacite.org/
[26] https://www.mendeley.com/
[27] https://opendefinition.org/
[28] https://www.crossref.org/services/funder-registry/
[29] https://www.openaire.eu/

3 FAIR

Projects financed by the European Commission must develop a Data Management Plan and deposit data in open access following the FAIR principles: Findable, Accessible, Interoperable and Reusable. The FAIR principles were generated to improve the practices for data management and data curation, and FAIR aims to describe the principles in order to be applied to a wide range of data management purposes, whether it is data collection or data management of larger research projects regardless of scientific disciplines. With the endorsement of the FAIR principles by H2020 and their implementation in the guidelines for H2020, the FAIR principles serve as a template for lifecycle data management and ensure that the most important components for lifecycle are covered.

These principals do not affect the implementation options and do not necessarily suggest any specific implementation technology, standard or solution. At the same time, there are datasets, or parts of datasets, generated in this project that cannot be shared in order to protect the privacy of voluntary participants in the pilots.

The FAIR principles were generated to improve the practices for data management and data curation, and FAIR aims to describe the principles in order to be applied to a wide range of data management purposes, whether it is data collection or data management of larger research projects regardless of scientific disciplines.

Table 9 - FAIR actions overview

Findable	Accessible	Interoperable	Re-usable
Discoverability of data	Data openly available	Standard vocabulary or mapping to commonly used ontologies	Data available for reuse
Data detailed in a specification document	Unshared data		Usability of data by third parties
Identifiability of data	Data availability	Data licensing for wide reuse	Restrictions on data re-use
Clear versioning approach	Methods or software tools for the data access		Quality assurance process
Search keywords approach	Software documentation		Length of time of data re-usability
Standards or procedures for metadata creation applied	Repository for depositing data, metadata, documentation and code		
Naming conventions used	Access restrictions		
	Data interoperability assessment of the high level		

FUDGE-5G will implement the FAIR principles through the following actions.

3.1 Findable

The data that will be generated in the project should be easily discovered by research communities. For this reason, one needs to ensure that data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

Making data findable includes provisions for metadata:

- the datasets will have very rich metadata to facilitate the findability (c.f. Section 2.6).
- all the datasets will have a Digital Object Identifiers (c.f. Section 2.3).
- the standards for metadata will be defined for each dataset (c.f. Section 2.6).

3.2 Accessible

The H2020 Open Access Mandate aims to make research data generated by H2020 projects accessible with as few restrictions as possible, but also accepts the protection of personal or sensitive data due to privacy concerns and/or commercial or security reasons.

All public datasets, scientific publications and deliverables will be made openly available, free of charge. Some datasets will not be shared due to privacy concerns or protection for commercial exploitation. If such cases arise during the project, this decision will be included in the final version of the DMP.

3.3 Interoperable

The project will evaluate data interoperability and specify which data vocabularies and metadata, standards or methodologies will be followed to facilitate interoperability, as referenced in Section 2.9.

3.4 Re-usable

The FUDGE-5G project will enable third parties to access, mine, exploit, reproduce and disseminate (free of charge for any user) all public data sets (c.f. Section 2.8).

4 Allocation of resources

The consortium will use free-of-charge repositories for making the datasets publicly accessible (c.f. Section 2.12). On the other hand, costs will incur, especially in personnel resources making the data available and maintaining the live data. The costs related to open access will be claimed as part of the Horizon 2020 grant. The costs to make the data FAIR in FUDGE-5G will be handled by each partner who generates the data. Costs are expected to be in the range of hundreds to a few thousands of Euros per annum. After the project, each partner will bear its own costs. ONE will support the post project costs for preserving the FUDGE-5G website and repository.

5 Ethics

This section describes an analysis of the application of GDPR and ePrivacy to FUDGE-5G project. FUDGE-5G ethics and privacy management are based on three main concepts:

- Confidentiality and anonymity – Confidentiality will be guaranteed whenever possible. The only exemption can be in some cases for the project partners directly interacting with a group of participants (e.g., focus group). The Consortium will not make publicly accessible any personal data. Anonymity will be granted through generalization.
- Informed consent – The informed consent policy requires that each participant will provide his/her informed consent prior to the start of any activity involving him/her. All people involved in the project activities (interviews, focus groups, workshops) will be asked to read and sign an Informed Consent Form explaining how personal data will be collected, managed and stored.
- Circulation of the information limited to the minimum required for processing and preparing the anonymous open data sets - The consortium will never pass on or publish the data without first protecting participants' identities. No irrelevant information will be collected; at all times, the gathering of private information will follow the principle of proportionality by which only the information strictly required to achieve the project objectives will be collected. In all cases, the right of data cancellation will allow all users to request the removal of their data at any time.

The next sections describe the project GDPR and ePD compliance.

5.1 GDPR

In May 2018, the new European Regulation on Privacy, the General Data Protection Regulation (GDPR) came into effect. In this section, it is described how the founding principles of the GDPR will be followed in FUDGE-5G.

According to the GDPR, the processing of personal data necessary for the exercise of public authority is lawful. That is, within the scope of its legal incubations, the Public Administration has the legitimacy to proceed with the processing of the personal data necessary for its activity without the need to obtain the consent of the data owner, which is a citizen. However, despite their legitimacy, the principles of data protection cannot be applied and ensure that they are treated only for the fulfilment of specific, explicit and legitimate purposes, guaranteeing the rights and freedoms of the data subject. In this way, the GDPR obliges entities that relate to natural persons in the EU to comply with the rules of information, transparency and loyalty, so that holders can understand who uses their data, for what purpose and for how long.

Personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject. In the FUDGE-5G project, all data gathering from individuals will require informed consent to individuals who are engaged in the project. Consent is an act of agreement given in a totally free, spontaneous and informed manner, must be provided for each of the purposes for which it is requested and cannot be given in aggregate. Consent must be unambiguous, so there can be no pre-filling of consent or assuming that it is given due to no response. If the data in question is sensitive, the GDPR reinforces that the consent must be explicit. The FUDGE-5G project, when requesting data from an individual, will inform the data subject about the purpose of the request, following the principles of legality, loyalty and transparency:

- Law.
- Conclusion of a contract.
- Defence of vital interests.
- Consent.
- Purpose limitation.
- Minimization of data (relevant data and extracts).
- Accuracy (eliminate those that are not correct).
- Conservation limitation (deadlines).
- Integrity and confidentiality (protect against loss).

In this way, informed consent requests to individuals will contain:

- Information letter and a consent form.
- Description of the specific causes for the activity.
- Description of how the data will be handled, safely stored, and shared.
- Information of individuals rights – have the data updated or removed.
- Information of project's policies on how rights are managed.

Thus, the following rules will be applied:

- The participation of all involved is voluntary.
- At any time, participants can refuse or cancel participation.
- All groups of individuals who are unable to freely give their consent will be excluded.

Personal data shall be adequate, relevant and limited to what is necessary for relation to the purposes for which they are processed. Only data that is relevant for the project purpose will be collected. However, since the involved stakeholders are free in their answers, this could result in them sharing personal information that has not been asked for by the project. These data will be treated according to all guidelines on personal data and will not be shared without anonymization or explicit consent of the stakeholder. The

FUDGE-5G consortium will try to anonymize the personal data as far as possible, however, this may not be possible in all instances. Therefore, additional consent will be requested to use the data for open research purposes, where qualitative and quantitative data from the user's research carried out on the project will be made available so that external parties can validate the results. These results may include presentations at conferences, publications in journals, as well as the deposit of a data set in an open repository at the end of the project.

Personal data shall be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed. All personal data that will no longer be used for research purposes will be deleted as soon as possible. All personal data will be made anonymous as soon as possible. At the end of the project, if the data has been anonymised, the data set will be stored in an open repository. If data cannot be made anonymous, it will be pseudonymised as much as possible and stored for a maximum of the partner's archiving rules within the institution.

Personal data shall be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures. All personal data will be handled with appropriate security measures applied. All people with access to the personal data files will need to sign a confidentiality agreement.

FUDGE-5G establishes the security and protection of personal data, which result in the rules established by the GDPR within the scope of the fundamental right to privacy and in the implementation of technical mechanisms that prevent unauthorized or illegal access to the holder of personal data. Security completes the protection of personal data with the implementation of secure rooms where data is generated, allowing access only to certain people.

FUDGE-5G treats personal data when carrying out an operation or a set of operations on personal data or on a set of personal data, either by automated or non-automated means. Such operations consist of the collection, registration, organization, structuring, conservation, adaptation or alteration, recovery, consultation, use, dissemination by transmission, diffusion or any other form of availability, comparison or interconnection, limitation and erasure or destruction. In other words, it is practically impossible to find an operation on personal data that is not handled by FUDGE-5G.

The GDPR will be applied to the same treatment of personal data, whether in digital or paper format. With this, data processing is done in a transparent and legal manner. Data are kept only as long as they are needed, stored safely (avoiding unauthorized and unauthorized use) and used for only for the specific purposes it was requested for (scientific purposes). The FUDGE-5G will collect some photos and videos of your activities. The

processing of these elements will comply with all best practices and legal obligations. In this way, also in these elements, no element containing personal information of an individual will be disclosed without giving his consent.

The FUDGE-5G shall be responsible for and be able to demonstrate compliance with the GDPR. At the project level, the DPO is responsible for the correct data management within the project.

The entire data protection process will be documented where the following requirements regarding privacy, data protection and security will be defined:

Minimization: the FUDGE-5G must treat only relative data (that is, the personal data that is provided for the conduct of the project) on the participants, where a complete mapping of the processing of personal data will be made.

Transparency: the project will inform data subjects about what data will be stored, to whom this data will be transmitted and for what purpose, and about the places where data can be stored or processed, defining priorities in the treatment of personal data.

Consent: consents must be handled allowing the users to agree with the transmission and storage of personal data. An applicant, who does not provide this consent for data necessary for the participation process, will not be allowed to participate.

Purpose specification and limitation: personal data must be collected just for the specified purposes of the participation process and not further processed in a way incompatible with those purposes. Moreover, FUDGE-5G partners must ensure that personal data are not (illegally) processed for further purposes. Thus, participants in the project activities should receive a legal note detailing this subject, where an assessment of the impact of data privacy will be carried out to assess the risks of processing personal data.

Erasure of data: personal data must be kept in a form that only allows for the identification of data subjects for no longer than is strictly necessary for the purposes for which the data were collected or for which they are further processed. Personal data that are not necessary any more must be erased or truly anonymized.

