



Funded by
the European Union

HORIZON EUROPE PROGRAMME – TOPIC HORIZON-CL5-2021-D3-03-05

*Wind energy in the natural and social environment
Research and Innovation action (RIA)*



wimby
WIND IN MY BACKYARD

WIMBY

Wind in My Backyard: Using holistic modelling tools to advance social awareness and engagement on large wind power installations in the EU

Grant Agreement No. 101083460

Starting date: 1st January 2023 – Duration: 36 months

**Deliverable D7.3
Data Management Plan**

DOCUMENT INFORMATION

Deliverable number	D7.3
Deliverable title	Data Management Plan
Work Package	WP7
Deliverable type	R – Document, report
Dissemination level	PU – Public
Due date	30.09.2023 (M9)
Pages	27
Document version¹	4.0
Lead author(s)	Luis Ramirez Camargo (UU)
Contributors	Maisam Mohammadi Dadkan (UU), Garret Speed (UU), Stella Arapoglou (VUB), Rebecca Hueting (DBL), Andrea N. Hahmann (DTU), Peter Burgherr (PSI), Dominik Huber (VUB), Etard Adrienne (IIASA), Chen Ruihong (ETH), James Price (UCL), Lien Bakelants (NAZKA), Claudio Moscoloni (POLITO), Monika Bucha (KIE), Neil Davis (DTU), Thomas Oeberseder (BOKU), Hshing-Hsuan Chen (UU), Marianne Zeyringer (UiO), Marta Cecconi (DBL)

The WIMBY project has received funding from the European Union’s research and innovation programme Horizon Europe under the grant agreement No. 101083460. This document reflects only the author’s view, and the Commission is not responsible for any use that may be made of the information it contains.

¹ First digit: 0: draft; 1: peer review; 2: peer review 3: coordinator approval; 4: final version

DOCUMENT CHANGE HISTORY

Version	Date	Author	Description
DRAFT			
0.1	24.02.2023	Maisam Mohammadi Dadkan (UU) Luis Ramirez Camargo (UU)	Creation
0.2	15.04.2023	Stella Arapoglou (VUB), Rebecca Hueting (DBL), Andrea Hahmann (DTU), Peter Burgherr (PSI), Dominik Huber (VUB), Chen Ruihong (ETH)	Feedback round 1
0.3	11.07.2023	Luis Ramirez Camargo (UU), Maisam Mohammadi Dadkan (UU), Garret Speed (UU), Stella Arapoglou (VUB), Rebecca Hueting (DBL), Andrea N. Hahmann (DTU), Peter Burgherr (PSI), Dominik Huber (VUB), Etard Adrienne (IIASA), Chen Ruihong (ETH), James Price (UCL), Lien Bakelants (NAZKA), Claudio	Changes made due to the DMP workshop



		Moscoloni (POLITO), Monika Bucha (KIE), Neil Davis (DTU), Thomas Oeberseder (BOKU), Hshing-Hsuan Chen (UU), Marianne Zeyringer (UiO), Marta Cecconi (DBL), Thomas Schauppenlehner (BOKU)	
0.4	12.09.2023	Luis Ramirez Camargo (UU)	Consolidated full version
FIRST PEER REVIEW			
1.0	28.09.2023	Eva Schöll (BOKU), Rebecca Hueting (DBL), Christian Mikovits (BOKU), Andrea N. Hahmann (DTU), Maisam Mohammadi Dadkan (UU)	Proofreading and peer review
1.1	29.09.2023	Stella Arapoglou (VUB)	Consolidation of input from reviewers
COORDINATOR APPROVAL			
3.0	29.09.2023	Luis Ramirez Camargo (UU)	Coordinator review, Coordinator approval
FINAL VERSION			
4.0	29.09.2023	Stella Arapoglou (VUB)	Format review, version ready for submission

SHORT ABSTRACT FOR DISSEMINATION PURPOSES

Abstract D7.3 Data Management Plan provides the guidelines and procedures to handle data, data collection, and open access in WIMBY. This deliverable is the result of an iterative co-creation process led by Utrecht University and supported by all WIMBY partners. The holistic approach of the project is reflected in the multitude of data that are required and that are generated throughout the project. These data are key for the tools and assessments to be developed in WIMBY as well as for the dissemination strategy and the project's expected impact. WIMBY strives to make data and other outputs open while following FAIR principles and INSPIRE guidelines. Data management will be the responsibility of a data management committee that will be supported by the data management teams of individual institutions. Multiple mechanisms are in place to ensure that data is managed securely and thorough ethical and privacy protocols are followed for personal data. This data management plan offers a common ground and a solid basis for all activities related to data in the project.



