

PROJECT ACRONYM:
PROJECT NUMBER:
PROJECT TITLE:
Supporting

CEANGAL
COALESCE/2021/41
Community Based Decentralised Renewable Energy Systems and
Structures for Improving Electricity Access in Low Income Countries



ceangalproject.com

CEANGAL Data Management Plan

Deliverable Information	
Deliverable name	CEANGAL Data Management Plan (DMP)
Deliverable number	7.1
WP	7
WP Lead	Institute of Technology, Sligo (ITS)
Author	Dr. Ehiازه Ehimen

Contributing Partners to Deliverable
1.

Revision Monitor		
Version No.	Date	Version Description
1	10/03/2022	Report prepared by Ehiازه Ehimen

About CEANGAL

The CEANGAL project takes us a few important steps closer to achieving sustainable electricity access for low-income communities currently with no accessibility.

Energy supply, and specifically electricity accessibility has been recognised as an important driver of economic growth and human development, and deemed essential to helping achieve global poverty reduction goals in low income countries globally. Goal 7 of the UN Sustainable Development Goals identifies the ‘universal access to affordable, reliable, sustainable, and modern energy for all’ as a target for the next 10 years. However, approximately 1 billion people globally have no access to electricity, the majority of whom are in Sub-Saharan Africa. Mechanisms which can be used to expand electricity access in such regions are therefore needed.

The CEANGAL project puts forward an ambitious adaptable and replicable model to support activities and know-how relevant to the selection, procurement, installation, and ownership of renewable energy systems (RES), as well as providing supporting structures to ensure their continuous local operation and maintenance.

The CEANGAL framework achieves this by addressing the major issues currently hindering RES adoption in developing communities, through the contribution of significant Irish expertise and knowledge in the community RES sector, and working collaboratively with local knowledge partners and stakeholders to develop vital longstanding local funding, ownership and operational support hubs for community driven RES projects. Through research partnerships with Malawian researchers, the project concept and outcomes will be developed, demonstrated and validated in four pilot case sites located in currently underserved rural areas of Malawi (a country having one of the lowest access to electricity globally with 82% of the population having no electricity access).

The CEANGAL outcomes directly contributes to the Global Ireland plan (Footprint to 2025) and addresses two of the four policy priorities of the Irish “A Better World” policy for international development (climate action and reduced humanitarian need), thus greatly highlighting Ireland’s leadership role in creating a more equal and sustainable world.

The project will through its activities aim to address the central questions:

- (a) How can community driven decentralised electricity systems be successfully implemented and replicated in LIC communities?*
- (b) How can we further ensure greater local buy-in and ownership of such systems in disadvantaged communities?*
- (c) What mechanisms should be put in place to ensure effective knowledge and expertise transfer, and facilitate continued operation, future expansion and replication potentials of RES projects post-installation and commissioning period?*

For more information on the CEANGAL project

Project Coordinator: Dr. Ehiaze Ehimen, ehimen.ehiaze@itsligo.ie

Dissemination Manger: Dr Unnikrishnan Brijitha-Madhavan, brijitha-madhavan.unnikrishnan@itsligo.ie

Executive Summary

This deliverable presents the first version of the CEANGAL Data Management Plan (DMP) and describes:

- **The guiding principles** for data management in the project
- **The legal framework** constituted by the General Data Protection Directive (GDPR)
- **Data Summary:** Overview of what data will be gathered and processed in the project
- How data will be stored and processed according to **FAIR Data Management principles**, making data: finable, accessible, interoperable, and reusable.
- **Resource allocation:** The costs of making data FAIR in this project
- **Data Security:** How we intend to keep the data secure
- **Ethical aspects:** A summary of the ethics and privacy strategy in CEANGAL

The purpose of the DMP is to contribute to good data handling through indicating what research data the project expects to generate and describe which parts of the data that can be shared with the public. Furthermore, it gives instructions on naming conventions, metadata structure, storing of the research data and how to make public data available.

During the 36 active months of the project, a Microsoft® SharePoint site will be used as the online working and collaboration platform. SharePoint will only be accessible to project participants and can provide further access control through establishing folders and sub-sites with stricter access granted than to the main site. During the project all *anonymised* (public) datasets will be uploaded to this site and stored in accordance with the ethics and privacy strategy of the CEANGAL project. *Non-anonymised* datasets (if they arise) for the pilot sites in Malawi will be stored locally by the designated Data Controller in Malawi University of Business and Applied Science (MUBAS), and not shared or distributed in any way to others. Metadata will be added to all datasets, and instructions on how to upload research data is provided.

The CEANGAL project will use the open research data repository *Zenodo* to comply with the Irish Research Council (IRC) Open Access recommendation. This recommendation applies to the underlying research data of publications, but project partners can also voluntarily make other datasets open. In CEANGAL, all deliverables, publications and the anonymous parts of the underlying datasets will be uploaded to the *CEANGAL Community* in Zenodo. Uploads will be done upon approval of the deliverables by the IRC, upon publication or acceptance of scientific publications, or, for underlying datasets, at the end of the project at the latest. Each dataset will be given a persistent identifier (Digital Object Identifier, DOI), supplied with relevant metadata and linked to the project name and project number. Publications and underlying research data will be linked to a Creative Commons license which **regulates** reuse. Data security arrangements are defined for the SharePoint site and Zenodo. Ethical aspects related to data collection, generation and sharing have been considered and nothing in this project shall be deemed to require a party to breach any mandatory statutory law under which the party is operating. This includes any national or European regulations, rules and norms regarding ethics in conducting research.

The DMP will be treated as a living document and will be updated at the end of the project to reflect the actual research data generated during the project and include updated instructions for how to access open data. Day-to-day data management and monitoring will be done using an online list in the SharePoint site that will be continuously updated to reflect actual data generation. The maintenance of this list is the responsibility of the Project Coordinator, supported by the designated data controller and Co-PI Dr Esther Phiri in MUBAS.