Anonymity: The consortium must guarantee anonymity by applying two strategies and applying the procedures described in FUDGE-5G to comply with all GDPR requirements. On the one hand, anonymity will be guaranteed by generalizing the data and on the other hand, the participation of stakeholders in the project will be anonymous, unless they voluntarily decide otherwise.

5.2 ePrivacy Directive (ePD)

The GDPR is complemented by Directive 2002/58/EC on privacy and electronic communications (ePrivacy Directive) [30], amended by Directive 2009/136[31], which concerns the protection of privacy in the electronic communications sector and covers some data not classed as "personal" such as some communications metadata. As a Directive, it is transposed into EU nations' laws rather than being imposed in a unified way as a Regulation is. FUDGE-5G project activities will design, develop, trial and showcase novel solutions that rely greatly on electronic communication. From these, new or improved products will emerge, as well as research outputs will be published and shared open and freely. Thus, the project will follow the ePD regulations regarding the protection of users' data and their informed consent.

[30] http://data.europa.eu/eli/dir/2002/58/oj
[31] http://data.europa.eu/eli/dir/2009/136/oj

6 Datasets

In this chapter we present a first version of the FUDGE-5G collected and generated datasets description. This section will be revised throughout the project duration and released publicly in M19 and M29.

6.1 Deliverables

The project will create several deliverables, some will be classified as confidential/restricted, but other will be made available to the public (open and free). The public project deliverables will be available for download on the FUDGE-5G website and in Zenodo. The confidential deliverables will be not available.

Table 10 - Dataset descriptor for FUDGE-5G confidential deliverables dataset

Field	Details
Identifier	FUDGE-5G_Deliverable_[UC]_[date]_[name]_[version]_RESTRICTED
Name	FUDGE-5G Confidential Deliverables
Description	This dataset will contain all restricted deliverables produced by the FUDGE-5G project
Policy	Restricted
Licence	-
Origin	Project research output
Responsibility	UPV (Project Coordinator)
Scale	100Mb - 10GB
Target	Internal usage
Format	PDF
Metadata	Documents and reports
Security & storage	FUDGE-5G repository

Table 11 - Dataset descriptor for FUDGE-5G public deliverables dataset

Field	Details
Identifier	*[provided by Zenodo]*
Name	FUDGE-5G Public Deliverables
Description	This dataset will contain all open deliverables produced by the FUDGE-5G project.
Policy	Open
Licence	CC BY
Origin	Project research output
Responsibility	UPV (Project Coordinator)
Scale	100Mb - 10GB
Target	Industry, Researchers
Format	PDF
Metadata	Documents and reports
Security & storage	Zenodo repository FUDGE-5G website

6.2 Scientific publications

FUDGE-5G research activities will result in scientific publications created by the project partners. These publications will be made available open and freely to the public (except when restrictions apply). These scientific publications will be available for download on the FUDGE-5G website and in Zenodo.

Table 12 - Dataset descriptor for FUDGE-5G scientific publications dataset

Field	Details
Identifier	*[provided by Zenodo]*
Name	FUDGE-5G Scientific Publications
Description	This dataset contains project journals and conferences papers/ posters, white papers, or other scientific publications.

Policy	Open
Licence	CC BY
Origin	Project research output
Responsibility	THA (WP5 leader)
Scale	100Mb
Target	Industry, Researchers
Format	PDF
Metadata	Documents and reports
Security & storage	Zenodo repository FUDGE-5G website

6.3 Other dissemination and communication publications

FUDGE-5G activities will result in dissemination and communication publications (non-scientific) such as website pages, promotional materials, press releases, website news, posts (e.g., Twitter, LinkedIn, blogs), etc. These publications will be made available open and freely to the public.

Table 13 - Dataset descriptor for dissemination and communications publications dataset

Field	Details
Identifier	*[provided by Zenodo]*
Name	FUDGE-5G Dissemination and Communication Publications
Description	This dataset contains all the results of research activities in terms of publications, as well as data collection on the website and social networks. These results involve white papers, promotional materials (brochures, flyers, newsletters, posters, etc.), press releases and videos produced during the FUDGE-5G project.
Policy	Open
Licence	CC BY
Origin	Project dissemination and communication activities

Responsibility	THA (WP5 leader)
Scale	10Gb
Target	Industry, Researchers
Format	PDF, JSON, MPEG-4
Metadata	Documents and reports
Security & storage	Zenodo repository FUDGE-5G website

6.4 Tracking data

Many of FUDGE-5G communication and dissemination platforms (e.g., website, Twitter, LinkedIn, YouTube) gather visitors tracking data that will be exploited by the project to correlate and assess the project visibility to the community in general and the targeted audience.