TABLE OF CONTENTS

DOCUMENT CHANGE HISTORY	3
SHORT ABSTRACT FOR DISSEMINATION PURPOSES	5
TABLE OF CONTENTS	6
LIST OF PARTNERS	8
ABBREVIATIONS	9
LIST OF TABLES	10
EXECUTIVE SUMMARY	11
1. Data Summary	12
1.1 Existing data to be used, their origin/provenance and their purpose 12	
1.2 Types and formats of data that the project will generate or re-use 16	
1.3 Purpose and relation of the data generation and re-use to the objectives of the project	17
1.4 Expected size of the data generated or re-used in WIMBY	19
1.5 Stakeholders that can make use of WIMBY's data	19
2. FAIR data	20
2.1 Making data findable	20
2.2 Making data accessible	20
2.3 Making data interoperable	22
2.4 Increase data re-use	22
3. Other research outputs	24
4. Allocation of resources	24
4.1 Direct and indirect costs related to storage, archiving, re-use and security	24
5. Data security	25
6. Ethics	26
7. Other issues	27



















8. Conclusions 27

9. References 27



LIST OF PARTNERS

No	Logo	Name	Short Name	Country
1		VRIJE UNIVERSITEIT BRUSSEL	VUB	Belgium
2		DANMARKS TEKNISKE UNIVERSITET	DTU	Denmark
3		INTERNATIONALES INSTITUT FUER ANGEWANDTE SYSTEMANALYSE	IIASA	Austria
4		UNIVERSITAET FUER BODENKULTUR WIEN	BOKU	Austria
5		UNIVERSITETET I OSLO	UiO	Norway
6		NAZKA MAPPS BVBA	NAZKA	Belgium
7		KELSO INSTITUTE EUROPE GEMEINNUTZIGE GMBH	KIE	Germany
8		DEEP BLUE SRL	DEEP BLUE	Italy
9		UNIVERSITEIT UTRECHT	UU	Netherlands
10		POLITECNICO DI TORINO	POLITO	Italy
11		UNIVERSITA DEGLI STUDI DI PALERMO	UNIPA	Italy
12		APREN-ASSOCIACAO PORTUGUESA DE ENERGIAS RENOVAVEIS	APREN	Portugal
13		MULTICONSULT NORGE AS	MCN	Norway
14		EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH	ETH Zürich	Switzerland
15		PAUL SCHERRER INSTITUT	PSI	Switzerland
16		UNIVERSITY COLLEGE LONDON	UCL	United Kingdom



ABBREVIATIONS

Acronym	Description
WP	Work Package
NEWA	New European Wind Atlas
DMP	Data management plan
DOI	Digital Object Identifier
ESA	European Space Agency
EC	European Commission
NIR	Near-Infrared
GDT	Geo Data Team
FAIR	Findability, Accessibility, Interoperability, and Reuse





LIST OF TABLES

Table 1. Potential data portals to retrieve geospatial data for the 3D environments of pilot cases.....	14
Table 2. Data required for the 3D environment of the pilot case in Pantelleria	15



EXECUTIVE SUMMARY

This Data Management Plan (DMP) provides the guidelines and procedures to handle data, data collection, data sharing and open access in WIMBY. It was developed in an iterative process led by the WIMBY UU team and the data stewards of the Geoscience Faculty of UU. Apart from interaction between project partners during the first six months of the project to consolidate the first version of the deliverable, a one-day workshop was organised by the data stewards of the Geoscience faculty of UU. This provided a common ground and understanding and defined how to deal with the data required and generated in the project. This DMP was constructed based on standard guiding questions for DMPs that were posteriorly transformed into a narrative.

WIMBY strives to make data and other outputs open while following FAIR principles and INSPIRE guidelines. Data management will be responsibility of a data management committee that will be supported by the data management teams of individual institutions. Multiple mechanisms are in place to ensure that data is managed in a secure way and thorough ethical and data privacy protocols are followed for personal data. To ensure data quality the project counts with dedicated tasks and the support of the data management teams of individual institutions. Moreover, WIMBY will make use of multiple channels for data sharing. These channels include a SharePoint cloud environment, the Yoda platform (Utrecht University data management system) for data sharing among the consortium, and the Zenodo repository and Yoda for publishing data sets.

