

COoperative Cyber prOtectiOn for modern power grids

D10.3 Data Management Plan - Version 1

Distribution Level	SEN
Responsible Partner	UCY
Prepared by	Irina Ciornei (UCY)
Checked by WP Leader	UCY
Verified by Reviewer #1	Jose Maria Maza Ortega (USE) 5/09/2024
Verified by Reviewer #2	Julio Alejandro González Mejías (ING) 9/09/2024
Approved by Project Coordinator	Angelos Marnerides (UCY) 16/09/2024



Co-funded by the European Union

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Directorate General for Communications Networks, Content and Technology. Neither the European Union nor the Directorate General for Communications Networks, Content and Technology can be held responsible for them.

Deliverable Record

Planned Submission Date	17/07/2024
Actual Submission Date	16/09/2024
Status and version	FINAL

Version	Date	Author(s)	Notes
0.1 (Draft)	19/07/2024	Irina Ciornei (UCY)	ToC, initial input
0.2 (Draft)	3/08/2024	Irina Ciornei (UCY), AUTH, SEL, ING, TUD	Input for all chapters and edits of datasets from partners
0.3 (Draft)	28/08/2024	Irina Ciornei (UCY)	Updates in all chapters, final draft ready for internal review
0.4 (Draft)	9/09/2024	Jose Maria Maza Ortega (USE), Julio Alejandro González Mejías (ING)	Comments to 1st draft
0.5 (Draft)	10/09/2024	Irina Ciornei (UCY)	Updates in all sections of the commented 1st fully integrated draft
1.0 (Final)	15/09/2024	Angelos Marnerides (UCY)	Final quality check

Table of contents

Definition of Acronyms 4
Executive Summary 6
1 Introduction 7
1.1 Scope of the deliverable 7
1.2 Structure of the deliverable 7
2 COCOON Data Management Methodology 9
2.1 Methodology for Data Management 9
2.2 DMP monitoring and update 9
2.3 Data Summary 9
2.3.1 Purpose of data collection / generation and relation with COCOON 10
2.3.2 Characterization of collected/generated data 10
3 FAIR Data 12
3.1 Making Data Findable 12
3.1.1 Naming conventions, persistent identifiers, and metadata 13
3.2 Making Data Accessible 13
3.3 Making Data Interoperable 14
3.4 Increase data re-use 15
3.5 Allocation of resources 16
3.6 Data Security 17
3.7 Ethical aspects 18
4 COCOON Datasets 19
5 Conclusions 23
6 Annex 24
7 References 49

Definition of Acronyms

API	Application Programming Interface
ASCII	American Standard Code for Information Interchange
BSS	Battery Storage System
CA	Consortium Agreement
CC BY	Creative Commons Attribution license
CERN	European Organization for Nuclear Research
CI/CD	Continuous Integration/Continuous Deployment
CPN	COCOON Programmable Node
CSV	Comma-separated values
DBF	dBase
DMP	Data Management Plan
DOI	Digital Object Identifiers
DOC/DOCX	MS Word Document
EU	European Union
EPES	Electrical Power and Energy Systems
ESS	Energy Storage Systems
FAIR	Findable, Accessible, Interoperable, and Reusable
GA	Grant Agreement
GDPR	General Data Protection Regulation
GOOSE	Generic Object Oriented Substation Events
GPL	General Public License
HTML	Hypertext Markup Language
IAB	International Advisory Board
IED	Intelligent Electronic Device
IIPM	Innovation & Intellectual Property Management
IPR	Intellectual Property Rights
JSON	JavaScript Object Notation
.m	Matlab files
MFA	Multi-Factor Authentication

MIT	Massachusetts Institute of Technology
MMS	Manufacturing Message Specification
MU	Merging Units
NDA	Non-Disclosure Agreement
ODT	Open Document Text
ODS	OpenDocument Spreadsheet
OpenAIRE	European Repository Network
PCAP	Packet Capture
PMU	Phasor Measurement Units
PNG	Portable Network Graphics
PV	Photovoltaics
RTDS	Realtime Digital Simulator
RTU	Remote Terminal Units
SCADA	Supervisory Control and Data Acquisition
SCL	Substation configuration Language files
SIEM	Security Information and Event Management
SV	Sample Value
TBD	To be defined later
TC	Technical Coordinator
TCP	Transmission Control Protocol
TXT	Text-Formatted Data
WP	Work Package
XLS/XLSX	MS Excel files
XML	eXtensible Mark-up Language

Executive Summary

This deliverable provides the first version of the Data Management Plan of the COCOON project. The DMP focuses on the strategy for the data management lifecycle, which includes the project's data collection, processing, and/or generation. Furthermore, this deliverable contains the initial description of all the data collected and created within the first year of the COCOON project implementation and how the HORIZON Program Guidelines for Findable, Accessible, Interoperable, and Reusable (FAIR) Data Management has been applied. Based on these criteria, a template was built and given to all the partners in the project to define, characterize, and manage the datasets that will be delivered, used, and implemented throughout the project.

The DMP will be updated during the project whenever substantial changes occur, such as new datasets, changes in the consortium's policy, new information, or due to external causes. The updates will be reported by the end of each reporting period, while the second version of the DMP will be also reported as D10.7 (M36).

The DMP highlights the use of the COCOON's internal repository as a collaborative platform, in which the COCOON datasets will be stored, modified, and reviewed, as needed, during the project execution. The datasets collected in the first year of the project are presented in the Annex.

1 Introduction

This deliverable defines and presents the first version of the Data Management Plan (DMP) of the COCOON project. The DMP follows the provisions and guidelines established in the "Horizon Europe Data Management Plan Template" version 1.0 (published by the European Commission on May 5th, 2021[1]). Thus, the COCOON datasets were defined, described, and managed in accordance with this process. The COCOON internal collaboration process was described as part of Deliverable D10.1 – Project Management Plan – version 1, which also introduced the collaborative COCOON SharePoint, based on Microsoft Office cloud services to be hosted and managed by the partner SEL. The COCOON SharePoint will be the central point to support the COCOON DMP presented herein.

All COCOON partners will adhere to this DMP which was prepared following the Findable, Accessible, Interoperable, and Reusable (FAIR) principles, as they are defined by the HORIZON Europe Program Guide[2]. More specifically:

- The project team members will have easy access to this deliverable and its updates through the project's SharePoint repository.
- Each organization which is part of the project's consortium that provides a dataset or an Intellectual Property (IP) Asset (e.g., documents, software, architectural diagrams, etc.) as part of the COCOON project is requested to fill the relevant template, which is also provided in the Annex of this report.
- As part of the Project Management Activities, and specifically during the General Assembly meetings of the project, a monitoring process for the adherence and application of the DMP will be addressed by the project coordinator (PC).

1.1 Scope of the deliverable

This deliverable presents the COCOON's DMP in its initial version (version 1), while an updated version (version 2) will be reported by the end of the project. This deliverable is prepared in the framework of Task 10.3 of WP10 which oversees the project's overall management. It is related to all technical Work Packages (WPs) since it introduces the COCOON's datasets that will be used throughout the project lifecycle. This deliverable is also connected to the WP9 responsible for the exploitation activities, where some of the IP assets and datasets exposed in this report might be further used as part of the COCOON's Exploitation Plan (D9.5).

1.2 Structure of the deliverable

This deliverable is structured as follows:

Section 1 – Introduction provides an introduction about this deliverable, highlighting its purpose and its relationship with other WPs, as well as the methodology and structure for composing this report.

Section 2 – ***COCOON's Data Management Methodology*** describes the projects' Data Management methodology, the monitoring procedures, and the time plan of the DMP update.

Section 3 – FAIR Data describes the process adopted by COCOON to make its research data findable, accessible and interoperable based on best practices and recommendations offered by the European Commission Guide.

[1] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx

[2] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf

Section 4 – COCOON's Datasets describes the project datasets, as available or identified so far by the time of the release of the current version of the DMP, and which are expected to be expanded and updated as part of version 2 of the DMP.

Section 5 – Conclusions of this deliverable.

Annex - Dataset Description Template: includes the template used to describe the COCOON datasets.

2 COCOON Data Management Methodology

2.1 Methodology for Data Management

The COCOON consortium based on past experience from other Horizon Europe and Horizon H2020 projects , has developed a template for defining, describing, and managing the various datasets and other research outputs of the COCOON project which are to be used by more than one partner in the consortium based on the Horizon Europe Data Management Plan Template" version 1.0. The method used to create this template was based on seven characteristics: (a) Dataset Summary, (b) Relationship with the tasks and deliverables of the COCOON Project, (c) Partners Services and Responsibilities, (d) the FAIR framework, (e) Allocation of Resources, (f) Data Security, and (g) Ethical aspects. The template built based on these seven characteristics is available as part of the Annex of this report, while Chapter 4 provides the available examples of the use of this template by the relevant partners to describe the current status of their generated/used datasets.