Table 14 - Dataset descriptor for communication and dissemination tracking data dataset

Field	Details
Identifier	*[provided by Zenodo]*
Name	FUDGE-5G Tracking data
Description	This dataset contains the gathered tracking data from the main communication and dissemination platforms used by the project.
Policy	Open
Licence	CC BY-ND
Origin	Website, social networks
Responsibility	THA (WP5 leader)
Scale	1Gb
Target	Researchers
Format	PDF, JSON, XML
Metadata	Documents and reports

Security & storage	Zenodo repository

6.5 Tests and trials

6.5.1 Media Showroom Vertical

Media showroom vertical tests and trials will create the following datasets:

- Radio Metrics gathered from 5G-VINNI
- Packet Capture and Logs generated inside FUDGE-5G Platform
- Video source content for the Remote Production and Concurrent Media Delivery Realizations
- Service Communication Proxy routing collected information
- User plane name-based routing collected information
- Service Function Virtualization Orchestrator collected information

Table 15 - Dataset descriptor for the Media Showroom Radio Metrics

Field	Details
Identifier	FUDGE-5G_RADIO_METRICS_UC1_20210601
Name	MediaShowroom_RadioMetrics
Description	This dataset contains several radio metrics gathered from the 5G modem and the gNB, related to the channel quality conditions, air latency, physical layer bitrate and error rate. It would be show to the public after being processed in graphics and tables.
Policy	Restricted
Licence	-
Origin	5G devices and equipment
Responsibility	NRK, TNOR
Scale	100MB to 1GB per trial
Target	Researchers
Format	XML, CSV, plain text

Metadata	Technical Metadata
Security & storage	FUDGE-5G repository

Table 16 - Dataset descriptor for the Media Showroom packet logs

Field	Details
Identifier	FUDGE-5G_NETWORK_METRICS_UC1_20210601_Restricted
Name	MediaShowroom_PacketLogs
Description	This dataset contains several transport network metrics gathered from the FUDGE-5G platform, such as packet captures (via tcpdump or wireshark). It also includes application data providing end to end data regarding the video transmission. It would be show to the public after being processed in graphics and tables.
Policy	Restricted
Licence	-
Origin	FUDGE-5G Platform and NRK application
Responsibility	NRK, TNOR, CMC, ATH
Scale	100MB to 1GB per trial
Target	Researchers
Format	Pcap, csv, json, xml
Metadata	Technical Metadata
Security & storage	FUDGE-5G repository

Table 17 - Dataset descriptor of the video source content for the Remote Production and Concurrent Media Delivery

Field	Details
Identifier	*[provided by Zenodo]*
Name	MediaShowroom_VideoSource

Description	This dataset is made of the video source, both after being captured in lossless quality and after being encoded, radiated, transported and stored over the FUDGE-5G platform.
Policy	Public
Licence	CC-BY
Origin	NRK equipment and application
Responsibility	NRK, UPV
Scale	Up to 20Gb
Target	Researchers
Format	Mp4, YouTube file
Metadata	Technical Metadata
Security & storage	Zenodo

Table 18: Dataset descriptor for the Service Communication Proxy

Field	Details
Identifier	FUDGE-5G_MEASURED_DATA_UC1_20210401_SCP_MAJOR_OPEN
Name	Service Communication Proxy implementing Name-based Routing
Description	The information collected to perform service routing among 5GC network functions is the MAC address as part of the system's ARP table, source IP address and port, the HTTP host and URI as well as all other HTTP request header fields. Once a 5GC is being removed, no SCP component has any information stored.
Policy	Restricted
Licence	InterDigital's FLIPS evaluation license (royalty free, non-commercial usage, closed source)
Origin	The data is being generated by the 5GC vendors.
Responsibility	IDE

Scale	One entry in the text-based database for each FQDN of producers and one entry for each consumer's URI.
Target	- SCP internal purposes to fulfil the functional objectives - Number of operations over the content identifiers (CID) are reported to the cross layer analytics component. But not the CIDs themselves
Format	string values stored in clear text
Metadata	Technical Metadata
Security & storage	The data will be stored on disk under /var/local and is readable by any user on that machine. Upon the deletion of a 5GC all information is remove from this database.

Table 19: Dataset descriptor for the Name-based Routing realisation on the user plane

Field	Details
Identifier	FUDGE-5G_MEASURED_DATA_UC1_20210401_5GC_NbR_MAJOR_OPEN
Name	5GC implementing Name-based Routing on the user plane
Description	The information collected to perform service routing among UPFs (infrastructure mode) or UEs and UPFs (UE mode). For both modes, the MAC address as part of the system's ARP table, source IP address and port, the HTTP host and URI as well as all other HTTP request header fields are being parsed. If the UE mode is being used, an Android APK must be installed to perform the service routing capabilities which does not read, process and store any other information as the UPF (infrastructure mode).
Policy	Restricted
Licence	InterDigital's FLIPS evaluation license (royalty free, non-commercial usage, closed source)
Origin	The data is being generated by the vertical applications.
Responsibility	IDE
Scale	One entry in the text-based database for each FQDN of producers and one entry for each consumer's URI.
Target	UE/UPF/SMF internal data collection and processing

Format	In memory state and a text-based database in the centralised SMF functionality to perform service routing.
Metadata	Technical Metadata
Security & storage	The data is stored either in memory (UE/UPF) or in text file (SMF).

