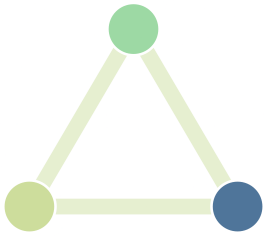
Resilient Energy System Infrastructure Layouts for Industry, E-Fuels and Network Transitions



Supported by:

Federal Ministry for Economic Affairs and Climate Action

on the basis of a decision by the German Bundestag

CET Partnership 2022 | RESILIENT

DATA MANAGEMENT PLAN

PROJECT	
Ecosystem:	RESILIENT is part of the The Clean Energy Transition (CET) Partnership initiative co-funded by the European Union
Number:	Cetp-2022-00109
Acronym:	RESILIENT
Name:	Resilient Energy System Infrastructure Layouts for Industry, E-Fuels and Network Transitions
Dissemination level:	Public
Coordinator:	Technical University of Berlin
Partners	Technical University of Berlin (TUB), Germany University of Pisa (UNIPI), Italy Fraunhofer ISI, Germany Chalmers University of Technology (Chalmers), Sweden Lappeenranta-Lahti University of Technology (LUT), Finland TransnetBW, Germany Électricité de France (EDF), France IN4climate.NRW, Germany Stockholm Exergi, Sweden ABB Finland, Finland
National funding organisations	Germany: Forschungszentrum Jülich GmbH (BMWK) - FZJ/PtJ Italy: Ministero dell'Università e della Ricerca – MUR Sweden: Swedish Energy Agency - SWEA Finland: Innovaatiorahoituskeskus Business Finland - BF France: Agence de la transition écologique - ADEME

DATA MANAGEMENT PLAN	
Date:	05/08/2024
Version:	v0.1
Deliverable:	D1.2
Authors:	Technical University of Berlin (TUB): Iegor Riepin, Bobby Xiong, Tom Brown
Comments/errata:	For any comments please contact: iegor.riepin@tu-berlin.de

1. Executive summary

The Data Management Plan (DMP) characterises the existing and planned data management, data security, and data access policies within the RESILIENT project. The project's data management plan is closely linked to its open science plan. The consortium will make every effort to ensure project outcomes adhere to the FAIR data principles, i.e., data has to be Findable, Accessible, Interoperable, and Reusable.

2. Data Summary

Will you re-use any existing data and what will you re-use it for?

Yes, RESILIENT database will incorporate and re-use any relevant data and information from repositories or official portals that might be useful for the project objectives. These sources include existing open-source data from energy and weather databases such as JRC-IDEES, eurostat, and ERA5 products. These datasets will be utilized to enhance the PyPSA-Eur model by providing historical data for energy modelling, as well as for calibration and validation purposes.

What types and formats of data will the project generate or re-use?

The project will generate and re-use data in various formats, including CSV, JSON, GEOJSON, XML, TIF, NC, and XLS. The data types collected and processed for energy system modelling include, but are not limited to, energy consumption data, techno-economical data for energy technologies, geospatial data, energy time-series data, industry transformation datasets, and global e-fuel trade data.

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

The data collected by RESILIENT will support the development of a multi-vector energy infrastructure planning model PyPSA-Eur. This directly relates to the project's objecttive: improving an ability of industry need-owners and broader scientific community to optimise energy infrastructure at a sub-national, national and European levels in a sustainable and resilient way. The detailed data on industry (e.g., fuel and process switching options) will help in modeling the industry transformation pathways and understanding the impact of uncertainties on the European energy system.

What is the expected size of the data that you intend to generate or re-use?

Output datasets will vary in size from a few MB to a couple of GB. The size of datasets is planned to be capped at a level allowing pulbication at Zenodo, which has a 60 GB maximum file size.

What is the origin/provenance of the data, either generated or re-used?

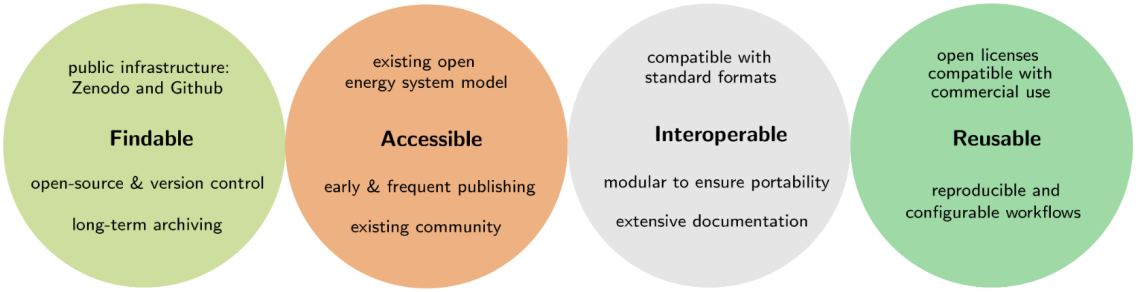
The data will originate from public databases (JRC-IDEES, eurostat), meteorological reanalysis projects (such as ERA5 carried out by the European Centre for Medium-Range Weather Forecasts), open collaborative project such as Open Street Map, as well as models such as FORECAST (FORecasting Energy Consumption Analysis and Simulation Tool) maintained by Fraunhofer ISI and PyPSA-Eur (A Sector-Coupled Open Optimisation Model of the European Energy System) maintained by TUB.

To whom might your data be useful ('data utility'), outside your project?