Table of Contents

Executive Summary	3
List of Abbreviations, acronyms and definitions	6
1 About this Deliverable	8
1.1 Why would I want to read this deliverable?.....	8
1.2 Intended readership/users.....	8
2 Introduction	9
2.1 Guiding principle.....	9
2.2 Legal Framework.....	10
2.3 Permissions for collecting and handling personal data.....	10
3 Data Summary	11
3.1 Purpose of data collection and generation.....	11
3.2 Data types, formats and size.....	11
3.3 Origin of data.....	12
3.4 SharePoint and metadata provision.....	12
3.6 Instructions for uploading datasets to SharePoint.....	13
4 FAIR Data Management	16
4.1 Making data findable.....	16
4.1.1 The CEANGAL Community in Zenodo.....	16
4.1.2 Metadata in Zenodo.....	16
4.1.3 Approach to search keywords.....	16

The research leading to these results has received funding from Irish research Council (IRC) COALESCE (Collaborative Alliances for Societal Challenges) programme, under Strand 2B: Better World Awards, in partnership with the Department of Foreign Affairs

4.1.4	Naming conventions	17
4.2	Making data accessible	17
5.1	Costs	19
5.2	Data Manager	19
6.1	Data security as specified for ITS Microsoft SharePoint	20
7	Ethical aspects	21
7.2	Summary of Ethics and Privacy Strategy	21
7.2.1	Commitment to ethical principles	21
7.2.2	Ethical Issues Checklist	22
7.2.3	Collecting personal data from pilots	22
7.2.4	Information letter and consent form	23
7.2.5	Protecting personal data collected from pilots	23
7.2.6	Using and sharing data from pilots	23
7.2.7	Managing contact information	24
8	Conclusions	25

The research leading to these results has received funding from Irish research Council (IRC) COALESCE (Collaborative Alliances for Societal Challenges) programme, under Strand 2B: Better World Awards, in partnership with the Department of Foreign Affairs

List of Abbreviations, acronyms and definitions

Table 1: List of abbreviations

Abbreviation, acronym definition	Explanation
Anonymous data	Items of information that do not allow the identification of individuals in the data material, neither directly through names or personal ID numbers, nor indirectly through background variables, a list of names, connection keys, encryption formula, or codes. However, a combination of background variables such as gender, age, location, etc. increases the risk of identifying individuals in the data material. At this stage CEANGAL assesses that withholding the person's name and contact information is sufficient to ensure the privacy of the informants, but we will need to reassess this continuously
BibTeX	A reference management software for formatting lists of references.
CA	Consortium Agreement
CSL	Citation Style Language An open XML-based standard to format citations and bibliographies.
Data controller	The institution/company/other legally responsible person who determines the purposes, conditions, and means of the processing of <i>personal data</i> . The Data Controller is a formal position and involves requirements for compliance with a number of duties in the Personal Data Act (<i>GDPR</i>).
Data processor	An external person or business that processes <i>personal data</i> on behalf of the Data Controller. The law requires that this relationship should be regulated by a data processing agreement
DMP	Data Management Plan
DOI	Digital Object Identifier
FAIR data	Findable, Accessible, Interoperable, Re-usable data
GDPR	General Data Protection Regulation EU regulation (regulation (EU) 2016/679) on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.
Gold Open Access	Open access publishing (gold open access) means that an article is immediately provided in open access mode on the publisher or journal's website. Some publishers charge Article Processing Chargers (APCs) to make articles open.
Green Open Access	Self-archiving (green open access) means that a published article or the final peer-reviewed manuscript is archived (deposited) in an online repository before, alongside or after its publication. In some cases, the author can choose to delay access to the article (embargo period).
JSON	JavaScript Object Notation An open-standard file format.
JSON	JavaScript Object Notation An open-standard file format.
MARCXML	An XML schema based on the common MARC21 standards.

Abbreviation, acronym definition	Explanation
Personal data	<p>Personal data is any information that relates to an identified or identifiable living individual. Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data.</p> <p>Personal data that has been de-identified, encrypted or pseudonymised but can be used to re-identify a person remains personal data and falls within the scope of the law.</p> <p>Personal data that has been rendered <i>anonymous</i> in such a way that the individual is not or no longer identifiable is no longer considered personal data. For data to be truly anonymised, the anonymisation must be irreversible.</p>
Processing personal data	Processing personal data entails collecting, registering, storing and/or putting together and handing out personal data.
RES	Renewable Energy Systems
SSL/TLS	Secure Sockets Layer / Transport Layer Security These are protocols offering secure communication on the internet.
Zenodo	Zenodo is a catch-all research data repository that enables researchers, scientists, EU projects and institutions to share research results, make research results citable, and search and reuse open research results from other projects. Zenodo is hosted by the CERN cloud infrastructure.

1 About this Deliverable

1.1 Why would I want to read this deliverable?

It provides an easy overview of research data the project is expected to generate, the types and formats of this data, and how this data is processed and stored to make them findable, accessible, inter-operable and re-usable, according to the principles of FAIR data management. The purpose of the DMP is to contribute to good data handling during the project's lifetime, and to describe how such data will be curated and preserved.

1.2 Intended readership/users

Internally in the project:

- All project participants who are responsible for, or in any way involved with, data collection and data handling can use this document, providing instructions on how to handle, store and process data.
- All project participants can use this document to get an overview of all data collected in the project and how this is processed and stored.

External audience:

- **Section 3, 4:** All relevant stakeholders who are interested in the CEANGAL project related activities and targeted outputs can use this document to get an overview of the data collected in the project, how to access this data, and, if applicable, how to re-use this data in their own activities.
- **Section 2, 4, 6, 7:** All persons who voluntarily participate in the pilots and contribute data to the project can use this document to learn how the project processes and store their data.

