



5G! PAGODA

D1.2 – Open Data Management Plan

Document Number	D1.2
Status	Issue 1.0
Work Package	WP 1.2
Deliverable Type	Report
Date of Delivery	30/December/2016
Responsible	AALTO
Contributors	All beneficiaries
Dissemination level	PU

This document has been produced by the 5GPagoda project, funded by the Horizon 2020 Programme of the European Community. The content presented in this document represents the views of the authors, and the European Commission has no liability in respect of the content.



This project has received funding from *the European Union's Horizon 2020 research and innovation programme* under grant agreement No 723172.

Change History

Version	Date	Status	Author (Company)	Description
1.0	30.Dec.2016	Issue 1.0	Timo Brander & Tarik Taleb (Aalto)	First approved version

Executive Summary

This deliverable covers the open data management principles for the EU/JP 5G!Pagoda project. The document outlines the data collection, storage, sharing and protection tools and processes applied by the consortium during and after the project. The document provides the guidelines for data handling by the project consortium, as well as for external stakeholders, and thus ensures systematic approach to data management and protection. Data management guidelines are discussed and referred to in the Grant Agreement, the Consortium Agreement, the Coordination Agreement, and the Project Management, Quality and IPR Guide.

The Open Data Management Plan is an evolving document, and will gain more precision and substance during the course of the project. The plan will be revised, as needed, in order to keep the information up to date, e.g. as per changes in the produced data.

Table of Contents

1. Introduction	5
2. Data Types	8
2.1. Data Types of the Project.....	8
2.2. Levels of Confidentiality and Flow of Data.....	9
2.3. Data Formats.....	10
2.4. Expected Data of the Project	11
3. Collection, Storage and Use of Research Data and Analysed Research Data.....	13
4. Storage and Use of Project Data.....	14
4.1. Storage of Data	14
4.2. Documenting of Data	15
5. Data Sharing.....	17
5.1. Depositing Research Data	17
5.2. Open Access Publishing.....	17
5.3. Other Dissemination	18
6. Opting Out Research Data.....	19
7. Summary and Perspectives	20

1. Introduction

This document outlines the principles and processes for data collection, annotation, analysis and distribution as well as storage and security of data within the EU/JP 5G!Pagoda project. The procedures will be adopted by all project partners throughout the project in order to ensure that all project related data are well-managed according to contractual obligations as well as applicable legislation both during and after the project.

5G!Pagoda is committed to openness with respect to research data and results, as manifested by its participation in the Pilot on Open Research Data in Horizon 2020. Therefore, as a further dissemination channel, 5G!Pagoda will share datasets on which its results are derived as explained in this Open Data Management Plan. Public research repositories are used to maximize the visibility and impact of the research carried out by partners by making such data available for reuse by 3rd parties and reproducibility of the project's results.

The Grant Agreement of the 5G!Pagoda project (as an Open Data Pilot participant) obligates the project to deposit [digital research data generated in the project] in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:

- (i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;
- (ii) other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan', i.e. this document.

The Grant Agreement contains an option to discard the obligation to deposit a part of research data in the case where the achievement of the action's main objective, described in Annex 1 of the Grant Agreement, would be jeopardised. In such case, the Open Data Management Plan must contain the reasons for not giving access.

As the obligation to deposit research data in a databank does not change the obligation to protect results, take care of confidentiality and security obligations, or the obligations to protect personal data, the Open Data Management Plan addresses these topics. This document details, how the seemingly contradicting commitments to share and protect are implemented within the project.

The Open Data Management Plan has, on the other hand, also served the purpose of acting as a tool to agree on the data processing of the 5G!Pagoda project consortium. The production of the Open Data Management Plan has helped the consortium to identify situations, where the practices were thought to be agreed upon and where a common understanding on practices was thought to have been achieved, but where such in fact did not exist.

Documents related to the Open Data Management Plan are the 5G!Pagoda project Grant Agreement, the Consortium Agreement, the Coordination Agreement, and the Project Management, Quality and IPR Guide. Some of the deliverables also contain information which link to the Open Data Management Plan. The relationships are described below in Table 1.

Table 1 Documents related to the Open Data Management Plan.