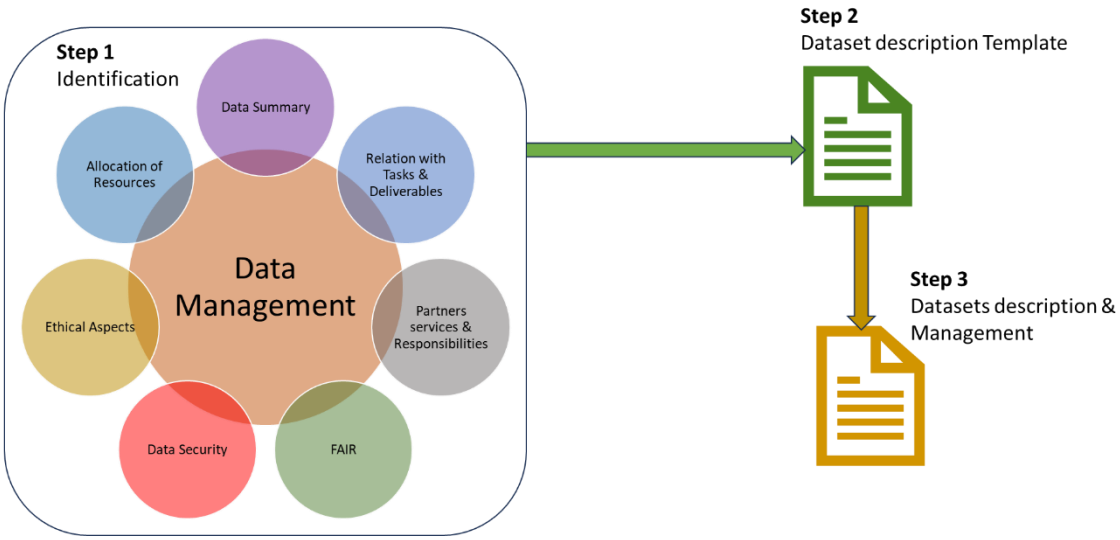


Figure 1: Data Management Methodology of the COCOON project

Thus, the first version of the DMP of the COCOON project is built on the aforementioned characteristics. Figure 1 summarizes the Data Management Methodology adopted in the COCOON project, while the following sections will detail each of the above characteristics in a greater detail.

2.2 DMP monitoring and update

The DMP will be updated during the project whenever there are significant changes such: new datasets, new information is available, changes in the policies of the consortium, changes in the composition of the consortium or external factors. COCOON partners (i.e., the data owners) will be responsible to update the information in the provided template. The DMP will be officially updated in the project's Periodic Reports at the end of the project, in M36, when the second version of the DMP will be released.

2.3 Data Summary

It is worth mentioning that most of the COCOON project deliverables are listed as public. This also ensures that the project consortium will do every effort needed to use open data whenever possible as well as generating open research data, while protecting personal information and privacy, including those related to business sensitive information. Thus, it is to be highlighted that data will not be made public if it could jeopardize someone's privacy or confidentiality and/or for security considerations,

especially due to the project application domain, particularly given the sensitive nature of the project and its relation to critical infrastructures such as Electrical Power and Energy System (EPES).

The data summary section of the DMP template includes the following information:

- State the purpose of the data collection/generation.
- Explain the relation to the objectives of the project.
- Specify the types and formats of data generated/collected.
- Specify if existing data is being re-used (if any).
- Specify the origin of the data.
- State the expected size of the data (if known).
- Outline the data utility: to whom will it be useful.

2.3.1 Purpose of data collection / generation and relation with COCOON

For the development, testing and validation of the various components of the COCOON architecture and its associated cybersecurity services, various data must be gathered. Further, the COCOON datasets are going to be used during the design and tuning phases related to the development of the COCOON tools, as well as component-to-component interfaces. The following are some of the COCOON-related activities and goals which entail data collection:

- Learn about end users' and stakeholders' demands, challenges, and requirements to better build and construct the COCOON's architectural components. The goal is to develop solutions that are applicable to the end users' actual needs, rather than relying on assumptions.
- Monitor and evaluate the application of the COCOON Programmable Node (CPN) and the COCOON cybersecurity services in the project's pilots, considering both the opinions of EPES stakeholders' part of the COCOON's consortium, as well as the technical partners in charge with the development of the COCOON's innovative solutions to improve the final version of these components.
- Monitor and evaluate the project's dissemination and communication results to measure their impact and fine-tune the project strategy in this regard, as well as fulfil the Commission's reporting requirements.

Based on the above, the following information is expected to be collected:

- Related to research, innovation, and communication (e.g., reports, publications, posters, interviews, etc.).
- Various types of technical data (network data, infrastructure data, etc.) in the development, testing and demonstration phases of the project.
- Key performance indicators for the progress of the project and for the dissemination and communication activities.

In case of any personal information in the collected data, the consortium will follow the procedures related to Ethics Requirements, as they are described in the Grant Agreement (GA) Guidelines.

2.3.2 Characterization of collected/generated data

Considerable data will be processed as part of the COCOON project. Type, format, origin, expected size, storage, purpose, and re-use are all indicated for each type of data. The "Type of data" defines the data category under which a given data element belongs to.

Types and formats of collected/generated data

The expected data types include (but are not limited) to:

- Existing datasets (a summary of which is detailed in Chapter 4).

- Direct input from partners or external stakeholders (e.g. network or grid diagrams, assets descriptions, IT network packets captures , power generation profiles, etc.).
- Data collected during pilots' deployment and testing.
- Data collected from project events such as workshops, conferences, EPES stakeholders' roundtables, etc.
- Data collected from social media, visits on the project website, etc.
- Data from open access databases.

The "Format of data" defines the type or form of support under which a given data falls. Thus, the expected data formats include (but are not limited) to: ASCII text-formatted data (TXT), Comma-separated values (CSV), dBase (DBF), eXtensible Mark-up Language (XML), JavaScript Object Notation (JSON), Hypertext Markup Language (HTML), MS Word (DOC/DOCX), MS Excel (XLS/XLSX), OpenDocument Spreadsheet (ODS), Open Document Text (ODT), Matlab files (.m), Packet Capture (.pcap), etc.

Origin of data and re-use of pre-existing data

The data set's origin must specify how the data was obtained and what type of equipment was utilized to obtain the data set. The origin of data includes (but are not limited) to: surveys, interviews, specifications, requirements, tests and demonstrations, paper or digital use case records, etc. Data re-use ability defines whether, or not, the data will be considered as confidential and if so, limited stakeholders who will have access to the data.

Data size

The following definitions apply to each data criteria, the "Expected size" defines the size in kB, MB or GB which can be expected from the data and the "Storage" is related to the repositories where the data is going to be stored, such as Zenodo[3] (in case of open access data) or the COCOON internal SharePoint repository.

Data utility

The "Data utility" property is related to the objective and goal of processing of the data and defines who outside of the COCOON project consortium could use the data.

[3] https://zenodo.org/

3 FAIR Data

3.1 Making Data Findable

Data will be stored, managed, and described in such a way to ensure a maximum likelihood of discoverability and identifiability. This will facilitate interoperability and data re-use, always according to data accessibility restrictions. As an example, each dataset must comply with the following provisions:

- Metadata and unique identifiers such as Digital Object Identifier (DOI) will be utilized to determine whether and how the data can be discovered.
- Naming Convention will be used to define the naming pattern of a dataset.
- Versioning will be used, to clarify if it is supported or not.
- Keywords will be used.
- History and other metadata will be used, not only the dataset's filename.

As part of the COCOON project, researchers are encouraged to find or work with open-access (OA) research data on top of the specific data available or generated by consortium partners. When it comes to scientific papers, the COCOON project uses a self-archiving strategy. To ensure that the final peer-reviewed manuscript complies with both the Consortium Agreement (CA) and the GA, COCOON partners who plan to make their works open access will deposit the manuscript to the internal COCOON SharePoint repository. The open access data and scientific papers will be deposited in an open access repository. COCOON uses Zenodo, a collaboration of the European Organization for Nuclear Research (CERN)[4] and the European Repository Network (OpenAIRE)[5] for the creation of an open-source repository of academic publications and data [1]. The COCOON partners will identify the bibliographic metadata while providing access to the publication in accordance with the following structure [3]:

- Funding grant: Horizon Europe
- COCOON, Grant Agreement No 101120221
- Publication date, persistent identifier

Even though open access datasets are made freely available, they must be used in accordance with the terms of a license, such as the Creative Commons Attribution (CC BY) license (Creative Commons Attribution 4.0 International), as recommended by the European Commission (EC) under Horizon Europe.