2 Introduction

2.1 Guiding principle

The guiding principle of the CEANGAL project is to be an *open* project, with almost all the deliverables produced in the project made publicly available. Only the periodic administrative and finance related reports produced by the project will not be made public during the course of the project. Figure 1 illustrates the main procedure used in the project to ensure open access to research data and publications.

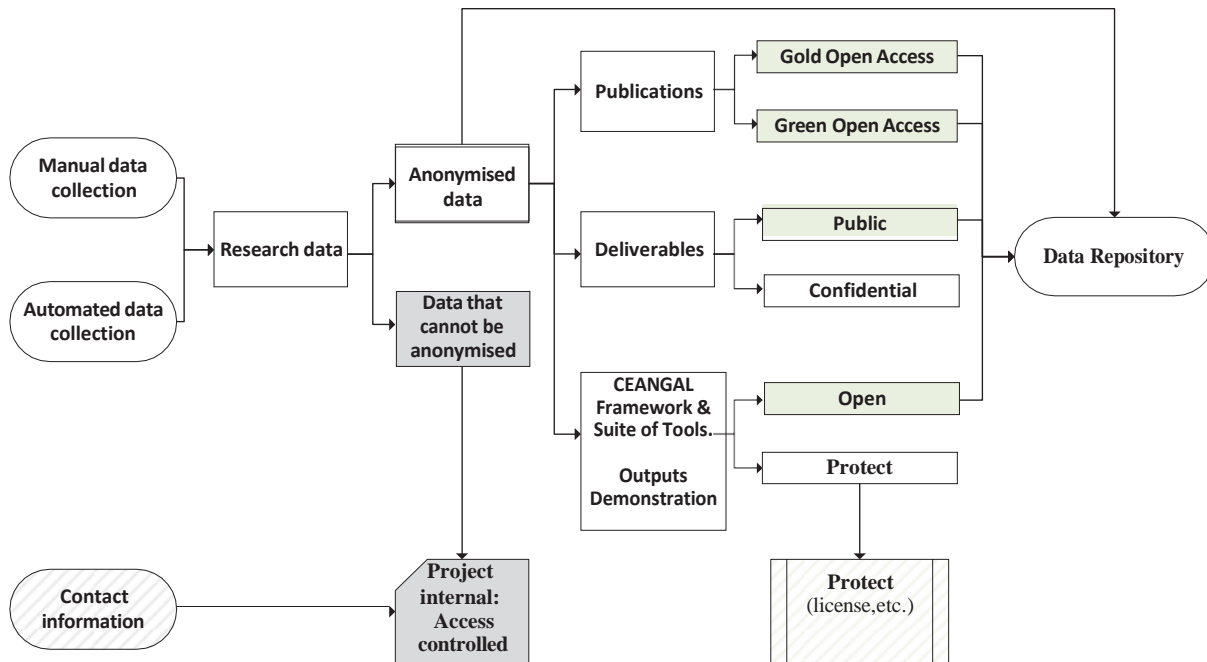


Figure 1: CEANGAL data management procedure

To protect the privacy of individual participants in the pilots, only data that can be irreversibly anonymised to the degree that it is impossible to identify individuals will be shared publicly. Non-anonymised data will be kept internally in the project and used as input to project work, but never shared publicly in its original format.

Both the anonymised and non-anonymised data will, in an aggregated format, feed into project work and provide basis for analysis in deliverables and scientific publications. If the editor of a deliverable is concerned that their deliverable contains personal information, they request a separate screening for privacy and ethics issues before submission to be sure that no personal data is included. The Ethics & Data Monitoring Manager (EDMM) employed by the project will be responsible for performing these screenings. Public deliverables, publications and anonymised datasets will be shared openly through an open research data repository.

2.2 Legal Framework

As of May 2018, the General Data Protection Regulation (GDPR)¹ is applicable in all Member States in the European Union, as well as in the countries in the European Economic Area (EEA). GDPR updates and modernises existing laws on data protection to strengthen citizens' fundamental rights and guarantee their privacy in the digital age. Even though the project is being implemented in an African country, the GDPR will be applicable to the project activities and data collection and processing.

GDPR regulates the processing by an individual, a company or an organisation of personal data relating to individuals in the EU². It does not apply to the processing of personal data of deceased persons or of legal entities. It sets down one set of data protection rules for all companies and organisations operating anywhere in the EU and European Economic Area (EEA), for two main reasons: 1) to give people more control over their personal data, 2) level the playing field for businesses and organisations operating in the EU and EEA. GDPR grant individuals a set of rights that must be protected by any actor who processes personal data. The individual rights include the right to:

- information about the processing of your personal data;
- obtain access to the personal data held about you;
- ask for incorrect, inaccurate or incomplete personal data to be corrected;
- request that personal data be erased when it's no longer needed or if processing it is unlawful;
- object to the processing of your personal data for marketing purposes or on grounds relating to your particular situation;
- request the restriction of the processing of your personal data in specific cases;
- receive your personal data in a machine-readable format and send it to another controller ("data portability"); and
- request that decisions based on automated processing concerning you or significantly affecting you and based on your personal data are made by natural persons, not only by computers. You can also have the right in this case to express your point of view and to contest the decision.

2.3 Permissions for collecting and handling personal data

No personal data is expected to be collected during the duration of the CEANGAL project. Any data that is generated during the course of the project will be continuously monitored, to ensure compliance, as part of WP7.

Nevertheless, all data collected from stakeholders in the project will be done in accordance with applicable ethical standards and requirements in the respective countries of the data collection, as well processed and handled securely and in line with applicable rules and regulations on privacy and data protection.

