Deliverable 5.2 Data Management Plan

Revision.................................. 2
Preparation date ...................... 2019-05-17 (M20)
Due date .............................. 2019-03-31 (M18)
Lead contractor......................... VTT

Author:
Erkka Rinne ....................... VTT
Yrsa Cronhjort........................ VTT

This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by-sa/4.0/.

<table>
<thead>
<tr>
<th>Dissemination level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PU Public</td>
<td>X</td>
</tr>
<tr>
<td>CO Confidential, only for members of the consortium (including the Commission Services)</td>
<td></td>
</tr>
</tbody>
</table>
## Deliverable administration

<table>
<thead>
<tr>
<th>No &amp; name</th>
<th>D5.2 Data Management Plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status</td>
<td>First update</td>
</tr>
<tr>
<td>Due</td>
<td>M18</td>
</tr>
<tr>
<td>Date</td>
<td>2019-05-28</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Erkka Rinne VTT</td>
</tr>
<tr>
<td></td>
<td>Yrsa Cronhjort VTT</td>
</tr>
<tr>
<td>Reviewer(s)</td>
<td>Terence O’Donnell UCD</td>
</tr>
<tr>
<td></td>
<td>Juha Kiviluoma VTT</td>
</tr>
<tr>
<td></td>
<td>Manuel Marin KTH</td>
</tr>
</tbody>
</table>

### Description of the related task and the deliverable. Extract from DoW

**T5.2 Data management plan for the Open Research Data Pilot (m1-m48)**

Participants: VTT (leader), UCD

Estimated effort: 2.0 person months.

This task is for the creation and maintenance of the project Data Management Plan (DMP) as required by the Pilot on Open Research Data. The DMP will describe which data collected during the project will be kept for a longer term, how the data is stored and documented. The first version of the DMP will be delivered within the first six months of the project (D5.2) and it will be updated at least annually. VTT will be responsible for the plan with support from WP leader UCD.

**D5.2 Data Management Plan**: (T5.2) Project data management plan for Open Research Data Pilot, due in M06 (first version), M18 (updated version at the end of the financial period), M30 (updated version at the end of the financial period), M48 (final version). Success criteria: acceptable data management plan for the open research data pilot.

<table>
<thead>
<tr>
<th>Planned resources PM of T5.2</th>
<th>VTT</th>
<th>UCD</th>
<th>KUL</th>
<th>KTH</th>
<th>ER</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.5</td>
<td>0.5</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>2.0</td>
</tr>
</tbody>
</table>

### Comments

<table>
<thead>
<tr>
<th>V</th>
<th>Date</th>
<th>Authors</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>2018-03-09</td>
<td>VTT</td>
<td>First draft</td>
</tr>
<tr>
<td>1.1</td>
<td>2018-03-22</td>
<td>VTT</td>
<td>Improved version</td>
</tr>
<tr>
<td>1.2</td>
<td>2018-04-03</td>
<td>VTT</td>
<td>Submission</td>
</tr>
<tr>
<td>1.3</td>
<td>2018-05-11</td>
<td>VTT</td>
<td>Additional comments for next version</td>
</tr>
<tr>
<td>1.4</td>
<td>2019-05-17</td>
<td>VTT</td>
<td>Updated version for internal review</td>
</tr>
<tr>
<td>1.5</td>
<td>2019-05-27</td>
<td>VTT</td>
<td>Updated version for submission</td>
</tr>
</tbody>
</table>

### Disclaimer

The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information as its sole risk and liability.

The document reflects only the author’s views and the Community is not liable for any use that may be made of the information contained therein.

---

# Table of contents

1. Introduction .......................................................................................................................... 4

2. Data Summary .......................................................................................................................... 4
   2.1 Data generated in the Spine project ................................................................................. 4

3. Fair Data .................................................................................................................................. 8
   3.1 Processing, standards and metadata ................................................................................ 8
   3.2 Data sharing and ownership, IPR management ............................................................... 9
   3.3 Interoperability of data .................................................................................................... 10
   3.4 Reuse of data ................................................................................................................. 10

4. Allocation of Resources ......................................................................................................... 11

5. Data Security .......................................................................................................................... 11

6. Ethics and Privacy .................................................................................................................. 11

7. Other Issues ............................................................................................................................ 11
1. **INTRODUCTION**

The Data Management Plan (DMP) describes the data management methods, and especially how the data will be generated, processed, collected, stored and documented. The DMP includes also a description of the methodology and standards to be followed and what data sets are exploitable or made accessible for verification and re-use. The DMP will be a dynamic tool that is updated regularly throughout the project duration. By these actions, the Spine project supports the European Commission’s goal of shared data and open science that serves innovation and growth.