Due to wide range of topics addressed in the project and the continuous effort of individual partners to make use of the most appropriate data sources it is expected that the list of data sets to be re-used will grow and therefore this data management plan will be updated during the project lifetime.





1. Data Summary

1.1 Existing data to be used, their origin/provenance and their purpose

In the WIMBY project, data from a wide and heterogenous range of sources is required. The project consortium is putting effort in using the most up-to-date, accurate and suitable sources available for individual tasks. This also means that the list of data sets to be used will grow and that replacement of particular data sets could be expected. The following are data sets that have been identified as required during the proposal writing stage and during the first six months of the project:

- **Energy-related Severe Accident Database (ENSAD):** In WIMBY, PSI's Energy-related Severe Accident Database (ENSAD) will specifically be updated for wind power accidents. For this purpose, different types of primary information sources will be surveyed, harmonized, and imported (e.g., industrial databases, news archives, and online portals, among others). The resulting data collection will also be made publicly available as part of the WIMBY dissemination and outreach activities.
- **Wind Climate (NEWA):** The European wind resources database from the NEWA project (DTU, NAZKA) will be the source of the meteorological time series for WP1 and WP2. We connect the raw wind resource data (NEWA microscale) via a series of coupling and optimization algorithms to estimate the annual energy production (AEP), considering wind turbine size, type, and position in the landscape.
- **Turbulence and siting (GASP)** We link the raw siting data (e.g., turbulence and extreme winds; GASP over land and GASPOC over the sea) to models that evaluate the type of turbine installed at a site given the terrain and wind climate (e.g., for turbulence and extreme winds).
- **Ecoinvent:** This state-of-the-art database provides data for the fore- and background modelling of the Life Cycle Assessment (LCA). The included datasets provide extensive life cycle inventories and multiple impact methods. The partners involved in the modelling possess a commercial license, whereas calculated results can be shared and disseminated.
- **Windpower.net:** Those commercial datasets will be acquired to obtain mass balance information about different wind turbine





technologies. The component weights are important to include the corresponding materials for their production, which determines the environmental impact of the turbines.

- **Enriched Data of Wind Farms (EDWIn)**

EDWIn is an open access data set based on OpenStreetMap that provides information about global wind farms. The data set includes location and basic contextual information of more than 350 thousand turbines that constitute over 20 thousand wind farms across the world. The data set was generated as of September 2022 and is available via Zenodo ([10.5281/zenodo.7558884](https://zenodo.org/record/7558884)).

- **Prospective databases for Life Cycle Assessment**

Prospective databases couples the ecoinvent database with projections from Integrated Assessment Models (IAM) and will be used in order to assess future environmental impacts of wind turbines installed in Europe. Those databases are published open-access and available in a public GitHub repository.

- **OpenStreetMap:** it is a free wiki map of the world that contains data on buildings and all sorts of infrastructure including e.g. transportation and energy. The data is available under <https://www.openstreetmap.org> and can be retrieved through a large list of specialized APIs/wrappers. This data will be used in multiple tasks across the project for e.g. indicators calculation about landscape quality and visualization in the developed tools.

- **Sentinel 2 data:** This is satellite derived data gathered and generated by the European Space Agency (ESA) in the framework of the Copernicus Programme of the EC. The data is imagery derived from optical sensors at resolutions between 10 m and 60 m. The data is made available by the ESA and can be retrieved through multiple channels (e.g., sentinelhub). This data will be mainly used for task 1.2. to assess land use change impact of wind turbines.

- **CORINE Land Cover:** It is an inventory of land cover classes with continental coverage. This is a data set generated and made available in the framework of the Copernicus Programme of the EC. In WIMBY this data set is in used by multiple tasks as for example the assessment of landscape quality and the 3D immersive environment.

For the four pilot cases and their respective 3D environments data will be retrieved and re-used from platforms such as the ones listed in Table 1.



Table 1. Potential data portals to retrieve geospatial data for the 3D environments of pilot cases.

