The project has been accepted for funding within the Key Digital Technologies Joint Undertaking (KDT JU), a public-private partnership in collaboration with the Horizon Europe (HORIZONI Framework Programme and National Authorities under grant agreement number 101096387.

Ref. Ares(2023)4535370 - 30/06/2023

PowerizeD

Digitalization of Power Electronic Applications within Key Technology Value Chains

Deliverable	**Data Management Plan**		
Involved WPs	WP 5	**Deliverable type**	Public
Project	PowerizeD	**Grant Agreement Number**	101096387
Deliverable File	D5.4	**Last Modified**	29.06.2023
Due Date	6	**Actual Submission Date**	6
Status	final version	**Version**	1.0
Contact Person	Heike Lepke	**Organisation**	OTH-AW
Phone	+49(0)9621/482-3621	**E-Mail**	h.lepke@oth-aw.de

Document history

V	Date	Author	Description
0.1	23.06.2023	Heike Lepke	Initial Version
0.2	27.06.2023	Heike Lepke	Ready to review
0.3	28.06.2023	Daniela Maier	Review
0.4	28.06.2023	Heike Lepke	Minor Additions
1.0			Final and reviewed Version

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

Table of Contents

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

1 Publishable Summary

Project acronym	***PowerizeD***
Project Logo	
Project full title	Digitalisation of Power Electronic Applications within Key Technology Value Chains
Project Coordinator	Mr. Jochen Koszescha, Jochen.koszescha@infineon.com Infineon Technologies AG
Coordinating Entity	Infineon Technologies AG

The PowerizeD project is committed to effective data management practices to ensure the responsible collection, storage and sharing of data. The Data Management Plan (DMP) outlines the framework for managing different types of data, including project communication and dissemination data, descriptive project data and research data.

Innovation and research projects such as PowerizeD often involve the generation and processing of a significant amount of data in various categories, some of which may be sensitive in nature. The data collected and processed in the course of the project may cover different objects and individuals, including natural persons.

Data security and privacy are paramount in the project. Measures such as encryption, access controls and user authentication will be implemented to protect sensitive data. Informed consent and compliance with data protection regulations are also prioritised.

The DMP emphasises the use of federated learning methods to aggregate data while maintaining privacy. This innovative approach enables collaborative model training without compromising individual privacy.

To promote open science, research datasets will be published in repositories such as Zenodo, following the FAIR data principles. This ensures proper metadata documentation, long-term preservation and citation of the datasets. The project leader takes responsibility for data management, including curation of the project website and coordination of data-related activities.

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

Financial aspects, such as costs for data storage and website maintenance, are addressed in the DMP. The project ensures that the data will be accessible after the project ends, thus promoting long-term data preservation.

To comply with legislative and ethical requirements, it is crucial for the PowerizeD project consortium to conduct the project in accordance with applicable national and international laws and regulations. The project should also adhere to specific sources of regulation relevant to projects supported by the KDT JU.

Overall, the PowerizeD project embraces ethical considerations, transparency and responsible data management practices. By adhering to the DMP, the project contributes to the advancement of knowledge in the field of federated learning and energy systems, while respecting data privacy and promoting open collaboration.

2 Introduction & Scope

Purpose and target group

The purpose of the PowerizeD Data Management Plan (DMP) is to provide a structured framework for the effective management of data throughout the project. It aims to ensure the responsible collection, storage, sharing and preservation of data in accordance with ethical considerations, data protection regulations and best practice.

The target audience of the DMP includes project team members, researchers, data scientists, project managers and stakeholders involved in the PowerizeD project. The DMP serves as a guide for these individuals to understand and implement data management practices that are consistent with the project's objectives and ethical standards.

In addition, the DMP may also be of interest to regulators, funders and other external parties involved in monitoring or evaluating the project. It provides transparency about data handling processes, privacy measures and compliance with relevant regulations.

By establishing clear guidelines and responsibilities, the DMP ensures that all project participants understand their roles and obligations in the effective management of data. It promotes collaboration, transparency and open access to data, fostering a culture of responsible data management within the project and contributing to the advancement of knowledge in the field.