Related document	Relationship to the Open Data Management Plan
The Grant Agreement	<ul style="list-style-type: none"> • Article 27 details the obligation to protect results • Article 29 details dissemination of results including open access to publications and research data • Article 36 details confidentiality obligations

	<ul style="list-style-type: none"> Article 39 details obligations to protect personal data, if applicable Annex 1, Part B, Chapter 2.4 and Chapter 2.5 detail open access publication and research data management principles.
Consortium Agreement	<p>Chapter 4.1 on the General principles: <i>“Each Party undertakes to notify promptly, in accordance with the governance structure agreed in the Coordination Agreement and directly to the Coordinator, any significant information, fact, problem or delay likely to affect the Project, submission of the deliverables or reports in accordance with the Grant Agreement and shall promptly provide all information reasonably required by the Coordinator to carry out its tasks.</i></p> <p><i>Each Party shall take reasonable measures to ensure the accuracy of any information or materials it supplies to the Coordinator or the Parties.”</i> This is a general declaration of the partners to abide by the rights and obligations set out in the Grant Agreement.</p>
Coordination Agreement	<ul style="list-style-type: none"> Chapter 4.1 on the General principles: <i>“Each Party undertakes to take part in the efficient implementation of the Coordinated Project, and to cooperate, perform and fulfil, promptly and on time, all of its obligations in the said project and under this Coordination Agreement as may be reasonably required from it and in a manner of good faith.</i> <p><i>Each Party undertakes to notify promptly, in accordance with the governance structure of the Coordinated Project, and as required for the coordination under this Coordination Agreement, any significant information, fact, problem or delay likely to affect the project(s) or the coordination.”</i> This is a general declaration of the partners to abide by the rights and obligations set out in the Coordinated Project, Project Plan, part of the Grant Agreement.</p> <ul style="list-style-type: none"> Chapter 8.4 on Publication defines publication principles.
Project Management, Quality and IPR Guide	The Project Management, Quality and IPR Guide defines the quality criteria for all work conducted in the 5G!Pagoda project.
Yearly report on standardization, dissemination and exploitation achievements	Contributions to standardization will enable 5G!Pagoda to achieve broader recognition of its results by a wide industry community, stimulate higher levels of interoperability and contribute to establishing economies of scale for 5G!Pagoda applications. Moreover, close coordination between research projects and standardization organizations, particularly through running and validated testbeds, is an important mechanism for exploitation of results and for inspiring and initiating innovation.
Report on innovation achieved and forthcoming industrial exploitation	This public report summarises the innovations achieved in the project and the activities and plans to exploit these innovations and other project results.

5G!Pagoda does not intend collecting any data that may expose private information (i.e., user context-related data). If during the course of the project, this becomes a necessity, the 5G!Pagoda consortium is committed to respect privacy and comply with the EU respective directives and national law: any data deemed private will not be made publicly available, and if it has to be made public, it will be appropriately anonymized. The project may generate quantitative models of various aspects of the 5G!Pagoda system (e.g., performance of specific network function, performance of a service over a specific network slice, network or application level traces, dependence of end-user experience on specific system and environment parameters, etc.), which will be made available via open-access publications, technical reports, and the project’s public deliverables. The partners will evaluate the applicability of

standards with respect to various aspects of data publication, indexing, storage, preservation, privacy, etc., and will report on them in this Open Data Management Plan (i.e. deliverable D1.2).

Where applicable, each contributor to a publication as an outcome of 5G!Pagoda will acquire a persistent identifier from the ORCID registry and each published data set or other research data object will be assigned a persistent identifier from DataCite, which builds on OIDs. The consortium will strive to publish data on repositories indexed by the Registry of Research Data Repositories, which provides an XML schema and an API for querying research data. The selected repositories should support the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) standard for repository interoperability and cross-archive search. The consortium will ensure that appropriate metadata are provided for the published data, including the acknowledgement to the funding of EU, H2020, and the Japanese Ministry of Internal Affairs and Communications, the name of the project, the grant number and the name of the action.

The ODMP may be re-evaluated during the project. On each updated version, the necessary corrective actions will be included, to ensure that research data remain discoverable, accessible, assessable and intelligible, interoperable and (re)usable beyond their original purpose. The European Commission as well as the Japanese Ministry of Internal Affairs and Communications will be accordingly notified.

2. Data Types

2.1. Data Types of the Project

In the 5g!Pagoda project there are four basic types of data (Figure 1): research data, analysed research data, project data and reports and communication data.

Research data covers the data collected on the project subject matter and it refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation.

In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form.