Data will be collected only upon receiving informed consent from the participants, and any participant providing data can at any time withdraw their participation and related data from the project.

Before any publications (e.g. scientific papers, public deliverables) is released to the public, it will go through an ethics and privacy screening to ensure that all data included is anonymised, aggregated and/or analysed in such a way as to ensure that none of the content can be traced back to an individual participant or respondent. The EDMM is responsible for these checks under WP 7.

¹ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679>

3 Data Summary

Appendix A provides a list of all datasets currently expected to be generated in the CEANGAL project and their planned accessibility. We recognise that this list will develop and grow as the project evolves.

3.1 Purpose of data collection and generation

The overall motivation for data collection in CEANGAL project is to facilitate knowledge generation, evaluations and facilitate learning. CEANGAL aims to use the solutions generated from the analysis of such data to empower communities to make the transition to implementing sustainable decentralised renewable energy systems with innovative business models, decision support tools and knowledge tools. This involves evaluating the proposed solutions in pilots to test their effect, as well as gathering input and feedback from citizens and users in order to improve solutions, increase user acceptance, encourage behavioural change, and achieve more optimal renewable systems uptake by such disadvantaged communities to afford energy accessibility.

Only data that is needed to perform project activities will be collected, and as far as possible, participants will not be asked to provide personal data unless absolutely necessary.

3.2 Data types, formats and size

Types of data

Some of the data the project could potentially collect and generate is classified as personal data, such as names, residence community of participants, family sizes and consumption patterns. This data must be irreversibly anonymised before being made public. The CEANGAL project will anonymise all of such data.

If such data cannot be irreversibly anonymised, it will remain confidential and only managed by designated Data Controllers in the project. Non-anonymous data, although not openly shared in the project or beyond, can still provide input to deliverables and publications, but only analysis of the aggregated data, which cannot be linked to individual participants, will be made public.

The data collected in CEANGAL can be split into the following three categories (details on datasets provided in appendix A):

1. Manually collected data:
 - A. Data on citizens and community potential energy consumption and energy use
 - B. Data on potential community user acceptance
 - C. Demographic data (e.g. gender, family size, RES ownership etc)
2. Data automatically collected through technology:
 - A. Data from RES (i.e. KWh generated from solar or wind systems)
 - B. Coz emissions reduction from potential use of RES
3. Contact Information
 - A. Project partner representatives
 - B. Project external communities and individuals who voluntarily participate in the project and pilot activities

Data will be organised in datasets relating to the category of the data and site of collection.

Data formats

A dataset can include different types of formats. As an example, a manually collected dataset concerning user acceptance can consist of both written interview notes, audio files from interviews, pictures from pilot sites, and survey responses. Some of this data cannot be anonymised within the scope of this project (e.g. audio files), so in most cases only parts of a dataset will be made openly available. Concerning the automatically

collected data the project expects to deliver these datasets anonymised as open research data.

The CEANGAL project will only use widely accepted formats for data generation, such as:

- Documents/Reports/Publications: .PDF/A, txt, doc/docx
- Spreadsheets: .xls/.xlsx
- Databases: .csv
- Audio files: .mp3, .wav, .wma, .ra
- Pictures: jpg, png
- Video: avi, flv, mov, mp4, wmv

3.3 Origin of data

CEANGAL will collect data initially mainly from selected collaborative potential pilot sites in Malawi, with the country forming the basis of the development of the project solutions, which is intended to be further expanded for use in other low income countries in the Sub-Saharan African (SSA) region. Desktop data will also therefore be generated for similar communities in SSA regions where applicable.

Depending on the type of data, there will be various methods and origins of data collection involved at each site.

For manually collected data the main origins will be:

- Interviews with groups and individual participants in the pilots at each site
- Feedback from participants at stakeholder workshops
- Survey responses
- Market survey
- Literature study/review and open data (re-use of existing data)

For automatically collected data the main origins will be:

- Results from modelling and simulation software's
- Electricity generation from RES
- Price of energy sources in the open market
- Current bank loan and interest rates

3.4 SharePoint and metadata provision

All anonymised datasets will be stored in an ITS SharePoint project site. This will be the project's online collaboration platform during the project lifetime, and for up to 6 months after the end of the project for final closing activities. Data Controllers at each pilot site will be responsible for uploading their public datasets to SharePoint. All datasets will use standard SharePoint version control.

The non-anonymous datasets will be stored locally by the Data Controllers and not shared with others, with the exception of project generated contact lists which will be stored in a strict access-controlled SharePoint folder.

The following list describes the metadata that will be provided for each dataset:

- | | |
|---------------|----------------------------------|
| • File name | WP (Work Package) number |
| • Date | • Responsible person |
| • Version | • Lead partner / Data Controller |
| • File type | • Dissemination level |
| • Description | |

3.5 Zenodo

The CEANGAL project will use the open research data repository *Zenodo* to comply with the IRC Open access recommendation, which is in line with the EU Horizon H2020 Open Access Mandate². All scientific publications, including public deliverables and public parts of underlying datasets will be uploaded to the CEANGAL *Community*³ in Zenodo.

Zenodo is a "catch-all" open research data repository which gathers research data across all disciplinary fields. It is for non-military purposes only, and the repository is hosted and managed by CERN. All data deposited to Zenodo is stored securely in the CERN Data Centre's cloud infrastructure⁴ (see section 6.2).

3.6 Instructions for uploading datasets to SharePoint

Table 2 and Figure 2 details the instructions to project participant on how to upload datasets to SharePoint and Zenodo.