Contributions of partners

Explain which partner were involved and their activities in their various sections:

TABLE 1: CONTRIBUTIONS

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

Chapter	Partner	Contribution
All	OTH	Setup of the DMP Initial collection of data sets Providing internal data share and communication channels on behalf of the project lead Infineon

Relation to other activities in the project

The initial collection of datasets for the PowerizeD project focused primarily on UC 1.7: System of Systems and CDT1.4: Federated Learning. These use case and Cross Topic Domain provided valuable data for developing and training models within a federated learning framework. However, as the project progresses, the data collection efforts will be extended to other Use Cases (UCs) and Cross-Domain Technologies (CDTs).

3 Data Summary

The PowerizeD project involves an ongoing process of data collection, which plays a crucial role in achieving the project's objectives. The data collected is categorised according to its type and use, with specific purposes identified during the project set-up and initial data collection phase. The three main purposes that have emerged are 'project communication and dissemination data', 'descriptive project data' and 'research data'.

These three types of data serve different purposes in the PowerizeD project. Project communication and dissemination data facilitate effective communication and outreach, descriptive project data provide insights into project progress and performance, and research data form the basis for research results and contribute to scientific knowledge. Proper management, organisation and documentation of these types of data is critical to project success, facilitates collaboration and enables future research and data sharing opportunities.

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

3.1 Collected Data Types

3.1.1 Project Communication and Dissemination Data

Within the PowerizeD project, there is a strong focus on effective communication and dissemination of project related information. To this end, data will be collected and organised to facilitate clear and effective communication with stakeholders and the wider public. This data includes public deliverables, project-related publications, media content (such as videos and images) and the project's presence on social media platforms. The aim is to engage, inform and raise awareness of the project's progress, outcomes and impact.

Publication channels for dissemination data

- **Research findings**: The PowerizeD project uses a variety of publication channels to disseminate project-related information and research results. These channels help to reach a wider audience and encourage collaboration and knowledge sharing. These channels typically follow rigorous peer review processes to ensure the quality and validity of the research being published. Some examples of traditional publishing channels are academic journals, conference proceedings, and project related technical papers.

In addition to traditional publication channels, the PowerizeD project uses several online platforms for communication and dissemination.

- **Project website**: The PowerizeD project maintains a dedicated website www.powerized.eu to share project information, updates, research findings and relevant resources. The website serves as a central hub for stakeholders to access project-related information and keep abreast of the project's progress.

  General project information and all publications will be listed with links for download. All media will also be linked or integrated into the website and the LinkedIn post will be mirrored in the news section.

- **Video platforms**: The project will use video sharing platforms such as YouTube and Vimeo to disseminate videos related to the project. These videos may include project presentations, demonstrations, interviews and promotional content. These platforms allow video content to be easily shared and viewed by a wide audience.

- **LinkedIn**: The PowerizeD project has a presence on LinkedIn, a professional networking platform. LinkedIn is used to share project updates, news, achievements and relevant articles or publications. It facilitates engagement with professionals, organisations and potential collaborators in the project's field.

More information about dissemination and the website is available in the deliverable “D4.1 Project website: PowerizeD web portal establishment and

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

operation". The dissemination process is described in the deliverable "D4.2 Draft Plan for Use, Dissemination, and Communication of the Foreground".

3.1.2 Descriptive Project Data

Descriptive project data refers to the information and statistics that describe various aspects of a project. This data provides insight into the characteristics, progress and performance of the project, allowing stakeholders to gain a comprehensive understanding of the status and outcomes of the project. Descriptive project data is typically collected and analysed throughout the project lifecycle to inform decision making, evaluate progress and facilitate effective project management.

Internal project storage:

For storage and access, the data described below are stored in a data portal that is accessible to all project partners, now 282 persons. Access to the data portal with access to the descriptive project data can be requested and will be granted to any project partner who has signed the grant agreement number 101096387.

Nextcloud can be used as a secure and collaborative platform for managing and sharing descriptive project data in the PowerizeD project. It provides access control features, version control and collaboration tools. By using Nextcloud for descriptive project data, the PowerizeD project can ensure secure and organised data management, facilitate collaboration between team members and streamline project processes. It provides a centralised and accessible platform for storing, organising and collaborating on project-related information, supporting effective project management and decision making.

The main parts of the descriptive data types that are stored as documents in the Nextcloud are described below:

Management data:

For efficient project management, communication, financial management and reporting in the PowerizeD project it includes various aspects related to project management, coordination and financial management.