Users can normally access, mine, exploit, reproduce and disseminate openly accessible research data free of charge. /H2020 Online Manual/

Based on discussion with the consortium members, it has become apparent that it is difficult at this stage to identify the kind of data to be collected on the project. However, once such data become identified, this Open Data Management will be accordingly updated.

Analysed research data means the reports composed of the research data. Analysed data also refers to qualitative and quantitative data analyses conducted on the data. Reviews of earlier published data and records will be utilised to some degree. This data will be considered as analysed research data for the purposes of this document. Project related workshops and stakeholder engagement events are public events and the workshop notes of project partners will be treated in the same way as analysed research data (i.e. the notes will be shared within the consortium).

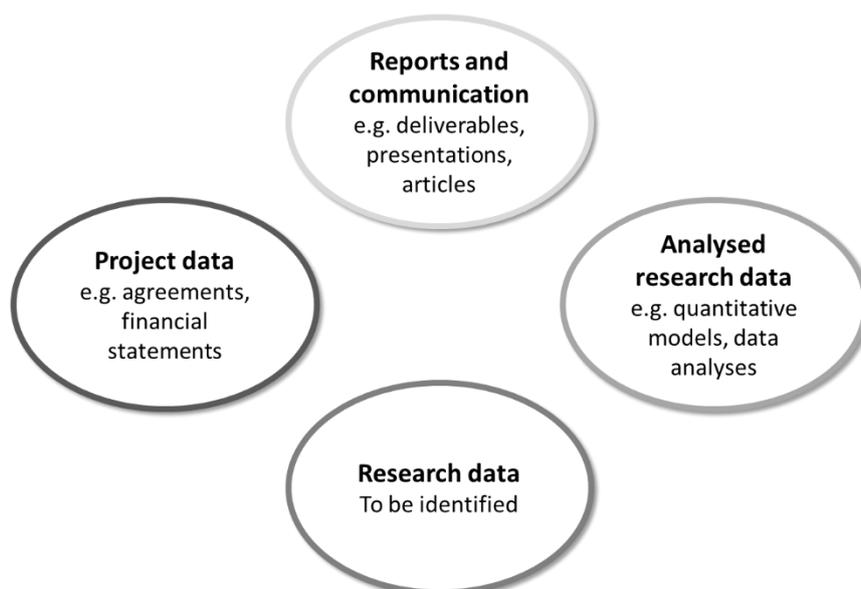


Figure 1 Data types.

Project data includes administrative and financial project data, including contracts, partner information and periodic reports, as well as accumulated data on project meetings, teleconferences and other internal materials. This data is confidential to the project consortium and to the European Commission.

Project data includes mainly MS Office documents, in English, which ensures ease of access and efficiency for project management and reporting. Most of the project data is stored in the password protected OwnCloud repository, administrated by Aalto University.

Reports and other communication data includes deliverables, presentations and for example articles. This data type also refers to the contents of the 5G!Pagoda project website.

Each data type is treated differently with regard to the level of confidentiality (see Chapter 2.2). Some of the data falls under the EU, Japan, and national laws on data protection and for this reason the project is obliged to seek necessary authorisations and to fulfil notification requirements.

The project will assume the principle of using commonly used data formats for the sake of compatibility, efficiency and access. The preferred means of data types is MS Office compatible formats, where applicable.

2.2. Levels of Confidentiality and Flow of Data

Overall, there are three basic levels of confidentiality, namely Public, Confidential to consortium (including Commission Services), and Confidential to the Partner.

