D6.2 – Data Management Plan

Deliverable 6.1. – Data Management Plan

TwInSolar

(Improving Research and Innovation to achieve a massive integration of Solar renewables)

Organisation: University of La Réunion

Main author: Mathieu David

Contributors: CPMR, Fraunhofer

This document has been developed as part of the project titled "**TwInSolar – Improving research and innovation to achieve a massive integration of solar renewables**"

The TwInSolar project (September 2022 - August 2025) has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 10107647

Document information

Deliverable:	D 6.2
Work Package:	WP6 – Project management
Task:	T6.3 – Data and Risk Management
Author:	Mathieu David
Document type:	Report
Dissemination level:	Public
Copyright:	Copyright © 2023, TwInSolar Consortium – All right reserved

Revision

Version	Date	Author	Description of changes
V1	2-12-2022	Mathieu David	Drafting
V2	3-12-2022	Nicolas Alarcon	Content update
V3	7-12-2022	Claire Helly	Typo corrections
V4	23-03-2023	Chloé Durif	Name, layout and minor modifications

Contents

Executive Summary

The Data Management Plan (DMP) is a living document that aims at providing an analysis of the main elements of the data management policy that will be used by the TwInSolar Consortium regarding the project data.

This document describes the methodological framework taking into account the Horizon Europe guidelines regarding Open Research Data with a twofold objective:

1. to clearly define how all public data assets issued by the project consortium will be further disseminated to a wider audience,
2. to outline the methodology which can make the research data generated in the context of TwInSolar, findable, accessible, interoperable and reusable (FAIR principles).

This DMP will evolve during the implementation of the project, when the project progresses and when significant changes occur, in order to keep an updated version of the guidelines and recommendations and therefore contribute to knowledge discovery and innovation.

The current version is the first iteration in which the envisioned data management strategy is presented and first effort is made to plan the definition of the types of research data that will be generated or re-used during the project, the standards that will be used, how the data will be preserved.

This document is open for further iterations and will be updated during the project lifecycle.

I.   Data Summary

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded. What types and formats of data will the project generate or re-use? What is the purpose of the data generation or re-use and its relation to the objectives of the project? What is the expected size of the data that you intend to generate or re-use? What is the origin/provenance of the data, either generated or re-used? To whom might your data be useful ('data utility'), outside your project?

The TwInSolar project is mainly focused on knowledge transfer, but it also includes a small research part dedicated to the study of innovative solutions for the microgrid of the Terre Sainte university campus. TwInSolar partners will thus re-use existing data and generate new data to support these two activities. After only 3 months, the project is not yet in the position to provide a detailed list of all the data sets of the project neither the details on the formats and specific standards to be used. As defined in the DoA, it is expected to have this information by Month 6 and after. Therefore, in this version of the Data Management Plan overviews of already identified data sets are only included. Next version of the DMP will include up-to-date decisions on datasets to be used.

The first re-used data identified in the project (see Table 1) will provide the information needed to estimate past, current and future energy transfers of 2 selected energy systems: Terre Sainte university campus and Reunion power grid.

Table 1: Summary of the re-used data

Description	WP	Type	Format	Size	Source
Electricity generation and consumptions by districts and end-uses of La Réunion	WP1	Observations	Text (*.csv)	≈10Mo	https://opendata-reunion.edf.fr/
Overall electricity consumption Terre Sainte campus	WP4	Observations	Text (*.csv)	≈10Mo	EDF
Cooling load and electricity consumption by sub-meters Terre Sainte campus	WP4	Observations	Text (*.csv)	≈1Go	UR
Onsite photovoltaic generation Terre Sainte campus	WP4	Observations	Text (*.csv)	≈100Mo	UR
Images from Meteosat geostationary satellite MSG for La Réunion	WP4	Observations	Images (HRIT)		eumetsat
Weather forecasts La Réunion	WP4	Simulations	GRIB or NetCDF	≈1Go	ECMWF

Similarly, to the re-used data, the generated data will support knowledge transfer and research activities planned in the project. They will provide additional information needed to size and to optimally schedule the operation of the selected energy systems.

Table 2: Summary of generated data

Description	WP	Type	Format	Size	Data utility
Images from All Sky Imagers (ASI)	WP4	Observations	Images		Generation of very short-term solar forecasts
Forecasts of sky imager derived maps	WP4	Forecasts	Images	Depends on forecast resolution	Development foreseen in the project, intermediate results towards solar generation forecasting
Forecasts of satellite-based cloud index images	WP4	Forecasts	Images	24 x times size of satellite images	Development foreseen in the project, intermediate results towards solar generation forecasting
Forecasts of solar power generation and energy load	WP4	Simulations	Text (*.csv)	≈10Go	Development foreseen in the project and dissemination

II.   FAIR data

A) Making data findable, including provisions for metadata

Will data be identified by a persistent identifier? Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how. Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use? Will metadata be offered in such a way that it can be harvested and indexed?