The management data category in PowerizeD includes:

- **Final versions of deliverables**: Results or outcomes of specific project tasks or milestones, shared public audience or with project partners to demonstrate progress and achievement.
- **Contact information**: Contains relevant contact information for project team members involved in the project. This information ensures efficient communication and collaboration between project participants.
- **Contracts**: Contracts refers to the legal agreements and documentation related to the PowerizeD project.

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

- **Project reporting**: Project reporting involves the collection and documentation of project-related information to track progress within the project.
- **Budget data**: It includes budget plans and financial reports.

PowerizeD will use a standardised naming convention for all the project files and folders at data repositories that will be constructed using the following characteristics:

1. A unique chronological number of the dataset in the project.
2. The name of the dataset.
3. The acronym of the project.
4. A version number for each new version of the dataset that will be incremental at each revision.

A detailed description of the project management can be found in deliverable “D5.1 Project handbook”.

Tooling and Overview:

The "Tooling and Overview" section of the PowerizeD project includes the "Project Handbook" and "How To's" as subcategories to provide guidance and instructions on project tools and processes.

Work package Data with Use Cases and Cross Topic:

The Work Package Data with Use Cases and Cross Topic section of the PowerizeD project includes the collection and documentation of data related to specific work packages, together with their associated use cases and cross topic analysis. This section aims to provide insights into the practical applications of the project's work packages and to identify potential links or overlaps between different topics.

The inclusion of work package data with Use Cases and Cross Topics in the PowerizeD project ensures that the data collected is not only used within the specific work packages, but also analysed for broader applications and connections. This approach increases the overall effectiveness of the project, facilitates knowledge sharing and promotes interdisciplinary collaboration.

3.1.3 Research Data

Research data sets are collections of data collected, curated and used specifically for research purposes. These datasets play a critical role in scientific

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

investigations, allowing researchers to analyse and derive insights from the data to answer research questions, test hypotheses, and make new discoveries.

The data sets used in this (and the reuse of existing data sets of other) projects are on the one hand usually highly sensitive data, e.g. in terms of design secrets or open patents, and on the other hand data needed to comply with data protection rules for privacy reasons. The approach promoted in PowerizeD of using new technologies, mainly federated learning methods to aggregate data, allows the data sets generated to be private in themselves. This helps to promote the publication of datasets by preserving the integrity of the data sources. With the methods of the Data Management Plan (DMP) described below as the FAIR Data Principle, the publication of such datasets will be initiated and continuously promoted.

Research data sets are vital assets for the advancement of knowledge and scientific discovery. They provide researchers with empirical evidence and opportunities for analysis, interpretation and the generation of new knowledge. Proper management, documentation and sharing of research datasets contribute to the transparency, integrity and reproducibility of research results.

3.1.3.1 Data sets for Federated Learning

Data sets in PowerizeD especially for Federated Learning purposes should include time series estimation and training of a joint time series forecasting model, and could be structured by data format, historical data, party-specific information, subsets of data, target variables and associated meta data.

3.1.3.2 Reuse of available Data sets for Federated Learning purposes

Re-using existing data sets for federated learning purposes offers several benefits, including increased efficiency, improved privacy and broader data representation. By reusing existing data sets, organisations can leverage existing data resources rather than starting from scratch.

By strategically reusing available data sets for federated learning, organisations can harness the power of collective intelligence while respecting privacy and confidentiality, ultimately leading to more robust and effective machine learning models.

We are currently collecting existing datasets for the project. These datasets will serve as the basis for our analysis and modelling efforts. The datasets collected may include historical data from a variety of sources, such as public databases, research studies or industry reports. We aim to collect a diverse range of data sets covering relevant variables and time periods to ensure comprehensive analysis and accurate time series estimation.

Once collected, existing data sets are carefully reviewed and curated. This includes assessing data quality, addressing missing or erroneous values and ensuring data consistency across different sources. We will also perform data pre-processing tasks such as data cleaning, normalisation and feature engineering where necessary to prepare the datasets for analysis.

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

To ensure data protection and privacy, we will adhere to data protection regulations and ethical considerations throughout the data collection process. Any sensitive or personally identifiable information will be anonymised or masked to ensure confidentiality.