Region	Name	URI	Content
Italy	Geoportale nazionale - National Data Portal	http://www.pcn.mi.nambiente.it/mattm/en/	Different geodata from Italy
Italy	Italian Data portal	https://dati.gov.it	Different geodata from Italy
Portugal	OGD Portal of Portugal	https://dados.gov.pt/en/	Different geodata from Portugal
Portugal	SNIG - National Data Portal	https://snig.dgterri.torio.gov.pt/	Ortho images 25 cm (as WMS) and Near-infrared (NIR) images
Norway	NIBIO Research institute	https://nibio.no/en	Research data
Norway	Geonorge map catalogue	https://www.geonorge.no/en	Catalogue to search for, look at or download official Norwegian geodata
Norway	Dataportal of the Norwegian Mapping Authority	https://kartverket.no/en	Access to different data from Norway
Norway	WebGIS of Norway	https://norgeskart.no/	WebGIS with data for Norway
Norway	Kartverket API and data (INSPIRE)	https://www.kartverket.no/api-og-data	Different data from Norway
Norway	Kartverket WebGIS	https://hoydedata.no/LaserInnsyn2/	Norwegian DTM/DSM in 1x1m resolution
Austria	Austrian Data Portal	data.gv.at	Austrian Data Portal with data from different authorities and organizations
Austria	INSPIRE Geodatenportal	https://geoportal.inspire.gv.at/meta-datensuche/inspire/ger/catalog.search#/home	Nationwide DTM, DSM and orthoimages provided by BEV

Austria	Austrian Geodata Portal	https://www.geoland.at/index.html	Austrian Geodata Portal with data from the Federal States
Europe/ Global	Copernicus Landmonitoring Service	https://land.copernicus.eu	Corine Landcover and other high resolution (10m) of Europe (Hydro-Network, urban areas, etc)
Europe/ Global	OpenStreetMap	www.openstreetmap.org	Road network, waterways and waterbodies, natural and landuse data, Point of interests
Europe	ForestMap of Europe	https://efi.int/knowledge/maps/forest	Forest coverage data
Europe	European primary forest database v2.0	https://www.nature.com/articles/s41597-021-00988-7	Forest coverage data

The data required for each pilot case will be equivalent to the one that has been already retrieved for the pilot case in Pantelleria (see Table 2)

Table 2. Data required for the 3D environment of the pilot case in Pantelleria

Dataset name	Short description	Region covered	Source
Road Network - Sicily	Road Network of Sicily	Local - Sicily	Sicilian Regional Technical Chart
Buildings floor plants - Sicily	Buildings floor plants of Sicily	Local - Sicily	Sicilian Regional Technical Chart
Line network and PP	Electric lines of Sicily	Local - Sicily	Sicilian Regional Technical Chart
Forest Data	Forest data in compliance with D.Lgs. 227/01	Local - Sicily	S.I.T.R. - Regional Geoportal
Water Surface	Sicilian Water surface	Local - Sicily	Sicilian Regional Technical Chart

Land Cover 1/2	Sicilian Habitat Chart - Corine Biotopes (HCB)	Local - Sicily	S.I.T.R. - Regional Geoportal
Land Cover 2/2	Sicilian Land Cover Use - Corine Land Cover	Local - Sicily	S.I.T.R. - Regional Geoportal
DSM	Digital Model Surface of Pantelleria	Local - Pantelleria	National Park - Island of Pantelleria
Orthophoto	Orthophoto of Pantelleria	Local - Pantelleria	National Park - Island of Pantelleria
DTM	Digital Terrain Model of Pantelleria	Local - Pantelleria	National Park - Island of Pantelleria
DTM	Digital Model Terrain of Sicily	Local - Sicily	S.I.T.R. - Regional Geoportal

1.2 Types and formats of data that the project will generate or re-use

The project will generate all data in open and non-proprietary formats to be used by other groups and also in the Web-GIS interactive forum. The major part of the data is spatial and spatiotemporal data that is available and will be made available in common raster and vector data formats such as GEOTIFF, GRIB, NETCDF, SHP and GEOJSON. The data will be exchanged among the consortium in the original file formats, but it will be transformed into formats that are optimal for the particular final uses i.e. the Web-GIS interactive forum, the immersive 3D-environment and final dissemination.

Tabular data gathered from data repositories, literature reviews, interviews, surveys and resulting from the energy system modelling tasks will be stored and re-distributed in CSV format.