Table 2: Instructions for uploading datasets

Upload instructions - CEANGAL Sharepoint Site
<ul style="list-style-type: none">• Please upload all public datasets to this folder in the CEANGAL Sharepoint site:<ul style="list-style-type: none">○ <i>Research Data</i><ul style="list-style-type: none">▪ There will be one sub-folder per pilot site• Use this naming convention (for details see 4.1.4):<ul style="list-style-type: none">○ <i>Descriptive text IRC_CEANGAL_DeliverableNumber_UniqueDataNumber</i>○ <i>Descriptive text IRC_CEANGAL_PublicationNumber_UniqueDataNumber</i>• Be sure to use the same file name when uploading later versions• Register mandatory metadata on your data set by adding a new item to this list, located in the same folder. This list will also generate a Unique Data Number for your dataset:<ul style="list-style-type: none">○ <i>CEANGAL Research Data</i>
Upload instructions - Zenodo
<ul style="list-style-type: none">• Scientific publications, public deliverables and public datasets must also be uploaded to the CEANGAL Community in Zenodo. To do this you must complete the following steps:<ul style="list-style-type: none">○ Create a profile in Zenodo to be able to upload files○ Click on the CEANGAL link above, or search for "CEANGAL" under the "Communities" tab at the top of the Zenodo site○ On the Community site, click the green "New upload" button in the top right corner○ Enter requested data and confirm the upload. The information requested is located in the metadata list on SharePoint (<i>CEANGAL Research Data</i>)• Uploading should be done as soon as possible and at the latest on article publication. Data Controllers are responsible for uploading datasets generated by them.

² http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

³ <https://zenodo.org/communities/ceangal/>

⁴ <https://zenodo.org>

The research leading to these results has received funding from Irish research Council (IRC) COALESCE (Collaborative Alliances for Societal Challenges) programme, under Strand 2B: Better World Awards, in partnership with the Department of Foreign Affairs

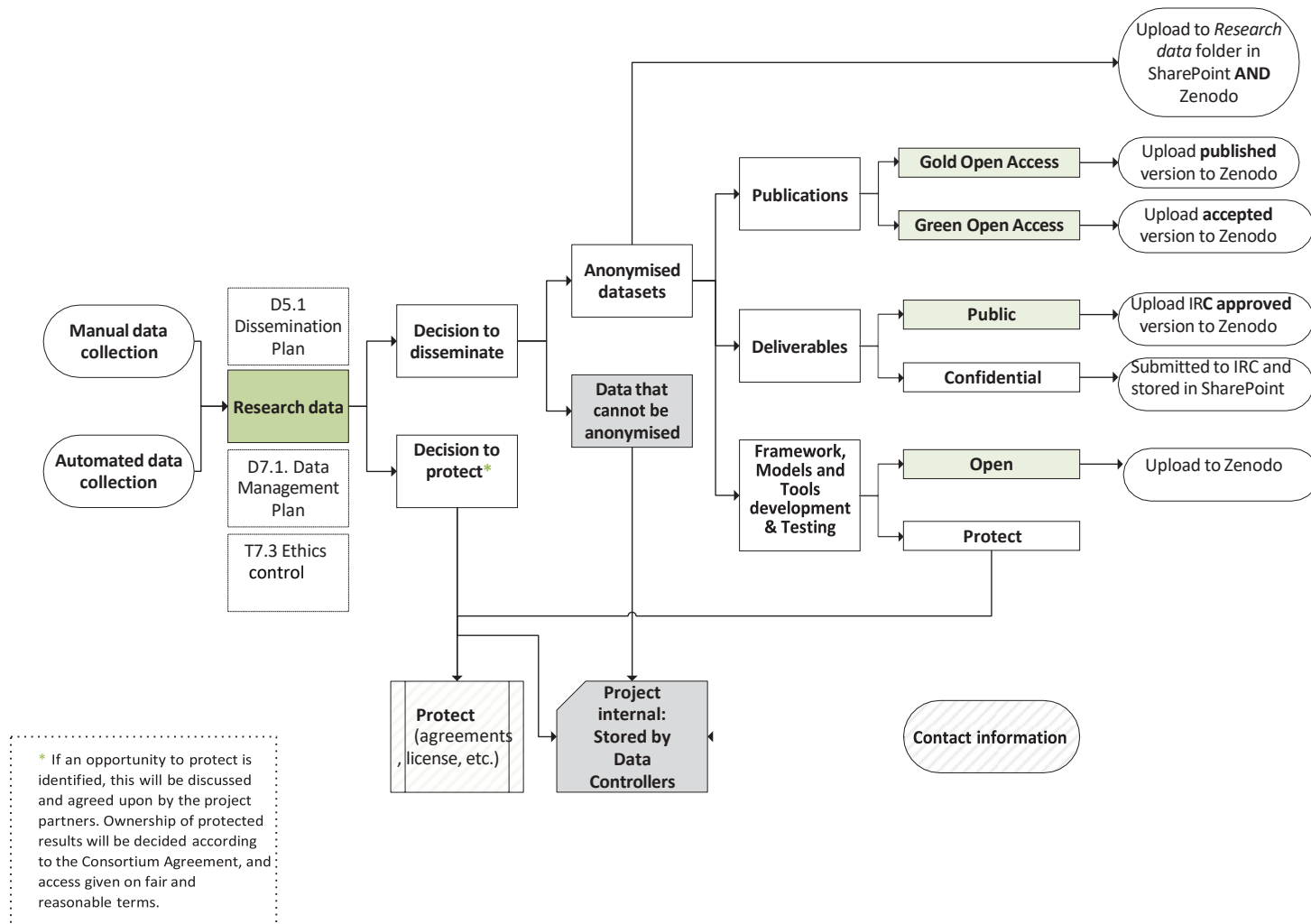


Figure 2: Process for uploading datasets

4 FAIR Data Management

CEANGAL will manage data in accordance with the principles of **FAIR data management**⁵ (Findable, Accessible, Interoperable and Re-usable data). The project aims to maximise access to, and re-use of research data generated by the project. At the same time, there are datasets, or parts of datasets, generated in this project that cannot be shared in order to protect the privacy of voluntary participants in the pilots.

