

Data Management for Preludium 24

1. Data description and collection or re-use of existing data

How will new data be collected or produced and/or how will existing data be re-used?

New data in this project will be produced through both **modelling-based simulations and empirical data collection**. Primary data will be collected via structured interviews with selected stakeholders (e.g. cluster managers, energy firms, and local policymakers) to support model calibration and contextual understanding.

In addition to original data collection, the project will re-use existing datasets from national statistical offices, European statistics, energy agency databases, and academic publications. These secondary sources will include regional industrial indicators, firm-level energy investment records, and policy intervention timelines.

All data collection and reuse will follow ethical standards and data protection regulations, including GDPR compliance.

What data (for example the types, formats, and volumes) will be collected or produced?

The project will generate **both qualitative and quantitative research data** in digital formats. The qualitative data will include audio recordings of interviews in .mp3 format and their corresponding notes in .doc format. Data collected via structured surveys, will be stored and processed using SPSS software, resulting in a .sav database file.

In addition to these primary data types, simulation outputs generated from modelling tools (e.g. lifecycle curve simulations, network structure data) may also be included in the dataset in structured formats such as .csv or .xlsx. Cluster lifecycle quantification using Stata will be exported in .dta and .csv formats for further use and reproducibility. Quantification of the time–factor relationships using Python-based models will be output as .xlsx and .json files.

The total volume of the analysed data is expected to be approximately 50 GB. All data will be stored securely, access control, and GDPR-compliant handling of sensitive information.

2. Documentation and data quality

What metadata and documentation (for example methodology or data collection and way of organising data) will accompany data?

The data in this project will be generated through expert interviews and structured database searches. All stages of the research will be carried out in accordance with established ethical

standards, ensuring informed consent, voluntary participation, and the anonymity of respondents. Personal data will be anonymised and handled in full compliance with GDPR regulations.

Selected datasets and research outputs will be made openly accessible via **The Bridge of Knowledge** platform maintained by Gdańsk University of Technology. Metadata will be documented following the **DDI (Data Documentation Initiative)** standard and stored in JSON-LD format to ensure interoperability across systems. Each dataset will be linked to the responsible researcher via a registered ORCID identifier.

To maximise transparency, reuse, and scientific impact, the published data will be made available under the Creative Commons Attribution 4.0 International (**CC BY 4.0**) licence. This allows others to freely copy, distribute, and build upon the data—even for commercial purposes—provided that appropriate credit is given to the original author.

This approach ensures compliance with the FAIR data principles (Findable, Accessible, Interoperable, Reusable) and contributes to open science within the EU research framework.

What data quality control measures will be used?

Data quality in this project will be ensured through a combination of methodological rigor, standardised procedures, and verification practices.

The data will be catalogued and made available in accordance with the requirements of FAIR standards. This means that the data will be: Findable, Accessible, Interoperable, Reusable, thus ensuring that high standards of scientific research and data management quality are met. The data, available in the open repository, will be assigned a DOI and positioned to ensure its accessibility, with metadata to support discoverability and use across platforms. In addition, the repository will implement long-term maintenance measures, maintaining the integrity and usability of the data over time. Users will have access to full documentation and support to facilitate the correct interpretation of data and its integration into new and ongoing research project.

3. Storage and backup during the research process

How will data and metadata be stored and backed up during the research process?

All data collected and processed during the development process is stored in appropriately named and traceable password-protected files, only on team members' storage disk with password and on external drives.

Researchers can also use secure and password-protected cloud platforms – MS OneDrive and MS Teams. Permissions and access to platforms are carefully monitored and managed by the Principal Investigator of the project to prevent unauthorized use or access, and accidental loss or corruption of data. The automatic synchronization of files and folders will be used to reduce the risk of accidental loss or corruption of research data. Datasets and metadata will be

backed up at regular intervals (depending on the stage of the project) by each member of the research team. The data shall be promptly transferred from the recording equipment (audio recorder) to a secure storage location at the earliest possible moment, with immediate deletion from the recording equipment upon transfer.

How will data security and protection of sensitive data be taken care of during the research?

All recordings will be stored on encrypted, password-protected servers accessible only to authorised members of the research team. Data will not be stored on unsecured local devices or transmitted via unprotected channels. Internal data transfers between team members will be carried out using a secure local network or encrypted file-sharing platforms. The solution for secure storage will be a password-protected cloud platform, in particular MS OneDrive and MS Teams.