Metadata will be used to communicate the most important rights and licensing terms. Metadata can include date of publication, version, publisher, keywords, GA number, project acronym, title, date of creation, copyright, license, relevant publication, last update, DOI, data format etc.

The technical infrastructure enforces the license and access limitations. To ensure consistency, COCOON partners are expected to follow discipline-specific metadata rules, such as the use of keywords and version control protocols. To identify datasets, there is no single standard method that can be used for all datasets in the COCOON project. However, when possible, partners utilize persistent and unique identifiers like DOIs, although alternative methods of unique identification may be used instead. Any exemptions to open access for the datasets generated by the project's partners will be explained and justified.

Three methodologies will be used to achieve the discoverability and identifiability of data: (a) Strict **naming conventions** and other persistent identifiers of each data-item, (b) Usage of **data management tools** (e.g., GitLab), and (c) Organization of data in a searchable **folder structure**.

[4] https://home.cern/
[5] https://www.openaire.eu/

Strict naming conventions are followed irrespective of the data location. For instance, deliverables can be named after their unique assigned number, while collected manuscripts can be identified by the year produced, the title and their authors. Possible sensor readings will be identified by the date and place of collection, and questionnaires by a unique identifier assigned to each study. Naming conventions will be refined and changed as data are collected and generated, and this existing DMP will be updated accordingly.

Data Management Tools are used to provide adequate data grouping and metadata descriptions to facilitate easier data searchability. For example, source code will be kept on ***GitLab*** source repository that is installed on an internal storage server. GitLab can organise the developed software in discrete projects, whereas the version-control of Gitlab mechanisms manages the versions of the software.

Finally, data will be sorted in searchable folder structures, according to their data type and the intended data usage. For example, different folders may be created for deliverables, administration, market research, publications etc. These folders will become more specific as the project develops and data are collected/generated, and this document will be updated accordingly.

3.1.1 Naming conventions, persistent identifiers, and metadata

The following table describes the information kept on the methods and metadata that make the collected or generated data discoverable and potentially reusable.

Table 1 Naming conventions, persistent identifiers, and metadata

Dataset	**Naming Convention / Persistent Identifier / Metadata**	**Description / Metadata**
Dataset Name	Describes the format of the naming convention, persistent identifier, or metadata (e.g., keywords) to make the dataset findable	Describes the metadata provided to allow discovery, the disciplinary or general standards to be followed. In case metadata standards, outlines the type of metadata to be created and how. Metadata will be offered in such a way that it can be harvested or indexed.

3.2 Making Data Accessible

Below, it is exemplified on how the COCOON data, which was approved to be openly available is made accessible and which means are used for the proper access. Thus, in the case of the COCOON's Dataset Openly Accessible, the data are retrievable by their identifier utilizing a standardized communications protocol which is open, free, and globally implementable and it enables for an authentication and authorization mechanism, as appropriate. To access the dataset, the method and/or program must be specified.

Methods/Tools for Accessing Data:

• Storage, searching and accessing functions are provided by the repository of the owner. An initial list of platforms where COCOON will be sharing its public data (deliverables, publications, etc.) is as follows:

- The COCOON project website [4].
- Zenodo, as open-source repository of the COCOON academic publications and data [1].

Non-sensitive and non-personally identifiable data will be shared. Unless consented to otherwise, data from the COCOON project will be stored in a secure SharePoint repository managed by partner SEL with the help of partners HEDNO and UCY. A consultation will be asked by the Project Management Board (PMB) and the Innovation and Intellectual Property Management (IIPM) Board of COCOON before the release of data. A SharePoint folder with a direct link to all the research datasets will be made available to COCOON consortium members over the course of the project. COCOON's internal SharePoint repository will remain online for a minimum of two years after the project's completion. Access to data will be different based on where it is stored at the time of this first iteration of the DMP's information. Access, reuse, analysis, exploitation, and dissemination of the data will be made available to other parties beginning with the existing datasets (bound by the license specifications). Sharing requirements are in accordance with open data models, and machine-readable data formats will be made available to accompany journal papers and presentations. Metadata will help search engines and other automated processes gain access to the data by enabling them to use conventional web techniques. Also, a machine-readable version of the dataset's metadata will be made available.

Project-related data (deliverables, reports, code, and databases) will be stored in the COCOON SharePoint with appropriate permissions for each user. Access to the repository is allowed by the coordinating partner to the partners actively involved in the development of the project, and it is not meant to be disseminated publicly or otherwise made available to third parties.

3.3 Making Data Interoperable

Data to be collected, generated or re-used will be stored in a format that will allow their usage in various computing platforms. To this end, formats that comply with open standards will be used such as pdf, CSV, or XML thus being able to be loaded into different software. Moreover, data is accompanied by adequate metadata. In cases where a dataset follows a non-standardized format, additional documentation will accompany the dataset to describe explicitly the parsing methodology. The use of open standards, metadata and documentation will enable data utilization at all levels, thus promoting and facilitating interoperability and re-use.

The following table provides an example on what type of information shall be provided for describing the file formats, methods, or actions to make the data in each dataset interoperable.

Table 2 Description of file formats, methods or actions to make data in datasets interoperable

Dataset	**Format, methods, or actions to make data interoperable**
Name of the dataset	Describes in detail either: • the file format together with the software that recognize it; or • methods, e.g., parsing, for automatic data interoperability; or • actions to take to interpret the data

COCOON's standards and methodology for data and metadata are geared toward facilitating the interoperability of project data. In order to facilitate the reuse of data from various sources, COCOON's will adhere to generic format standards which are compatible with freely available (open source) software programs. An important aspect of data interoperability is the usage of common terminology across all data types. In case of any rare data type, while the project is progressing, the

corresponding mapping will be provided in order to achieve commonality with the regularly used terminology.

3.4 Increase data re-use

For data re-use, COCOON will adopt a policy which attempts to make generated data as widely available as possible, after a scrutinization process to ensure that sensitive information and privacy are duly preserved. As a result, particular data licenses, sharing agreements, and deadlines are laid down in writing (e.g., Non-Disclosure Agreements, where necessary). Several considerations will be utilized, such as:

- if a specific license (e.g., Creative Commons CC BY 4.0) is enforced, the license specifies how the data is accredited.
- a field called "Data Sharing Terms" will be specifying the terms under which open data is shared. When no specific license is in place, the sharing terms are used; dataset sharing after the conclusion of the project indicates whether the data will be made accessible to the public once the project has finished.
- data preservation and updating after the project indicates whether the dataset will be updated after the project finishes.
- using the re-use timescale, the owner of the data may enforce how long the dataset to be available.

During the project's development, pilot related data will be linked to findings that could be used to safeguard commercial or industrial interests. As a result of intellectual property protection procedures, this pilot data cannot be made generally available for verification and re-use.

Open access to pertinent data may be granted on a case-by-case basis in accordance with the COCOON CA and the project GA. Upon the approval of the Consortium members, the data will either be published in gold or green open access. For example, the project website will be up and running for at least two years after the project concludes. Repository policies (such as the Zenodo repository policies) will apply to data in open access on other platforms. The data owner/producer who releases the data will assure its quality.

Reusability and searchability will be made possible using data in standard formats. Licensing is the most common way dataset owners offer access to their data. Where possible, the preferred and encouraged license for offering the datasets openly available will follow the same Creative Commons Attribution 4.0 International CC BY4.0 license[6], which is recommended also for the publications or software. There are no ownership rights transferred when a license is issued, but it establishes conventional constraints and permissions for a piece of work. When working with unstructured data, the best option is to utilize a CC0 license[6], which absolves someone of all responsibility for the content of the database. CC licenses exclusively cover copyright and copyright-related issues. Personal information is not included in CC.

The current DMP ensures several re-usage mechanisms for the collected or generated data. Specifically, for the purposes of software reuse, the GNU[7] General Public License (GNU-GPL or GPL) is considered as the most appropriate and widely used free software license, which guarantees end users (individuals, organizations, companies) the freedoms to run, study, share (copy),

[6] https://creativecommons.org/licenses/by/4.0/deed.en
[7] https://www.gnu.org/licenses/gpl-faq.html

and modify the software. An alternative to be considered is the Massachusetts Institute of Technology (MIT) license[8].

With regards to restricted or confidential data, sharing between the COCOON's partners and International Advisory Board (IAB) members will be facilitated by means of contractual agreements (e.g., Non-Disclosure Agreement (NDAs)), legal and bilateral binding documents, or other written protocols.