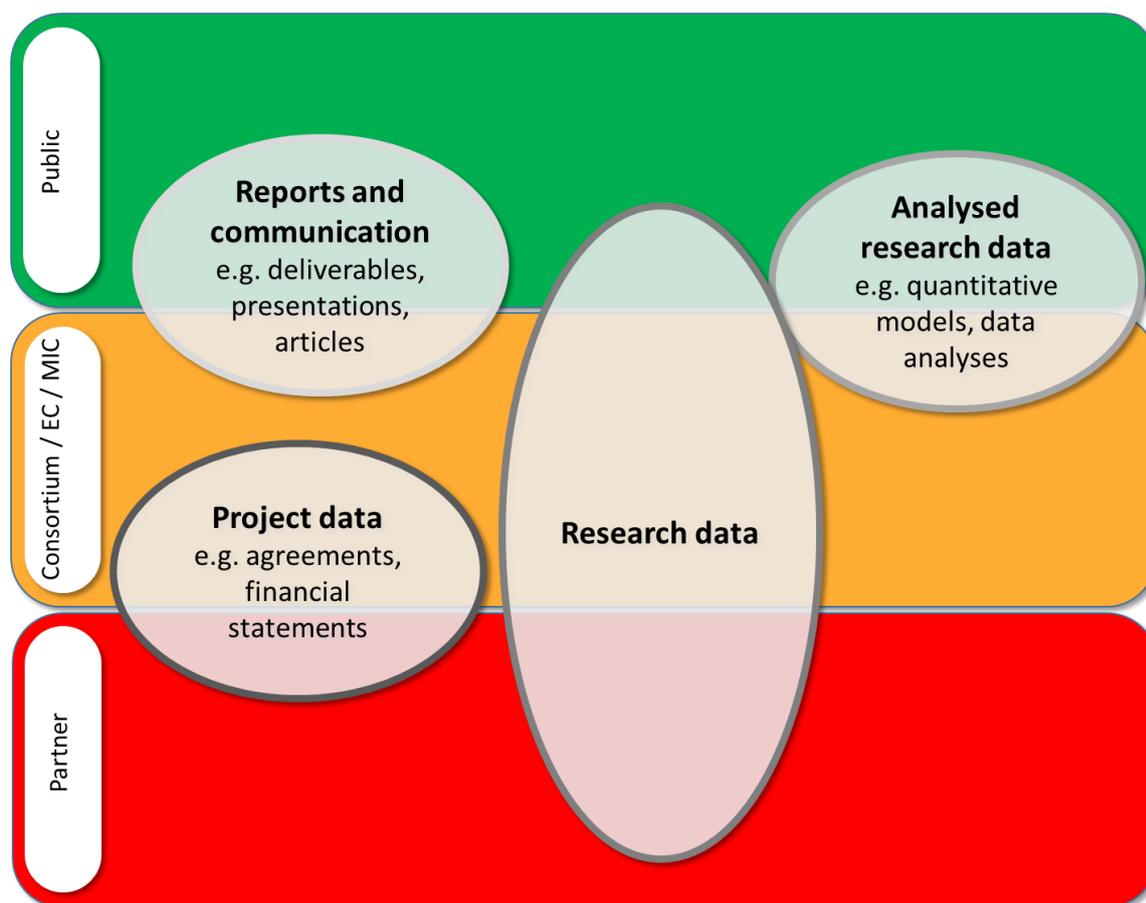


Figure 2 Data types displayed in three levels of confidentiality.

Figure 2 displays how the previously mentioned data types are positioned in the level of confidentiality context. Figure 3 displays the data in more granularity.



Figure 3 Data distributed into the three levels of confidentiality in more detail.

2.3. Data Formats

Recommended file formats are presented in the following table. See best practices for file formats from [Stanford University Library](#).

Table 2 Recommended file formats.

Text format	File extension
Acrobat PDF/A	.pdf
Comma-Separated Values	.csv
Open Office Formats	.odt, .ods, .odp
Plain Text (US-ASCII, UTF-8)	.txt
XML	.xml
Image / Graphic formats	File extension
JPEG	.jpg
JPEG2000	.jp2
PNG	.png
SVG 1.1 (no java binding)	.svg
TIFF	.tif, .tiff
Audio formats	File extension
AIFF	.aif, .aiff
WAVE	.wav
Motion formats	File extension
AVI (uncompressed)	.avi
Motion JPEG2000	.mj2, .mjp2

2.4. Expected Data of the Project

Table 3 presents the data expected to be created during the project.

Table 3 Expected data of the project.

