D1.2 Data Management Plan Year 1

Organisation: MicroEnergy International GmbH

October 2024

Settings	Value
Milestone Title	D1.2 Data Management Plan – Year 1
Work Package Title	WP1 – Project management and coordination
Description	This deliverable describes the project data management policy and specifies any data sets that may endure embargo or be kept from publication, taking into consideration any issues of ethics, privacy, confidentiality, and IPR protection. MEI will ask partners to prepare and treat datasets produced in accordance with the DMP and the standards for metadata, repositories, archiving, and presentation identified therein. The DMP will be delivered first in M6, and updated versions will be submitted in M24 and M48. This deliverable corresponds to Task 1.3 Open Research Data Management.
Lead Beneficiary	MEI
Lead Authors	Dr. Raluca Dumitrescu (MEI)
Contributors	Carlos Guerrero Lucendo (VITO), Emilie Martin (WUP)
Doc. Version (Revision number)	0.5
Date:	16.10.2024

Document Approver(s) and Reviewer(s):

NOTE: All Approvers are required. Records of each approver must be maintained. All Reviewers in the list are considered required unless explicitly listed as Optional.

Name	Role	Action	Date
Shadrack Omwenga (MEI)	approver	< Approve>	18.10.2024
Emilie Martin (WUP)	reviewer	< Review>	14.10.2024
Carlos Guerrero (VITO)	reviewer	< Review>	14.10.2024

Document history:

The Document Author is authorized to make the following types of changes to the document without requiring that the document be re-approved:

- Editorial, formatting, and spelling
- Clarification

To request a change to this document, contact the Document Author or Owner.

Changes to this document are summarized in the following table in reverse chronological order (latest version first).

Revision	Date	Name	Short Description of Changes
1	08.10.2024	v0.4	Feedback on storage and sharing clarified; spellcheck corrections; summary of deliverable added; data summary added; preliminary data overview added
2	16.10.2024	V0.5	Final formatting and spellcheck

Configuration Management: Document Location

The latest version of this controlled document is stored in here.

Nature of the deliverable		
DMP	Data Management Plan	

Dissemination level		
PU	Public	

ACKNOWLEDGEMENT

SWARM-E is a trans- and multi-disciplinary approach for sustainable, affordable and modern energy access and well-being for Sub-Saharan Africa, aligned with the AU-EU Agenda 2063.

SWARM-E consists of several layers: 1) an innovative renewable electricity infrastructure, the SWARM grid, a circular and cyber-smart network where end-users exchange electricity of their solar home systems and form the nodes of a smart grid which can dynamically grow to meet demand; 2) unlocking unutilised renewable energy for productive uses in the water energy food nexus – cold storage, water purification, water pumping and irrigation, carpentry; 3) transfer and decentralisation of Global North energy transformation innovations – decentralised hydrogen production for cleaner cooking, bi-directional charging of light electric vehicles (two- and three-wheelers) to transport goods and people. SWARM-E builds on network effects generated through the inclusion of localised economies with strong producer-consumer linkages embedded within larger systems of trade and exchange for the creation of bottom-up energy communities.

SWARM-E will operate and replicate 5 pilots in Rwanda and Tanzania, under which 5 SWARM grids are installed, delivering 6.9 GWh of renewable electricity while generating income through the trading of electricity and avoiding the discard of 3,200 batteries; 5 water purification applications deliver 101.M L of clean water; 15 light electric vehicles deliver farmers' produce, power mobile productive uses and cold storage, increasing the yields of 1,000 farmers and reducing the food losses of more than 5,000; 700 kg of H_2 blended with LPG for cleaner cooking, and more than 500 jobs for women and youth to be created.

The balanced participation of EU and AU private, public and civil society organisations in the consortium will ensure the knowledge transfer North-South and South-South, and the sustainability of value chains based on local value creation and entrepreneurship.

More information on the project can be found at: www.swarm-e.eu

COPYRIGHT

Contents

Summary of the deliverable

The Data Management Plan (DMP) is a key element for appropriate data management. It includes producing, collecting, or processing of research and commercial data. Under SWARM-E, several types of data are going to be collected and processed. Guidelines on FAIR data management under Horizon programme are followed. Accessibility, interoperability and reusability of data will be assured within the SWARM-E project. The DMP describes the data management life cycle for the data to be collected, processed and/or generated under SWARM-E.

