



Data Management Plan Version 1.0, Oct 2019

DELIVERABLE NO. 4.1



This report is licensed under a Creative Commons Licence



Attribution 4.0 International License



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 835896



Deliverable No.	D4.1	Work Package No. 4	WP4
Work Package Titel	Open modelling platform		
Status	Submitted		
Dissemination level	PU		
Due date deliverable	2019.10.31	Submission date	2019.10.30
Deliverable version	Final for submission as deliverable D4.1 (M6) To be updated throughout the project as required and resubmitted in M48		

Deliverable Contributors:	Name	Organisation	Date
Work Package Leader	Daniel Huppmann	IIASA	
Reviewer	Sandrine Charouset	EdF	
Final review and approval	Ingeborg Graabak	SINTEF Energy Research	2019.10.30

History of Change

Release	Date	Reason for Change	Status
Version 1.0	2019.10.30	Version submitted D4.1 (M6)	

DISCLAIMER / ACKNOWLEDGMENT

The content of this deliverable only reflects the author's views. The European Commission / Innovation and Networks Executive Agency is not responsible for any use that may be made of the information it contains.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 835896.

This report is licensed under a Creative Commons Licence

Attribution 4.0 International License



For more information about the Creative Commons Licence, including the full legal text, please visit:
<https://creativecommons.org/use-remix/cc-licenses/>



Table of Contents

History of Change	2
Table of Contents.....	4
Preamble.....	5
1. Data Summary	5
1. 1. Scope of data	5
1. 2. Size of data “blocks”	6
1. 3. Data template and file format.....	6
1. 4. Overview of scenarios generated from case studies.....	7
2. FAIR data.....	8
2. 1. Making data findable, including provisions for metadata.....	8
2.2. Making data openly accessible	8
2.3. Making data interoperable.....	8
2.4. Increase data re-use (through clarifying licences)	9
3. Allocation of resources.....	9
4. Data security.....	9
5. Ethical aspects	9

Preamble

The project ‘Open ENergy TRansition ANalyses for a low-carbon Economy’ – openENTRANCE develops an open, transparent and integrated modelling platform for assessing low-carbon transition pathways. The platform will gather a suite of state-of-the-art modelling tools and data for covering the multiple dimensions of a green and clean energy transition.

openENTRANCE is a project funded by the European Commission under the Horizon 2020 framework and implemented by a consortium of 14 leading European research institutions, Universities and civil society organisations.

1. Data Summary

The openENTRANCE project will produce a number of datasets concerning historic development and scenarios of future energy transitions in Europe. We are dedicated to the goal of the European Commission and the Horizon 2020 programme to make all academic work FAIR. All historical and scenario data sets will be made available via the “open platform” (developed in WP4)¹, with references to supporting literature and hyperlinks (where possible) to the open-source modelling frameworks used to generate the scenarios. All academic publications and policy briefs resulting from this project will be made available either following the green or gold open access standard and will be published on a dedicated project website.

The data will be useful for policymakers and researchers to assess quantified scenarios of the energy transition and related policy measures. The open platform will facilitate the re-use of results from openENTRANCE as reference in scientific work by other modelling teams and for science communication activities by NGOs and other organizations working on related topics.

1.1. Scope of data

The data created, compiled and curated during the project can be categorized into four groups:

- 1) Indicators of the historical development of the energy system (and related sectors) in Europe and at the global level provided by national/European statistics offices and international organizations such as the International Energy Agency (IEA). This data will be used for reference, calibration and validation of the quantitative scenarios developed across the work packages.

¹ Throughout this document, the term “open platform” refers to the interactive website to access and view scenario timeseries data. The part of the open platform in this project that refers to sharing/using open-source software and related tools is not discussed in this Data Management Plan.

