Document Number**:** H2020-ICT-52/RISE-6G/D8.6

Project Name:
Reconfigurable **I**ntelligent **S**ustainable **E**nvironments for **6G** Wireless Networks (RISE-6G)

Deliverable 8.6

Data Management Plan

Date of delivery:	09/07/2021	Version:	1.0
Start date of Project:	01/01/2021	Duration:	36 months

Deliverable D8.6
Data Management Plan

Project Number:	H2020-ICT-52-101017011
Project Name:	**R**econfigurable **I**ntelligent **S**ustainable **E**nvironments for **6G** Wireless Networks

Document Number:	H2020-ICT-52/RISE-6G/D8.6
Document Title:	Data Management Plan
Editor(s):	NKUA
Authors:	Kyriakos Stylianopoulos (NKUA), George Alexandropoulos (NKUA)
Dissemination Level:	PU
Contractual Date of Delivery:	30/06/2021
Security:	Public
Status:	Final
Version:	1.0
File Name:	RISE-6G_D8.6.docx

Abstract

The present document consists the first version of the Data Management Plan of the RISE-6G project. It describes the data management life cycle for the data to be collected, processed, and/or generated by the project. Since making research data Findable, Accessible, Interoperable, and Re-usable (FAIR) is one of the primary goals of the project in compliance with the relevant guidelines from the European Commission, this document includes information on the handling of 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. The RISE-6G project is at its early stages, hence, not all aspects of data management have been yet established. Nevertheless, for such cases, principles and guidelines for the forthcoming actions are presented in this deliverable.

Keywords

Data management, FAIR data, accessibility, interoperability, re-use, data security, ethical aspects

Contents

DISCLAIMER

This document has been produced in the context of the project Reconfigurable Intelligent Sustainable Environments for 6G Wireless Networks (RISE-6G).

The Research leading to these results has received funding from the European Union's Horizon 2020 research and innovation program under the grant agreement ID 101017011.

1 Introduction

The RISE-6G vision capitalizes on the latest advances on the technology of Reconfigurable Intelligent Surfaces (RISs) for radio wave propagation control, in order to substantially improve them and to conceive and implement intelligent, sustainable and dynamically programmable wireless environments that go well beyond the 5G capabilities, developed under 3G PPP releases 16 and 17.

To accommodate the aforementioned objectives, RISE-6G aims to: (i) define novel network architectures and operation strategies incorporating multiple RISs; (ii) characterize its fundamental limits, capitalizing on the proposed by the project realistic and validated radio wave propagation models; (iii) design solutions to enable online trade between high-capacity connectivity, energy efficiency, Electro-Magnetic Field (EMF) exposure, and localization accuracy based on dynamically programmable wireless propagation environments, while accommodating specific legislation and regulation requirements on spectrum use, data protection, and EMF emission; and (iv) prototype-benchmark proposed innovation via two complementary trials with vertical industries.

As a result, the project is expected to use, modify, create, and publicize various collections of data. To address the issue of the management of data, the present document constitutes the first version of the Data Management Plan (DMP) for the RISE-6G project. The aim of this document is to describe the existing and planned data management procedures, data access, and the data security policies established by the consortium. Since the project is at its earlier stages, the current report is a preliminary one. In its current form, it concerns the decisions and strategies adopted so far, and provides an estimate of the foreseen changes and amendments to the DMP. To attain relevant and appropriate data management policies, this document is to be updated concurrently to the project deliverables, with new versions being finalized by the periodic management reports.

2 Data Summary

This section specifies the purpose of the data, the formats and origin, the size, and to whom it will be useful.

There are various kinds of data to be generated in the project. Foremost, through physical simulations of wireless propagation experiments, experimental data will be generated which will constitute the simulated measurements. In parallel, actual simulations will be carried out in prototyped metasurfaces that will result in corresponding physical measurements. The format of such data is numeric. Accompanied descriptive data/metadata are to be stored and distributed alongside. Accordingly, another set of produced data will correspond to the results and evaluation processes of the proposed algorithms, methodologies, and pipelines utilized and developed in the project. Such data are mainly expected to be of figure-like format (e.g. diagrams, plots, and charts) or numerical (e.g. metric and KPI values). The relative code also constitutes part of the generated data, as well as the technical descriptions of the prototyping of the various forms of metasurfaces. Lastly, it is evident that throughout the project, a plethora of miscellaneous data will have to be accumulated and appropriately communicated between the partners. Those mainly concern managerial and dissemination actions mainly on document and multimedia formats, such as publications, slides, documents, talks, tutorials, etc.