The DMP will be updated over the course of the project whenever significant changes arise. The next update or evaluation of the DMP will be conducted in month 18, at the time of the periodic evaluation and reporting of the project.

1. Data summary

The data collected in scope of the SWARM-E project was to a large extent pre-defined in the working program of the Grant Agreement (Work Package Description). The data collection is focused and limited to data required to fulfil project aims and objectives. Below follows an overview of the types of data collected, the methodology for collection, dissemination level and utility within the scope of the project. This overview will be updated whenever significant changes arise.

WP	Name dataset	Description	Format	Origin/Collection methods	Dissemination level	Comments/Utility
1	Financial and administrative data	Financial and technical reports containing data for the project management, including involved staff members.	.xls, .pdf, .dox	Filling in technical and financial reports	SEN	-
2	Health facility electrification	Information on existing electricity sources in health facilities in pre-selected sites, including need for energy based on existing appliances, provision of services and size of facility. Also data on basic technical specifications of devices for energy generation, storage, and usage.	.xls	Survey with health facilities via SurveyCTO and in-person questionnaires	SEN	Raw data from questionnaires; to establish if the health facilities could be interconnected to the SWARM grid and under which operational and financial model
2	Education facilities electrification status	Information on existing energy generation and storage, uses and needs in pre-selected sites. Also data on basic technical specifications of devices for energy generation, storage, and usage.	.xls	Survey with education facilities via SurveyCTO and in-person questionnaires	SEN	Raw data from questionnaires; to establish if the education facilities could be interconnected to the SWARM grid and under which operational and financial model
2	Households energy needs and demand	Socio-economic, energy access, use and needs information of households in pre-selected sites. Also data on basic technical specifications of devices for energy generation, storage, and usage.	.xls	Survey with households via SurveyCTO and in-person questionnaires	SEN	Raw data from questionnaires to identify and quantify sites' energy access and demand, define baseline and determine if they can be interconnected.
2	Small and Medium Sized Businesses energy needs and demand	Socio-economic, energy access, needs and use information of SME in pre-selected sites. Also data on basic technical specifications of devices for energy generation, storage, and usage.	.xls	Survey with SME via SurveyCTO and in-person questionnaires	SEN	Raw data from questionnaires to identify and quantify sites' energy access and demand, define baseline and determine if they can be interconnected.
2	Suppliers of electric appliances, solar systems	Information about distribution and prices of energy generation equipment and electrical appliances, as well as better understanding of suppliers' general approach in pre-selected villages.	.xls	Survey with technology suppliers via SurveyCTO and in-person questionnaires	SEN	Raw data from questionnaires to identify available energy-related equipment and appliances.
2	Village leader	High level socio-economic, geographical and energy access information of pre-selected sites	.xls	Survey with the village leader via SurveyCTO and in-person	SEN	Raw data from questionnaires identifying socio-economics and energy-related information of pre-

WP	Name dataset	Description	Format	Origin/Collection methods	Dissemination level	Comments/Utility
				questionnaires		selected sites, to define baseline indicators.
2	Cooperatives	General information of operational cooperatives and their relation to energy projects in pre-selected sites.	.xls	Survey with cooperatives via SurveyCTO and in-person questionnaires	SEN	Raw data from the questionnaires to identify the productive groups that have clusters, to find out about activities and to be able to carry out further actions through them, such as working groups.
2	NGO	General information of operational NGO and their relation to energy projects in pre-selected sites.	.xls	Survey with NGO via SurveyCTO and in-person questionnaires	SEN	Raw data from the questionnaires to identify the NGOs on the ground in the pre-selected sites, to know their activities, especially those related to energy or training, in order to be able to align further project activities.
2	Mobile Money	Identification and identification of providers of Mobile Money services in pre-selected villages.	.xls	Survey with mobile money agents via SurveyCTO and in-person questionnaires	SEN	Raw data from questionnaires to identify the existence of mobile money services in pre-selected villages and assess their potential adaptation into project operations
3	Barriers and CB needs - Consortium partners	Qualitative data on anticipated barriers, threats, opportunities to the pilot's implementation, and capacity-building needs	.docx	Online interviews with selected consortium partners	SEN	-
3	Barriers and CB needs - External partners	Qualitative data on anticipated barriers, threats, opportunities to the pilot's implementation, and capacity-building needs	.docx	Online interviews with selected external partners	SEN	Tbd with WP3 partners if replaced or completed by surveys applying a Delphi method to rank barriers and CB needs
3	Focus groups	Qualitative data on 1) perceptions on the swarm approach 2) CB needs 3) gender-inclusive pilot & CB activities 4) youth	.docx	Focus groups with the community, female representatives of the community (women-only FG) and youth	SEN	Information needed to establish the operational model, ownership and capacitation of end-user

