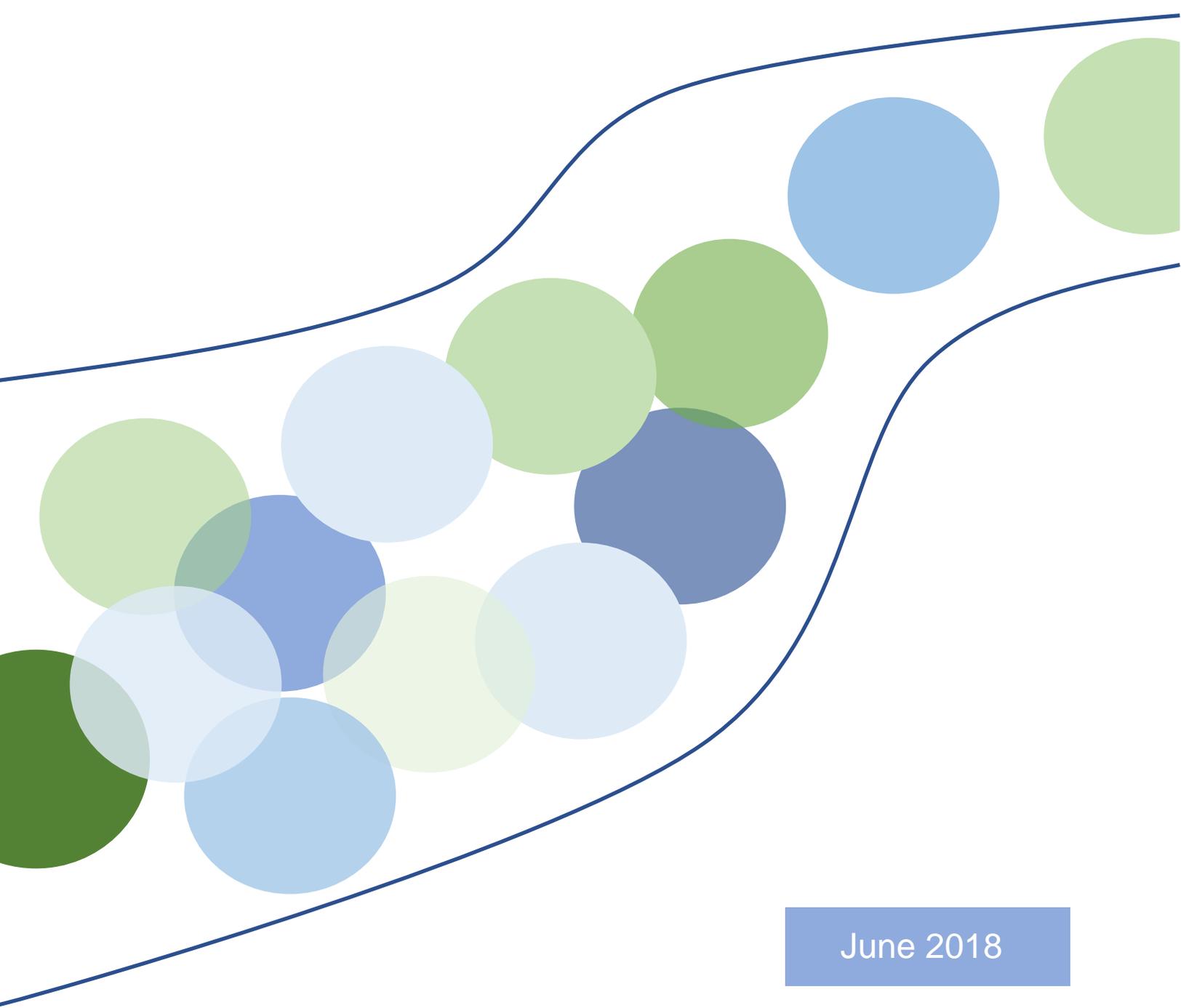


## D8.2 REEEM Data Management Plan (DMP)

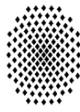
Collection, processing and dissemination of data



June 2018



## Project Partners



Universität Stuttgart

TOKNI



## Project Information

**Project Number:** 691739

**Project Acronym:** REEEM

**Project title:** Role of technologies in an energy efficient economy – model based analysis policy measures and transformation pathways to a sustainable energy system

## Authors

Ludwig Hülk (RLI), Berit Müller (RLI), Steve Pye (UCL), Francesco Gardumi (KTH)

## History of changes

Version	Publication date	Change
0.1	28.06.2018	Internal version for review
0.2	06.07.2018	Reviewed versions
1.0	23.07.2018	Initial version for publication



