D1.2 - Data Management Plan

Editor:	Halid Hrasnica, Eurescom
Deliverable nature:	ORDP: Open Research Data Pilot
Dissemination level:	Public (PU)
Date: planned | actual	30 June 2017	25 July 2017 (ver. 1.0) 3 July 2018 (Ver. 2.0 – resubmission)
Version | No. of pages	2.0	14
Keywords:	Open access, Datasets, Data availability

Abstract

This deliverable, the Data Management Plan, covers overarching issues concerning data management; all qualitative and quantitative data generated by the project. It outlines the measures that 5GINFIRE has put in place in order to publish its results namely: public deliverables, scientific publications, contributions to standards, software and APIs. Furthermore the document describes the data backup policy that is implemented for the main data of the project, both during the project execution as well as for the period after the end of the project.

Disclaimer

Impressum

Full project title: Evolving FIRE into a 5G-Oriented Experimental Playground for Vertical Industries

Short project title: 5GINFIRE

Number and title of work-package: WP1 – Project Management

Number and title of task: Task 1.1 - Project planning and coordination

Document title: D1.2 - Data Management Plan

Editor: Halid Hrasnica, Eurescom

Work-package leader: Halid Hrasnica, Eurescom

Copyright notice

Executive summary

This document covers overarching issues concerning data management. It concerns all qualitative and quantitative data generated by the project. For the creation of this Data Management Plan, the corresponding EC guidelines have been consulted.

This Data Management Plan (DMP) outlines the measures that 5GINFIRE has put in place in order to accommodate for the requirements set for projects contributing to the Horizon 2020 pilot action on open access to research data.

Pertaining to the data that the project will produce, the DMP initially identifies the types of datasets that will be outcome of the project, namely: public deliverables, scientific publications, contributions to standards, software and APIs. The DMP outlines the nature of the data that will be collected and how these data will be sanitized and stimulates the project partners for early dataset preparation helpful for future usage of open access data. The process also includes collection of relevant data from external experimenters using the 5GINFIRE infrastructure.

Furthermore the DMP describes the data backup policy that is implemented for the main data of the project, both during the project execution as well as for the period after the end of the project. Data that will be made publicly available will receive a Digital Object Identifier (DOI) issued by Eurescom and will be made available through Eurescom servers for at least 5 years after the end of the project. In addition the long term availability of DOI marked data complies with the related policies of the DOI handling system.

List of authors

Company	Author	Contribution
EURESCOM	Halid Hrasnica	Initial document
UoP	Christos Tranoris	Inputs on software provision
Ivan Vidal	UC3M	Inputs on software provision
Diego Lopez	TID	Inputs on software provision

Table of Contents

List of figures and tables

1   Introduction

This document covers overarching issues concerning data management. It concerns all qualitative and quantitative data generated by the project. For the creation of this Data Management Plan the EC guidelines on "Open Access to Scientific Publications and Research Data in Horizon 2020" and "Guidelines on Data Management in Horizon 2020" have been consulted.

During the lifetime of the project several results in form of documents, publications, specifications, software, and APIs will be produced. Following the guidelines of the EU for open access to scientific knowledge produced within the European funded projects, the members of the 5GINFIRE consortium are establishing mechanisms for allowing open access to the project results and outputs, which are lied down in this document.

This document provides an analysis of the main elements for the data management policy that will be used as guideline for all project partners with regard to the datasets that will be generated by the project. Most of these results will be publicly available, which induces the need to provide a plan for the long term availability of these results.

The second version of the document reflects additional aspects of the 5GINFIRE Data Management Plan, as identified during the first interim project review, related to data gathering from experimentation, described in Sec. 2.6:

- to be project specific and gather data from experiments
- to explain how different data will be gathered from the experiments
- to ensure scientific approach and reproducibility of experiment results

It is obvious that this document is a living document that will be continuously adapted during the lifetime of the project and information about corresponding updates will be provided in the project management reports.

2   Datasets to be produced

A list of planned and expected data sets to be collected and generated in the 5GINFIRE project is presented below:

- Project deliverables
- Scientific publications
- Other publications and outputs
- Contribution to standards
- Software and applications
- Data generated by the experimentation and testing

In addition, the project of course analyses available results from other research activities, publications, and further relevant information available. This information will be mainly used for internal project analysis and will be provided in respective project deliverables with appropriate references to origins of the gathered information. However, as the analysed information has not been created by 5GINFIRE, the project is not considering provision of these data as public data sets because the project does now own these information and results.