Collected existing data sets will be organised and stored in a secure and accessible manner. Appropriate documentation and metadata will be produced to provide clear descriptions of data sources, variables and any transformations applied. This documentation will facilitate data sharing, collaboration and reproducibility of our research results.

By collecting existing datasets, we aim to leverage valuable information and insights from previous research and industry sources. This approach allows us to capitalise on existing knowledge and trends, while ensuring that our analysis and modelling efforts are based on robust and reliable data.

3.1.3.3 Input Data Types

The data sets are used as input to train joint models that uses historical data and party-specific information to predict future values or patterns in the time series. The parties can jointly train the model using their respective subsets of the data set, contributing to a common understanding and predictive capability for the time series.

The following datasets are currently under audit:

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

Name/Autor	Usage	Download	Access
NASA Battery Dataset	Battery Lifetime Estimation (BLE)	https://www.nasa.gov/content/prognostics-center-of-excellence-data-set-repository	Public
CALCE Battery Dataset CALCE Center for Advanced Life Cycle Engineering 1103 Engineering Lab Building University of Maryland	Battery Lifetime Estimation (BLE)	https://calce.umd.edu/battery-data	Public
SiCWell Dataset Daniel Weber, Marius Gentejohann, Erik Goldammer, Michael Schlüter, October 28, 2021, "SiCWell Dataset", IEEE Dataport, doi: https://dx.doi.org/10.21227/mnmw-8c42.	Battery Lifetime Estimation (BLE)	https://ieee-dataport.org/open-access/sicwell-dataset	Public
RISE	Lifetime Data of MOSFETs		Project Internal
MBAG	Power Demand Estimation (PDE)		Project Internal
Further PDE data sets		ongoing	

TABLE 2: REUSE OF AVAILABLE DATA SETS FOR FEDERATED LEARNING

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

3.1.3.4 Outcome Data Types

The data sets used in federated learning systems can vary depending on the specific application and domain. However, the general principle of federated learning is to train machine learning models using data distributed across multiple devices or data sources without aggregating the data centrally.

The results produced by federated learning systems can be twofold:

- **Locally trained models**: Local models represent the trained models on each device or data source that capture the patterns and insights specific to their local data. These local models can be used for local inference or prediction without sharing the data externally.
- **The final global model**: This model is the result of aggregating the local model updates or gradients from all participating devices or data sources. This global model captures the collective knowledge and patterns learned from the distributed data sets. The global model can be used to make predictions or inferences about new data instances, taking advantage of the diversity and scale of the distributed data while maintaining privacy.

3.2 Keeping privacy through Federated Learning

It's important to note that federated learning is often used in scenarios where there are privacy, data security or data ownership concerns. By using federated learning, organisations and data owners can collaborate to improve the performance of their models without compromising privacy, and the results can be shared or used in a way that respects privacy and data protection regulations.

Overall, federated learning systems enable collaborative model training using decentralised datasets, generating both local models and a final global model that balance privacy, security and performance for a range of applications and domains.

3.3 Purpose of reusing and generating data in a FL System

To use existing data sets for training and to generate new data to improve the performance and generalisation of the FL model.

- Reuse of existing data: FL systems can benefit from reusing existing data sets that are relevant to the learning task. By incorporating diverse and representative data from different sources, FL models can capture a wider range of patterns and improve their predictive capabilities. Reusing existing data sets also makes better use of available resources and reduces the need to collect data from scratch.
- Generate new data: In addition to reusing existing data, FL systems can generate new data to augment the training process. Data generation techniques, such as data augmentation, can artificially increase the size and diversity of the training data set. This helps to address issues such as data scarcity or class imbalance, and improves the model's ability to

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

generalise to unseen data. By generating new data, FL systems can improve the robustness and performance of trained models.

3.4 Publishing of generated research data sets

Publishing generated research datasets is an essential aspect of promoting open science, fostering reproducibility and enabling the wider research community to benefit from and build on the data generated.

Publishable data generated by the project will be stored in a repository that meets the requirements of the European Open Science Cloud (EOSC[1]) and contributes to the vision of a "Web of FAIR Data and Services for Science in Europe"[2]. One such repository that fulfils these criteria is Zenodo[3], which is part of OpenAIRE[4]. Zenodo serves as a repository for researchers to deposit their publications and data, and also provides long-term data storage facilities.