Concerning the inventories for the LCA, the generated data is made available in common data formats, e.g. as csv or xlsx. In the same format, used inventories are made available. The reviewed literature will be described in form of a report in pdf or docx format. Information on future

end-of-life treatment will be obtained by a combination of semi-structured interviews and literature data. Regarding evaluation of the new foundation technologies, scientific literature and expert opinions will be gathered. To include this data in LCA, it needs to be translated into life cycle inventories. Data format is standard csv or xlsx and the size of such inventories is comparable small (below one MB each).

The data generated from semi-structured interviews (conducted to e.g. identify job creation potential and assess end-of-life treatment of turbines) includes mp3 audio files and transcripts in txt or docx formats. The audio files will be deleted after transcription. Transcripts will be anonymised and only be used for evaluating and not be published.

The energy system modelling and optimization tasks will also produce data bases in e.g. gdx and associated formats. Such files will be kept for internal use and key results will be transformed into CSV files for dissemination.

In the case of the dissemination tasks (WP6) most data collected and produced will be in the form of written and/or layouted documents. Some additional data formats specifically related with the Communication, Dissemination and Exploitation activity will be stored, such as audio-visual recordings, interactive documents, infographics and code files. Data formats will include:

- Documents: .doc, .pdf, .ppt, .txt, .xlsx, .csv
- Images: .jpg, .png, .svg, .gif
- Audio-visuals: mp3, .mp4, mpeg
- Code: .html, .py, and other relevant code file formats

1.3 Purpose and relation of the data generation and re-use to the objectives of the project

Data is key for the achievement of all WIMBY objectives. WP1 and WP2 are the WPs that will re-use, gather and produce most data in the project. These data is the one populating and allowing the analysis and assessments of WP4, the development of the tools of WP5 and the implementation of the workshops of WP3. The data that is reused, collected, and generated in WIMBY serves to address the seven specific objectives of WIMBY:



- **SO1. Understand and develop ways to value cumulative impacts of onshore and offshore wind-power on the environment in Europe** by estimating wind resources availability and quality, their consequences on land use and sea use and the assessment of biodiversity risk including detailed ecological modelling.
- **SO2. Understand and find ways to value cumulative impacts of onshore and offshore wind-power on society in Europe** by assessing impacts on the local communities and other sectors (e.g. tourism, fisheries) and addressing issues concerning governance, regulations, health impacts and intangibles, such as scenery or landscape in a spatially explicit way.
- **SO3. Develop an assessment of onshore and offshore wind-power deployment potential in high spatial resolution for Europe that goes beyond state-of-the-art** by including estimations of the quality of the wind resources availability, a comparison of alternative technologies and wind park shapes as well as valuations of impacts on landscape, flora, fauna and humans and a location-based Life Cycle Assessment (LCA).
- **SO4. Understand the system impact of large scale integration of wind power on the European energy system** by iteratively coupling a range of planning models across local to continental scales that capture the implications of social and environmental factors on resource availability and spatially optimise the deployment of the system
- **SO5. Co-create and validate with wind power projects stakeholders a Web-GIS interactive platform**, which merges the results of the environmental, technical, economic and social assessments. This platform serves as a forum where interested audience, affected industry, specialized users, ancillary business can visualise environmental, economic, ecologic and socially relevant trade-offs and exchange information and provide input to one another to help understand the impacts and trade-offs of new wind turbines and parks at a medium to large scale.
- **SO6. Develop an immersive 3D environment** that allows stakeholders within dedicated workshops visualise and understand the impacts and trade-offs of wind energy development in their specific neighbourhood and enhance the participatory processes.



- **SO7. Develop and validate an improved methodological framework for wind farm planning including guidelines for participatory processes to raise public understanding and engagement and boost energy citizenship** based on the interaction with stakeholders in the case studies, additional working groups and the forum.

1.4 Expected size of the data generated or re-used in WIMBY

By the end of WIMBY, we expect that several terabytes (TB) of data will be generated as intermediate results. Most part of it will be related to the spatiotemporal data sets to be created in T1.1, T1.2 and T2.1. and to sensitivity analysis of models run in WP4.