Table 20: Dataset descriptor for the Service Function Virtualisation Orchestrator

Field	Details
Identifier	FUDGE-5G_MEASURED_DATA_UC1_20210401_SFVO_MAJOR_OPEN
Name	Service Function Virtualisation Orchestrator
Description	For orchestrating vertical applications or 5GCs, the SFVO is offering a unified interface that is authenticating genuine users based on arbitrary username and passwords that are stored in a single database. The information provided to describe the service chain and its service functions is being used to orchestrate and lifecycle manage the service chain and is kept in the SFVO components for the lifetime of a service chain.
Policy	Restricted
Licence	InterDigital's FLIPS evaluation license (royalty free, non-commercial usage, closed source)
Origin	The data is being generated by 5GC or vertical application providers
Responsibility	IDE
Scale	The amount of data kept for each service chain is in the range of kBs.
Target	The SFVO layer itself only.
Format	Descriptors in JSON, YML or TOML format
Metadata	Technical Metadata
Security & storage	The data will be stored in memory or in a MySQL database component inside the SFVO which is password protected

6.5.2 PPDR Vertical

PPDR vertical tests and trials will create the following datasets:

- Slice orchestrator event log
- End user devices call record and data transmission logs
- Mobitrust platform recorded data
- Mobitrust platform usage
- Service Communication Proxy routing collected information
- User plane name-based routing collected information
- Service Function Virtualization Orchestrator collected information

Table 21 - Dataset descriptor of the slice orchestrator event log dataset

Field	Details
Identifier	FUDGE-5G_Trial_PPDR_[date]_slice_orchestrator_events
Name	FUDGE-5G orchestrator events data
Description	This dataset will be gathered from the PPDR vertical tests and trials. It contains the timestamped set of events received and the actions taken by the orchestrator. The dataset includes both network and events as monitored by probes, NF and application lifecycle events from the virtualization infrastructure, and the interactions of the orchestrator with the different elements composing the network infrastructure.
Policy	Restricted
Licence	-
Origin	PPDR vertical tests and trials
Responsibility	THA
Scale	100MB to 1GB per trial
Target	Internal usage
Format	CSV, JSON
Metadata	Test/trial infrastructure and scenario description
Security & storage	FUDGE-5G repository

Table 22 - Dataset descriptor of the end user devices event log from UC2

Field	Details
Identifier	*[provided by Zenodo]*
Name	FUDGE-5G end user interactions data
Description	This dataset will be gathered from the PPDR vertical tests and trials. It contains a timestamped set of aggregated events recorded at the end user devices. The dataset includes metadata about call record, and data transmission from the end user perspective.
Policy	Open
Licence	CC BY
Origin	PPDR vertical tests and trials
Responsibility	THA
Scale	100MB to 1GB per trial
Target	Industry, Researchers
Format	CSV, JSON
Metadata	Test/trial infrastructure and scenario description
Security & storage	Zenodo

Table 23 - Dataset descriptor of the Mobitrust platform recorded data restricted dataset

Field	Details
Identifier	FUDGE-5G_Trial_PPDR_[date]_Mobitrust_[version]_RESTRICTED
Name	FUDGE5G Mobitrust platform recorded data
Description	This dataset will be gathered from the PPDR vertical tests and trials. The data will be collected from the Mobitrust platform and will include field sensors data, communications tracking data and vertical KPIs related measurements.
Policy	Restricted

Licence	-
Origin	PPDR vertical tests and trials
Responsibility	ONE
Scale	100MB to 1GB per trial
Target	Internal usage
Format	CSV
Metadata	Test/trial infrastructure and scenario description
Security & storage	FUDGE-5G repository

Table 24 - Dataset descriptor of Mobitrust Platform usage data on UC2

Field	Details
Identifier	*[provided by Zenodo]*
Name	FUDGE5G Mobitrust platform usage data
Description	This dataset will be gathered from the PPDR vertical tests and trials. The data collected from the Mobitrust platform will be converted to usage indicators and recordings of the recorded field sensors data, communications tracking data and vertical KPIs related measurements.
Policy	Open
Licence	CC BY
Origin	PPDR vertical tests and trials
Responsibility	ONE
Scale	10MB to 100MB
Target	Industry, Researchers
Format	CSV
Metadata	Trial infrastructure and scenario description

Security & storage	Zenodo repository

Table 25: Datasets of the Service Communication Proxy

Field	Details
Identifier	FUDGE-5G_MEASURED_DATA_UC2_20210301_SCP_MAJOR_OPEN
Name	Service Communication Proxy implementing Name-based Routing
Description	The information collected to perform service routing among 5GC network functions is the MAC address as part of the system's ARP table, source IP address and port, the HTTP host and URI as well as all other HTTP request header fields. Once a 5GC is being removed, no SCP component has any information stored.
Policy	Restricted
Licence	InterDigital's FLIPS evaluation license (royalty free, non-commercial usage, closed source)
Origin	The data is being generated by the 5GC vendors.
Responsibility	IDE
Scale	One entry in the text-based database for each FQDN of producers and one entry for each consumer's URI.
Target	- SCP internal purposes to fulfil the functional objectives - Number of operations over the content identifiers (CID) are reported to the cross layer analytics component. But not the CIDs themselves
Format	string values stored in clear text
Metadata	Technical Metadata
Security & storage	The data will be stored on disk under /var/local and is readable by any user on that machine. Upon the deletion of a 5GC all information is remove from this database.