By publishing generated research datasets researchers contribute to the advancement of knowledge, enable the reproducibility and validation of research results, and facilitate collaboration and innovation within the research community. Openly sharing generated datasets promotes transparency, accelerates scientific progress, and fosters opportunities for interdisciplinary research and data-driven insights.

3.5 Further collection of data sets

In the ongoing process of data collection within the project, partners will be encouraged to share their data with other partners as much as possible. The distribution of data will be facilitated through two channels:

- **Public download:** Data sets that can be openly shared without confidentiality or privacy restrictions will be made available for public download. This allows researchers and stakeholders outside the project to access and use the datasets for their own purposes. Public download ensures transparency and promotes data sharing within the wider research community.
- **Project internal data sharing**: In cases where datasets contain sensitive or confidential information that cannot be shared publicly, a project internal data sharing mechanism will be established. This data sharing platform via Nextcloud is accessible only to the project partners and will ensure the necessary confidentiality and privacy controls. It provides a secure environment for sharing and collaborating on sensitive datasets within the project team.

The choice of data distribution channel (public download or project internal data sharing) will depend on the nature and sensitivity of the data sets. Open and non-

[1]https://eosc-portal.eu/

[2]https://www.nature.com/articles/sdata201618

[3]https://zenodo.org/

[4]https://explore.openaire.eu/

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

sensitive datasets can be shared via public download, while sensitive or confidential datasets are restricted to project internal data sharing.

By facilitating data sharing between project partners, both through public download and project internal data sharing, the project aims to promote collaboration, knowledge sharing and synergistic use of data. This ongoing process of data collection and sharing will enhance the collective intelligence of the project, enable cross-validation of results, and promote the generation of new knowledge through the shared data resources.

The result will be collected for deliverable “D5.5 Updated Data Management Plan”.

4 FAIR Data Management

FAIR data management refers to the principles and practices that ensure data are findable, accessible, interoperable and reusable. These principles aim to maximise the value and impact of research data by making it more discoverable, understandable and usable by both humans and machines.

4.1 Making data findable

The PowerizeD project data will be hosted in a designated data repository that adheres to the FAIR principles and provides open access to the research community. The specific repository where the data will be hosted will depend on the practices and policies of the PowerizeD project.

Publishing Data Sets:

One repository for publishing and sharing research data is Zenodo. Zenodo is an open access repository that allows researchers to deposit and publish their datasets, ensuring long-term preservation, discoverability and citation[5].

- **Rich Meta Data:** By using a repository such as Zenodo, which supports EOSC requirements and provides DOI assignment and metadata collection capabilities, the project will ensure that publishable datasets are securely stored and can be accessed, cited and integrated within the wider research community. This contributes to the principles of open science, facilitates data sharing and reuse, and promotes transparency and reproducibility in research.

  A detailed description of the meta data structure is described at https://schema.datacite.org/[6]

  Keywords are used to search with the project or related keywords on the Zenodo platform to browse the available datasets, explore the accompanying documentation and metadata, and download the data for further analysis and research.

[5]https://about.zenodo.org/principles/

[6]https://schema.datacite.org/meta/kernel-4.4/doc/DataCite-MetadataKernel_v4.4.pdf

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

In addition, the PowerizeD project will show the published data sets on its own project website. The website will include information on how to access and download the data.

By providing clear information on where to find the data, researchers interested in the PowerizeD data can access and use it for their own research, fostering collaboration, knowledge sharing and the advancement of scientific understanding.

- **Harvesting and Index:** Importantly, the chosen repository should support the collection of metadata from the datasets provided by the project partners. This can be achieved using protocols such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The OAI-PMH protocol allows the repository to harvest and expose the metadata associated with the deposited datasets, enabling better discoverability, interoperability and integration with other research data infrastructures.

  Each data set will be assigned a persistent identifier (PID) such as a Digital Object Identifier (DOI). PIDs ensure that the data sets have a unique and permanent reference, enabling easy and reliable identification and citation.

4.2 Making data accessible

Access to PowerizeD data can be facilitated through various mechanisms to ensure wide availability and appropriate use. The specific access procedures may vary depending on the sensitivity of the data and any legal or ethical considerations.

The access is granted at least through public downloads on the website, through a Zenodo repository and the internal data share.