The data of final results per task would range from few MB (from e.g. the regulatory information coming from T2.5) to few TB (that will be the case e.g. for the spatiotemporal data sets from T1.1 and T2.1). Data related to LCA inventories is expected to do not be larger than 5GB in total and all the communication, dissemination and exploitation data (including layouted graphics and audio-visuals) should not exceed 10 GB. In contrast, tasks producing spatial data as e.g. the land use change task will produce dozens of files that could easily require several GB each. Data will be tiled and stored in multiple files so that computational requirements for accessing, reading and writing is optimized for the tools and software in used in the project.

1.5 Stakeholders that can make use of WIMBY's data

The data generated in WIMBY can be used by the following groups:

- Researchers working in different fields e.g. energy system modelers, energy economists, energy lawyers
- Wind power engineers
- Wind power project developers
- Local and regional policymakers in Europe
- Local (planning) authorities
- Local communities

- Interested Citizens

2. FAIR data

2.1 Making data findable

All the WIMBY's publishable data outputs and results will be stored in repositories that provide a persistent Digital Object Identifier (DOI) upon publishing. Several options including [Yoda](#) and [Zenodo](#) will be used. Yoda (Utrecht University data management system, with options for public access and only WIMBY-internal access) will be used for any type and size of data. Zenodo (publicly accessible) will also be used for datasets or outputs less than 50 GB.

The data made available will be accompanied by standard metadata and documentation to ensure they are findable. WIMBY also includes a dedicated task led by Utrecht University (T5.1) to ensure that generated and provided spatially explicit data and metadata are aligned with [INSPIRE technical guidelines \(INSPIRE Directive 2007/2/EC\)](#) and fulfil the FAIR principles. T5.1 is responsible to decide which format of data files and metadata standards will be used for the whole project. However, for data to be stored on Yoda or Zenodo, the metadata standard is DataCite v4.

The metadata includes keywords based on the INSPIRE guidelines and is offered in a way that it can be harvested and indexed. This is done to optimize the possibility of discovery and then potential re-use. The metadata will be mostly available as a [JSON Schema](#) as the internal representation of metadata for Yoda and Zenodo, where most of the data will be stored.

2.2 Making data accessible

All the data repositories that will be used in WIMBY are trusted repositories and are being selected in consultation with the data team at the Faculty of Geo-Science at Utrecht University. We considered several options to deposit data depending on the size of the data such as Yoda and Zenodo. Both Yoda and Zenodo assign DOI to the data upon publishing.

We aim to make the data of the WIMBY project openly available except for the following:

- Part of data that includes personal data and cannot be deidentified or anonymized
- The data under a non-disclosure license agreement with a third party

WIMBY has the ambition to make all data and outputs that are not strictly proprietary available under a permissive license to facilitate quality control and community-based improvement, cumulative science, and rapid availability to stakeholders. Most of the WIMBY data is accessible via HTTP/HTTPS protocols on the web.

Moreover, data with restricted access are held by the research groups that hold the licenses for these data, and they underly the archiving regulations at the different groups, typically a storage period of a minimum of ten years. The license under which the project's results will be published does not require any user tracking. For sensitive data, the data are stored in the platforms like Yoda which is protected by the UU authentication system with 2-Factor-Authentication (2FA).

All datasets will be accompanied by rich metadata (adhering to DataCite V4 metadata standard), supplemented by discipline-specific metadata as specified by Pauliuk et al. (2019). In addition, keywords describing the datasets will be added. The metadata is stored persistently along with the data. For datasets for which the license or institutional arrangement does not allow publication, we will still publish empty datasets without actual data but full metadata, including information to enable the user to submit a request to access the data.

The documentation and references of any software that are needed to access or read the data will be included. The data handling software, database structure, and data formats are open-source and will be archived on GitHub and a link to these tools will be provided along with the data.

After the project is completed, the data will be archived for at least 10 years persistently on Zenodo or Yoda. This will ensure that the data is securely retained for the required period of time according to Horizon Europe, and is

easily findable through its metadata. 10 years will be calculated from the end of the project, or later if required according to relevant requirements (e.g. Netherlands Code of Conduct for Scientific Practice, Horizon Europe).

Additionally, a Data Management Committee is foreseen in WIMBY that is responsible for managing all aspects of research data including data access and sharing. The Committee includes at least one institutional Data Protection Officer (DPO) and another two Consortium experts.

