

Venue: Brussels

Date: 18.05.2017

Focus Group: Publishing in open databases and increasing transparency and reproducibility

Introduction:

Scientific approaches to investigate technical, economical and social interactions of the energy system are always based on data. One of the major challenges of projects is the exchange, versioning and publishing of relevant data. With an increasing number of models and participating scientists, the need arises for functioning data management systems and defined exchange formats. After looking at typical and well known barriers we want to present and discuss developed tools and methods to increase transparency, to decrease double work and to avoid data confusion.

Objective of the focus group:

- To exchange about problems and solutions about publishing and opening up
- Overview on methods and standards for transparent energy system analyses (database, data platform, data warehouse,...)

Key questions for discussion:

- What are barriers to publish open data?
- What methods are commonly used to publish data and code?
- How to build an active community that knows and follows standardized methods to increase transparency and reproducibility

Schedule:

Input speeches

- Publishing code and data - crucial points and barriers (Paul Dean (UCC))
- Input along the chart "reproducibility of modelling processes",
Role and Importance of a commonly used database
- Development and features of the oedb and OEP

Structured discussion about key questions

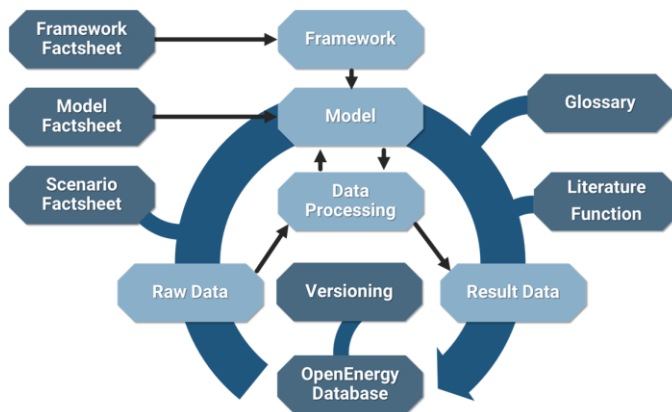
Contact:

Berit Müller: Berit.Mueller@rl-institut.de

Concept:

Definition of relevant subgroups and goals for each group:

- Different ways of publishing (platform, database, file storage)
- Use of suitable and machine readable metadata
- Utilisation of suitable open licenses



Goals:

1. To know barriers to publish open data and possible solutions

Barrier	Possible Solutions
understand database concepts and possibilities	
understanding the importance of licenses	
[Data] No facilities to upload data when submitting a paper	Create a Dropbox Link
No time to care about opening up data	Make it part of the project

2. Best practice collection how to publish data

Criteria	Example
Publish data online	Accessible without restriction Long term (10 years) management
Choose a suitable license / Utilisation of open source licenses	Closed by default!! Compliance with copyright legislations With or without copyleft
Follow the Linked Open Data (LOD) criterias	Five stars of open data <ol style="list-style-type: none"> 1. (OL) Available on the web (whatever format) but with an open licence, to be Open Data 2. (RE) Available as machine-readable structured data (e.g. excel instead of image scan of a table) 3. (OF) Non-proprietary format (e.g. CSV instead of excel) 4. (URI) Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff 5. (LD) Link your data to other people's data to provide context

3. Link all the open data of energy systems

=> How can we work together to find a common standard

Workshop Minutes:

Participants:

Berit Müller
Ludwig Hülk
Paul Dean
Kenneth Karlsson
Wided Medjroubi

..

..

- introductory round
- introduction of a model (electricity and gas) that was made available and a discussion round about this experience.
- slides by Ludwig Hülk about metadata and standards of metadata
- difference between openness and transparency and how to properly document data especially the aspect of the source of the data, i.e. does the data comes from the model or from another source?
- how do we proceed further? what about the IDEES database? how to interact with the project especially with the metadta issue?
- options to exchange data: eudat: <https://www.eudat.eu> and PRACE: <http://www.prace-ri.eu>
- data licenses are very important and they need to be considered