Partner	Publications		Reports & Deliverables		Software		Open Research Data			Opted Out Research Data			Contribution to Open Sources			Contribution to Standards		
	Gold	Green	Consortium	Public	Consortium	Public	Yes 1)	No	May be	Yes 2)	No	May be	Yes 3)	No	May be	Yes 4)	No	May be
AALTO	X	X	X	X	X			X			X				X			X
UT	Δ	X	X	X	X			X			X				X	FGIMT2020,5GMF		
Ericsson	X	X	X	X	X			X			X			X				X
Orange	X	X	X	X		X		X			X				X	3GPP, ETSI NFV, ITU-T		
FOKUS																		
EI	X	X	X	X	X	X		X			X		X					X
MI	X	X	X	X												ITU-T		
DG	X	X	X	X	X			X			X			X				
KDDI			X	X				X			X			X		3GPP, 5GMF		
HITACHI			X	X	X			X			X			X				X
NESIC			X	X	X			X			X			X			X	
WU	X	X	X	X	X			X			X				X	ITU-TFG IMT2020, SG13		
If you answered yes to the following items, please give detailed comments/explanations below!																		
Yes 1)	Describe the data, give also format and expected amount (MB, GB)																	
Yes 2)	If you are opting out data (keep it confidential), describe the data and give reasons for opting out																	
Yes 3)	If you are contributing to open sources, describe the contributing data and the open source instance																	
Yes 4)	If you are contributing to standards preparation, describe the instance to contribute to																	
Detailed comments																		
Partner	Comment																	
AALTO	4) Aalto may contribute to ITU-T IMT 2020 Network Softwarization Working Group along with The University of Tokyo																	
UT	4) UT will contribute to ITU-T IMT-2020 and 5GMF Network Architecture WG																	
Ericsson																		
Orange																		
FOKUS																		
EI	3) EURECOM is interested to contribute the code around Open Air Interface (OAI) to the OAI Software Alliance																	
MI	4) MI will contribute to ITU-T SG20																	
DG																		
KDDI	4) KDDI will contribute to 3GPP SA2 WG and 5GMF Technology Promotion Group																	
HITACHI																		
NESIC																		
WU	4) Waseda will contribute to ITU-T FG IMT-2020 and may be to SG13																	

3. Collection, Storage and Use of Research Data and Analysed Research Data

The research data will be collected following jointly agreed guidelines and principles in order to guarantee the achievement of sound research data and to make the reusability of the research data possible. Possible data flows are presented in Figure 4.

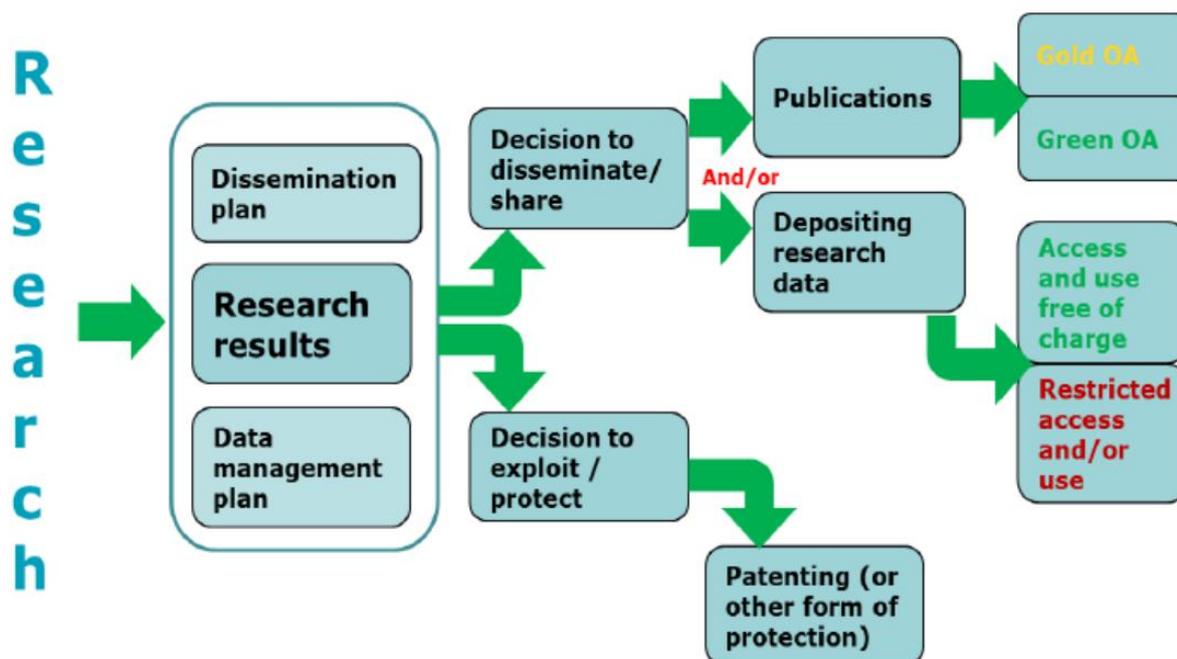


Figure 4 Open access to scientific publication and research data in the wider context of dissemination and exploitation (Source: H2020 Programme, Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, Version 3.1, 25 August 2016).