2.3 Making data interoperable

The dataset resulting from WIMBY will be provided in open format to facilitate reusing by others now and in the future. There is also a dedicated task in the project, T5.1, that will ensure all the data are aligned with the INSPIRE guidelines. If during the project we find it necessary to generate new ontology or terminology, we will document, share, and if possible, integrate them with the commonly used one in the community based on the INSPIRE rules.

We will also include qualified references to the data. Cross-references will be made clear by supplying appropriate metadata which data form input data into which model, which data are intermediate results, and which are model results (and from exactly which model run and with which input and intermediate data).

2.4 Increase data re-use

All the datasets contain all information needed to re-use the data (variable definitions, units of measurement, License, etc.) in a form of README or other documentation forms.

The documentation for tracing the data to their original sources, for validating them, and for assessing their information content is compiled and provided by the participating research teams according to the state of the art of the scientific fields involved.

The software/codes that are developed during the project and used to process data will be openly available under an open-source licence and will be linked to the associated dataset (e.g. using GitHub and Zenodo integration).

Survey data will be fully documented, including the data sampling methodology, codebooks and survey items in original languages and English, data cleaning procedures applied (syntax for that will be documented), and variable definitions.

All the data and other outputs of the project will be openly available using open-access licenses like CCBY V4 or GPL taking into account the "As open as possible and as close as necessary" rule.

We expect that the data generated in the project can be used by a wide range of communities. The projects include partners from a large range of scientific disciplines and results and data will be produced for each one of them. Moreover, these data might be useful for even more research disciplines related to wind power research and is a clear goal of the project to reach a wide group of stakeholders (see section 1.5.). Also, WIMBY adopts and promotes citizen science activities and deliberations as part of its multi-actor and co-creation approach, specifically for the generation of new scientific knowledge and understanding of enabling and hindering factors for wind power projects.

Moreover, during the project we will establish a data provenance model to provide a formal and actionable definition of provenance and enable sharing of provenance information between systems and project entities by supporting tool development and implementation.

UU is responsible to define the required processes to assess the data quality during and after the project. However, it has been already agreed to use the following guidelines to ensure the data and software quality:

- Using agile and open collaborative software development principles to make sure that the data processing tools are peer-reviewed and decentralize these activities.
- A data management working group (DMWG) that includes people who generate data, analyse data, and data managers will plan and define a data dictionary (including validation rules), and metadata templates before collecting data.
- Training of study participants and researchers, surveyors, or other staff involved to implement best practices in the data lifecycle. Geo

Data Team (GDT) at UU will provide support to the researchers involved in the project.

Further to the FAIR principles this DMP also addresses research outputs other than data, and carefully consider aspects related to the allocation of resources, data security and ethical aspects. These are addressed in sections 3.-6.

3. Other research outputs

WIMBY also considers making a software management plan for the codes and models that will be developed in the project. The FAIR principles will be also applied to large parts of the **modelling software** developed during this project (model code, data checks, interfaces, visualisation tools).

In particular:

Findability: The software will be archived in public repositories where mandated by the funding agreement or where in line with institutional policy, and also registered in software dashboards/catalogues, such as GitHub.

Accessibility: Software typically is archived in version management repositories (e.g. GitHub) that have a requirements file that lists all dependencies to other libraries to allow for running the software. In some cases, we will provide the software with all the dependencies in a portable format like Docker containers.

Interoperability: Interfaces (either software pieces or data files) will be archived in the same fashion to allow different software pieces to be run together or consecutively.

Reusability: We strive to develop modular, version-managed, documented, and unit-tested software under a proper open-source license to facilitate re-use both within and beyond the project.

4. Allocation of resources

4.1 Direct and indirect costs related to storage, archiving, re-use and security

The costs of open-access publishing:



All peer-reviewed scientific publications resulting from the project are published under “Gold” OA. i.e. immediately available and with associated costs cover either by individual OA agreements of partners (e.g. BOKU, UiO2, and UU) with particular publishers (e.g. Elsevier) or by the budget that has been allocated for this purpose by the partners.

Costs of data management:

The personal costs for making WIMBY’s outputs FAIR are part of the costs calculation of the project. Additionally, to the effort of individual partners and the dedicated tasks in WP5, the data teams of individual institutions such as UU’s Geoscience data team, are contributing to this effort while being financed by the respective institutions.