During the project lifetime, additional Information on the following aspects will be elaborated for all data sets on case by case base before making consortium decision on handling of the particular data generated or collected;

- Nature and scale of the data in consideration,
- To whom it could be useful / targeted audience and its size / level of interest,
- Information on the existence of similar data and possible synergies,
- Possibility for integration and reuse of the provided data by external users / researchers, and
- Any further related issue

In order to follow up status of the data generated and its provision for external audience, audience outside of the project, the project established respective processes, as described in Chapter 4.

2.1   Public Deliverables

The list of the project deliverables is available for the public at the 5GINFIRE project website https://5ginfire.eu/deliverables/. The public project deliverables will be provided for download on the website after their approval by the consortium and submission to EC. The confidential deliverables will be not available through the website and they might be requested by external parties, in which case the consortium will make decision to disseminate corresponding deliverables or specific parts of the deliverables to particular external parties.

The project deliverables on the website will be provided in the widely adopted PDF format.

2.2   Scientific Publications

The scientific publications, mainly scientific papers, created by the consortium members, where the publications contain results from the 5GINFIRE project, will be usually made available for the wide public audience. Restricted access to the publications will be accepted only if there are serious reasons expressed by the consortium members or publishers of the scientific papers.

Beside provision of the 5GTINFIRE scientific publications on the project website for download, the following will be ensured for all scientific publications:

- A Digital Object Identifier - DOI and
- Open Access to the publication.

2.3   Other Publications

Besides the scientific publications mentioned above, e.g. in journals or conference proceedings, it is expected that the project will generate further publications and other project outcomes, such as:

- Promotion material (brochures, flyers, posters, etc.),
- Press releases and further project announcement,
- White / position papers created by the consortium on particular subjects, and
- Any further publication generated by the project.

Depending on nature of other publications, the audience targeted by them, and potential for usability of the publications, they will be handled accordingly. Thus, all non-confidential publications will be provided on the project website and the project consortium will decide on case by case base if the publications will be also published through the Open Access, including definition of a particular data set if appropriate, and if a DOI will be assigned to the publication.

2.4   Contribution to Standards

Standardization constitutes an important dissemination activity in 5GINFIRE project. It aims to contribute to the activities in major international standardization bodies, as defined in the 5GINFIRE Description of Action (Part of the project Grant Agreement) and the deliverable D8.2 "Initial exploitation, dissemination and standardization plan". For deployment of the new architecture a joint standardization initiative between and within the targeted standardization and regulation bodies is planned to be initiated.

Once, a contribution to a standardization/regulation body from the 5GINFIRE project is in preparation, appropriate publication means for the contribution (e.g. its availability in Open Access) will be discussed among the consortium members, to make a corresponding decision.

2.5   Software

2.5.1   Contributions to OSM framework

Contributions to the development of OSM software authored by UC3M, which may be produced as part of the deployment of the 5GinFIRE MANO platform, will be available as open source under the OSM license (https://osm.etsi.org/about/osm-license). In addition, UC3M and TID are working on a transport layer protocol to support the secure dissemination of information, which may be applicable to MANO environments. An implementation of this protocol, along with an API for developers, will be made publicly available as open source.

2.5.2   The 5GinFIRE portal web frontend

This software will be provided as Open Source via the 5GinFIRE public GitHub repository under the Apache 2.0 License. The service called eu.5ginfire.portal.web Web front UI is an AngularJS based UI to be used with eu.5ginfire.portal.api Web Service.

2.5.3   The 5GinFIRE portal API backend

This software will be provided as Open Source via the 5GinFIRE public GitHub repository under the Apache 2.0 License. The service called eu.5ginfire.portal.api Web Service is a RESTful service for 5GinFIRE project that allows to view and administer 5GinFIRE assets (i.e. experiments, VxFs, etc).

2.6   Data generated by the experimentation and testing

Depending on experiments and testing which will be performed on the top of the 5GINFIRE infrastructure by Third Parties, which will join the project through the planned Open Calls, further data sets might be generated as well. As for the time being is not known which kind of experiments will be executed and, accordingly, which data sets might be generated, corresponding decision on provision of these data will be made by the consortium, once the experiments are known and completed.