3.5 Allocation of resources

The DMP of the COCOON project will be monitored and updated in accordance with the roles identified in the following table:

Table 3: Roles and Responsibilities for the COCOON DMP

Partner	Roles	Responsibilities
SEL and HEDNO	Data Storage and Management	Responsible for adhering to the technical details of the DMP implementation with respect to the offered internal project repository, the COCOON SharePoint and the hosting and managing of the project website. Responsible for the dataset quality and integrity. Responsible for monitoring the open access datasets.
UCY	DMP Manager and Project Manager	Leader of the DMP. Responsible for preparing and updating the DMP, including updates in the periodic reports, as well as the updates for the Dataset template available in Annex. Responsible to inform the consortium partners about the DMP. Provide support to the partners in the implementation of the DMP.
SEL	Dissemination Manager	Responsible for the scientific publications for open access.
USE	Intellectual/Industrial Property Manager	Responsible for guaranteeing that the published datasets do not affect the industrial protection of the technologies developed within the COCOON project.

The expenditures associated with making the data collected/generated during the COCOON project activities FAIR are included in the project budget. These expenses cover a certain set of data

[8] https://joinup.ec.europa.eu/licence/mit-license

processing and data management tasks. The data processing and management activities which are considered are the following: collection, documentation, storage, access and security, preservation, availability and reuse, and overall data management. The DMP task leader is responsible for coordinating the collection of data in accordance with this data management plan. The project's indirect costs cover the cost of data storage on COCOON SharePoint and the cost of making data available after the project ends. An individual researcher or team of researchers engaged in COCOON will be responsible for handling their own data in accordance with the DMP, which applies to all consortium participants (universities, multi-actors and SME partners). While working with collaborators outside the COCOON Consortium, the researchers will inform them of the COCOON's DMP and ensure that primary research data are stored in accordance with HORIZON Europe criteria. In this case, the main author who is part of the COCOON consortium will be responsible for this compliance.

3.6 Data Security

One or more partners will have access to the COCOON datasets, which may be made publicly available or kept confidential. The importance of data security is paramount in these circumstances. Two options are adopted to describe the methods taken to ensure the data security:

- A security measure is one that is put in place to protect sensitive data while it is being stored or exchanged.
- Personal information is protected in accordance with the General Data Protection Regulation (GDPR) [5].

Members of the COCOON project take privacy and data protection in the highest regard and adhere to a strict code of conduct in this regard. The level of confidentiality of the data has an impact on data security, which includes how the data is stored and transferred. As such, the following guidelines will be followed by the COCOON consortium to ensure the security of the data:

- Type: *Confidential data of a member*. Guideline: *Store data in their own local storage.*
- Type: *Confidential data between two members*. Guideline: *Store data in their local storage.*
- Type: *Confidential data between all members*. Guideline: *Store data in the shared COCOON SharePoint, in the corresponding WP folder.*
- Type: *Public access data*. Guideline: *Store and share data utilizing Zenodo under the COCOON project community.*

As a preventive measure, the following guidelines should be followed:

- To avoid data loss, save data in at least two different locations when using local storage.
- If you're using a cloud storage service, make sure it includes a data policy that prevents accidental deletions, such as a recycling bin.
- Ensure that the final dataset is consistent by labelling files in a methodical manner.

The current security measures in place for the COCOON SharePoint as project internal repository are listed below:

- **Multi-Factor Authentication (MFA):** We require every external user to set up MFA before they can access SharePoint, adding an extra layer of security.
- **Controlled Access:** Access is strictly controlled; users can only gain entry after we manually approve them. This process ensures that only authorized individuals can interact with our SharePoint data.

- **SIEM Integration:** We utilize a Security Information and Event Management (SIEM) system that actively monitors for unauthorized access attempts. If someone not on our pre-approved list tries to log in, the system immediately alerts us, allowing us to respond quickly.
- **User Management:** Our initial list of authorized users was created in collaboration with the project coordinator. We add new members only upon request from existing team members, ensuring that all users are vetted.
- **Behavior Monitoring:** We are in the process of implementing an advanced tool that will analyze user behavior within SharePoint. This tool is designed to detect and alert us to any unusual or potentially malicious activities, helping to respond proactively to security threats.

These measures, combined with Microsoft's built-in security features such as data encryption and compliance tools provide up to date industry standards to maintain a secure and efficient SharePoint environment for the scope of the COCOON project and aligned with the EU regulations.

There will be an initial review of the COCOON project's assessment tools, data collecting, and analytic processes by its Technical Coordinator (TC) and the IIPM board and with the full support and contribution from all consortium partners. The TC and IIPM oversee the project's Intellectual Property Rights (IPR) including their security aspects. Project deliverables which are listed as public will be reviewed by the IIPM to determine if they contain any security sensitive information, and if so, provide guidelines on how to avoid the misuse of it. Both TC and the IIPM include representatives with a good knowledge of the issues raised by the security concerns.

External parties should be able to verify scientific integrity by examining datasets and records, and researchers should be confident that their findings are founded on reliable evidence. The following information will be kept on file:

- Data used in publications: for example, records used to gather data: questionnaires, complete questionnaires, photographs, videos, a database of analyzed research data (such as spreadsheets and midterm reports), as well as the accompanying supporting documents (date of the collected information, source of the information, location, etc.).
- If relevant, data sharing agreements.
- Original data (e.g., video and audio recordings, hard copy surveys, paper or digital experimental records, raw data such as unprocessed digital data, network traffic records, etc.).
- Processed datasets used in publications, including information used to understand the processing, in particular coding steps, data cleaning logs, missing values, metadata for each dataset, to guarantee transparency and scientific integrity of data processing.
- Documentation of actions (protocol documents, application for regulatory approvals and approvals granted, risk assessments etc.).
- Records pertaining to the project's financial management (PDF invoices, any other supporting account records etc.).
- Data that might be used in future research: interim reports, final reports, and manuscript drafts in literature are all examples of data that could be used in future research, subject to any limitations imposed by GDPR and ethical permission procedures.

3.7 Ethical aspects

All members of the project consortium agree that tackling legal and ethical issues is an important part of the project. Ethical considerations include any legal or ethical difficulties that may affect data sharing. Thus, to ensure that ethical aspects are well regarded, the following fields are included in each dataset:

- Ethical and Legal Aspects shows if there are specific ethical and/or legal aspects that should be taken into consideration. In addition, any steps taken to assure the accuracy of this information should be stated here.
- For data exchange and long-term preservation, a consent form signed by all parties involved in the respective process indicates whether the dataset can be used. This also shows whether the parties involved have agreed to share their data publicly if a consent form has been signed by all of them in this respect.

Personal information will be protected in accordance with the EU's Data Protection Laws, which include the Regulation (EU) 2016/679 of the European Parliament & Council of 27 April 2016 on the protection of natural persons regarding the processing of personal data, and on the free movement of such data, and repealing Directive 95/46/EC, also known as the "General Data Protection Regulation". All partners must ensure that EU regulations on ethics and data management, including the GDPR, are adhered to.

4 COCOON Datasets

Several datasets are expected to be generated as part of the COCOON project. For their storage and internal access, the COCOON consortium will use its internal repository (i.e., COCOON SharePoint hosted and managed by SEL), which will also support the continuous monitoring of the dataset's status. Each dataset will be converted to an appropriate SharePoint file that can be updated on demand. Additionally, each dataset's status will be further reviewed at General Assembly meetings and the fields of the COCOON Dataset Description Template will be updated accordingly.

The following table summarizes the current available information in terms of datasets that will be used or generated in the COCOON research and innovation project in line with the technical developments and the COCOON pilots, as they were identified in the first six months of the project and revised by the time of the publication of this report. The table ignores the datasets related to the COCOON SharePoint hosted and managed by partner SEL, which are, however, provided as part of the Annex. Further, it is to be highlighted that the filled templates for each of those datasets (which includes the DMP related information specified in Section 2.2 of the present deliverable) can be found in the Annex. The datasets can be updated and revised during the project's life cycle. The updated status will be available at the COCOON internal repository.