While storing data in Zenodo does not represent costs, storing data in Yoda has a cost of 4euro/month per TB. A basic contingent (the first TB) is financed by UU’s Geoscience department the rest of the storage costs are calculated in UU’s budget in WIMBY. The total cost will depend on the final amount of produced/stored data.

Responsibilities for data management in WIMBY:

Data Management Committee is responsible for managing all aspects of research data. The Committee includes at least one institutional DPO and another two Consortium experts. UU is the partner that monitors and manages the data management activities and provides reports to the data management committee as defined in T5.1.

Long term preservation:

All the infrastructures that are being selected for data storing and archiving have the policy to keep data at least for 10 years. However, data on Zenodo can be persistently stored for a longer period.

For datasets for which the license or institutional arrangement does not allow publication, we will still persistently publish on Zenodo the empty datasets without actual data but full metadata, including information to enable the user to access the data.

For later archiving, free public infrastructure (such as Zenodo) will be used as well as the participating institutions’ data archiving infrastructure will be used, e.g., [DAG](#).

5. Data security



All the data storage solutions that we are planning to use have a high-standard backup and data recovery policy. Furthermore, the research data will be subject to UU's "Information security policy". We will use data classification; each dataset will be classified according to the CIA triad (Confidentiality, Integrity, Availability). Procedures to store and get access to data sets will be set according to the CIA classification.

For data exchange among consortium partners, SharePoint hosted at the VUB, the Mattermost hosted by DBL and Yoda hosted by UU will be used. SharePoint, which is under a Data Processing Agreement with the VUB, enables cloud-based long-term storage of data, which will be used to store the raw interview data (audio files) during the project. SharePoint also contains data backup and recovery features as well as long-term preservation beyond the project's lifespan. Members of the research team will be granted access to one or more particular folders on a need-to-know basis. As described above, regular backups of all datasets of the project will be made and stored in Microsoft data centres in Europe. Mattermost is the main interaction platform of the project consortium, which also allows files exchanges. Files of several MB can be exchanged there. The platform is hosted and backed-up by DBL and has been set up in collaboration with DTU support. The configuration of Mattermost includes high levels of security including encryption of messages and data sent through it. Moreover, Yoda will be used for large files. As explained previously Yoda complies with a high-standard backup and data recovery policy.

All the online surveys will be designed in a secure and GDPR-compliant format to make sure the personal data are well protected. UU has a license to use [Qualtrics online survey management tool](#) so all the researchers can make and manage the survey via this platform which is under the IT-security umbrella of the UU.

6. Ethics

UU has a mechanism in place called [privacy scan](#) that is used in all projects that the university is participating. This can be understood as part of the ethics procedures where personal data is involved. This is developed and reviewed with the support of UU's privacy officer and serves to ensure that there are no ethical issues in the project.

Concerning questionnaires including personal data a notice on personal data at the very beginning will be included. Respondents need to check a box stating that they have understood and agreed with the processing of



their personal data. Checking this box will be mandatory to save their answers.

7. Other issues

Surveys and other personal data collection, processing, storage, sharing, and publishing will follow GDPR regulations and in addition the regulations of data protection in the countries they are collected in.

Finally, the materials supporting the Workshop that contributed to the creation of this DMP are available in Zenodo (See Dadkan et al. 2023).

8. Conclusions

The WIMBY project will (re-)use and produce a wide variety of data. While data for re-use and intermediate results will be made available among the project partners, project results will be made available to the public following INSPIRE guidelines, when applicable, and FAIR principles in all cases. Data and other research outputs will be made “as open as possible as close as necessary”. A data management committee will ensure that the data in the project will be handled in a secure and ethical way. Due to the wide range of topics addressed in the project and the continuous effort of individual partners it is expected that the list of data sets to be re-used will grow and therefore this data management plan will be updated during the project life-time.

9. References

Dadkan, M. M., Speed, G., & Pastrana, F. R. (2023, Juli 13). *Research Data Management Workshop for WIMBY project*.

<https://doi.org/10.5281/zenodo.8142292>

Pauliuk, S., Heeren, N., Hasan, M. M., & Müller, D. B. (2019). A general data model for socioeconomic metabolism and its implementation in an industrial ecology data commons prototype. *Journal of Industrial Ecology*, 23(5), 1016–1027. <https://doi.org/10.1111/jiec.12890>