# TABLE OF CONTENTS

Project Partners .....	2
Project Information .....	2
Authors .....	2
History of changes .....	2
Project summary .....	4
About this report .....	4
Principles & Summary .....	4
1. DMP Checklist (DMPonline) .....	5
A. Data Collection .....	5
B. Documentation and Metadata .....	5
C. Ethics and Legal Compliance .....	5
D. Storage and Backup .....	6
E. Selection and Preservation .....	7
F. Data Sharing .....	7
G. Responsibilities and Resources .....	7
H. Data Protection Impact Assessment .....	8
2. Definition and Matter .....	10
3. Links .....	11



## Project summary

REEEM aims to gain a clear and comprehensive understanding of the system-wide implications of energy strategies in support of transitions to a competitive low-carbon EU energy society. This project is developed to address four main objectives

- (1) to develop an integrated assessment framework
- (2) to define pathways towards a low-carbon society and assess their potential implications
- (3) to bridge the science-policy gap through a clear communication using decision support tools
- (4) to ensure transparency in the process.

This deliverable includes considerations on how the data is handled both during and after the completion of the project.



The REEEM project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 691739. This publication reflects only the views of its authors, and the European Commission cannot be held responsible for its content.

## About this report

The Data Management Plans (DMP) constitutes a deliverable of the project. It aims at providing a clear picture of how the project manages data and metadata. The DMP describes the data management life cycle for the data to be collected, processed and/or generated by this Horizon 2020 project.

## Principles & Summary

One of the grand challenges of data-intensive science is to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of, scientific data and their associated algorithms and workflows. The [FAIR data principles](#) are a set of guiding principles to make data **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (FAIR). This research project uses these principles to improve the scientific work during the project and the dissemination of the project results.

# 1. DMP Checklist (DMPonline)

## A. Data Collection

- **What data will you collect or create?**

Achieving the interdisciplinary and wide-ranging project objectives requires the use of a wide range of methods, including mathematical modelling of different parts of the energy system and analysis of current systems. Most of these methods necessitate data collection to populate models, calibrate them, as well as allow for data exchange between different types of models and different partners. Given the wide scope and interdisciplinary approach of the project, a wide range of types and formats of data will be collected and generated, including but not limited to:

- **Quantitative data**
  - Data that describes snapshots of current or potential future systems and their performance, including characteristics of current infrastructure, electricity generation facilities, their efficiencies, social and economic data for European countries and worldwide (e.g., tabular data)
  - Time series data describing the performance of historical and potential future energy systems, including energy demand and supplies (e.g., tabular time series data)
  - Geospatial data (e.g., shapefiles, maps)
- **Qualitative data**
  - Interview transcripts (e.g., survey data)
  - Photos, audio and video recordings of events (e.g., conference material)
  - Project related works (e.g., reports, graphics, plots)

- **How will the data be collected or created?**

Most data collected and generated will be stored using open formats (e.g., CSV, TXT), or at least, formats that can be accessed by others without significant expense (e.g., CSV, XLSX, DOCX). Project internal documents and data are stored and made available for the project partners using a web-based collaborative platform (SharePoint). Modelling data (input and output data tables) will be processed and stored in an internal project database (PostgreSQL) hosted by one of the project partners (DTU).

## B. Documentation and Metadata

- **What documentation and metadata will accompany the data?**

Within the project, data will be made interoperable by publishing data in common machine readable open formats (e.g., CSV, PostgreSQL), which can be accessed through open and free software (e.g., Python, LibreOffice). The accompanying metadata is gathered in a human and machine readable JSON format. The project uses the [OEP metadata version v1.3](#). This selected metadata set follows the requirements of established metadata standards (e.g., frictionless data, Dublin Core, DCAT-AP). Database tables have the metadata string as a “comment on table” and data files have an additional metadata file.

## C. Ethics and Legal Compliance

- **How will you manage any ethical issues?**

Since some investigations in this project involve the participation of human participants, we will take the following steps towards safeguarding of any data that is being generated through and from them. The project will follow strict guidelines on ethical conduct involving human participants and this includes following the standards set for



European researchers through "[The European Code of Conduct for Research Integrity](#)" that applies to research in all scientific and scholarly fields.

All those investigations (e.g., surveys, questionnaires) that involve human participants (internal and external) will be handled accordingly to current data privacy protection regulations (privacy policy). This requires that everybody is being asked for their consent in relation to their involvement and all potential risks and benefits of their voluntary participation. This information will also include assurances regarding protection of their privacy and that they can withdraw their participation. Participants are being handed an information sheet for future reference.