WP	Name dataset	Description	Format	Origin/Collection methods	Dissemination level	Comments/Utility
3	Value chains and PUE	Qualitative data on market access of existing and new value chains for PUE and SWARM grid components	.docx	Interviews with farmers and SMEs of the selected sites	PU	Impact assessment and economic value creation through increased energy productivity
4	OGB Energy	Data from Victron (VRM Portal): energy produced & consumed, uptime	VRM Portal	Information from the Cerbo that is connected to the charge controller and inverter.	PU	Expectation of energy data - including load curves, feedback data, charging data, etc.
4	Vehicle Specification and Usage Characteristics	Technical Specifications of LEVs intended for VPP use. Also, vehicle usage patterns and characteristics	.xls, .pdf, .docx	Survey with fleet and vehicle operators	SEN	Technical specs of vehicle, batteries, motor, etc. Usage characteristics like mileage, charging habits, driving missions, etc.
4	Grid System Modeling	Output Data from system simulation	.xls	System Model	SEN	modeled/calculated load curves, energy flow data, charging and demand data
4	Pre-grid Installation	Qualitative and quantitative household/microbusiness data required to design the grid (energy demand, microutilities needed etc.)	.xls	Collected using the Mobile Data Collection (MDC) App	CO	Data is needed to understand end user energy needs + aspirational needs so that the designed grid can meet the needs required by the community
5	Pilot operation	Number of clients (by gender, age and business type), Revenues	appsheet/ google sheet	Depends on the technology used; if manual, use appsheet going into Google Sheets	CO	Monitoring and troubleshooting the energy infrastructure
5	Post Grid Installation	Quantitative data points on all grid components	.xls	Collected using the Mobile Data Collection (MDC) App and checklists	SEN	Data is required to ensure that all the grid components are working and the grid is online for smooth operations
5	Remote Monitoring	Qualitative and quantitative datapoints on energy trading, buy and sell values, mobile money amounts on the SWARM controller, system details and end user information	.xls	Data is monitored on SOLshare's backend the SOLweb, through cloud services from AWS.	CO	Data is used to monitor all the SWARM controller to make sure they are working and online, energy flows, if a SWARM controller is

WP	Name dataset	Description	Format	Origin/Collection methods	Dissemination level	Comments/Utility
						running low of mobile money and needs to be topped up and overall maintenance and troubleshooting
6	Focus groups	Discussion on potential impacts and how to enhance positive and avoid/mitigate negative impacts of the SWARM-E intervention	.docx	Online focus group discussions with experts and stakeholders	PU	Impact monitoring and assessment
6	Social impact survey	Towards the end of the project to measure actual impacts (survey would need to be conducted by local partners)	.xls	Survey with households, SMEs and village leaders via SurveyCTO and in-person questionnaires	PU	Impact monitoring and assessment
7	Finance	Financial model and business plan incl. local business ecosystem, commercial value chain and drivers	.xls, .pdf	Survey, local stakeholder interviews	SEN	The sensitive dissemination level is driven by D7.1, D7.4, D7.5 dissemination levels
8	Events materials	Presentations, statistics materials, fact sheets, pictures from partners' attending public events as guest lecturers or networking activity.	.jpeg; .tiff; .gif; .png; .pdf	Distribution of materials from the events' organizations	PU	For pictures, general rule should be that events' participants have already agreed to be taken pictures which then could be published.
8	Interviews with partners	Interviews conducted to involve consortium partners in the communication activities and to promote consortium organizations	video	Online interviews	PU	
8	Live events participation lists	Data collected during live events to keep track for reporting purposes	.xls	Collection during events	CO	Communication and dissemination KPIs
8	Online events participation lists	Data collected during online events to keep track for reporting purposes	.xls	Platforms used for e-meetings	CO	Communication and dissemination KPIs
8	Mailing lists	Mail contacts collected for internal/external communication activities (newsletters, members brief etc.)	.xls	Mailing subscriptions	SEN	Communication and dissemination KPIs
8	X analytics	Analytics of the web traffic on the project page	.xls	Collected through social channels	CO	Communication and dissemination KPIs
8	LinkedIn analytics	Analytics of the web traffic on the project	.xls	Collected through social	CO	Communication and dissemination