2. **DATA SUMMARY**

Research data will be generated and collected manually and automatically by using programming scripts from a variety of different sources. The data sources that are considered to be used are documented in the Deliverable 5.1 [D5.1]. The generated data sets will cover the 13 case studies of the project, and thus contribute to the project objectives O6: *Improve, validate and deploy the Spine Toolbox together with the industry in 13 case studies* as well as O7: *Initiate, grow and support a user community for the deployment and future development*. Formats used for the datasets will be common formats used for data sharing. The size of the data packages will vary between studies and their sizes will be estimated when the case studies are under way. The data will be useful in future studies as well as to help new users to learn the Spine Toolbox and in training and teaching related to energy systems.

2.1 **Data generated in the Spine project**

Table 1 lists data generated or collected in the Spine project per Work Package including data set identifier (ID), description (origin and nature), potential use/users, standards and metadata, data sharing including how the data is shared, access, embargo periods, dissemination means, required software/tools for using the data, restrictions including the motivation and repository to be used.
Table 1 Datasets created in the Spine project

<table>
<thead>
<tr>
<th>ID</th>
<th>WP</th>
<th>Dataset description</th>
<th>Origin</th>
<th>Nature &amp; size</th>
<th>Use/ users</th>
<th>Metadata</th>
<th>Sharing</th>
<th>Dissemination</th>
<th>Software / Tools</th>
<th>Access</th>
<th>Restrictions</th>
<th>Repository</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>Dataset for case study A1 <em>Replicate an Irish dispatch study with power flows</em>. Input data for the case study simulation model, relates to objectives O6 and O7.</td>
<td>Collected during the project</td>
<td>Data package format, size TBA</td>
<td>Research, training</td>
<td>Following DCMES(^5)</td>
<td>Dataset: ODbL(^5), metadata: CC0(^6)</td>
<td>[TBD]</td>
<td>Spine Model</td>
<td>Online</td>
<td>(some data might be proprietary)</td>
<td>Zenodo</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>Dataset for case study A2 <em>Replicate a Belgian gas grid study</em>. Input data for the case study simulation model, relates to objectives O6 and O7.</td>
<td>Collected during the project</td>
<td>Data package format</td>
<td>Research, training</td>
<td>Following DCMES</td>
<td>Dataset: ODbL, metadata: CC0</td>
<td>[TBD]</td>
<td>Spine Model</td>
<td>Online</td>
<td>(some data might be proprietary)</td>
<td>Zenodo</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>Dataset for case study A3 <em>Replicate a district heating study of Stockholm</em>. Input data for the case study simulation model, relates to objectives O6 and O7.</td>
<td>Collected during the project</td>
<td>Data package format</td>
<td>Research, training</td>
<td>Following DCMES</td>
<td>Dataset: ODbL, metadata: CC0</td>
<td>[TBD]</td>
<td>Spine Model</td>
<td>Online</td>
<td>None</td>
<td>Zenodo</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>Dataset for case study A4 <em>Cost optimization study with building heat physics</em>. Input data for the case study simulation model, relates to objectives O6 and O7.</td>
<td>Collected during the project</td>
<td>Data package format</td>
<td>Research, training</td>
<td>Following DCMES</td>
<td>Dataset: ODbL, metadata: CC0</td>
<td>[TBD]</td>
<td>Spine Model</td>
<td>Online</td>
<td>(some data might be proprietary)</td>
<td>Zenodo</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>Dataset for case study A5 <em>Hydro power study with river systems</em>. Input data for the case study simulation model, relates to objectives O6 and O7.</td>
<td>Collected during the project</td>
<td>Data package format</td>
<td>Research, training</td>
<td>Following DCMES</td>
<td>Dataset: ODbL, metadata: CC0</td>
<td>[TBD]</td>
<td>Spine Model</td>
<td>Online</td>
<td>None</td>
<td>Zenodo</td>
</tr>
</tbody>
</table>
## Dataset description