2.6.1   Data gathering from experimenters

In order to collect relevant data from the experimenters – 3rd parties accepted after the project Open Calls – a process of mandatory deliverables/reports from the experimenters during their planned involvement in the project of six months have been put in place (and is also part of the contracts with the 3rd parties) as follows:

- D1 (M2) - Detailed experiment description and implementation / execution plan
- D2 (M6) - Report on experiment implementation and execution – Experience on experiment implementation and execution
- D3 (M6) – Experiment results

Whereas the formats of the reports D1 and D3 are rather free, where the experimenters can present their approach and results in the best way fitting to an experiment scope, the D2 reports will have to follow a particular structure (provided by the project), in order to gather all needed feedbacks on usage of the 5GINFIRE infrastructures and tools and to be able to

adapt and further improve them in accordance with the received inputs from the experimenters.

2.6.2   Updates of Data Management Plan

In respect to the data from experimenters, needed to update the above sections of this chapter (Data sets to be provided by the project - scientific publications, other publications and outputs, contribution to standards, Software and applications – as part of the Data Management Plan), the 5GINFIRE consortium will consider relevant inputs from the experimenters together with updates (if any) to be received from the project beneficiaries. Accordingly, during the final reporting phase for the 3rd party experimenters, a request to corresponding contributions will be sent to both consortium members and the 3rd parties, which beside experimenters also include the 3rd party infrastructures also joining the project after the open calls.

2.6.3   Reproducibility of experiment results

In order to perform an experiment, it is needed to provide (from the side of experimenters) a formal description of experiments, which will be carried out in the known 5GINFIRE infrastructure. The formal experiment description will be provided through experiment descriptors which will be stored in the related project repository, allowing an experiment to be deployed and executed in the 5GINFIRE infrastructure. In this way, the reproducibility of the experiments at any time is ensured as well.

3   Data availability

3.1   Digital Object Identifier Repository

The data availability will be implemented by means of an own instance of a Digital Object Identifier (DOI) repository.

A digital object identifier (DOI) is a character string (a "digital identifier") used to uniquely identify an object such as an electronic document. Metadata about the object is stored in association with the DOI name and this metadata may include a location, such as a URL, where the object can be found. The DOI for a document remains fixed over the lifetime of the document, whereas its location and other metadata may change. Referring to an online document by its DOI provides more stable linking than simply referring to it by its URL, because if its URL changes, the publisher need only update the metadata for the DOI to link to the new URL. A DOI name differs from standard identifier registries such as the ISBN and ISRC. The purpose of an identifier registry is to manage a given collection of identifiers, whereas the primary purpose of the DOI system is to make a collection of identifiers actionable and interoperable.

Organizations that meet the contractual obligations of the DOI system and are willing to pay to become a member of the system can assign DOIs. The DOI system is implemented through a federation of registration and agencies coordinated by the International DOI Foundation, which developed and controls the system. The DOI system has been developed and implemented in a range of publishing applications since 2000; by late April 2011 more than 50 million DOI names had been assigned by some 4,000 organizations. By April 2013 this number had grown to 85 million DOI names assigned through 9,500 organizations.

For 5GINFIRE deliverables and publications, Eurescom (the Project Coordinator) developed so called landing pages which are stored on the 5GINFIRE shared and secured project workspace server. Of course, the public documents and data will be accessible for external users as well, which will have access to the repository from the project website. The URL of the landing pages is stored in the associated DOI and the actual download link is on the landing page, so that the archived publications and data can be reached through other websites as well. The actual storage of the results is managed by the Eurescom technical infrastructure team and Eurescom guarantees the storage and availability of the results for at least 5 years after the end of the project.

3.2   Open Access Repository

The rules and principles of the European Commission's Horizon 2020 Framework Programme clearly establish that scientific results generated within H2020 projects will be made available as open access publications, i.e. freely available online to any user. Following these rules, enforced in the 5GINFIRE Grant Agreement, open access will be ensured to all peer-reviewed scientific publications regarding 5GINFIRE and its composite solutions by the consortium. Furthermore, other project outputs and results, such as collected data and software, will be provided in the Open Access upon respective decision of the project consortium.