Table 4: Summary of the COCOON technical datasets in the first six months of the project

Dataset Name	Dataset Description	Type/Format of Data
UCY EPES measurements from actual or virtual power system setups	EPES measurements and network traffic data from actual or virtual power system setups related to Phasor Measurement Units (PMUs), power electronics inverters for Photovoltaics (PV) or Energy Storage Systems (ESS) or to power protection relays. The data is captured from actual equipment connected in several laboratory setups or from virtual equipment emulated within a real time digital simulator for power systems, all from the power system laboratory of KIOS CoE (UCY). The dataset contains PMU measurements following the IEEE C37.118 protocol,	.csv, .txt, .xlsx,.pcap

	measurements from PV/BSS inverters and from smart meters using Modbus/TCP protocol, and data from protection relays at digital substations using IEC 61850-GOOSE/MMS/SV protocols, The network traffic data was captured using a Linux Machine via a WireShark software.	
CUE dataset with measurements from HGI PV Plant	Electrical data from the photovoltaic plant Hoyas Grandes I (power, voltage, current, frequency, etc.). The data comes from the actual power equipment, such as Remote Terminal Units (RTUs), meters, inverters, etc. Inverters model: Huawei SUN2000-330KTL-H1 RTU model: BCO-1010-1900J, Server: Dell R450, Firewall: 01 Fortinet 60F.	csv, excel, PDF, png
ING dataset with measurements from virtual power system setups	The ING dataset with measurements from actual or virtual power system setups includes logs (.log) files, GZIP compressed logs (.log.gz) files, Comma-Separated Value (CSV) files, GZIP compressed TAR (tar.gz) files and Portable Network Graphics (PNG). The data is generated by a PPC module called *Ghostwriter*. This module is activated during the commissioning process in order to validate the inverters' time response and as part of a quality report for the client.	csv, .log, .log.gz, .tar.gz, .png
ENCS dataset with network traffic from the cyber range	The ENCS dataset with network traffic from the cyber range as Packer CAPture files (.pcap). The data is captured from virtual equipment emulated within a lab setup at ENCS.	.pcap
TUD dataset with EPES measurements from actual and virtual grid substation setups	The TU Delft dataset includes measurements from simulated power system setups and actual power system devices. It contains Common Separate Value (CSV) files related to Intelligent Electronic Devices (IEDs), Realtime Digital Simulator (RTDS) or Merging Units (MU), and network traffic data as Packer Capture files (.pcap). The data is captured from actual equipment connected in several laboratory setups, as well as from virtual and real equipment. The dataset contains substation measurements following the IEC 61850 protocol, measurements from RTDS simulated	.csv, .txt,. xlsx, .pcap, SCL files

	devices and IEDs using MMS/SV/GOOSE protocols, and data from network devices using TCP/IP-based protocols. The dataset would also contain IED configuration files, including Substation configuration Language files (SCL).	
HEDNO dataset with measurements from actual Distribution Network	The dataset includes measurements that conform to the DLMS/COSEM protocol (IEC 62053-22[9] and IEC 62053-23[10]). For the HEDNO dataset, the metering is conducted using the ITRON SL7000 [11]measuring device, combined with a 400/5 A current transformer and a 30/5 V voltage transformer to ensure accurate readings. The measured values are three phase active power, three phase reactive power, line to line voltage for all three phases and line current for all three phases. The measurements have a resolution of 15 minutes and correspond to the period 01/05/2024-31/05/2024.	.xlsx
SEL datasets for DevOps	The SEL dataset includes logs and documentation files (JSON, XML, YAML) related to the management of project data on GitHub and within the DevOps pipeline. This encompasses commit histories, pull request logs, continuous integration/continuous deployment (CI/CD) pipeline logs, and issue tracking data. The logs capture developer activities, code changes, and automated build/test/deployment processes	.json, .xml, .yaml
SEL datasets for Firewall Logs	This SEL dataset includes log files generated by firewalls in a replicated setup environment. The logs capture network traffic, firewall rule matches, and potential intrusion attempts. This includes data from different layers (network, transport, application) and log entries related to rule bypass attempts by threat models.	.log, .pcal, .txt

[9] https://webstore.iec.ch/en/publication/29987
[10] https://webstore.iec.ch/en/publication/29239
[11] https://emea.itron.com/products/ace-sl7000

SEL dataset for SharePoint	This SEL dataset encompasses data related to document and file management on the SharePoint platform. This includes user access logs, document version histories, and metadata about file uploads and downloads. The logs capture activities such as file creation, modification, deletion, and access permissions.	.docx, xlsx, .pptx, .pdf
USE datasets from the smart grid laboratory data	The datasets of USE come from different origins: computer simulations, real-time simulations and laboratory testing of IT and power components. The datasets may have different formats depending on the data: (i) Text files (TXT) for static information related to power networks; (ii) Portable Network Graphic (PNG) files for one-line diagrams of power networks and schemes of laboratory setups; (iii) Common Separate Value (CSV) files related to time series of electrical magnitudes; (iv) network traffic data as Packet Capture files (.pcap).	.txt, .png, .csv, .pcap

5 Conclusions

This report addressed at a comprehensive level the initial version of the DMP to be followed within the COCOON project's implementation. The first step was to establish the methodology, and based on that methodology, to prepare and fill the required templates. This structured approach ensures consistency and thoroughness in managing both data and intellectual property assets which are in direct relationship, despite the latter not being the focus of this report but of later deliverables of WP9 (D9.5: Exploitation Plan).

Electronic data and IP assets can be monitored, controlled, and adjusted as needed by the COCOON project partners using this DMP. The plan took full advantage of the internal collaboration platform/repository, such as the COCOON SharePoint, to facilitate seamless and efficient management. In the first six months of the project, nine datasets were defined, and they were fully documented as part of the Annex of this report. This initial setup provides a solid foundation for managing the COCOON dataset as well as other related IP assets (e.g., open access publications, etc.). Project partners have the flexibility to propose new datasets, in sync with possible relevant IP assets or modify existing ones based on the template provided. This dynamic approach allows the project to adapt to evolving needs and ensure that all relevant data and IP assets generated as part of the COCOON project are accounted for.

At the project plenary meetings, the status of the datasets will be updated. These regular updates ensure that all stakeholders are informed about the progress and any changes made to the data and their subsequent IP assets (if needed). By maintaining transparency and regular communication, the project can effectively manage its data, ensuring that they are utilized to their full potential. This collaborative and iterative process is crucial for the success of the COCOON project, as it enables continuous improvement and alignment with the project's goals and objectives.

6 Annex

The following table provides the example Template about the description of a dataset generated by UCY for the tests needed in COCOON WP1/Tasks 1.2 and 1.3. Each partner is kindly requested to fill in Table 1 example for each dataset that they produce as part of the COCOON Project, with as much information as possible.

Table A-1: UCY dataset with measurements from actual or virtual power system setups

Dataset Name	**UCY dataset with measurements from actual or virtual power system setups**
Dataset Summary	
Dataset Description	The UCY dataset with measurements from actual or virtual power system setups includes Common Separate Value (CSV) files related to Phasor Measurement Units (PMUs), power electronics inverters for Photovoltaics (PV) or Energy Storage Systems (ESS) or to power protection relays, and network traffic data as Packer CAPture files (.pcap). The data is captured from actual equipment connected in several laboratory setups or from virtual equipment emulated within a real time digital simulator for power systems, all from the power system laboratory of KIOS CoE (UCY). The dataset contains PMU measurements following the IEEE C37.118 protocol, measurements from PV/BSS inverters and from smart meters using Modbus/TCP protocol, and data from protection relays at digital substations using IEC 61850-GOOSE/MMS/SV protocols.
Dataset Purpose	Testing and validation of plethora of power system applications, including cyberthreat analysis related to Tasks 1.2 and 1.3
Dataset Type/Format	.csv, .txt, .xlsx,.pcap
Re-use of existing data	Yes, dataset generated as part of previous projects such as ELECTRON[12] and Smart5Grid[13].
Dataset Origin	This dataset is generated using actual or virtual equipment from the KIOS Cyber-Physical Power System Laboratory.
Dataset Collection	This dataset is collected using: • A Phasor Data Concentrator connected to actual or virtual PMUs. • A PC/controller connected in the lab for receiving all the measurements /set-points from actual or virtual MODBUS /TCP compatible devices (i.e., smart meter, inverter, etc) • A device connected at the local network to receive the GOOSE traffic from the intelligent electronic devices (i.e., protection relays). • All the network traffic is collected using WireShark[14] software.
-	2-8 GB
COCOON Work Packages, Tasks and Deliverables	
Relevant Work Packages	WP1: Cyber security Algorithmics
Relevant Tasks	T1.2 and T1.3: Threat Models & Vulnerability Assessment and Cyber Risk Profiling
Relevant Deliverables	D1.2: Threat Models, Vulnerability Assessment and Cyber Risk Profiling
Partners Services and Responsibilities	
Partner Owner	UCY
Partners responsible for data collecting	UCY
Partners responsible for data analysis	UCY and maybe others to be defined