WP	Name dataset	Description	Format	Origin/Collection methods	Dissemination level	Comments/Utility
		page		channels		KPIs
8	Website Analytics	Analytics of the web traffic on the project page	.xls	Collected through website	CO	Communication and dissemination KPIs
8	Photographic material	Photographic evidences of activities implementation and for communication purposes	.jpeg	Collected during the implementation of project activities, prior consent	PU	-

2. FAIR data

The FAIR (Findable, Accessible, Interoperable, and Reusable data) data management principles will be ensured by SWARM-E. While some closed data will be protected and thus not publicly available, the project will maximize access and re-use of research data generated by the project.

2.1 Making data findable, including provisions for metadata

The data used in SWARM-E is anonymised, pseudonymised and de-identified. To make data findable and re-usable, the following measures are put in place:

- Location: all documents are stored in relevant folders/sub-folders in Project Folder on Google Drive;
- Naming of files: folder names are standardised to make it easier to find documents; files are stored in Archive sub-folders when out of date or when new versions are created/updated;
- Naming of deliverables: deliverables produced under SWARM-E use the following file name convention: Project acronym; Deliverable number (based on DoA); Deliverable name (based on DoA); Deliverable version (starting from 0.1);
- Reports and documents: all reports and documents will contain information on Year, Author, Subject, Version number, Key Words;
- Search functionality: Google Drive has a search functionality that enables users to navigate files, folders, documents based on date of upload, version number, type of file, location.

To maintain the accuracy and the safety of data, quality control measures will be taken. Discipline complaint metadata elements will be used describing the data to aid data discovery and potential re-use. List of metadata elements and metadata standards used are provided in a separate spreadsheet. Metadata of opened data will be made available via FAIR compliant repository for research and re-use after project closure. Persistent identifiers provided by the repository will be used in linking to datasets.

SWARM-E makes data accessible as follows:

- Public dissemination: via scientific publications and on the project's website;
- Internal data management: via an internal repository to internally collect, process, share, and make available open research data. This will allow knowledge exchange and transparency of results for all partners;
- Sharing datasets on the Zenodo repository.

Metadata

Metadata is information that makes new data usable. The metadata for the different identified datasets will be generated through manual content annotation. For published data, the metadata will feature:

- Grant Agreement number
- Abstract/description
- Access and licensing information
- Associated project and community
- Associated publications and reports

- Bibliographic information
- Digital Object Identifiers (DOI)
- Grant information
- Key words
- Language
- Version number.

2.2 Making data openly accessible

The raw datasets will only be available for the partners during the project. The data will be shared using Google Drive as an internal repository, which is web browser and application accessible. Only people with whom the repository has been shared can access the data. Logging in requires an e-mail address and password. The raw data from the research will not be publicly available.

Decisions concerning the sharing of (selected) datasets will be taken by SWARM-E General Assembly. Before sharing any datasets, the DC (Data Controller) and the PM (Project Manager), as established under D1.1 Project Management Plan, will ensure that no disclosure information is included.

As per clause 15 of the SWARM-E Grant Agreement (GA), each SWARM-E beneficiary will ensure open access (free of charge online access for any user) via gold open access routes to all peer reviewed scientific publications relating to its results. Project partners will discuss governance of open access requirements and their implementation further. Data sharing in the open domain can be restricted as a legitimate reason to protect results that can reasonably be expected to be commercially or industrially exploited. Strategies to limit such restrictions will include anonymising or aggregating data, agreeing on a limited embargo period or publishing selected datasets.