4 Data management process

Data collected or generated by the project, as described in Chapter 2, are stored in the BSCW based project repository described in the 5GINFIRE Deliverable D1.1. The repository consists of a project internal area, not possible to be accessed by external users, and a public area with direct access to it from the project website (Figure 1).
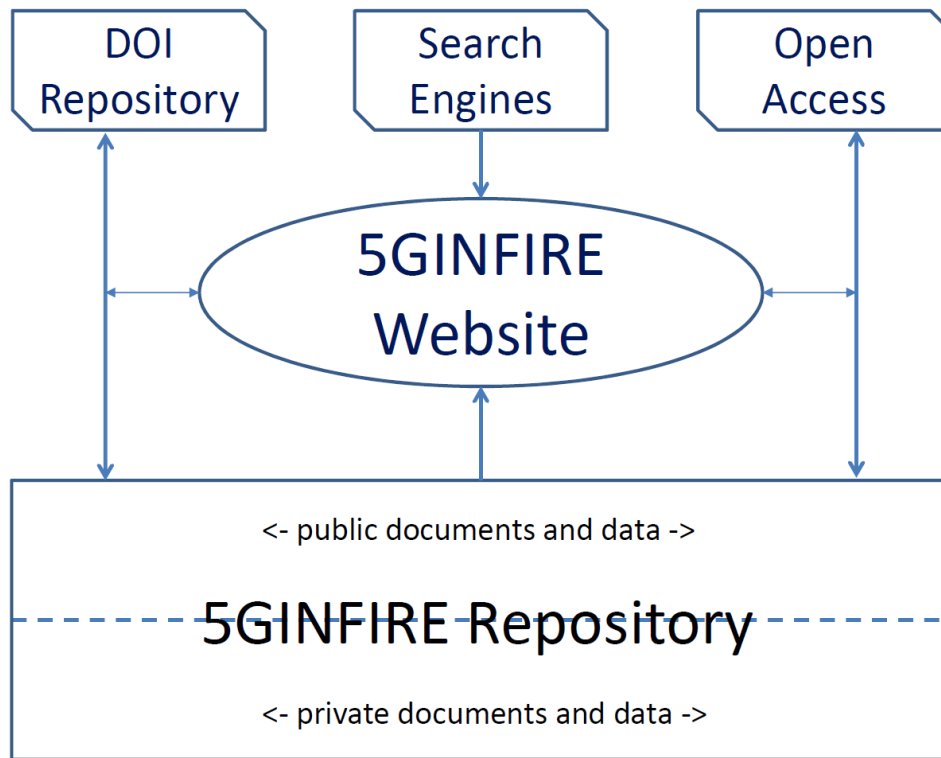
Figure 1: 5GINFIRE Data Management Concept

This approach ensures that all the relevant data is stored in a single repository and that it can be found various search engines as well as through the DOI and Open Access repositories.

4.1 Data Back-up

Eurescom hosts a number of 5GINFIRE support services on its servers. The hosted services currently include:

- A shared project workspace server based on Only Office
- A project website based on Wordpress
- A mail list service implemented internally by Eurescom based on Exim
- A project reporting system based on Reporter which is internally developed by Eurescom
- A project dissemination tracking system based on Tracker which is internally developed by Eurescom
- An audio conferencing service implemented internally by Eurescom based on the Asterisk soft switch

All these services run on different virtualized Linux servers. The disks of each VM node are mirrored to protect against initial hardware failures. In addition following backup procedures are implemented:

- Full system backup of each VM node is made on a daily basis and stored on dedicated backup servers in Frankfurt and Heidelberg. The full system backups are on a per-file basis and are uncompressed to ensure fast turnaround times for restore jobs.
- A 30 days backlog of full system backups is kept.
- Latest daily backup is also copied offsite.
- Since external backups are run on a 24 hours interval, a file saved on the BSCW server is only guaranteed in the backup after maximum of one day.
- SQL Databases are dumped to plain serialized files before backup to ensure there is no corruption.

If some of the services are used from external providers, which might be necessary and therefore decided by the consortium, copies of all public repositories are stored on the BSCW server which is hosted by Eurescom and to which the aforementioned backup procedures apply.

4.2   Responsibilities and Decision Making

As indicated in Chapter 1 of the document, the Data Management Plan presented in this deliverable presents its first version and the related consortium discussions will be continuously carried out, to identify the relevant project outputs and the collected data for publication as well as to decide on way and means of their publication. To ensures it, a dedicated time slot will be reserved at each of the project plenary meeting and, if needed, at selected consortium audio conferences. EC and project reviewers will be informed about related work done and publications provided in the project management reports.

Individual responsibilities on data management in the project consortium are:

- Project Coordinator – to prepare and lead related discussions at the relevant project meetings and to maintain the project document repository, including DOI records
- Scientific and Technical Project Manager – to identify data collected by the project and technical project outcomes eventually suitable for publication
- Dissemination (WP7) leader – to identify publications suitable for publication in the considered repositories and maintain 5GINFIRE inputs for the Open Access
- Each individual partner – to identify own project results suitable for publication