The data will be useful to energy planners, policymakers, academic researchers, and industry stakeholders interested in energy system optimization, infrastructure planning, and industry decarbonization pathways.

3. FAIR data

The project's data management plan is closely linked to its open science plan. The consortium will make every effort to ensure project outcomes adhere to the FAIR data principles, i.e., data has to be Findable, Accessible, Interoperable, and Reusable.



3.1. Making data findable

Will data be identified by a persistent identifier? Will rich metadata be provided to allow discovery? What disciplinary or general standards will be followed? How long will the data remain available and findable?

The consortium will exclusively use public infrastructure to publish data and code. This assures long-term archiving of all project outcomes beyond the project lifetime and avoids commertial solutions for data management.

- ✔ Code for energy system model PyPSA-Eur will be open-source and managed on Github with versioned releases: https://github.com/PyPSA/pypsa-eur
- ✔ Input and output data, as well as tool releases, will be published and versioned on Zenodo & Github with persistent and unique DOIs.

  Data bundle for PyPSA-Eur: https://zenodo.org/records/12760663
  Code for power plant data tool: https://github.com/PyPSA/powerplantmatching
  Code for weather & spatial data tool: https://atlite.readthedocs.io/en/latest/
  Code for technology data tool: https://github.com/PyPSA/technology-data|
  Code for modelling framework PyPSA: https://github.com/PyPSA/PyPSA
- ✔ The project website will not host data itself but link to the respective persistent repositories where tools, datasets and reports are stored: https://resilient-project.github.io/
- ✔ Research outputs will be posted on preprint servers, like arXiv, first in order to dispense insights quickly, before open-access publishing in academic journals is approached.
- ✔ Keywords will be included in the metadata of GitHub repositories, and research deliverables (reports, woking papers, journal articles) to optimize discoverability.

3.2. Making data accessible

Will the data be deposited in a trusted repository? Will the data be accessible through a free and standardized access protocol?

As specified in section 3.1, data will only be stored in free and trusted repositories such as Zenodo and GitHub. Data will be accessible through standardized protocols, ensuring ease of access and interoperability. This includes using web interfaces and APIs provided by repositories like Zenodo and Github. Data will remain available and findable for many years after the project ends. As of now, data is guaranteed to remain available for the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.

The consortium will ensure accessibility through early and frequent publishing. Building on the foundations of

an existing open model allows releases from day one with continuous improvements, rather than an all-encompassing release at the project's end. The fast and regular release rhythm with exhaustive version control expedites the diffusion of new features and datasets to the work of need-owners and other energy models. Intermediary data for sharing among project partners is stored in internal cloud storage by the coordinating partner (TUBcloud).

3.3. Making data interoperable

What data 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? Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)?

The consortium will ensure interoperability through modularity, documentation and compatibility with standard formats. It is our our priority to avoid model lock-in by modular design so that other energy system models can leverage selected parts of this project's tools and dataset effortlessly and according to their needs. We will continue to develop extensive documentation of data and software and further facilitate portability. This includes machine-readable and open formats (e.g., CSV, XML, JSON). The data will include qualified references to raw data sources, supporting integration and interoperability with related research and datasets. A machine-readable output of the pypsa/technology-data project is an example of portable data interface that supports version control: https://github.com/PyPSA/technology-data/tree/master/outputs

For all project outcomes we will use approved open licences that are compatible with commercial use. This maximises the applicability of the tools developed. It allows industry players to scope business opportunities, public authorities and network operators to evaluate infrastructure plans, and academics to focus on research rather than model development. We will deploy community-endorsed lisensing tools and standards, such as REUSE, to ensure that data can be exchanged easily and without restrictions: https://reuse.software/

3.4. 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?

The consortium will provide and continuously update documentation of the PyPSA-Eur model and related tools that will be hosted through no-cost solutions like Readthedocs or Github Pages:

Docs for PyPSA-Eur: https://pypsa-eur.readthedocs.io/en/latest/
Docs for power plant data tool: https://powerplantmatching.readthedocs.io/en/latest/
Docs for weather & spatial data tool: https://github.com/PyPSA/atlite
Docs for technology data tool: https://technology-data.readthedocs.io/en/latest/
Docs for modelling framework PyPSA: https://pypsa.readthedocs.io/en/latest/

To lower the bar for getting started with the modelling tools developed in the project, we will advance documentation in a way that will help both new and experienced users improve their understanding of the models. We will extend our portfolio of tutorials and illustrate how the tools can be applied in practice. The tutorials will be recorded and made available on YouTube so that they can be accessed asynchronously. An interactive website will also target a more general audience without modelling expertise to explore sustainable energy futures.

4. 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? Who will be responsible for data management in your project?

The project budget will cover the costs associated with making data FAIR. These costs include personnel time spent managing data and open access publishing fees. The code and data will be stored exclusively in free-of-charge repositories, such as Zenodo and GitHub.

5. 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?

Data security measures will include using trusted and inherently reliabile repositories such as Zenodo and Github, using version control systems for code and documentation (git), and controlled access to sensitive data (TUBcloud storage).

6. Ethics

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?

Ethical issues have low potential impact on data sharing in RESILIENT. It is anticipated that all data and information that might contain sensitive information (i.e. personal data) will be already aggregated and de-individualised by providers of the raw data (JRC-IDEES, eurostat).

7. 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)?

No other procedures will be applied.

HISTORY OF CHANGES

VERSION	PUBLICATION DATE	CHANGE
v0.1	05.08.2024	Initial version.