- 2) Timeseries data and indicators of quantitative scenarios of the energy transition to a low- or no-carbon economy in Europe. This data will be developed with the suite of modelling frameworks (partly open-source, partly proprietary) in the consortium. The scenarios will be used to provide policy-relevant advice on pathways to achieve the strategic aims of energy sector decarbonisation in the European Union.
- 3) Timeseries data and indicators similar to (2) but developed and submitted by modelling teams that are not part of the consortium. The open platform will be made accessible for both passive (retrieving openENTRANCE data from the platform for reuse) and active use (using the platform to publish and disseminate own scenario data related to the scope of openENTRANCE).
- 4) Indicators based on aggregations of survey data at the household or consumer level gathered from previous work provided by Energie Institute (Austria), in particular from the Horizon 2020 projects ECHOES and PEAKapp. The data will be anonymized before they are made available to the openENTRANCE consortium to side-step any ethical issues (see section 5).

1. 2. Size of data “blocks”

A “block” of data can be (1) all reference/historical indicators and timeseries data compiled from one source or (2/3) all indicators and timeseries data pertaining to one scenario (a quantitative implementation of a qualitative narrative) developed with a specific version of a modelling framework (e.g., using the same narrative with different versions of the same modelling framework software should be seen as multiple blocks; the same logic applies to using the same narrative and the identical software but different quantifications of relevant input parameters). Alternatively, one block can be all indicators derived from a particular survey (4).

Each block usually contains hundreds to thousands of indicators and timeseries variables; dozens to hundreds of countries and subregions; and decadal to yearly to sub-hourly temporal resolution. Indicators may also be reported at global or continental scale where relevant. A block can therefore comprise millions of data points.

1. 3. Data template and file format

All data created and published/distributed via the open platform will adhere to the data template developed by the Integrated Assessment Modeling Consortium (IAMC)². Where required, the template will be extended to facilitate specific use cases within openENTRANCE. This is discussed in more detail in Deliverable 4.2.

² For more information, refer to <https://data.ene.iiasa.ac.at/database/>

In previous model comparison projects, *xlsx* and *csv* file formats were used to exchange data in the IAMC format. Given the expected size of data blocks, database formats such as *hdf5* or *netcdf* will be considered for data upload to and download from the platform.

The following limitations of the existing data template were identified.

- The “region” column cannot be used directly to report trade data between regions.
Solution: If the timeseries in question is related to trade data from “region A” to “region B”, the entry in the region column should be written as “region A>region B” to indicate a directional flow. For bi-directional flows, the data should be reported as two directional timeseries.
- Subannual time slides cannot be reported in the existing data template.
Solution: A new column “subannual” will be added where necessary.

1. 4. Overview of scenarios generated from case studies

The list below provides an overview of the quantitative scenario data (2) to be generated within openENTRANCE, along with the work packages and tasks that are responsible for these data sets (in parenthesis). Reference data (1) will not be listed explicitly but will be compiled and curated as necessary throughout the project.

Scenario data and case studies created during the project (as expected at the current time):

- Quantification of scenarios at pan-European level based on pathway storylines developed in WP7 (Task 3.1)
- Validation and verification of the open source GENeSYS-MOD model (Task 3.3)
- Development of consistent scenarios for low carbon futures of the electricity, heating/cooling and transport sector (Task 3.4)
- Reconciling first-best model worlds and the real world in increasingly decentral energy systems (case study in WP5)
- Demand-Response from household consumers (Task 6.1, case study in WP6)
- Behaviour of communities of actors (Task 6.2, case study in WP6)
- Flexibilities and storage (Task 6.3, case study in WP6)
- Cross-sector integration, with a specific focus on the flexibilities provided by electric vehicle owners to the electricity system (Task 6.4, case study in WP6)
- Geographic coordination for investment decisions (Task 6.5, case study in WP6)
- Innovative technology in terms of underground rocks for seasonal storage of heat (Task 6.6, case study in WP6)
- Flexibilities from the heating sector at different time scales (Task 6.7, case study in WP6)
- Role of natural gas storage (Task 6.8, case study in WP6)
- Comparative Macro-economic analyses of the energy transition (Task 7.3)

2. FAIR data

2.1. Making data findable, including provisions for metadata

Each block of data (see above) used in scientific publications or reports will be assigned its own Digital Object Identifier (DOI) via a service such as zenodo. This data will also be accessible via the open platform (developed in WP4), with cross-references to ensure that the work is findable. Both zenodo (or similar services) and the platform provide metadata tagging features. Metadata entries for each block will include at least the underlying modelling framework (including version number), work package, case study, and storyline/narrative identifier (where applicable).