Table 26: Datasets of the Service Function Virtualisation Orchestrator

Field	Details
Identifier	FUDGE-5G_MEASURED_DATA_UC2_20210301_SFVO_MAJOR_OPEN

Name	Service Function Virtualisation Orchestrator
Description	For orchestrating vertical applications or 5GCs, the SFVO is offering a unified interface that is authenticating genuine users based on arbitrary username and passwords that are stored in a single database. The information provided to describe the service chain and its service functions is being used to orchestrate and lifecycle manage the service chain and is kept in the SFVO components for the lifetime of a service chain.
Policy	Restricted
Licence	InterDigital's FLIPS evaluation license (royalty free, non-commercial usage, closed source)
Origin	The data is being generated by 5GC or vertical application providers
Responsibility	IDE
Scale	The amount of data kept for each service chain is in the range of kBs.
Target	The SFVO layer itself only.
Format	Descriptors in JSON, YML or TOML format
Metadata	Technical Metadata
Security & storage	The data will be stored in memory or in a MySQL database component inside the SFVO which is password protected

6.5.3 5G Virtual Office Vertical

5G vertical office vertical tests and trials will create the following datasets:

- Vertical Application collected data
- 5G modem metrics and event log
- Service Communication Proxy routing collected information
- User plane name-based routing collected information
- Service Function Virtualization Orchestrator collected information

Table 27 - Dataset descriptor of the Vertical Application remote monitoring data

Field	Details
Identifier	*[provided by Zenodo]*

Name	FUDGE-5G Remote Monitoring data
Description	This data set will be generated from the 5G Virtual Office trials. The data will be collected from the vertical application, which include biosensors data, environmental data and other vertical related data.
Policy	Open
Licence	CC BY
Origin	5G Virtual Office trials
Responsibility	ONE (vertical champion)
Scale	100Mb to 1Gb per trial
Target	Researchers
Format	CSV
Metadata	Trial infrastructure and scenario description
Security & storage	Zenodo repository

Table 28 - Dataset descriptor of the 5G Modem collected data and event log

Field	Details
Identifier	*[provided by Zenodo]*
Name	5G Modem recorded data
Description	This dataset will be gathered from the 5G Virtual Office vertical tests and trials. The data will be collected from the 5G modem (Quectel RM500Q-GL) and will include Signal-to-Noise Ratio (SINR) and Reference Signal Received Quality (RSRQ) values, among others. Information about the connection status will be also recorded.
Policy	Open
Licence	CC BY
Origin	5G modem

Responsibility	ONE
Scale	50MB to 1GB per trial
Target	Internal usage and to the public
Format	CSV
Metadata	Technical Metadata
Security & storage	Zenodo repository

Table 29: Dataset descriptor for the Service Communication Proxy

Field	Details
Identifier	FUDGE-5G_MEASURED_DATA_UC3_20210401_SCP_MAJOR_OPEN
Name	Service Communication Proxy implementing Name-based Routing
Description	The information collected to perform service routing among 5GC network functions is the MAC address as part of the system's ARP table, source IP address and port, the HTTP host and URI as well as all other HTTP request header fields. Once a 5GC is being removed, no SCP component has any information stored.
Policy	Restricted
Licence	InterDigital's FLIPS evaluation license (royalty free, non-commercial usage, closed source)
Origin	The data is being generated by the 5GC vendors.
Responsibility	IDE
Scale	One entry in the text-based database for each FQDN of producers and one entry for each consumer's URI.
Target	- SCP internal purposes to fulfil the functional objectives Number of operations over the content identifiers (CID) are reported to the cross layer analytics component. But not the CIDs themselves
Format	string values stored in clear text
Metadata	Technical Metadata

Security & storage	The data will be stored on disk under /var/local and is readable by any user on that machine. Upon the deletion of a 5GC all information is remove from this database.

Table 30: Dataset descriptor for the Service Function Virtualisation Orchestrator

Field	Details
Identifier	FUDGE-5G_MEASURED_DATA_UC3_20210401_SFVO_MAJOR_OPEN
Name	Service Function Virtualisation Orchestrator
Description	For orchestrating vertical applications or 5GCs, the SFVO is offering a unified interface that is authenticating genuine users based on arbitrary username and passwords that are stored in a single database. The information provided to describe the service chain and its service functions is being used to orchestrate and lifecycle manage the service chain and is kept in the SFVO components for the lifetime of a service chain.
Policy	Restricted
Licence	InterDigital's FLIPS evaluation license (royalty free, non-commercial usage, closed source)
Origin	The data is being generated by 5GC or vertical application providers
Responsibility	IDE
Scale	The amount of data kept for each service chain is in the range of kBs.
Target	The SFVO layer itself only.
Format	Descriptors in JSON, YML or TOML format
Metadata	Technical Metadata
Security & storage	The data will be stored in memory or in a MySQL database component inside the SFVO which is password protected

6.5.4 Industry 4.0 Vertical

Industry 4.0 vertical tests and trials will create the following datasets:

- 5G Modem recorded data

- 5G RAN recorded data
- 5G Core recorded data
- Service Communication Proxy routing collected information
- Service Function Virtualization Orchestrator collected information

Table 31: Dataset description for the 5G modem

Field	Details
Identifier	*[provided by Zenodo]*
Name	5G Modem recorded data
Description	This dataset will be gathered from the Industry 4.0 vertical tests and trials. The data will be collected from the 5G modem and will include Signal-to-Noise Ratio (SINR) and Reference Signal Received Quality (RSRQ) values, among others. Information about the connection status will be also recorded.
Policy	Open
Licence	CC BY
Origin	5G modem
Responsibility	5CMM
Scale	50MB to 1GB per trial
Target	Internal usage and to the public
Format	CSV
Metadata	Test/trial infrastructure and scenario description
Security & storage	Zenodo repository

Table 32: Dataset description for the 5G RAN equipment.