<table>
<thead>
<tr>
<th>ID</th>
<th>WP</th>
<th>Description and relation to the project objectives</th>
<th>Origin</th>
<th>Nature &amp; size</th>
<th>Use/ users</th>
<th>Metadata</th>
<th>Sharing</th>
<th>Dissemination</th>
<th>Software / Tools</th>
<th>Access</th>
<th>Restrictions</th>
<th>Repository</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>6</td>
<td>Dataset for case study B1 Spatial aggregation of nodal systems. Input data for the case study simulation model, relates to objectives O6 and O7.</td>
<td>Collected during the project</td>
<td>Data package format</td>
<td>Research, training</td>
<td>Following DCMES</td>
<td>Dataset: ODbL, metadata: CC0</td>
<td>[TBD]</td>
<td>Python, Julia, Spine Model</td>
<td>Online</td>
<td>(some data might be proprietary)</td>
<td>Zenodo</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>Dataset for case study B2 Temporal aggregation of time periods. Input data for the case study simulation model, relates to objectives O6 and O7.</td>
<td>Collected during the project</td>
<td>Data package format</td>
<td>Research, training</td>
<td>Following DCMES</td>
<td>Dataset: ODbL, metadata: CC0</td>
<td>[TBD]</td>
<td>Python, Julia, Spine Model</td>
<td>Online</td>
<td>(some data might be proprietary)</td>
<td>Zenodo</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>Dataset for case study B3 Planning a low emission transport sector. Input data for the case study simulation model, relates to objectives O6 and O7.</td>
<td>Collected during the project</td>
<td>Data package format</td>
<td>Research, training</td>
<td>Following DCMES</td>
<td>Dataset: ODbL, metadata: CC0</td>
<td>[TBD]</td>
<td>Python, Julia, Spine Model</td>
<td>Online</td>
<td>(some data might be proprietary)</td>
<td>Zenodo</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>Dataset for case study B4 Biomass resource constraints. Input data for the case study simulation model, relates to objectives O6 and O7.</td>
<td>Collected during the project</td>
<td>Data package format</td>
<td>Research, training</td>
<td>Following DCMES</td>
<td>Dataset: ODbL, metadata: CC0</td>
<td>[TBD]</td>
<td>Python, Julia, Spine Model</td>
<td>Online</td>
<td>(some data might be proprietary)</td>
<td>Zenodo</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>Dataset for case study B5 Industrial energy use. Input data for the case study simulation model, relates to objectives O6 and O7.</td>
<td>Collected during the project</td>
<td>Data package format</td>
<td>Research, training</td>
<td>Following DCMES</td>
<td>Dataset: ODbL, metadata: CC0</td>
<td>[TBD]</td>
<td>Python, Julia, Spine Model</td>
<td>Online</td>
<td>(some data might be proprietary)</td>
<td>Zenodo</td>
</tr>
<tr>
<td>11</td>
<td>6</td>
<td>Dataset for case study C1 Market design for integrated energy systems. Input data for the case study simulation model, relates to objectives O6 and O7.</td>
<td>Collected during the project</td>
<td>Data package format</td>
<td>Research, training</td>
<td>Following DCMES</td>
<td>Dataset: ODbL, metadata: CC0</td>
<td>[TBD]</td>
<td>Python, Julia, Spine Model</td>
<td>Online</td>
<td>(some data might be proprietary)</td>
<td>Zenodo</td>
</tr>
<tr>
<td>ID</td>
<td>WP</td>
<td>Description and relation to the project objectives</td>
<td>Origin</td>
<td>Nature &amp; size</td>
<td>Use/ users</td>
<td>Metadata</td>
<td>Sharing</td>
<td>Dissemination</td>
<td>Software / Tools</td>
<td>Access</td>
<td>Restrictions</td>
<td>Repository</td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>---------------------------------------------------</td>
<td>--------</td>
<td>---------------</td>
<td>------------</td>
<td>----------</td>
<td>---------</td>
<td>---------------</td>
<td>----------------</td>
<td>--------</td>
<td>-------------</td>
<td>-----------</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>Dataset for case study C2 Power grid investments under uncertainty. Input data for the case study simulation model, relates to objectives O6 and O7.</td>
<td>Collected during the project</td>
<td>Data package format</td>
<td>Research, training</td>
<td>Following DCMES</td>
<td>Dataset: ODbL, metadata: CC0</td>
<td>[TBD]</td>
<td>Python, Julia, Spine Model</td>
<td>Online</td>
<td>(some data might be proprietary)</td>
<td>Zenodo</td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>Dataset for case study C3 Integrated energy system planning with high operational detail. Input data for the case study simulation model, relates to objectives O6 and O7.</td>
<td>Collected during the project</td>
<td>Data package format</td>
<td>Research, training</td>
<td>Following DCMES</td>
<td>Dataset: ODbL, metadata: CC0</td>
<td>[TBD]</td>
<td>Python, Julia, Spine Model</td>
<td>Online</td>
<td>(some data might be proprietary)</td>
<td>Zenodo</td>
</tr>
</tbody>
</table>