If, after project closure, permission to re-use the data is required, all requests for further use of data will be considered carefully and whenever possible approved by the Project Manager or the person mandated with the task. Permission for data use will be granted providing there are no IPR or confidentiality issues involved or any direct overlap of research questions with the primary research. Permission will be provided by contacting the Project Manager. Contact information and appropriate procedure will be provided in connection with other metadata.

In SWARM-E there is no need for a Data Access Committee, as we do not use personal data for publications, reports and deliverables and the data we get are anonymised. As the open research datasets will be included in the Zenodo repository, no methods or software tools are needed to access them and there are no restrictions on use.

2.3 Making data interoperable

Sensitive data will not be used during SWARM-E. While knowing that not all the research datasets that are being collected and that will be generated during the project's duration are public, some confidential datasets will be kept confidential and therefore will not be available for access or inter-operability. Moreover, files such as simulation models cannot be used by other partners that do not have the software license. The public results of the project will be interoperable, i.e., allow for the exchange and reuse of data between researchers, institutions, organisations, and countries. SWARM-E ensures data interoperability through the consistent use of common, standardised file formats. Using standard Office Software will enable data exchange and re-use between researcher, institution,

organizations, stakeholders and any interested parties. Nevertheless, some of the data that will be publicly available and put in open access repositories like Zenodo, can be interoperable. JSON (JavaScript Object Notation) scheme is used by Zenodo for metadata. It is possible to export research data into other formats such as: BibTex1, CSL2, DataCite, Dublin Core, DCAT, JSON3, JSON-LD, GeoJASON, MARCXML4, and export to Mendeley, Citavi.

2.4 Increase data re-use

Ownership of datasets will belong to the project consortium after the project completion. Creative Commons licence CC-BY-SA or CC-BY will be used for any opened datasets unless there are compelling reasons to select a more restricted type of CC-licence. Creative commons licences will by default also include a disclaimer of liability for the re-use of opened data.

No definite period or time limit is planned for access or re-use of the data. However, the opened data will be deposited in a repository that guarantees data integrity on the bit level. At this point, no continuous data curation policy to guarantee full long-term digital preservation of datasets is planned. Justification for possible case-specific embargo for published data will be decided by project consortium. Embargo will be sought primarily in connection with any potential patent application based on project results or any commercially sensitive data.

Some of project's work may re-use (aggregate, synthesise or analyse) existing materials, publicly available datasets (e.g., figures, tables, quotations) from existing literature (academic, policy or other documents); in particular from other projects' results. In such cases, they will be properly referenced and acknowledged, and any necessary permissions for re-use will be obtained. SWARM-E will use literature (both academic, industry stakeholders and press articles) relevant to the tasks.

Public deliverables developed during the project will be accessible via SWARM-E website and the institutional open access repositories. Intellectual property issues will be re-visited (if needed) in the next update to this plan and the final version of the DMP.

3. Allocation of resources

The responsibility for adhering to the guidelines set out by this Data Management Plan lies in particular with the Project Coordinator but also with Work Packages Leaders and the individual partners in the SWARM-E consortium.

To ensure that the guidelines in this document are clear and sufficient, and that they are followed, there will be frequent consultations with the Project Coordinator, Project Manager and Work Packages Leaders. The SWARM-E DC provides professional advice on data protection and educates staff in the field of data protection compliance requirements. The DC also responds to the questions and inquiries of data subjects and is the point of contact for the data protection supervisory authority.

During the project implementation, the consortium partners will be responsible for managing and curating datasets at their possession. At the project ending, the project General Assembly will mandate the Project Manager or the Data Controller to take care of long-term preservation and sharing of datasets. The costs of data management are allocated in the project budget, as most partners have foreseen budget for Open Access fees, so no further costs are envisaged for such activity.

4. Data security

Project data is safely stored in online repository for long-term preservation and curation. Google Drive Repository is administrated by the Project Coordinator who will manage access rights and monitor folders and file names to ensure the data repository is consistent. The data from the project will be also stored for long-term in the institutional repositories of the partners. The final version of the DMP will further describe how data from the SWARM-E project will be managed in a sustainable manner.