4.3 Making data interoperable

Making PowerizeD data interoperable ensures that data sets from different sources within the project can work together seamlessly and be shared, integrated and understood by different systems, tools and stakeholders. It involves the implementation of standards, formats and protocols that enable effective communication, collaboration and integration of data across different platforms, applications and domains.

Zenodo uses JSON Schema as internal representation of metadata and offers export to other popular formats such as Dublin Core or MARCXML.

PowerizeD ensures that data from different sources within the PowerizeD project are harmonised, i.e. transformed, mapped and aligned to a common representation or ontology. This enables data integration and facilitates cross-referencing and analysis of heterogeneous data sets.

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

5 Allocation of Resources

In the PowerizeD project, the duration of the storage and curation, storage and preservation costs are allocated as follows:

5.1 Project internal data storage

Project partners can store non-public datasets free of charge in the project's internal data storage. The costs associated with this storage will be covered by the project leader. This will ensure that partners have a dedicated and secure space to share and access data within the project for at least three years after the project has ended.

The access for the internal data storage via Nextcloud will be available 6 Months after the project ends. The partners are able to download the internal published data and use it under the GA conditions further on.

5.2 Project website

Infineon, as the project leader, will bear the costs of maintaining the project website. This includes costs related to hosting, domain registration and other technical aspects. The website will serve as a central platform for the dissemination of project information and may include access to public datasets, publications and other project-related resources.

A section will be added to the website with links to the collected datasets and project related publications. The website and the corresponding links to the data sets will be available 4 years after the project ended.

5.3 Public data sets

Publication of data sets in Zenodo, a repository for long-term storage of research data, is usually free of charge. This allows the project to openly publish and share datasets in accordance with the FAIR Data Principles. Zenodo assigns a Digital Object Identifier (DOI) to each uploaded dataset, ensuring its long-term accessibility and citation. It is designed to be a reliable and sustainable repository for the preservation and sharing of research results. Zenodo's mission is to provide a permanent home for research data, ensuring its long-term availability and accessibility.

5.4 Responsibility of publishing partners

The responsibility for accessing and disseminating other publishable data, such as scientific publications in the form of journal articles and conference papers, lies with the respective publishing partner. This may involve following appropriate publishing models, such as Green Open Access, where the publishing partner ensures the accessibility and availability of the publications.

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

6 Data Security

6.1 Internal Data Store and Website

The PowerizeD project data share uses Nextcloud, as a well-known self-hosted file sync and sharing platform. This offers various features and mechanisms to ensure data security.

Key aspects of data security in Nextcloud:

1. Encryption: Nextcloud provides client-side encryption, which means that data is encrypted on the client device before it is uploaded to the server. This ensures that data remains encrypted during transit and at rest, minimizing the risk of unauthorized access.
2. Access Controls: Nextcloud allows administrators to define fine-grained access controls and permissions for users and groups. This ensures that only authorized individuals have access to specific files or folders within the Nextcloud instance.
3. Authentication and User Management: Nextcloud supports various authentication methods, including username/password, LDAP, SAML, and two-factor authentication (2FA). Strong authentication mechanisms help prevent unauthorized access to the Nextcloud instance. User management features enable administrators to manage user accounts, assign roles and permissions, and enforce password policies.
4. Secure Communication: Nextcloud uses secure communication protocols such as HTTPS/TLS to encrypt data during transmission between clients and the server. This safeguards data from interception or tampering during transit.
5. Auditing and Logging: Nextcloud provides auditing and logging capabilities to track user activities, file modifications, and system events. This helps identify and investigate any security incidents or unauthorized access attempts.
6. Server Security: Nextcloud encourages best practices for server security, such as keeping the server software up to date, applying security patches, and following secure configuration guidelines. Regular security updates and proactive monitoring help protect the Nextcloud server from potential vulnerabilities.

It's worth noting that while Nextcloud provides robust security features, the overall security of the Nextcloud deployment also depends on factors such as the server infrastructure, network configuration, and adherence to security best practices by administrators and users. Regular security audits, proper user education, and ongoing monitoring are essential to maintaining a secure Nextcloud environment.

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

The data server and the website are located in Germany and therefore under full General Data Protection Regulation (GDPR)[7] responsibilities.

The data has a nightly backup accessible through the Nextcloud and Website provider[8].