4.1 Making data findable

4.1.1 The CEANGAL Community in Zenodo

The CEANGAL project will use the Zenodo repository as the main tool to make our research data findable in accordance with the IRC Open Access recommendation.

A CEANGAL community has been established on the Zenodo website, and the project will upload all our public datasets and deliverables as well as scientific publications to this community. All uploads will be enriched with standard Zenodo metadata, including Project Acronym and number. Zenodo provides version control and assigns DOIs to all uploaded elements.

4.1.2 Metadata in Zenodo

Metadata associated with each published data set in Zenodo will by default be as follows:

- Digital Object Identifiers
- Version numbers
- Bibliographic information
- Keywords
- Abstract/description
- Associated community
- Associated publications and reports
- Language

In addition, we will add the project name and number

4.1.3 Approach to search keywords

The Data Controllers will be responsible for uploading public datasets that they have generated and assigning specific keywords relevant to these datasets. Dataset specific keywords must be descriptive to the content of the dataset. E.g., a dataset containing information on RES installation must be tagged with corresponding keywords such as, "*RES community installation*". In addition, the project has defined a set of general keywords that should apply to all public datasets, scientific publications and public deliverables. These are as follows:

- Electricity Accessibility
- Decentralised systems
- Renewable Energy Systems

⁵ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
The research leading to these results has received funding from Irish research Council (IRC) COALESCE (Collaborative Alliances for Societal Challenges) programme, under Strand 2B: Better World Awards, in partnership with the Department of Foreign Affairs

4.1.4 Naming conventions

Datasets will be named using the following naming conventions:

DS_DataCategoryNr_DataController_Description_Acronym_UniqueDataNr

Explanation of the naming convention:

- "DS" stands for dataset
- "DataCategoryNr" refers to the list of data categories described in section 3.2:
 - 1 = Manually collected data
 - 2 = Automatically collected data
 - 3 = Contact information
- "DataController" refers to the short name of the partner/Data Controller who is responsible for the dataset. CEANGAL has two Data Controllers, one in ITS and the other in MUBAS.
- "Description" refers to a *short* description of the content of the dataset (see example)
- "UniqueDataNr" is the number automatically generated by the research metadata list in SharePoint (

Example of dataset name: *DS_2_MUBAS_User-RES-AcceptanceE_CEANGAL_0003*

4.2 Making data accessible

The Open Access Mandate complied with by the CEANGAL project aims to make research data generated during the project implementation accessible with as few restrictions as possible, but also accept protection of personal or sensitive data due to privacy concerns and/or commercial or security reasons.

All public datasets, scientific publications and deliverables will be uploaded to Zenodo and made openly available, free of charge. Publications and underlying data sets will be linked through use of persistent identifiers (DOI versioning).

4.3 Making data interoperable

Zenodo uses JSON schema as the internal representation of metadata and offers export to other formats such as Dublin Core, MARCXML, BibTeX, CSL, DataCite and export to Mendeley. The data record metadata will utilise the vocabularies applied by Zenodo. For certain terms, these refer to open, external vocabularies. Reference to any external metadata is done with a resolvable URL.

4.4 Reusable data

The CEANGAL project will enable third parties to access, reproduce and disseminate (free of charge for any user) for all *public* data sets, and regulate this by using Creative Commons Licences.

4.4.1 Longevity of the CEANGAL research datasets

Public (anonymous) data

For data published in scientific journals, the underlying data will be made available no later than by journal publication. The data will be linked to the publication. Data associated with public deliverables will be shared once the deliverable has been approved and accepted by the IRC. For other public datasets not directly linked to a scientific publication or deliverable, such datasets will be made available upon assessment by the Data Controllers that it is ready for publishing, and in the final month of the project at the latest.

The public data will remain reusable via Zenodo for at least 20 years. This is currently the lifetime stated by the host laboratory CERN. In the event that Zenodo has to close their operations, they have provided a

PROJECT ACRONYM:
PROJECT NUMBER:
PROJECT TITLE:
Supporting

CEANGAL
COALESCE/2021/41
Community Based Decentralised RenewAble Energy Systems and
Structures for Improving Electricity Access in Low Income Countries

guarantee that they will migrate all content (including metadata) to other suitable repositories.

Confidential (non-anonymous) data

All non-anonymous data will be deleted at the end of the project. In case permission is given by the party providing and owning the data, some non-anonymous data will be kept for a maximum of 6 months after the contractual end date of the project. The additional 6 months is to keep the underlying datasets available to allow the completion of any scientific publications being prepared towards the end of the project.

An exemption is pictures and videos, taken with consent from voluntary pilot participants, that are used for communication purposes. If consent is *not* withdrawn at an earlier time, such data will be kept for up to 4 years after the end of the project in order to comply with any contractual obligation to continue dissemination and exploitation activities after the project ends. If a party withdraws the consent to use this material (pictures, videos), it will be deleted without delay.

5 Allocation of resources

5.1 Costs

CEANGAL uses standard tools and a free of charge research data repository. The costs of data management activities are limited to project management costs and will be covered by allocated resources in the project budget.

Long-term preservation of the public data is ensured through Zenodo. Other resources needed to support reuse of data after the project ends will be solved on a case-by-case basis.

5.2 Data Manager

The overall responsibility for data management lies with the project coordinator, Dr Ehiازه Ehimen from ITS.

Supporting the coordinator is a data management team consisting of the Data Controllers for each partner country (Ireland and Malawi), overseen by the EDMM.

6 Data security

In this chapter, the security features of the research data infrastructure used to store and handle data in the CEANGAL project are described.