The produced data sets are expected to be of moderate sizes (e.g. in the order of a few TBs), but this specification is due to modifications upon the first versions of the output data that will be produced. We envision the data to be useful to researchers in wireless communications, electrical engineering, physics, data science, mobile and communication network operators, and regulating agencies.

3 FAIR data

3.1 Data Findability, including provisions for metadata

This section outlines the discoverability and identifiability of the data and the use of persistent and unique identifiers.

Document-formatted data used for collaborating and preparation of the deliverables will be circulated through the Sharepoint platform, for which the partners have been granted access. At the current stage of the project, not enough amount of experimental data has been yet generated, as a result, a central repository has not been established, thus far. Nevertheless, the use of a standard OpenAIRE-compatible repository, such as Zenodo has been proposed. The details of the migration are to be discussed in the forthcoming managerial meetings. For dissemination-related data, such as publications, they are currently circulated through a specific email list, monitored by the responsible team of Work Package 8 and are listed in special documents within the Sharepoint platform. Upon the adoption of the data deposit, a detailed record of those will be kept there.

Up to now, the shared data within the project are indexed per Work Package and/or deliverable. As a result, there is no special need for detailed naming convention protocols and or additional metadata. Again, when more detailed data (e.g. experimental results) are to be produced, such policies will be agreed between the partners that will specify commonly used identifiers, keywords, conventions, etc.

3.2 Data Accessibility

This section specifies the extent of open access, how the data is made available, what methods and tools are used.

Experiment and model data might be openly accessible only after the end of the projects as soon as the results have been published. Publications will be open access. Where relevant and possible with regards to property rights, developed software will be made available through the open source repository Github or similar using a GPL license. We have not established a concrete policy on data access for all possible data categories, since it is possible that certain datasets cannot be shared for legal and contractual reasons. Because of such provisions, data access is to be evaluated per case/type of data.

3.3 Data Interoperability

This section covers what data and metadata vocabularies, standards, and methodologies are used to facilitate interoperability.

Since the procedures for data-sharing have not yet been established, there is no detailed plan on interoperability. Nevertheless, the future interoperability modules will be designed so as to

allow for inter-disciplinary integration with community and confrontation with commonly used standards, practices, and methods of operations.

However, due to the novel technology developed by the project, it is expected that new ontologies and terminology will be adopted. Those will be foremost described in detail in the relevant deliverables and resulting publications, but brief descriptions in dictionary forms will be given alongside with the relevant data when made public.

3.4 Data Re-use

This section specifies data licensing, availability, and length of time for re-use.

Wherever possible the data will be shared right after production following the Creative Commons 4.0 International License with Attribution (CC4BY). Experimental data will, in some cases, only become available after the end of the project or after the publication of the results, whatever comes first, and will be shared by the same CC4BY license.

The CC4BY licenses guarantees maximum re-use (and redistribution), while maintaining the traceability of the use and credit to the data providers and their sponsors. Since the data mainly concern experimental findings on an emerging technology with great potential applications, it is expected for the published data to be useful to the community for a period over five years.

4 Allocation of resources

Up to present, the cost of the data management has been incorporated to the budget of the respective institutions. When the central platform will be launched, the majority of the budget is expected to be allocated there and will depend on the storage costs of the chosen platform. The cost of long-term preservation of the data is at this moment impossible to estimate. While some subsets of data can be indefinitely stored publicly (e.g., code on Github), the decision on the permanent storage of them will be taken by the later stages of the project.

5 Data security

RISE-6G produces non-sensitive data. Personal information is to be processed and stored according to the respective platforms' policies. In case of a foreseen event of requiring personal, sensitive data, the DMP will be updated accordingly.

6 Ethical aspects

The project primarily makes use of synthetic data during its run-time. Such instances do not constitute personal data, thus they are not expected to pose ethical issues. In case of using real-world measurement data (for instance during the trial phase), such datasets will be anonymised and processed only during the lifetime of the project. In such cases, only the necessary information for the intended applications will be collected and will be used exclusively for that purpose.