6.2 Public data store security

Zenodo takes security very serious and does our best to protect your data.[9]

CERN Data Centre: The data centre is located on CERN premises and all physical access is restricted to a limited number of staff with appropriate training and who have been granted access in line with their professional duties (e.g. Zendo staff do not have physical access to the CERN Data Centre).

Servers: The servers are managed according to the CERN Security Baseline for Servers, meaning e.g. remote access to our servers are restricted to Zenodo staff with appropriate training, and the operating system and installed applications are kept updated with latest security patches via our automatic configuration management system Puppet.

Network: CERN Security Team runs both host and network based intrusion detection systems and monitors the traffic flow, pattern and contents into and out of CERN networks in order to detect attacks. All access to zenodo.org happens over HTTPS, except for static documentation pages which are hosted on GitHub Pages.

Data: Zenodo stores user passwords using strong cryptographic password hashing algorithms (currently PBKDF2+SHA512). Users' access tokens to GitHub and ORCID are stored encrypted and can only be decrypted with the application's secret key.

7 Ethics

In the PowerizeD data management plan, several ethical considerations will be taken into account to ensure the responsible and ethical use of data.

7.1 Ethical considerations

Privacy and confidentiality: Privacy and confidentiality are protected throughout the data management process. Relevant data protection regulations, such as the General Data Protection Regulation (GDPR), when collecting, storing and sharing personal data are applicated.

Informed consent: Informed consent from individuals whose data is collected, ensuring that they understand the purpose and scope of the data collection, use, and sharing is obligated. Also, clear information about their rights, including the right to withdraw consent and request deletion or modification of their data.

[7]https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_en

[8]www.oth-aw.de

[9]https://about.zenodo.org/infrastructure/

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

Data ownership and intellectual property: Data ownership and intellectual property rights within the project have to be clarified.

Responsible data sharing and collaboration: Responsible data sharing and collaboration within the project and with external stakeholders are promoted. Data sharing policies, including data access agreements, data sharing protocols, and appropriate data de-identification or anonymisation techniques are used.

7.2 Legal aspects

All legal aspects for privacy and sharing of data are handled in the PowerizeD Project Consortium Agreement (PCA), especially in

Section 8: Results and there 8.1. Ownership of Results and

Section 11: Privacy and data protection:

Where necessary, the Parties shall cooperate in order to enable one another to fulfil legal obligations arising under applicable data protection laws (the 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 relevant national data protection law applicable to said Party) within the scope of the performance and administration of the Project and of this Consortium Agreement.

In particular, the Parties shall, where necessary, conclude a separate data processing, data sharing and/or joint controller agreement before any data processing or data sharing takes place.

This assures that personal data is under the GDPR.

8 Conclusion

In conclusion, the Data Management Plan (DMP) for the PowerizeD project establishes a comprehensive framework for the collection, storage, sharing, and preservation of data. The DMP considers various considerations, including project communication and dissemination data, descriptive project data, research data, and the reuse of existing data sets for federated learning purposes.

The DMP emphasizes the importance of data security, privacy, and ethical considerations throughout the project. It outlines measures to protect sensitive data, ensure informed consent, and comply with relevant data protection regulations. The use of technologies like federated learning helps maintain data privacy while allowing for collaborative data analysis and model training.

To promote open science and data sharing, the DMP encourages the publication of research data sets in repositories like Zenodo, following the FAIR data principles. This enables long-term storage, proper metadata documentation, and cutability of the data sets. It also highlights the responsibility of the project lead

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

in managing data, curating the project website, and coordinating data-related activities among partners.

Moreover, the DMP addresses the financial aspects of data management, including costs associated with data storage, preservation, and website maintenance. It highlights the need for long-term data storage and accessibility even after the project's completion.

Overall, the PowerizeD Data Management Plan sets a solid foundation for responsible and ethical data management practices within the project. By adhering to the outlined guidelines, the project can ensure data integrity, promote collaboration, and contribute to the advancement of knowledge in the field of federated learning and energy systems.

9 References

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

10 List of Figures

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

11 List of Tables

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.

- Last page of the document is intended to be blank! -

This document and the information contained may not be copied, used or disclosed, entirely or partially, outside of the PowerizeD consortium without prior permission of the partners in written form.