A glossary of common terminology with regard to modelling approach and metadata nomenclature will be compiled over the course of the project. For example, a list of region and sub-region names will be compiled and will then be mandatory for any data reporting.

2.2. Making data openly accessible

All scenario data created in the course of the project (see item 2 in Section 1.1) will be published under the Creative Commons CC-BY 4.0 International License³. The data will be accessible via the open platform and via the DOI-generating service (if applicable, see Section 2.1).

Reference data (see item 1 in Section 1.1) will be made available under an appropriate open-source license if this is legally possible, depending on the source of the underlying data.

2.3. Making data interoperable

By following the IAMC data template (see Section 1.3) and extending it where necessary, openENTRANCE ensures that the data is interoperable with standard software tools such as Microsoft Excel. Dedicated open-source packages for scientific programming languages exist or under development for working with this data template, e.g., the Python package pyam.⁴

The glossary and nomenclature (see Section 2.1) developed and used in openENTRANCE will be consistent with the conventions in other model comparison projects and the work of the IPCC. The glossary and nomenclature will also be available via the open platform. Related tools to facilitate the creation of mappings to naming conventions used in other projects will also be publicly released, should we encounter instances where a strict harmonization or consistency is not possible.

³ See <https://creativecommons.org/licenses/by/4.0/> for the details on this license.

⁴ See <https://pyam-iamc.readthedocs.io> for the documentation of this package.



2.4. Increase data re-use (through clarifying licences)

As stated in Section 2.2, all data pertaining to quantitative scenarios created in openENTRANCE will be distributed under the CC-BY 4.0 license. The data will be made available for re-use at the latest upon publication of any scientific work using the data.

While the CC-BY 4.0 license implies that the data can be re-used indefinitely, it is worth pointing out that quantifications of *future* energy transition scenarios have a limited relevance and usability after a few years as real-world developments make projections and underlying assumptions during the project outdated. In any case, IIASA as the host of the open platform is committed to hosting the openENTRANCE data platform for at least 10 years following the end of the project. This ensures that there is an easy entry point for policy-makers and researchers to the quantitative results of the project for an extended period, facilitating re-use.

3. Allocation of resources

This Data Management Plan will be reviewed and updated regularly throughout the project. IIASA as the host of the open platform will ensure that all modelling teams obtain DOIs and add the relevant metadata to their data prior to the data being accessible via the platform.

4. Data security

IIASA as the host of the open platform will ensure that the underlying database and web servers are managed according to state-of-the-art data security standards.

5. Ethical aspects

The execution of Task 6.1 will exploit human data collected in previous research projects, in particular data collected in the Horizon 2020 projects PEAKapp (grant no 695945) and ECHOES (grant no 727470). Both datasets are made anonymous before use in openENTRANCE. i.e., these datasets do not contain information making it possible to identify the individual source of the data.

In Task 7.4. professionals will be interviewed to identify driver and barriers for investment in technology for transforming the energy sector. The professionals will be selected among the many stakeholders in the project.



The ethical aspects related to these data are described in Deliverable 9.1 and 9.2 of the Grant Agreement, copied here for reference:

D9.1

The informed consent procedures that will be implemented for the participation of humans must be submitted as a deliverable. Templates of the informed consent/assent forms and information sheets (in language and terms intelligible to the participants) must be kept on file. The procedures and criteria that will be used to identify/recruit research participants must be submitted as a deliverable. Copies of opinions/approvals by ethics committees and/or competent authorities for the research with humans must be submitted as a deliverable.

D9.2

The host institution must confirm that it has appointed a Data Protection Officer (DPO) and the contact details of the DPO are made available to all data subjects involved in the research. For host institutions not required to appoint a DPO under the GDPR a detailed data protection policy for the project must be submitted as a deliverable.

A description of the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants must be submitted as a deliverable. In case personal data are transferred from a non-EU country to the EU (or another third state), confirmation that such transfers comply with the laws of the country in which the data was collected must be submitted as a deliverable. In case of further processing of previously collected personal data, an explicit confirmation that the beneficiary has lawful basis for the data processing and that the appropriate technical and organisational measures are in place to safeguard the rights of the data subjects must be submitted as a deliverable.