a Data package format [WP]
b Dublin Core Metadata Element Set [DCMI]
c Open Database License [ODC]
d Creative Commons CC0 [CC0]
3. Fair Data

3.1 Processing, standards and metadata

Datasets will be processed and analysed mainly using scripting in Python and Julia programming languages. Variables and value names will be constructed to be clear, compact and descriptive. A common metadata schema (see below) following the Dublin Core Metadata Element Set [DCMI] will be used in describing the data to aid data discovery and re-use.

The datasets will be versioned so that it will be possible to distinguish which version was used with which particular Spine Toolbox and Spine Model version. Where possible, the datasets will be associated with digital object identifiers (DOIs).

Possibility for follow-up utilization of the data will be ensured by documenting data collection methods as well as the contents of the datasets. Data collection and processing scripts will published in an open repository.

3.1.1 Spine metadata description

This is the metadata description for Spine, modified from Frictionless data Data Package format [DP]. The current version of this description is available at https://gitlab.vtt.fi/spine/toolbox/wikis/Metadata-description.

3.1.1.1 Required properties

title
One sentence description for the data

sources
The raw sources of the data. Each source must have a title and optionally a path property. Example

"sources": [{
  "title": "World Bank and OECD",
  "path": "http://data.worldbank.org/indicator/NY.GDP.MKTP.CD"
}]

contributors
The people or organisations who contributed to the data. Must include name and may include path, email, role and organization. Example

"contributors": [{
  "title": "Joe Bloggs",
  "email": "joe@bloggs.com",
  "path": "http://www.bloggs.com",
  "role": "author"
}]

Role is one of author, publisher, maintainer, wrangler, or contributor.

created
The date this data was created or put together, in ISO8601 format (YYYY-MM-DDTHH:MM)

3.1.1.2 Optional properties

keywords
An array of keywords

homepage
A URL for the home on the web that is related to this data package.

name

Name of the data package, url-usable, all-lowercase string.

id

Globally unique id, such as UUID or DOI

licenses

Licences that apply to the data. Each item must have a name property (Open Definition license ID) or a path property and may contain title. Example

"licenses": [{
  "name": "ODC-PDDL-1.0",
  "path": "http://opendatacommons.org/licenses/pddl/",
  "title": "Open Data Commons Public Domain Dedication and License v1.0"
}]

temporal

Temporal properties of the data (if applicable). Example using DCMI Period Encoding Scheme

"temporal": {
  "start": "2000-01-01",
  "end": "2000-12-31",
  "name": "The first year of the 21st century"
}

spatial

Spatial properties of the data (if applicable). Example using DCMI Point Encoding Scheme

"spatial": {
  "east": 23.766667,
  "north": 61.5,
  "projection": "geographic coordinates (WGS 84)",
  "name": "Tampere, Finland"
}

### 3.2 Data sharing and ownership, IPR management

The project coordinator, in collaboration with the project Steering Committee (SC) will take all the appropriate measures to make the data openly available and usable for third parties for educational and research purposes. The primary focus in data sharing will be on data underlying prospective scientific publications ensuring the validation of results presented in publications. Data sharing will also help new users of the Spine Model to learn through examples and the case studies can also serve as training sets in energy systems education. During the project, the data will be used by the research consortium members.

A common repository service (Zenodo [ZENO]) will be used to enhance long-term data accessibility. The data can be directly used with Spine Toolbox and Spine Model but future usage is not limited to these. Spine Toolbox and Spine Model will be available as open source software and deposited also in an open repository.