6.1 Data security as specified for ITS Microsoft SharePoint

The ITS SharePoint is the online collaboration platform used the CEANGAL project. A dedicated project site has been established on this platform, accessible only by the partner representatives in the consortium. Furthermore, a dedicated folder for research datasets is set up, allowing for stricter access control than the main project site. Only anonymous datasets will be uploaded to this SharePoint folder.

The CEANGAL SharePoint site has the following security settings:

- Access level: Restricted to persons (project members only). Further access restrictions on specific folders is enabled.
- Encryption with SSL/TLS protects data transfer between partners and the ITS SharePoint site.
- Threat management, security monitoring, and file-/data integrity prevents and/or registers possible manipulation of data.

Documents and elements in the ITS SharePoint sites are stored in Microsoft's cloud solutions, based in Ireland and the Netherlands.

Nightly back-ups are handled by ITS's IT department. As a baseline, all project data will be stored for 5 years according to ITS's ICT policy, unless otherwise agreed in contracts and data processing agreements.

6.3 Data security as specified for Zenodo

The following list describes the security settings for Zenodo:

- Versions: Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and records are preserved.
- Replicas: All data files are stored in the CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- Retention period: Items will be retained for the lifetime of the repository. The host laboratory of Zenodo CERN, has defined a lifetime for the repository of the next 20 years minimum.
- Functional preservation: Zenodo makes no promises of usability and understandability of deposited objects over time.
- File preservation: Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
- Fixity and authenticity: All data files are stored along with an MD5 checksum of the file content.
- Files are regularly checked against their checksums to assure that file content remains constant.
- Succession plans: In case of closure of the repository, a guarantee has been made from Zenodo to migrate all content to suitable alternative institutional and/or subject based repositories.

7 Ethical aspects

7.1 Legal aspects

The proposed work in CEANGAL even though also operational in an African country, will fully comply with the regulations set out in the GDPR. In addition, CEANGAL will comply with the principles of the European Charter for Researchers, the European Code of Conduct for Research Integrity, including ethical standards and guidelines, regardless country in which research is carried out.

Nothing in this project shall be deemed to require a party to breach any mandatory statutory law under which the party is operating. This includes any national or European regulations, rules and norms regarding ethics in conducting research.

7.2 Summary of Ethics and Privacy Strategy

7.2.1 Commitment to ethical principles

All project partners are obliged by European (GDPR) and national laws to protect personal data.

The project consortium is aware that, because of the importance of the primary subject matter, due regard must be given to ensuring that the project and its outputs comply with the highest ethical standards.

However, as previously outlined, it is important to note that, following an extensive review, the project consortium has identified no ethical issues, and it is the considered opinion that there will be no ethical issues identified as the project progresses. The ethical measures implemented by the project is therefore to ensure that best practice is complied with, and that any ethical issue that may potentially arise can be effectively dealt with.

While community human participants involvement will be key in the realisation of the community electrification objectives of the project, the information which is derived from interactions with such participants does and is not expected to represent an ethical issue. Any information gathered will be anonymised, and no information of a personal nature will be aimed for collected. Full compliance in terms of data protection, informed consent and professional standards will be ensured.

The coordinator of the CEANGAL project, will ensure the project follows ethical guidelines in its work, and *all* work conducted is subject to the clearance granted by the ITS Ethics Council.

At the start of the project implementation period, the project coordinator formally applied for a research ethical clearance from ITS informing the board of the project objectives, methodologies and the potential ethical issues related to the project. No ethical concerns were identified. *The outcomes of the ITS ethics council will be appended to this document once received.*

The CEANGAL project will therefore comply with all ITS ethics regulations, and all aspects associated with such guidelines. Important aspects with respect to this are:

- The ethical guidelines are based on the vision of using science and technology to create a better society and are reviewed continuously to ensure they stay up to date with developments in society and the challenges of today. They generally fall into these categories: research ethics, business ethics, and ethics in interpersonal relationships.
- All ITS's employees are expected to act in accordance with the ethical guidelines and principles. As coordinator of the CEANGAL project, ITS will ensure that any ethical issues, which may arise, will be handled appropriately and in a transparent and fair manner

7.2.2 Ethical Issues Checklist

- A.** Human embryo and foetuses: **Not relevant, no issues.**
- B.** Humans
- B.1. Does your research involve human participants? **The project deals mainly with data generated from community groupings, no specific individual participants are targeted, but will be involved mainly in association with larger groupings.**
- B.2. Are they volunteers for social or human sciences research? **NO**
- B.3. Are they persons unable to give informed consent (including children/minors)? **NO**
- B.4. Are they vulnerable individuals or groups? **NO**
- B.5. Are they children or minors? **NO**
- B.6. Are they patients? **NO**
- B.7. Are they healthy volunteers for medical studies? **NO**
- C.** Human cells/tissues: **Not applicable, no issues**
- D.** Personal data
- D.1. Does your research involve personal data collection and/or processing? **No, only aggregated anonymised data is collected**
- D.2. Does your research involve further processing of previously collected personal data ('secondary use')? **No, no personal data is to be used.**
- E.** Animals? **No issues.**
- F.** Environment and Health and Safety
- F.1. Does your research involve the use of elements that may cause harm to the environment, to animals or plants? **No**
- F.2. Does your research deal with endangered fauna and/or flora and/or protected areas? **No**
- F.3. Does your research involve the use of elements that may cause harm to humans, including research staff? **No**
- G.** Dual Use:
No issues
- H.** Exclusive focus on civil applications
No issues
- I.** Misuse
No issues
- J.** Other ethics issues
No issues

7.2.3 Collecting personal data from pilots

Data collection activities (interviews, surveys, etc.) will be designed to maintain privacy. Personal data will not be requested unless this is absolutely necessary. Vulnerable groups like minors and individuals unable to freely provide an informed consent will be excluded. Participation is voluntary. Participants will be given the possibility to decline and withdraw their participation at any time.