The dataset will be identified by DOIs generated by DataCite (https://datacite.org). The XML Schema of DataCite will be used to the metadata discovery. This format includes research keywords facilitating the discovery and it is also harvested and indexed. The Dublin Core Elements will serve as standard for creating the metadata.

B) Making data accessible: Repository

Will the data be deposited in a trusted repository? Have you explored appropriate arrangements with the identified repository where your data will be deposited? Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

During the project, the data will be shared between the partners through a collaborative repository created on Microsoft TEAMS by the coordinator of the project (UR). This repository will allow a secured access to the data within the project.

For external dissemination, once the date will be consolidated, a DOI will be assigned with DataCite and the data will be uploaded to a trusted and open access repository like zenodo.org or recherche.data.gouv. So, the assigned DOI will provide a direct access to download the data.

C) Making data accessible: Data

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement. If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible. Will the data be accessible through a free and standardized access protocol? If there are restrictions on use, how will access be provided to the data, both during and after the end of the project? How will the identity of the person accessing the data be ascertained? Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

During the project, the raw data will be accessible by all TwInSolar partners. The appropriate access rights to the TEAMS repository (i.e. login and password) will be granted by the coordinator.

We plan to use a trusted repository like zenodo.org or recherche.data.gouv to disseminate the consolidated dataset. These depositories are open access and free of charge. They also provide assigned identifiers and direct access to data.

D) Making data accessible: Metadata

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data? How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available? Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

Through DataCite, metadata will be freely available and licensed under a CC0 public domain dedication. The metadata and corresponding data will remain accessible and findable in the long term. No specific software or code will be required to read the data.

E) Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones? In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them? Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?

As no commonly accepted ontologies exist in the field of energy, the metadata and data will follow the best practices of the scientific community in the domains of energy and solar forecasting.

When proposed by the provider, our re-used data will include qualified references.

F) Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g., readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)? Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement? Will the data produced in the project be usable by third parties, in particular after the end of the project? Will the provenance of the data be thoroughly documented using the appropriate standards? Describe all relevant data quality assurance processes. Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

The data will come with readme files with information on methodology, data cleaning, analyses, variable definitions, units of measurement, etc. The consolidated dataset will be available to the public after the end of the project in the public domain and licensed using standard reuse licenses. The reuse of data by third parties will be encouraged.

The quality check procedure defined in the "Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: Third Edition" (Technical Report NREL/TP-5D00-77635, April 2021) will be applied to the solar irradiance data.

III. Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

The other research outputs of the project are a methodology to size a smart micro-grid and numerical models to blend different sources of forecasts. They will be distributed in open access.

IV. Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g., direct and indirect costs related to storage, archiving, re-use, security, etc.)? How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions) Who will be responsible for data management in your project? How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

The costs for making data or other research outputs FAIR are:

- The Microsoft 365 License needed to manage a TEAMS repository,
- The secured storage of the raw data of the Terre Sainte university campus,
- The publication fees in open access scientific journals.

The costs related to the data repository used during the project (TEAMS) and the storage of the data of the Terre Sainte university campus will be directly supported by the partners as they are part of existing services. However, the publication fees will be supported by the TwInSolar project.

UR, as leader of the Work Package 6 "Project management", will be responsible for data management. A secured database maintained by UR will allow a long-term preservation of the raw data collected and used in the project. A repository like zenodo.org will assure the long-term preservation and accessibility to the consolidated dataset.

V. Data security

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)? Will the data be safely stored in trusted repositories for long term preservation and curation?

Microsoft Teams, which is used to store and share the data within the project, provides high data security, i.e. two-factor authentication, single sign-on through Active Directory, encryption of data in transit and at rest, data loss prevention (DLP), etc. Microsoft TEAMS complies with the following standards: ISO 27001, ISO 27018, SSAE18 SOC 1 and SOC 2, HIPAA, and EU Model Clauses (EUMC).

For the long-term preservation, the consolidated data will be safely stored in trusted depository like zenodo.org or recherche.data.gouv.

VI.   Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?

No ethics or legal issues related to the data are identified in the TwInSolar project.

VII.   Other issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

We will not use national/funder/sectorial/departmental procedures for data management.

Index