---

2 [http://licenses.opendefinition.org/](http://licenses.opendefinition.org/)


3.2.1 Publishing Policy

Deliverables defined as publications will be published in green or gold open access, peer-reviewed scientific journals if possible. Journal submission will be reviewed by the SC before the submission according to the procedures defined in the Consortium agreement as follows:

During the Project and for a period of 1 year after the end of the Project, the dissemination of own Results by one or several Parties including but not restricted to publications and presentations, shall be governed by the procedure of Article 29.1 of the Grant Agreement subject to the following provisions.

Prior notice of any planned publication shall be given to the other Parties at least 20 calendar days before the publication. Any objection to the planned publication shall be made in accordance with the Grant Agreement in writing to the Coordinator and to the Party or Parties proposing the dissemination within 10 calendar days after the date of the abovementioned prior notice. Notwithstanding the foregoing, a notice regarding a presentation to be held in a conference or a seminar shall be given to the other Parties at least 7 days prior to the presentation. Any objection regarding the planned presentation shall be given within 3 days prior to the presentation. If no objection is made within the time limits stated above, the publication is permitted.

3.3 Interoperability of data

The datasets will be stored in the Data Package format [WP]. It is a packaging specification to include both metadata and description of data resources and is used for example in the Open Power System Data project [OPSD]. There is support for working with Data Packages in most major programming languages. The datasets will be readily usable in a Spine Toolbox project. The example projects files will also be published along with the case study datasets.

Rich metadata will be stored along with the data. Standard file formats such as comma separated values (csv), Microsoft Excel workbooks and common binary data formats, e.g. Network Common Data Form (netCDF) [UNID] or Hierarchical Data Format (HDF) [HDF] will be used. Metadata and descriptions will be published with the data.

If a suitable metadata vocabulary is found, it will be used for the description of the data.

3.4 Reuse of data

The datasets created in the project will be licensed using Open Data Commons Open Database License (ODbL) [ODC]. This license is well suited for databases and allows for reuse with an attribution requirement to keep any derivative works open under the same license. All metadata will be licensed with a public domain license, CC0 [CC0]. The datasets will be published as soon as possible, and no embargo periods are expected.

Dataset contents will be licensed depending on the source requirements. For newly created data, Database Contents License [ODC] waiving all rights to the contents themselves will be used. In case of proprietary data for the case studies, this data will not be shared or it will be anonymised. In some cases it may not be possible to re-publish the data due to licensing issues – in these cases access to the original data sources will be documented. If the data is crucial to running the case study, a dummy data set may be established in order to enable running the case study.

All material other than datasets and their contents will be licensed using Creative Commons Attribution-ShareAlike 4.0 license [CC4].

The quality of the data will be checked within the case studies in order to produce good quality results.

The datasets will remain accessible as long as possible depending on the availability of chosen repositories.
4. **Allocation of Resources**

At the beginning of the research project, the research consortium will decide and agree on the tasks, roles, responsibilities, and rights relating to data collection, dataset management, and data use. No extra costs in addition to the project budget will be generated in making the data FAIR. Long-term preservation of the data in free repositories incurs no extra cost, but has potentially a large value to the energy modelling community.

5. **Data Security**

Generated datasets will be archived at an established open data repository Zenodo. By decision of the project steering group, some data can also be deposited elsewhere depending on data security reasons, for example if the data is proprietary.

All project partners are responsible for curating, preserving, disseminating, and deleting the datasets in its possession. Retention time for curated datasets is the same as other project materials at VTT, by default twenty years.

6. **Ethics and Privacy**

The Spine project has not identified any ethical issues.

Privacy of the project participants will be secured. No person or organisation involved (e.g. in surveys) will be unintentionally identifiable directly or indirectly in the datasets. Besides storing separately from the data all direct identifiers of any respondents or subjects (e.g. names and contact information of persons and organisations) – also indirect references to e.g. lines of businesses, branches or industries - will be removed and destroyed after the anonymised dataset has been checked and validated.

7. **Other Issues**

No additional issues in this version of the deliverable.

8. **References**


[CC0] Creative Commons. ‘CC0 – “No Rights Reserved”’. [https://creativecommons.org/share-your-work/public-domain/cc0/](https://creativecommons.org/share-your-work/public-domain/cc0/) [Accessed 2019-05-17].

[CC4] Creative Commons. ‘Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)’ [https://creativecommons.org/licenses/by-sa/4.0/](https://creativecommons.org/licenses/by-sa/4.0/) [Accessed 2019-05-17].