No individual will be identified by name in any published materials. To this end, survey data will be stored securely in the server of the partner conducting the survey, following the national and international data privacy regulations. Since this project is part of the binding research collaboration, selected material may be shared across the project team, but only to be viewed and analysed by researchers, for research purposes only. The project will use pseudonymisation for sharing sensitive data internally that has been generated from and through human participants. Any material that is shared externally or is used for the dissemination of data will not utilise any information that could identify individuals. Participants will be informed of this. Researchers involved in this project are aware of the importance of data protection and privacy issues in this kind of research.

- **How will you manage copyright and Intellectual Property Rights (IPR) issues?**

Each data set is assessed as to whether intellectual property rights are affected. The copyright ownership of generated data follows national and international legislation. The generated data will be made available as early as possible during the project. This can include publishing data at the same time open access articles or other study results are published. At the minimum, data will be made available at the end of the project.

Data for use by third parties will be facilitated by appropriate open licensing of data (following the [Open Definition](#)). For example, the ODC-BY license (a permissive open license) allows commercial re-use of data with appropriate attribution of the copyright owner. However, this will only apply if there is no conflicting license on used input data or unsolved IPR. Additionally, sensitive data may not be re-used if restricted. As data will be published on different repositories (as described in the next section) with long-term support, the data will be re-usable for future research or any other purpose.

## **D. Storage and Backup**

- **How will the data be stored and backed up during the research?**

All data will be backed up on the server infrastructure of the responsible institution as well as in a central backup of the SharePoint system. All open data will be copied to the [OpenEnergyPlatform](#) (OEP). The OEP has additional regular backups as well as a data versioning system. Long term preservation, monitoring, and curation are ensured. In addition, each partner is responsible for creating independent and physically separate backups. All data will be stored in the internal project database, which is backed up regularly.

- **How will you manage access and security?**

As most data will be published as open data, data security issues do only apply for sensitive data. The internal project database has a multi-level user management with "admins" who take care of user rights, "users" who have unlimited access to the database content, and "visitors" with read access to selected parts. All user accounts are password protected. The permission rights of each data set are also stored in the metadata.

The access to pseudonymized sensitive data will be restricted to the project partners and the supervisors from the European Commission. In the case that sensitive data without anonymization needs to be shared across partners within the project, individual arrangements must be made.



## **E. Selection and Preservation**

- **Which data are of long-term value and should be retained, shared, and/or preserved?**

All data sets that are licensed with an open license will be preserved in decentralised open data portals as mentioned above. As data will be published on different repositories with long-term support, the data will be available for use in the future. Sensitive data will be preserved in the internal project database, which has a long-term support of 5 years after the project ending.

- **What is the long-term preservation plan for the dataset?**

All data will be stored for at least 5 years in the internal project database. All open data will be made public on different platforms and thus will be preserved for as long as it is used.

## **F. Data Sharing**

- **How will you share the data?**

The generated data will be made available as early as possible during the project to increase use of data. This includes publishing data along with scientific publications. All publications must be available as open access articles and accompanying material should be published with suitable open licenses. At the latest, data will be made available at the end of the project. The data sharing between the project partners is done using the internal database. The main platform for data publication is the OpenEnergyPlatform (OEP). The developed web application software (REEEMBrowser) can be used to explore, visualise and locate the modelling data. While the OEP is the primary data hub, other platforms (e.g., Zenodo, energydata.info) can be used. In addition, the modelling data is available on each model website (e.g., OSeMBE data on osemosys.org).

- **Are any restrictions on data sharing required?**

The project aims at a maximal degree of data re-use for third parties using appropriate open data licenses. However, open licenses can only be granted if the copyright ownership is clarified without doubt. Possibly used original data must also be considered during this process. Since open data is not mandatory, all project partners are free to decide whether certain data sets are of particular commercial interest and cannot be published.

## **G. Responsibilities and Resources**

- **Who will be responsible for data management?**

Due to the different work packages in the project, the responsibility for the data management is split between the partner institutions. The work package leaders will be responsible for the data management in the work package that they oversee. The overall coordination of the data management will be undertaken by the project leader.

- **What resources will you require to deliver your plan?**

There will be no project costs associated with the process of publishing data as open data via the OpenEnergyPlatform or other platforms, except the researcher time spent for manual-, documentation- and metadata-creation. These resources are covered by the institutional budgets and research funding.

## H. Data Protection Impact Assessment

This section describes the collect and use of personal data and ensures compliance with the relevant data protection legislation (GDPR).

- What do you require this personal data for? What is the purpose of using the personal data?