Field	Details
Identifier	FUDGE-5G_Trial_Industry40_[date]_5GRAN_[version]
Name	5G RAN recorded data

Description	This dataset will be gathered from the Industry 4.0 vertical tests and trials. The data will be collected from the 5G Radio Access Network (RAN) and will include RAN counters relevant for the Industry 4.0.
Policy	Restricted
Licence	-
Origin	5G RAN from Telenor
Responsibility	TNOR
Scale	From 100Mb to 1Gb per trial
Target	Internal Usage
Format	CSV
Metadata	Test/trial infrastructure and scenario description
Security & storage	FUDGE-5G Repository

Table 33: Dataset description for the 5G RAN equipment.

Field	Details
Identifier	*[provided by Zenodo]*
Name	5G Core recorded data
Description	This dataset will be gathered from the Industry 4.0 vertical tests and trials. The 5GC generates log information that includes registration and session setup information.
Policy	Open
Licence	CC BY
Origin	5GC from Cumucore
Responsibility	CMC
Scale	10GB/hour
Target	Researchers

Format	Syslog data structure
Metadata	Test/trial infrastructure and scenario description
Security & storage	Zenodo repository

Table 34: Dataset description for the Service Communication Proxy

Field	Details
Identifier	FUDGE-5G_MEASURED_DATA_UC4_20210301_SCP_MAJOR_OPEN
Name	Service Communication Proxy implementing Name-based Routing
Description	The information collected to perform service routing among 5GC network functions is the MAC address as part of the system's ARP table, source IP address and port, the HTTP host and URI as well as all other HTTP request header fields. Once a 5GC is being removed, no SCP component has any information stored.
Policy	Restricted
Licence	InterDigital's FLIPS evaluation licence (royalty free, non-commercial usage, closed source)
Origin	5GC vendors
Responsibility	IDE
Scale	One entry in the text-based database for each FQDN of producers and one entry for each consumer's URI.
Target	SCP internal purposes to fulfil the functional objectives
Format	String values stored in clear text
Metadata	Technical Metadata
Security & storage	The data will be stored on disk under /var/local and is readable by any user on that machine. Upon the deletion of a 5GC all information is remove from this database.

Table 35: Dataset descriptor for the Service Function Virtualisation Orchestrator

Field	Details
Identifier	FUDGE-5G_MEASURED_DATA_UC4_20210301_SFVO_MAJOR_OPEN
Name	Service Function Virtualisation Orchestrator
Description	For orchestrating vertical applications or 5GCs, the SFVO is offering a unified interface that is authenticating genuine users based on arbitrary username and passwords that are stored in a single database. The information provided to describe the service chain and its service functions is being used to orchestrate and lifecycle manage the service chain and is kept in the SFVO components for the lifetime of a service chain.
Policy	Restricted
Licence	InterDigital's FLIPS evaluation license (royalty free, non-commercial usage, closed source)
Origin	5GC or vertical application providers
Responsibility	IDE
Scale	In the range of kBs per service chain
Target	The SFVO layer itself only
Format	Descriptors in JSON, YML or TOML format
Metadata	Trial infrastructure and scenario description
Security & storage	The data will be stored in memory or in a MySQL database component inside the SFVO, which is password protected

6.5.5 Interconnected NPNs Vertical

Interconnecting private NW vertical tests and trials will create the following datasets:

- User Equipment Authentication and Authorisation event log
- Service Communication Proxy routing collected information
- Service Function Virtualization Orchestrator collected information

Table 36 - Dataset descriptor of the authorization and authentication dataset from UC5

Field	Details
Identifier	*[provided by Zenodo]*
Name	FUDGE5G Authentication and Authorization dataset
Description	This data set will be generated from the Interconnected NPN vertical trials. The data will be collected from the authentication and authorization of the various equipment in the 5G network.
Policy	Open
Licence	CC BY
Origin	Interconnected NPNs vertical trials
Responsibility	FHG (vertical champion)
Scale	100MB to 1GB per trial
Target	Researchers
Format	CSV
Metadata	Trial infrastructure and scenario description
Security & storage	Zenodo repository

Table 37: Dataset descriptor of the Service Communication Proxy

Field	Details
Identifier	FUDGE-5G_MEASURED_DATA_UC5_20210401_SCP_MAJOR_OPEN
Name	Service Communication Proxy implementing Name-based Routing
Description	The information collected to perform service routing among 5GC network functions is the MAC address as part of the system's ARP table, source IP address and port, the HTTP host and URI as well as all other HTTP request header fields. Once a 5GC is being removed, no SCP component has any information stored.
Policy	Restricted

Licence	InterDigital's FLIPS evaluation license (royalty free, non-commercial usage, closed source)
Origin	The data is being generated by the 5GC vendors.
Responsibility	IDE
Scale	One entry in the text-based database for each FQDN of producers and one entry for each consumer's URI.
Target	- SCP internal purposes to fulfil the functional objectives Number of operations over the content identifiers (CID) are reported to the cross layer analytics component. But not the CIDs themselves
Format	string values stored in clear text
Metadata	Technical Metadata
Security & storage	The data will be stored on disk under /var/local and is readable by any user on that machine. Upon the deletion of a 5GC all information is remove from this database.