4. Storage and Use of Project Data

4.1. Storage of Data

The data accumulated in the 5G!Pagoda project will be analysed and stored according to the principles detailed in this plan. Overall the detailed research data will be stored by the organisations which has collected it, and the project data will be stored by Aalto University. Both the central repository and the databases of the individual organisations will be secured using the latest security protocols, and access to data will be granted only for persons nominated by the project partners.

All project administrative data will be stored at a dedicated database for the 5G!Pagoda project. The project uses the OwnCloud repository (<https://5g-pagoda.aalto.fi/owncloud/>), which is a secure, password protected document repository and archive system. Access to the database is managed by the coordinator and provided for project consortium and other parties as deemed necessary by the project team. The project data is stored on Aalto University servers, not in the cloud, for added security. The project data management structure and categories are as follows.

Workpackages

- WP1
- WP2
- WP3
- WP4
- WP5
- WP6

Templates

Standards-Activities

Project Meetings

- Physical Meetings
- ConfCalls

Market Analysis

Deliverables

- WP1
- WP2
- WP3
- WP4
- WP5
- WP6

Admin Relevant

- Original DoA
- Grant Agreement
- Consortium Agreement
- Coordination Agreement

The 5G!Pagoda Project Management, Quality and IPR Guide details the project internal management structure and processes, as well as quality and reporting practices. Related to project data management, best practices for data generation and sharing have been applied. This includes set rules for version control, whereby the partners are encouraged to use unified methods for naming the documents by the Task or Deliverable name, with a corresponding version number. The documents are stored in the database and preferably shared within the consortium via a link to the database rather than e-mail attachments. All deliverables have a unified look and feel with a unified template which helps the reviewers in their project evaluation.

The project coordinator assumes the responsibility for timely documentation and sharing of project management related documents and materials. Each Work Package (WP) leader monitors the timely documentation of WP related requirements within the consortium. Each task leader ensures the timely production of the deliverable for which he/she is responsible. With the strongly inter-related and intertwined Tasks in the different Work Packages, the same previously described principles regarding e.g. confidentiality levels and data types will be applied.

4.2. Documenting of Data

The stored data needs to be provided with metadata that complies with an international metadata standard. To facilitate this, the following questions should be answered already during the research work:

- Who are the creators and what are their affiliations
- Where the data is located and is there a persistent identifier
- What is the license chosen to allow reuse
- How, when and by whom the data has been collected/ created
- How the data has been prepared for analysis
- What kind of data manipulations have taken place
- How and what methods have been used to analyse the data
- What instruments and devices have been used
- Which scientific publications are based on this data
- What is the software used to process and analyse the data

Available metadata standards are listed e.g. in these sites:

- [General Metadata standards](#) listed by the Digital Curation Center (DCC) UK
- [Discipline-specific metadata standards](#) listed by the Digital Curation Center (DCC) UK
- [Metadata standards by topic](#) listed by the Research Data Alliance (RDA)

It is imperative to get a persistent identifier for data so that it is findable and citable. The most common identifier is Digital Object Identifier (DOI). Appropriate data repositories provide persistent identifiers for data sets.

The 5G!Pagoda project intends to use [DataSite](#) digital identifier for research data. For more details about DOI, the interested reader may refer to the [International DOI Foundation \(IDF\)](#).

To license a dataset requires either 1) that all creators agree to release the data they have created using the same license; or 2) the ownership of datasets is transferred to one legal entity. One straightforward and effective way of doing this is to attach [Creative Commons Licences \(CC BY or CC0\)](#) to the data deposited. The [EUDAT B2SHARE](#) tool includes a built-in license wizard that facilitates the selection of an adequate license for research data.

Table 4 presents the planned Open Access Data of the project. For repositories see Chapter 5.

Table 4 Curation of stored open data.

Partner	Data type	Repository	Metadata standard	License standard
All partners	PU deliverables	Participant Portal		
AALTO	Publications	TBD	TBD	TBD
UT	Publications	TBD	TBD	TBD
Ericsson	Publications	TBD	TBD	TBD

Orange	Publications	TBD	TBD	TBD
Orange	Software	TBD	TBD	TBD
EI	Publications	TBD	TBD	TBD
EI	Software	TBD	TBD	TBD
MI	Publications	TBD	TBD	TBD
DG	Publications	TBD	TBD	TBD
WU	Publications	TBD	TBD	TBD

5. Data Sharing

All parties have signed the Coordination Agreement and the European Partners have signed/accessed to the project Grant Agreement and Consortium Agreement, which all together detail the parties' rights and obligations, including – but not limited to – obligations regarding data security and the protection of privacy. These obligations and the underlying legislation will guide all of the data sharing actions of the project consortium.