Personal data (name, affiliation, gender, business phone, email address, and expertise) will be stored for the necessary stakeholder dialogue and workshop organisation. In order to effect outreach and dissemination of the research, the project partners need to communicate with people outside the project.

- How are you making people aware of how their personal data is being used? Do you need to update your privacy notice?

All people need to actively grant rights to use their personal data (Double -Opt-in). The written approval will be stored on the project database.

- Please select all that apply and provide any additional details.

### Section 1: Conditions for Personal Data

✓ The data subject has given consent to the processing (please provide the consent wording and where it is stored)

- Is all the personal data you are using necessary? Are you collecting enough to carry out the work, is there any you could do without to limit the risks to the individuals?

Only information that is used to fulfil the project goals will be stored and used.

- How are you ensuring that personal data obtained from individuals or other organisations is accurate? How will you keep it updated?

To ensure data accuracy, the project uses the so called Double -Opt-in. When personal data is stored, the people are given a copy to correct errors and mistakes. The data is updated continuously. Outdated records are deleted.

- How long will you keep the data and how will you dispose of it?

The personal data is stored only for the project duration. Before the projects ends, all data will be permanently deleted. Check will be undertaken to ensure that no backups contain the personal information. All persons will be informed that the data has been deleted and will not be used anymore.

- Where will the data be stored? If storage is in the cloud, where is the physical server?

The contact list is stored on a DTU-server located in Denmark. It is password protected and can only be accessed by the project partners.

- Will you be able to meet all the Data Subject Rights? Can you provide copies of data if requested? Are you able to fully delete the data (not just archive)?

As far as we know, we follow all data protection legislation that applies in the European Union. In the event of a violation of the rights of others, immediate action will be taken to correct these omissions.

- Please briefly document below any risks with the use of personal data and how you will control such risks. Include technical controls (IT security, encryption etc), physical controls



(location, locked room etc), personnel controls (training, access control etc), and procedural controls (contract, polices etc).

All project partners are informed that personal data is classified as sensitive data and needs special protection. The access to the server is password protected.



## 2. Definition and Matter

Name	Description
Existing data	A survey of existing data relevant to the project and a discussion of whether and how these data will be integrated.
Format	Formats in which the data will be generated, maintained, and made available, including a justification for the procedural and archival appropriateness of those formats.
Metadata	A description of the metadata to be provided along with the generated data, and a discussion of the metadata standards used.
Storage and backup	Storage methods and backup procedures for the data, including the physical and cyber resources and facilities that will be used for the effective preservation and storage of the research data.
Security	A description of technical and procedural protections for information, including confidential information, and how permissions, restrictions, and embargoes will be enforced.
Responsibility	Names of the individuals responsible for data management in the research project.
Intellectual property rights	Entities or persons who will hold the intellectual property rights to the data, and how IP will be protected if necessary. Any copyright constraints (e.g., copyrighted data collection instruments) should be noted.
Access and sharing	A description of how data will be shared, including access procedures, embargo periods, technical mechanisms for dissemination and whether access will be open or granted only to specific user groups. A timeframe for data sharing and publishing should also be provided.
Audience	The potential secondary users of the data.
Selection and retention periods	A description of how data will be selected for archiving, how long the data will be held, and plans for eventual transition or termination of the data collection in the future.
Archiving and preservation	The procedures in place or envisioned for long-term archiving and preservation of the data, including succession plans for the data should the expected archiving entity go out of existence.
Ethics and privacy	A discussion of how informed consent will be handled and how privacy will be protected, including any exceptional arrangements that might be needed to protect participant confidentiality, and other ethical issues that may arise.
Budget	The costs of preparing data and documentation for archiving and how these costs will be paid. Requests for funding may be included.
Data organization	How the data will be managed during the project, with information about version control, naming conventions, etc.
Quality Assurance	Procedures for ensuring data quality during the project.
Legal requirements	A listing of all relevant federal or funder requirements for data management and data sharing.



### 3. Links

<https://dmponline.dcc.ac.uk/>

Online tool to create a DMP

[http://www.dcc.ac.uk/sites/default/files/documents/data-forum/documents/docs/DCC\\_Checklist\\_DMP\\_v3.pdf](http://www.dcc.ac.uk/sites/default/files/documents/data-forum/documents/docs/DCC_Checklist_DMP_v3.pdf)

Checklist for DMP

<http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP-checklist-post-consultation-v2.pdf>

Content Checklist for DMP

<https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/resources.html>

DMP resources and examples

<https://libraries.mit.edu/data-management/plan/guides/>

Guides to DMP