The project will collect and process personal data only if it is necessary for its research and engagement activities i.e., research, consultations, interviews and events, and to share its findings and results with stakeholders via mailings, the website and newsletters. Primary legal basis for processing personal data will be an individual's consent. Individuals will have the right to withdraw consent at any time without any negative consequences. Personal data may be collected from members of the consortium, members of external organisations or individuals in their capacities as experts, respondents or participants. Use of such data will be in line with legal and ethical standards described in D1.3 Risk Management and Ethics Plan.

SWARM-E project will store personal data securely on password-protected computers. Personal data will only be used for the specific purpose for which it was collected (e.g., workshop management, travel arrangements) and will be deleted immediately after that purpose is fulfilled, unless legally required to be retained. Published interviews, podcasts, reports will not contain any personal data or reference to personal data. As part of each communication the participant receives, SWARM-E will give them the opportunity to opt out of further communications and have their data deleted from the project's records. If partners consider a plan to re-use personal data, they will give participants information about this as soon as it becomes available and give them the opportunity to consent or withdraw their data.

Recorded information (audio and/or visual) will be given special consideration to ensure that privacy and personal identities are protected. Participants will be provided with a consent form to read and sign if they will be photographed or recorded visually (e.g., video) during SWARM-E activities. The signed forms will be kept on file for inspection.

All partners of the consortium will adopt good practice data security procedures. This will help avoid unforeseen usage or disclosure of data, including the mosaic effect[1]. Measures to protect data include access controls via secure log-ins, installation of up-to-date security software on devices, regular data backups, etc. In addition to the Google Drive repository, partners may store local copies of research data on their institutional servers and or business cloud-based servers with access controls, encryption or password protection. Partners will follow their institutional security safeguards.

All SWARM-E partners will at least:

- o Ensure SWARM-E research data stored with them on their institutional servers is regularly backed up;
- o Ensure devices and data are safely and securely stored, and access controls are defined (e.g., via encryption, password protection, restriction of number of persons with access) at the user level;

---

[1] https://centre.humdata.org/glossary-2/mosaic-effect/#:~:text=Mosaic%20effect%20is%20when%20disparate,with%20other%20types%20of%20information.

o   Support good security practices by protecting their own devices and installing and updating antimalware software, anti-virus software and enabling firewalls;

o   In case personal data is processed, ensure appropriate security and confidentiality of the personal data, including for preventing unauthorised access to or use of personal data and the equipment used for the processing (GDPR);

o   Where necessary, the controller or processor of personal data will evaluate the risks inherent in the processing and implement measures to mitigate those risks (e.g., encryption) and ensure an appropriate level of security, including confidentiality, taking into account the state of the art and the costs of implementation in relation to the risks and the nature of the personal data to be protected (GDPR).

After the project ends, the responsibility concerning data security of the SWARM-E datasets will lie with the owners/managers of the repositories where these are stored.

5. Ethics

5.1 Legal issues, personal data and privacy aspects

Processes that handle personal data have been designed and built with the GDPR principles taken into account. Processes provide safeguards to protect research data (e.g., using pseudonymization or full anonymization where appropriate), and use the highest-possible privacy settings by default. No person or organisation involved will unintentionally be identifiable directly or indirectly in the datasets. Any indirect reference to sensitive personal information or e.g., lines of businesses, branches or industries will be removed and destroyed after the anonymised dataset has been checked and validated. After curation no person-related data is available publicly without explicit, informed consent, of the data subject and – if no anonymization is required – publicly available data cannot in any circumstances be used to identify a subject without additional information stored securely in a separate place (more information on the procedure in D1.3 Risk Management and Ethics Plan).

During and after closure of the project, the Project Coordinator will clearly disclose any datasets which have been collected during the project and declare the lawful basis and purpose for their processing. In addition, the Project Coordinator will state how long the data will be retained and also unambiguously declare, if it is being shared with any third parties or outside of the EEA. Data subjects of the project will have the right to request a portable copy of the data collected in a common format, and the right to have their data erased under specified circumstances.

5.2 Ethical issues

Research integrity and ethical principles related to data collection and use are covered in detail in the D1.3 Risk Management and Ethics Plan.

List of Acronyms and Abbreviations

Abbreviation	Definition
DC	Data Controller
EEA	European Economic Area
GA	Grant Agreement
GDPR	General Data Protection and Regulation
IPR	Intellectual Property Rights
PM	Project Manager