The 5G!Pagoda project has committed to participate in the Pilot on Open Research Data in Horizon 2020, which is an expression of the larger Open Access initiative of the European Commission¹. Participation in the pilot is manifested on two levels: a) depositing research data in an open access research database or repository and b) choosing to provide open access to scientific publications which are derived from the project research. At the same time, the consortium is dedicated to protect the privacy of the informants and companies.

5.1. Depositing Research Data

Following the principles of the European Commission Open Data pilot, the applicable research data gathered in the project will be made available to other researchers through open access database or repositories, in order to increase the potential exploitation of the project work.

To make research data openly accessible, the following matters shall be considered:

- Specify which data will be made openly available? If some data is kept closed, provide rationale for doing so
- Specify how the data will be made available
- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?
- Specify where the data and associated metadata, documentation and code are deposited
- Specify how access will be provided in case there are any restrictions
- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

There are online research data archives, which may be subject-based/thematic, institutional or centralised. Useful listings of repositories include the Registry of [Research Data Repositories](#) and [Databib](#). The Open Access Infrastructure for Research in Europe (OpenAIRE) provides additional information and support on linking publications to underlying research data. Some repositories like [Zenodo](#) (an OpenAIRE and CERN collaboration), allows researchers to deposit both publications and data, while providing tools to link them. Zenodo and some other repositories as well as many academic publishers also facilitate linking publications and underlying data through persistent identifiers and data citations.

5.2. Open Access Publishing

All peer-reviewed scientific publications relating to results are published so that open access (free of charge, online access for any user) is ensured. Publications will either immediately be made accessible online by the publisher (Gold Open Access), or publications are available through an open access

¹ "The European Commission's vision is that information already paid for by the public purse should not be paid for again each time it is accessed or used, and that it should benefit European companies and citizens to the full." Source: Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020, Version 1.0 11 December 2013, p. 4. Link: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

repository after an embargo period, which according to the Grant Agreement can be six months at maximum (Green Open Access). Possible Gold Open Access journals include IEEE Access Magazine. For all other articles, the researchers aim at publishing them in a Green Open Access repository. The coordinator, Aalto University, has a Green Open Access repository the 5G!Pagoda consortium can use, at <https://aaltodoc.aalto.fi/?locale-attribute=en>.

A repository for scientific publications is an online archive. Institutional, subject-based and centralised repositories are all acceptable choices.

The [Open Access Infrastructure for Research in Europe \(OpenAIRE\)](#) is the recommended entry point for researchers to determine what repository to choose. It also offers support services for researchers, such as the National Open Access Desks. Other useful listings of repositories are:

- [Registry of Open Access Repositories \(ROAR\)](#)
- [Directory of Open Access Repositories \(OpenDOAR\)](#)

A machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication will be available in a repository for scientific publications. Electronic copies of publications will have bibliographic metadata in a standard format and will include "European Union (EU)" and "Horizon 2020", the name of the action, acronym and grant number; publication date, and length of embargo period if applicable, and identifier.

[ORCID](#) identifier is used to identify the authors of publications in the 5G!Pagoda project.

5.3. Other Dissemination

In addition to the above, the project will release yearly a public report on standardization, dissemination and exploitation achievements (deliverables D6.2, D6.3, and D6.4). At the end of the project, a report on innovation achieved and forthcoming industrial exploitation will be also released (D6.5).

6. Opting Out Research Data

The 5G!Pagoda project can opt out at any stage and so free itself retroactively from the obligations associated with the conditions – if:

- participation is incompatible with the obligation to protect results that can reasonably be expected to be commercially or industrially exploited
- participation is incompatible with the need for confidentiality in connection with security issues
- participation is incompatible with rules on protecting personal data
- participation would mean that the project's main aim might not be achieved
- the project will not generate / collect any research data or
- there are other legitimate reasons

Alternatively, the 5G!Pagoda project can also choose to keep selected datasets or even all data closed for any of the reasons above, via this Data Management Plan.

At this phase of the project no research data is identified and thus opting out is not relevant.

7. Summary and Perspectives

The Open data management plan will be updated during the project lifetime if new practices for data management are introduced.