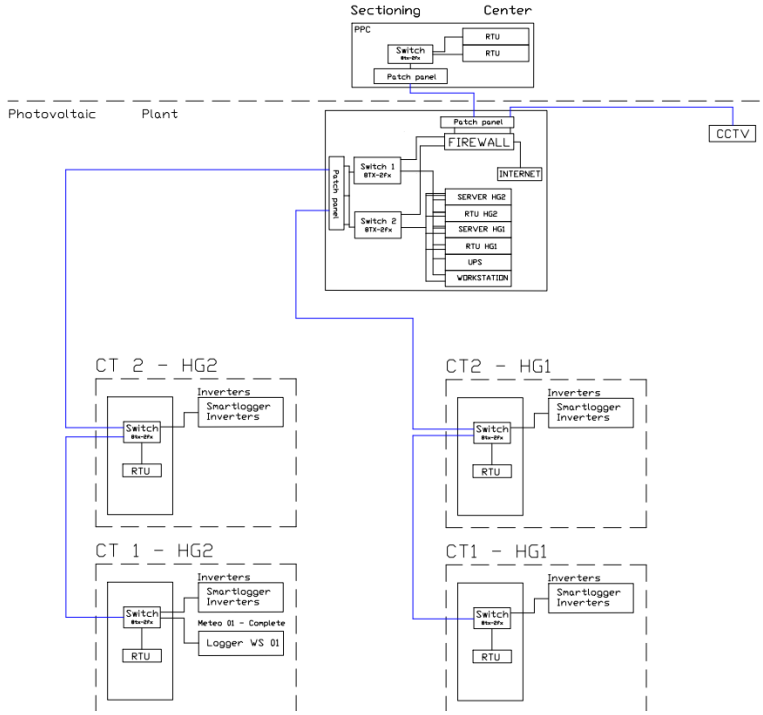
[12] https://electron-project.eu/
[13] https://smart5grid.eu/
[14] https://www.wireshark.org/

Partners responsible for data storage	UCY
FAIR Data - Making data findable, including provisions	
Discoverable and Identifiable Data	To be provided
Naming Convention	UCY_Measurements_EPES_Operation_Dataset_v1
Versioning	v1
Search Keywords	measurements, coordination signals, operation, PMU, C37.118, Modbus, IEC 61850-GOOSE/MMS/SV, Supervisory Control and Data Acquisition (SCADA).
Metadata	• Relevant Publication: TBD • Last Update: TBD • DOI: TBD • Data Format *.csv, *.pcap, .xlsx*, .txt*
FAIR Data - Making data openly accessible	
Dataset Openly Accessible	TBD
Methods/Tools for accessing the data	Wireshark, Tshark[15], Pandas[16], Python[17], Matlab[18], Excel[19]
Access Restrictions	TBD
Repository	Not available yet.
FAIR Data - Making interoperable	
Interoperability	To be provided
Standards	To be provided
FAIR Data – Increase data re-use	
License	TBD
Data sharing terms	Please cite the corresponding paper. More specific terms will be provided soon
Dataset sharing after the end of the project	Maybe
Preservation and update of the dataset after the project	Yes
Re-use timeframe	-
Allocation of Resources	
Cost for making the dataset FAIR	To be determined at a later stage in the project
Costs for the preservation (including backup) and the update of the dataset	No cost
Data Security	
Security Measures	The dataset does not refer to any actual entity participating in a real industrial environment.
Privacy Measure	-
Ethical Aspects	
Ethical and Legal Aspects	To be provided
Consent form for data sharing and long	To be provided

[15] https://www.wireshark.org/docs/man-pages/tshark.html
[16] https://pandas.pydata.org/
[17] https://www.python.org/
[18] https://www.mathworks.com/products/matlab.html
[19] https://www.microsoft.com/en-us/microsoft-365/excel

preservation within the COCOON consortium	
Consent form for sharing the dataset publicly.	To be provided
Other Issues	No

Table A-2: CUE dataset with measurements from HGI PV Plant

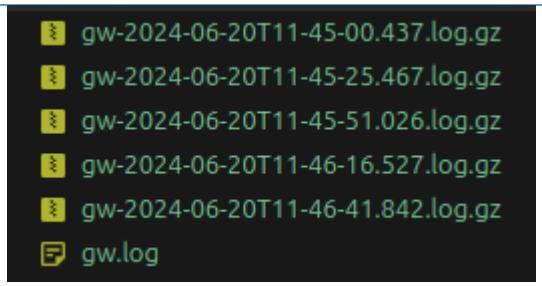
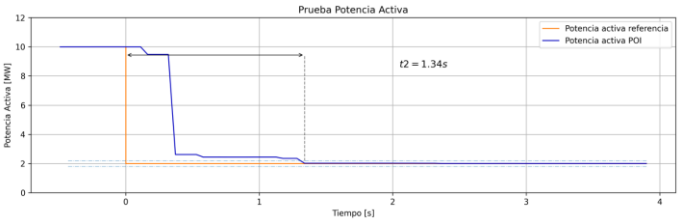
Dataset Name	**CUE dataset with measurements from Hoyas Grandes I (HGI) PV Plant**
Dataset Summary	
Dataset Description	Electrical data from the photovoltaic plant Hoyas Grandes I (power, voltage, current, frequency, etc.). The data comes from CT1-HG1 and CT2-HG1, as shown in the following diagram:  The data comes from the actual power equipment, such as RTUs, meters, inverters, etc. Inverters model: Huawei SUN2000-330KTL-H1 RTU model: BCO-1010-1900J Server: Dell R450 Firewall: 01 Fortinet 60F Inverter and smart meter measurements using Modbus/TCP protocol. The data that can be collected from the SCADA system includes both historical and real-time data. There are approximately 274 variables collected from various devices, such as: -I/O: 23 variables. -Inverters: 46 variables

	-Meteo Station: 18 variables -Meter PV: 36 variables -Plant: 51 variables -PPC: 44 variables -RCBs: 35 variables -Switch: 21 variables The monitoring and control system is EMPRO Monitoring System: -Historical and real-time data -The granularity and precision of the data allow for values ranging from 1 second to 1 day. The data can be displayed in a table or graph and can be downloaded in a CSV file. -The system has the capability to generate automatic reports based on historical information stored from each device in the installation, producing an Excel or PDF file.
Dataset Purpose	Provide real data to enhance and develop the COCOON cybersecurity tools for photovoltaic plants. This data is crucial for identifying potential security threats, validating cybersecurity protocols, and improving the overall resilience of the photovoltaic infrastructure against cyber-attacks.
Dataset Type/Format	csv, excel, PDF, png
Re-use of existing data	No
Dataset Origin	This dataset is generated by the operation in HG1 PV plant.
Dataset Collection	The dataset is collected mainly using inverters, smart meter measurements (via the Modbus/TCP protocol), and other devices connected to the SCADA and the monitoring and control system.
Data size	It will depend on the time and the variables selected (whether it is historical or real-time data. 20 KB – 1MB
COCOON Work Packages, Tasks and Deliverables	
Relevant Work Packages	WP8
Relevant Tasks	T8.1, T8.2, T8.3
Relevant Deliverables	D8.1, D8.2, D8.3
Partners Services and Responsibilities	
Partner Owner	CUE
Partners responsible for data collecting	CUE, ING, USE
Partners responsible for data analysis	ING, USE
Partners responsible for data storage	CUE, ING, USE
FAIR Data - Making data findable, including provisions	
Discoverable and Identifiable Data	To be provided
Naming Convention	Measurements
Versioning	v1

Search Keywords	measurements, FV Plant, variables, operational data
Metadata	• Relevant Publication: TBD • Last Update: TBD • DOI: TBD • Data Format *.csv, *.xlsx *.png
FAIR Data - Making data openly accessible	
Dataset Openly Accessible	TBD
Methods/Tools for accessing the data	Excel, Python
Access Restrictions	TBD
Repository	Not available yet.
FAIR Data - Making interoperable	
Interoperability	To be provided
Standards	To be provided
FAIR Data – Increase data re-use	
Licence	to be determined
Data sharing terms	-
Dataset sharing after the end of the project	Maybe
Preservation and update of the dataset after the project	Yes
Re-use timeframe	-
Allocation of Resources	
Cost for making the dataset FAIR	To be determined at a later stage in the project
Costs for the preservation (including backup) and the update of the dataset	No cost
Data Security	
Security Measures	The dataset refers to the operational data of a PV plant
Privacy Measure	-
Ethical Aspects	
Ethical and Legal Aspects	To be provided
Consent form for data sharing and long preservation within the COCOON consortium	To be provided
Consent form for sharing the dataset publicly.	To be provided
Other Issues	No

Table A- 3: ING dataset with measurements from actual or virtual power system setups

Dataset Name	**ING dataset with measurements from actual or virtual power system setups**
Dataset Summary	
Dataset Description	The ING dataset with measurements from actual or virtual power system setups includes logs (.log) files, GZIP compressed logs (.log.gz) files, Comma-Separated Value (CSV) files, GZIP compressed TAR (tar.gz) files and Portable Network Graphics (PNG). The data is generated by a PPC module called *Ghostwriter*. This module is activated during the commissioning process in order to validate the inverters' time response and as part of a quality report for the client. The way this module works is by recording data from the PPC. When a setpoint is sent, the *Ghostwriter* records the following parameters: ● Date ● Active Power POI ● Reactive Power POI ● Voltage POI ● Power Factor POI ● Frequency POI ● Reference Active Power ● Reference Reactive Power ● Reference Voltage ● Reference Power Factor ● Reference Active Power PFR ● Voltage Reference with Droop ● Reference Active Power AGC ● Active Power Ramp Slope ● Active Power Control Loop Time ● Plant Nominal Power ● Plant Nominal Voltage ● QV Setpoint Voltage ● Time taken to reach setpoint These parameters are recorded over a period of time, showing a representative behavior of the control type we are testing. The controls tested include active power control, reactive power control, QV control, power factor, and frequency. The following image shows an example of the data generated by the *Ghostwriter* module:

	 Once we have obtained these log files, a Python script is used to convert the files into .csv format and then into .png graphics, allowing us to analyze the data. The following image show an example of each type of file: - .csv:  - .png:  The dataset contains measurements from inverters and from smart meters using Modbus/TCP protocol.
Dataset Purpose	Testing and validation of the time response of inverters according to the country's grid code requirements when the PPC sends a setpoint to them.
Dataset Type/Format	.csv, .log, .log.gz, .tar.gz, .png
Re-use of existing data	No
Dataset Origin	This dataset is generated using a PPC module called *Ghostwriter*.
Dataset Collection	This dataset is collected using: A controller connected to receive all the measurements /set-points from actual or virtual MODBUS /TCP compatible devices (i.e., smart meter, inverter, etc).
Data size	~20 KB
COCOON Work Packages, Tasks and Deliverables	
Relevant Work Packages	WP8
Relevant Tasks	T8.1
Relevant Deliverables	D8.1
Partners Services and Responsibilities	
Partner Owner	ING
Partners responsible for data collecting	ING
Partners responsible for data analysis	ING
Partners responsible for data storage	ING
FAIR Data - Making data findable, including provisions	

Discoverable and Identifiable Data	To be provided
Naming Convention	results.tar.gz
Versioning	v1
Search Keywords	measurements, Modbus
Metadata	Relevant Publication: TBD Last Update: TBD DOI: TBD Data Format *.csv, *.log, *.log.gz, *.tar.gz, *.png
FAIR Data - Making data openly accessible	
Dataset Openly Accessible	TBD
Methods/Tools for accessing the data	Python
Access Restrictions	TBD
Repository	Not available yet.
FAIR Data - Making interoperable	
Interoperability	To be provided
Standards	To be provided
FAIR Data – Increase data re-use	
Licence	To be determined
Data sharing terms	-
Dataset sharing after the end of the project	Maybe
Preservation and update of the dataset after the project	Yes
Re-use timeframe	-
Allocation of Resources	
Cost for making the dataset FAIR	To be determined at a later stage in the project
Costs for the preservation (including backup) and the update of the dataset	No cost
Data Security	
Security Measures	The dataset does not refer to any actual entity participating in a real industrial environment.
Privacy Measure	-
Ethical Aspects	
Ethical and Legal Aspects	To be provided
Consent form for data sharing and long preservation within the COCOON consortium	To be provided
Consent form for sharing the dataset publicly.	To be provided
Other Issues	No

Table A-4: ENCS dataset with network traffic from the cyber range

Dataset Name	ENCS dataset with pcap files from traffic from the cyber range
Dataset Summary	
Dataset Description	The ENCS dataset with network traffic from the cyber range as Packer CAPture files (.pcap). The data is captured from virtual equipment emulated within a lab setup at ENCS.
Dataset Purpose	Testing and validation of plethora of power system applications, including cyberthreat analysis related to Tasks 1.2 and 1.3
Dataset Type/Format	.pcap
Re-use of existing data	No
Dataset Origin	The data is captured from virtual equipment emulated within a lab setup at ENCS'cyber range.
Dataset Collection	This dataset is collected using: - All the network traffic is collected using WireShark software.
-	To be determined
COCOON Work Packages, Tasks and Deliverables	
Relevant Work Packages	WP3 Cyber physical security integration and coordination, WP7: Digital Substation Security
Relevant Tasks	T3.4 and T3.5: Testing the cyber-physical anomaly diagnosis for OT networks in digital substation, testing the cyber secure, real-time protection schemes for digital substations T7.4 Demonstration of cyber security and resilience of digital substations
Relevant Deliverables	D3.3 Cyber-physical anomaly diagnosis for OT networks and Cyber secure, real-time protection schemes in digital substations D7.3 Demonstrator Final Report: Secure Digital Substations
Partners Services and Responsibilities	
Partner Owner	ENCS
Partners responsible for data collecting	ENCS
Partners responsible for data analysis	ENCS and maybe others to be defined
Partners responsible for data storage	ENCS
FAIR Data - Making data findable, including provisions	
Discoverable and Identifiable Data	To be provided
Naming Convention	ENCS_cyberrange_traffic_Dataset_v1
Versioning	v1

Search Keywords	Network traffic, cyber-range, pcap
Metadata	● Relevant Publication: TBD ● Last Update: TBD ● DOI: TB ● Data Format *.pcap
FAIR Data - Making data openly accessible	
Dataset Openly Accessible	No
Methods/Tools for accessing the data	Wireshark, Tshark, Pandas, Python
Access Restrictions	TU Delft only
Repository	Not available yet.
FAIR Data - Making interoperable	
Interoperability	To be provided
Standards	To be provided
FAIR Data – Increase data re-use	
Licence	To be determined
Data sharing terms	Please cite the corresponding paper. More specific terms will be provided soon
Dataset sharing after the end of the project	No
Preservation and update of the dataset after the project	No
Re-use timeframe	-
Allocation of Resources	
Cost for making the dataset FAIR	To be determined at a later stage in the project
Costs for the preservation (including backup) and the update of the dataset	No cost
Data Security	
Security Measures	To be provided
Privacy Measure	-
Ethical Aspects	
Ethical and Legal Aspects	To be provided
Consent form for data sharing and long preservation within the COCOON consortium	To be provided
Consent form for sharing the dataset publicly.	To be provided
Other Issues	No

Table A-5: TUD dataset with measurements from actual or virtual substation setups

Dataset Name	TUD dataset with measurements from actual or virtual substation setups
Dataset Summary	
Dataset Description	The TU Delft dataset includes measurements from simulated power system setups and actual power system devices. It contains Common Separate Value (CSV) files related to Intelligent Electronic Devices (IEDs), Realtime Digital Simulator (RTDS) or Merging Units (MU), and network traffic data as Packer Capture files (.pcap). The data is captured from actual equipment connected in several laboratory setups, as well as from virtual and real equipment. The dataset contains substation measurements following the IEC 61850 protocol, measurements from RTDS simulated devices and IEDs using MMS/SV/GOOSE protocols, and data from network devices using TCP/IP-based protocols. The dataset would also contain IED configuration files, including Substation configuration Language files (SCL).
Dataset Purpose	Testing and validation of plethora of power system applications, including Anomaly diagnosis, Hash based mac authentication analysis related to Tasks 1.4, 2.4, 3.3, 3.4, 3.5 and 7.1
Dataset Type/Format	.csv, .txt, .xlsx,.pcap, SCL files
Re-use of existing data	No
Dataset Origin	This dataset is generated using actual or virtual equipment from the TU Delft, ESP Laboratory.
Dataset Collection	This dataset is collected using: - A PC/controller connected in the lab for receiving all the measurements /set-points from actual or virtual TCP/IP compatible devices (i.e., RTDS, IED, PC, etc) - A device connected at the local network to receive the GOOSE traffic from the intelligent electronic devices (i.e., protection relays). - All the network traffic is collected using WireShark software.
-	10-50 GB
COCOON Work Packages, Tasks and Deliverables	
Relevant Work Packages	WP1: Cyber security Algorithmics WP2: Cybersecurity assessment considering power system characteristics.