The research leading to these results has received funding from Irish research Council (IRC) COALESCE (Collaborative Alliances for Societal Challenges) programme, under Strand 2B: Better World Awards, in partnership with the Department of Foreign Affairs

The CEANGAL project will collect pictures and video for use in communication activities (website, newsletter, social media). Pictures and video can contain personal data if an individual is the focus of the image or video. Examples include: 1) pictures/video of individuals; 2) pictures/video of individuals or community members posted on the project website along with biographical details i.e. names; 3) individual images published in a newsletter.

Examples of pictures and video that is unlikely to contain personal data are: 1) pictures/video where people are incidentally included in an image or are not the focus (e.g. at a big conference/workshop); 2) images of people who are no longer alive (the GDPR only applies to living people).

When collecting pictures and video CEANGAL will follow established guidance and best practice on collecting and processing such data to ensure that we adhere to the legal requirements (e.g. guidance established by the University of Reading, UK⁶). Under no circumstances will pictures containing personal information be publicly shared without the subject's explicit consent.

7.2.4 Information letter and consent form

All community participants in the project will be given an information letter and a consent form (on paper or electronically). The information letter will provide information about:

- The type of data that will be collected during the study.
 - How the data will be collected (interview, automatic data collection, etc.)
 - What the data will be used for. The information letter will explain the purpose of the project and the expected results. It will also be explained that published information always will be anonymous, and that no personally identifiable information will be published in any way.
 - How the data collected will be handled. The information letter will explain that personal data will be treated in full confidentiality and will be registered and stored in a secure manner. The data will be de-identified before it is processed (name or other characteristics serving to identify person will be replaced by a number and the list of identifiers will be kept separate from the data).
 - Who will have access to the data. The information letter will state that data will be handled by a very limited number of authorised personnel and that confidentiality will be regulated by legal agreements. The data will be de-identified before it is discussed and processed within the project.
 - The rights of the participants. The information letter will state that participation is voluntary and that participants have the right to see the data collected about them and that they can withdraw from the study at any time without any obligation to explain their reasons for doing so (contact information for such requests will be provided).

7.2.5 Protecting personal data collected from pilots

Personal data will be handled in accordance with European legislation on privacy (GDPR). Under no circumstance will the deliverables or processes compromise the individual right to privacy and satisfactory handling of personal data.

It is the intention of the project that no personal data is collected during the course of the project's activities, but in the cases where personal data is acquired, all personal data will be stored on secure servers with access control managed by the Data Controllers. Such personal data will be handled by authorised personnel, and no one will have access to the data unless this is necessary to carry out the project work.

7.2.6 Using and sharing data from pilots

Analysis of the data (e.g. in evaluations) will be carried out only on de-identified and anonymised data.

At the end of the project, all personal data (audio and video files included) will be deleted, and the de-

⁶ <https://www.reading.ac.uk/internal/imps/DataProtection/DataProtectionRequirements/imps-d-p-photographic.aspx>
The research leading to these results has received funding from Irish research Council (IRC) COALESCE (Collaborative Alliances for Societal Challenges) programme, under Strand 2B: Better World Awards, in partnership with the Department of Foreign Affairs

identified data will be completely anonymised, meaning that the links to the lists of keys will be deleted. No personal data will be stored after the end of the project, unless explicit consent to do this is given by the provider/owner of the data. If such permission is given, non-anonymous data will be stored for a maximum of 6 months after the contractual end of the project (to allow for finalisation of scientific publications).

For other non-anonymous data, such as pictures and videos used for project communication activities, these will be kept for up to 4 years after the end of the project. Such data will be disseminated and shared, upon explicit consent only, through the project website, newsletters, and social media. If a party withdraws the consent to use this material (pictures, videos), it will be deleted immediately without delay.

The anonymous data will be documented and archived in a research data repository as open research data, and thus placed at the disposal of colleagues who want to replicate the study or elaborate on its findings.

Any publications, including publications online, neither directly nor indirectly will lead to a breach of agreed confidentiality and anonymity.

The research outcomes will be reported without contravening the right to privacy and data protection.

7.2.7 Managing contact information

Some of the contact information to external parties will be totally curated and preserved by one partner (MUBAS) located in the pilot country for the project. The project partners ITS and MUBAs will use and further develop their pre-existing contact lists for dissemination and communication purposes. These contact lists will not be shared within the project, but they will be managed according to GDPR by these partners.

Contact information for other external actors established during the course of the project i.e. the database of potential expansion research institutions and communities as targeted in D6.2. List of future potential CEANGAL SSA regional hubs and adoption communities, will be managed within the project in accordance with GDPR. All project generated contact lists will be stored in the CEANGAL SharePoint project site hosted by ITS. Access control will be implemented to ensure that only those who require this information to perform their activities can access it. Access will be managed by ITS.

8 Conclusions

Formal approval and release of this deliverable (D7.1 CEANGAL Data Management Plan) within the consortium constitutes a formal commitment by partners to adhere to the data management strategy and the procedures it defines. When the deliverable is formally approved by the Irish Research Council (IRC), this will constitute confirmation that the procedures are considered by the Council to be adequate.

As coordinator of the CEANGAL project, ITS will ensure that any data management issues which may arise during the project will be handled appropriately and in a transparent and fair manner.

The DMP is a living document that will expand as the project evolves and new information on data collection, generation and handling arise. Day to day data management will happen through the online tools described in this document, and through continuous collaboration between the coordinator, the Malawian co-PI, Data controllers and the EDMM . A revised and extended version of this DMP will be prepared towards the end of the project to reflect the current status of data management in the project.