Wind-farm co-design in the North-Sea basin given climate and market uncertainty

Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Dataset name / ID	Description	New or reuse	Digital or Physical data	Data Type	File format	Data volume	Physical volume
		Indicate: ***N****(ew data) or* ***E****(xisting data)*	Indicate: **D**(igital) or **P**(hysical)	Indicate: **A**udiovisual **I**mages **S**ound **N**umerical **T**extual **M**odel **SO**ftware Other (specify)		Indicate: <1GB <100GB <1TB <5TB >5TB NA	
	Simulation results from OpenFast	N	D	N	OpenFast	<1TB	
	ML model for turbine lifetime consumption	N	D	SO	Python	<1GB	
	Wind-farm codesign software	N	D	SO	Python	<1GB	
	wind-resource–economic-market scenarios	N	D	N		>5TB	
	COSMO-CLM North Sea Model integrations	E	D	NetCDF		>5TB	

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Model integrations available from the group of NvL performed with the COSMO-CLM model

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.

- No

Will you process personal data? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).

- No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)? If so, please comment per dataset or data type where appropriate.

- Yes

It is our intention to share data sets that we generate open source to the community. Similarly, also software developments are published open source. Therefore, possible commercialization/tech transfer options are not related to licensing of access to data/software, but rather to consulting or further development towards commercial applications. In that sense, data and software provide a credible portfolio on which further business development strategies can be built.

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material or Data transfer agreements, Research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- No

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g. in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, codebook.tsv etc. where this information is recorded).

We require that each data file is accompanied by instructions on how to store, open and read it:

- Data and scripts should be stored together in the same folder.
- All data fields should have meaningful names.
- All data files should be accompanied by a README file that describes the goal of the experiment, the data format and the meaning of all stored quantities.
- Each README file should contain the names of the data files and script files, as well as the (version of) the software that generated it.
- There should be a relation between the content of the experiment and the name of the script.
- There should be a relation between the script name and the name of the derived data files and figures. For figures, ensure that the plotted quantities are in the file name.
- If the script takes input parameters, these should be used in the name of the data files.

We are currently migrating to a MANGO setup (KU Leuven's in-house data management system), which will be integrated with the workflow of the researchers to automate the above procedures and produce "computational experiment" records. These records can then easily be shared via KU Leuven's RDR system upon finalizing publications.

Will a metadata standard be used to make it easier to find and reuse the data?
If so, please specify which metadata standard will be used.

If not, please specify which metadata will be created to make the data easier to find and reuse.

- No

Climate simulation data are available in NetCFD
Software developments are preformed in Python, and documented using Readme files
Although no formal metadata standards are applicable to much of the research, the standardized way in which all steps in the research will be documented (see previous question) will ensure findability and reusability of the data.

Data Storage & Back-up during the Research Project

Where will the data be stored?

- ManGO
- Other (specify below)

Research data will be stored

- personal laptop of researcher
- Data generated using HPC are stored on staging
- All data will be stored on Mango (or RDR when published)

All software developments are stored on the gitlab server of KU Leuven

How will the data be backed up?

- Standard back-up provided by KU Leuven ICTS for my storage solution

Is there currently sufficient storage & backup capacity during the project?

If no or insufficient storage or backup capacities are available, explain how this will be taken care of.

- Yes

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

All data is stored on KULeuven devices. Access is only possible for authorized users.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

The costs for data storage result from university policy and are documented on the pages of ICTS. We have explicitly foreseen a working budget with this IDN proposal to cover these costs.

Data Preservation after the end of the Research Project

Which data will be retained for 10 years (or longer, in agreement with other retention policies that are applicable) after the end of the project?

In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

- All data will be preserved for 10 years according to KU Leuven RDM policy

Where will these data be archived (stored and curated for the long-term)?

- KU Leuven RDR
- Other (specify below)

KULeuven archive, which will become available in 25/26
All software will remain stored on the gitlab server of KULeuven

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

Smaller data sets can be stored for free at KU Leuven RDR
Larger data sets on KULeuven RDR cost 35€/TB/year. Also data sets on Mango cost 35€/TB/year. Costs for KU Leuven archive (being developed) are not yet clear. Currently KULeuven lacks a proper bookkeeping approach to cover these costs using the project funds allocated to this project. At the moment, it is implied (and it is our intention) that all costs for preservation after the project ends will be covered through funds of the PIs of the projects, but this cannot be guaranteed.

Data Sharing and Reuse

Will the data (or part of the data) be made available for reuse after/during the project?
Please explain per dataset or data type which data will be made available.

- Yes, as open data

Whenever appropriate, all data and code required to reproduce figures and tables in papers will be made openly accessible.
In cases where this would not be possible, data will be made available upon request to the scientific community.

If access is restricted, please specify who will be able to access the data and under what conditions.

Members of the scientific community, upon request. Specific licence agreements will be discussed when such requests occur.

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

Please explain per dataset or data type where appropriate.

- No

Where will the data be made available?

If already known, please provide a repository per dataset or data type.

- KU Leuven RDR (Research Data Repository)

When will the data be made available?

- Upon publication of research results

Which data usage licenses are you going to provide?

If none, please explain why.

- CC-BY 4.0 (data)
- GNU GPL-3.0 (code)

Do you intend to add a persistent identifier (PID) to your dataset(s), e.g. a DOI or accession number? If already available, please provide it here.

- Yes, a PID will be added upon deposit in a data repository

What are the expected costs for data sharing? How will these costs be covered?

There are no additional costs for data sharing other than the costs for storage itself (see above)

Responsibilities

Who will manage data documentation and metadata during the research project?

During the project, management of data documentation and metadata is performed by the different PhD students and their main supervisors

Who will manage data storage and backup during the research project?

Management of data storage and backup is performed by the different PhD students and their main supervisors

Who will manage data preservation and sharing?

During the project, management of data storage and backup is performed by the different PhD students and their main supervisors. After the project the responsibility is shared between the PIs of the project

Who will update and implement this DMP?

The PIs have regular project meetings during which updating and implementation of this DMP are discussed.