In cases where data is stored on mobile devices (e.g., laptops or phones), these devices will be encrypted to ensure data security. The research team maintains up-to-date security software on all devices used to collect and access data. In addition, it will be ensured that security features are actively enabled and working on all devices at all times. So the 3-2-1 rules are applied.

4. Legal requirements, codes of conduct

If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

The interviews and field investigations conducted as part of the study are designed to focus on institutional perspectives, expert insights, and thematic content relevant to the development of EBIC, without recording any personally identifiable information. Participation will be entirely voluntary and anonymous. In cases where references to job positions or organisational affiliations are relevant to the research context, such information will be used only with the explicit consent of the interviewee.

~~This consent allows us to share data with authorized scientists. By ensuring compliance with GDPR regulations, we protect personal identity, promote privacy, and maintain data security.~~ Adherence to the GDPR guidelines ensures that all data processing practices are lawful, transparent and in line with the rights and expectations of participants. The consent process is crucial to enable legitimate research while maintaining strict confidentiality and data protection standards.

How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

Gdańsk University of Technology is the owner and manager of all intellectual property created during the implementation of the project. Some of the data is published in scientific

journals in accordance with license agreements. In addition, some of the data is made available in the open access model under one of the Creative Commons licenses. Selected data are available in the MOST Wiedzy Open Research Data Catalogue.

5. Data sharing and long-term preservation

How and when will data be shared ? Are there possible restrictions to data sharing or embargo reasons?

Some of the data **is** published in the home university's open research data repository, MOST Wiedzy Open Research Data Catalogue, until the date of publication of the related results.

Depending on the requirements of the journal, some datasets may be submitted with publications and published by journals, which may also require the publication of raw data. Potential users of shared datasets have extremely easy access and the ability to obtain the information they need.

All data **is** securely stored for at least ten years from the date of completion of the project. This period is justified by the typical length of the publication process and the need to ensure that publications are based on current and up-to-date information.

How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?

The **MOST Wiedzy Open Research Data Catalogue** (<https://mostwiedzy.pl/en/>) is the main repository. The data made available in the repository meet the requirements of FAIR standards, guaranteeing their discoverability, accessibility, interoperability and reusable. The repository provides comprehensive technical support to both the research team and data users, facilitating the processes of uploading, retrieving and consuming data. In addition, it provides robust data security measures and regular updates to maintain data integrity and availability over time.

What methods or software tools will be needed to access and use the data?

The data used in this project will be processed using a combination of licensed and open-access software, depending on the specific nature of each research task. For quantitative modelling and statistical analysis, Stata and Python will be the primary tools. Python will be used for lifecycle simulation and factor interaction modelling, while Stata will be used for both linear and nonlinear regression tasks. For qualitative data coding and analysis, ATLAS.ti will be used. General documentation and report preparation will be managed using standard office software (e.g. MS Office).

Once processed, the resulting data—whether qualitative coding frameworks, time series

outputs, or model simulation results—will be exported and uploaded to an Open Access Research Data Repository, such as MOST Wiedzy Open Research Data Catalogue. All shared datasets will be formatted in open, machine-readable formats, such as CSV (.csv) .

How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

The MOST Wiedzy Open Research Data Catalogue allows ~~you to~~ assign unique Digital Object Identifiers (DOIs) to each uploaded dataset, ensuring precise and permanent identification. The data stored in this project is systematically linked to unique DOIs, facilitating accurate citation, tracking, and long-term access. This integration of DOIs increases the discoverability and interoperability of datasets, in line with best practices for research data management, and provides a robust framework for academic recall and data sharing.

6. Data management responsibilities and resources

Who (for example role, position, and institution) will be responsible for data management (i.e the data steward)?

The Open Science Competence Center (<https://pg.edu.pl/en/openscience>), established by Gdańsk University of Technology, is responsible for supervising the Data Management Plan (DMP), as well as for storing and sharing data. The project manager (PI) is responsible for evaluating procedures and ensuring overall data quality. The Centre implements best practices in data management and the Project Manager ensures that all protocols are followed, maintaining the highest standards of data integrity and reliability throughout the duration of the research project

What resources (for example financial and time) will be dedicated to data management and ensuring the data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Not applicable