Table 38: Dataset descriptor for the Service Function Virtualisation Orchestrator

Field	Details
Identifier	FUDGE-5G_MEASURED_DATA_UC5_20210401_SFVO_MAJOR_OPEN
Name	Service Function Virtualisation Orchestrator
Description	For orchestrating vertical applications or 5GCs, the SFVO is offering a unified interface that is authenticating genuine users based on arbitrary username and passwords that are stored in a single database. The information provided to describe the service chain and its service functions is being used to orchestrate and lifecycle manage the service chain and is kept in the SFVO components for the lifetime of a service chain.
Policy	Restricted
Licence	InterDigital's FLIPS evaluation license (royalty free, non-commercial usage, closed source)
Origin	The data is being generated by 5GC or vertical application providers
Responsibility	IDE

Scale	The amount of data kept for each service chain is in the range of kBs.
Target	The SFVO layer itself only.
Format	Descriptors in JSON, YML or TOML format
Metadata	Technical Metadata
Security & storage	The data will be stored in memory or in a MySQL database component inside the SFVO which is password protected

6.6 Demos and showcases

The datasets that will result from demos and showcases will follow the same description as the ones detailed in the previous section (6.5 Tests and trials). Specific personal data concerns may be required in demonstrations and showcases events, but the defined methodology already includes GDPR compliance.

Annex A: Information sheet template

You are asked to participate in a FUDGE-5G (Horizon 2020 grant agreement Nº 957242) trial conducted by [insert the use case champion name] from [insert partner name]. The results will contribute to FUDGE-5G innovations and technology validation and will be published and presented. You may participate in this research study if you are age 18 or older.

The research aims to validate and demonstrate [describe what is being validated or demonstrated].

It is your decision to participate or not in this trial. If you volunteer to take part on this trial, you have the possibility to withdraw from the trial at any time without any consequences and without the need to provide a justification.

Withdraw

To withdraw the consent, you must provide an unambiguous indication by a statement or a clear affirmative action that you withdraw the consent for the use of your data in the context of the trial. You must send your withdraw statement to [data protection officer email]. If you choose to withdraw your individual data will not be used and will be deleted from all FUDGE-5G storage facilities, but it will not affect the lawfulness of the processing up to the withdraw point.

Procedures

If you volunteer to participate on this trial, you will be asked to [Provide a detailed description of what participants will perform/encounter in the trial]

Benefits of Participation

The trial is designed to demonstrate the FUDGE-5G innovations mentioned earlier. There may be no benefits to you as a participant on this trial. However, we hope that the technologies and innovations showcased on the trial improve [improvements expected from the innovations being trialled].

Risks of Participation

[List the risks of participating in the trial]. In addition, there may be previously or uncommon risks that may happen. You should report any problems to the trial responsible.

Cost/Compensation

The participation on this trial will not have any financial cost for you. The trial will take [time amount] of you time. There will be no financial compensation for the participation on the trial.

Data Collected Usage
[Usage of the data during research, dissemination and storage]

Personal Information
(for example their name, where they live, information that can disclose their identity)

All of your personal information gathered on this trial will be kept confidential. No written or oral reference will be made that could link you to this trial.

All of your personal data will be stored in FUDGE-5G data repository, until the end of the project, February 2023. The repository is self-hosted and monitored by a security team. The repository is password protected and only accessible to authenticated and authorized users, the FUDGE-5G research team members.

The signed consent form will be stored [location] and it is subject to [security and access measures].

You have the right to request access to your personal data and to request the rectification or removal of such data.

Future Publishing, Archiving and Reuse of the Data
Data sharing brings benefits that impact the scientific community and untimely the society in general. It allows for the maximization of the data utility and makes best use of the contributions made by trial participants. By allowing the reusage, the data can be used in secondary analyses, that fill research gaps without enrolling more participants.
The data collected in the trial will be completely anonymized and then made publicly available in Zenodo, the repository developed under the European OpenAIRE program.

Contacts
You can contact the trial responsible partner, [partner name], data protection officer [DPO name] to the email [DPO email] or you can contact the FUDGE-5G data protection officer, Luís Cordeiro to the email cordeiro@onesource.pt.

Annex B: Informed consent template

Please tick the appropriated boxes

1. Participating in the trial

	Yes	No
I have read and understood the trial information dated [DD/MM/AAAA], or it has been read to me. I have been able to ask questions and my questions were answered clearly and to my satisfaction.	☐	☐
I consent to voluntary participate in this trial and I understand that I can refuse to answer questions and withdraw from the trial, with having to provide a justification.	☐	☐
I understand that taking part in this trial involves [what the participant has to do, will be subject to].	☐	☐
I understand that taking part in this trial has [risks] as potential risk.	☐	☐

2. Use of information

	Yes	No
I understand that the information and data that I will provide will be used for [list the planned outputs].	☐	☐
I understand that personal information collected that can identify me will only share among the trial team.	☐	☐
I agree that my information can be quoted in research outputs.	☐	☐

3. Future use or reuse of the information by others

	Yes	No
I give permission to [specify the data] that I provide to be deposit in Zenodo so it can be used for future research and learning.	☐	☐
I understand that all my data that will be publicly available is anonymized, thus I cannot be identified as the data provided.	☐	☐

4. Signatures

_____

Name of the participant [IN CAPITALS]

_____

Signature of the participant

_____

Date

I have read, in detail, the information sheet to the potential participant, and ensured that the participant understands to what he is giving consent.

_____

Name of the trial responsible [IN CAPITALS]

_____

Signature of the trial responsible

_____

Date

5. Trial contact details for further information

Trial:
[Trial DPO name]
[Trial DPO email]

Project:
Luís Cordeiro
cordeiro@onesource.pt