	WP3: Cyber-physical security integration and coordination. WP7: Digital Substation Security.
Relevant Tasks	T1.4: Anomaly Diagnosis. T2.4: Cyber secure, real-time protection schemes for digital substations. T3.3, T3.4, T3.5: Developing a real-time digital twin of the power grid for cyber security of digital substations, Testing the cyber-physical anomaly diagnosis for OT networks in digital substations, Testing the cyber secure, real-time protection schemes for digital substations. T7.1: Pilot design and configuration.
Relevant Deliverables	D1.3, D1.4, D2.3, D3.3, D7.1
Partners Services and Responsibilities	
Partner Owner	TUD
Partners responsible for data collecting	TUD
Partners responsible for data analysis	TUD and maybe others to be defined
Partners responsible for data storage	TUD
FAIR Data - Making data findable, including provisions	
Discoverable and Identifiable Data	To be provided
Naming Convention	TUD_Measurements _Dataset_v1.
Versioning	v1
Search Keywords	measurements, Anomaly diagnosis, substation, Configuration, IEC 61850-GOOSE/MMS/SV, SCADA
Metadata	● Relevant Publication: TBD ● Last Update: TBD ● DOI: TB ● Data Format *.csv, *.pcap, .xlsx*, .txt*
FAIR Data - Making data openly accessible	
Dataset Openly Accessible	TBD
Methods/Tools for accessing the data	Wireshark, Tshark, Pandas, Python, Matlab, Excel
Access Restrictions	TBD
Repository	Not available yet.
FAIR Data - Making interoperable	
Interoperability	To be provided
Standards	To be provided
FAIR Data – Increase data re-use	
Licence	to be determined
Data sharing terms	Please cite the corresponding paper. More specific terms will be provided soon
Dataset sharing after the end of the project	Maybe

Preservation and update of the dataset after the project	Yes
Re-use timeframe	-
Allocation of Resources	
Cost for making the dataset FAIR	To be determined at a later stage in the project
Costs for the preservation (including backup) and the update of the dataset	To be determined
Data Security	
Security Measures	The dataset does not refer to any actual entity participating in a real industrial environment.
Privacy Measure	-
Ethical Aspects	
Ethical and Legal Aspects	To be provided
Consent form for data sharing and long preservation within the COCOON consortium	To be provided
Consent form for sharing the dataset publicly.	To be provided
Other Issues	No

Table A-6: HEDNO dataset with measurements from actual Distribution Network

Dataset Name	HEDNO dataset with measurements from actual Distribution Network
Dataset Summary	
Dataset Description	The HEDNO dataset with measurements from the distribution network in excel (.xlsx) files. The dataset includes measurements that conform to the DLMS/COSEM protocol (IEC 62053-22[20] and IEC 62053-23[21]). For the HEDNO dataset, the metering is conducted using the ITRON SL7000[22] measuring device, combined with a 400/5 A current transformer and a 30/5 V voltage transformer to ensure accurate readings. The measured values are three phase active power, three phase reactive power, line to line voltage for all three phases and line current for all three phases. The measurements have a resolution of 15 minutes and correspond to the period 01/05/2024-31/05/2024. A zip file named “Historical Measurements” contains two subfolders: one named “P, Q” and a second one named “V, I”. The “P, Q” subfolder contains twelve (12) .xlsx files.

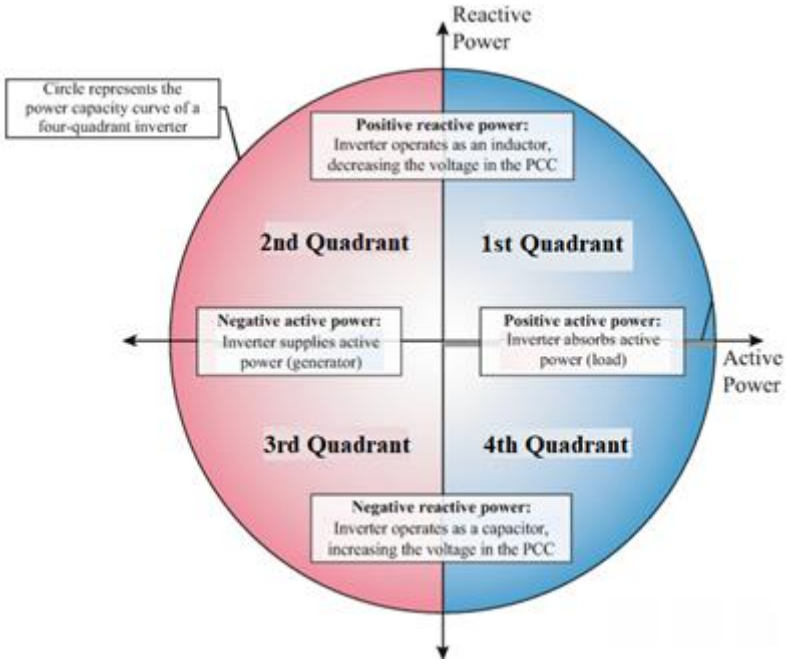
[20] https://webstore.iec.ch/en/publication/29987
[21] https://webstore.iec.ch/en/publication/29239
[22] https://emea.itron.com/products/ace-sl7000

Six (6) files correspond to the substation measurements (head of the feeder). The naming of these files begins with "82580075_Αίτημα15λέπτων_...". The head of the feeder contains two measuring devices. Another six (6) files contain the measurements of the PV systems. Their naming begins with "Αίτημα15λέπτων_...".

Each of the 6 files (either the ones corresponding to the substation or the PVs) has a different naming depending on the active or reactive power value that is measured.

- "…LpAct+…" is the active power that is consumed by the PV auxiliary loads (lighting, auxiliary devices)
- "…LpAct-…" is the active power that is produced by the PV systems
- "…LpRe…" is the reactive power in the 1st quadrant (reactive power absorbed by the PV system when the PV system loads consume active power)
- "…LpQ2…" is the reactive power in the 2nd quadrant (reactive power absorbed by the PV system when the PV system produces active power)
- "…LpQ3…" is the reactive power in the 3rd quadrant (reactive power injected by the PV system when the PV system produces active power)
- "…LpQ4…" is the reactive power in the 4th quadrant (reactive power injected by the PV system when the PV system loads consume active power)

The convention of the reactive power quartiles is depicted here:



The "V, I" subfolder contains two (2) .xlsx files, one with the PV node voltages and one with the PV line currents. Each .xlsx file contains

	measured values at the nodes of the PV systems and at the head of the feeder. Since there have been some inaccuracy issues in the voltage and current measurements, another similar dataset named "Historical Measurements (Corrected Values).zip" contains the correct values of current and voltages. Moreover, each .xlsx file described above contains two sheets, "…_measured" and "…_real". The first one corresponds to the measured values of the measuring devices and the second one to the real value considering that there is a transformer in the measuring device. The datasets are accompanied with the distribution network topology that is provided by HEDNO in .pfd format (DigSilent PowerFactory[23]). This .pfd file is used for simulations with respect to the state estimation algorithm of WP2/Task 2.3. Based on that topology, a Simulink model has been built to serve as a basis for the lab set up of WP3/Task 3.2.
Dataset Purpose	Testing and validation of the state estimation algorithms of energy communities developed in Task 2.3, the appropriate testing in Task 3.2 and the Pilot Validation in WP5
Dataset Type/Format	.xlsx
Re-use of existing data	No
Dataset Origin	This dataset is generated using HEDNO measuring equipment (ITRON SL7000) at the local distribution network.
Dataset Collection	Data from each PV plant is collected using a Remote Terminal Unit (RTU). The RTU functions as an intermediary, connecting the PV plant to the HEDNO Supervisory Control and Data Acquisition (SCADA) system, which oversees and manages the entire PV network. This connection is established via an industrial gateway (namely the Moxa OnCell 3120-LTE-1 Series), which is responsible for transmitting PV telemetry data to the HEDNO SCADA.
Data Size	16 MB
COCOON Work Packages, Tasks and Deliverables	
Relevant Work Packages	WP2: Cybersecurity assessment considering power system characteristics WP3: Cyber-physical security integration and coordination WP5: Secure Energy Communities
Relevant Tasks	T2.3: Cybersecurity assessment in energy communities T3.2: Testing of cyber-physical security integration in energy communities All Tasks of WP5: T5.1: Pilot design and configuration T5.2: Frequency-related Ancillary Services T5.3: Voltage-related Ancillary Services

[23] https://www.digsilent.de/en/powerfactory.html

Relevant Deliverables	D2.2 Report on the refinement of the generic power grid state estimation for PV plants and energy communities D3.2 Cyber-physical security integration in renewable power plants and energy communities D5.1 Secure Energy Communities Pilot Design, AS orchestration and infrastructure Configuration D5.2 Report on COCOON Toolset monitoring, vulnerability assessment and risk profiling evaluation for secure energy communities pilot D5.3 Demonstrator final report: Secure Energy Communities
Partners Services and Responsibilities	
Partner Owner	HEDNO
Partners responsible for data collecting	HEDNO, I&K
Partners responsible for data analysis	AUTH
Partners responsible for data storage	AUTH, UCY, Selene-CC
FAIR Data - Making data findable, including provisions	
Discoverable and Identifiable Data	To be provided
Naming Convention	Historical Measurements.zip
Versioning	v1
Search Keywords	measurements, energy community, Photovoltaic Plants
Metadata	● Relevant Publication: TBD ● Last Update: TBD ● DOI: TBD ● Data Format *.xlsx*, .txt*
Geographic location of data collection	The data were generated from measuring devices ITRON SL7000 in the local distribution network of the energy community.
Repository	Microsoft SharePoint
Path	SharePoint Paths: ● for the Distribution Network in pfd (DigSilent PowerFactory) format: Documents\COCOON\WP2\T2.3\Energy community data, files “EC_Network.pfd” and “EC_Network_reduced.pfd” ● for the Dataset in .xlsx format: Documents\COCOON\WP5\HEDNO Data, files “Historical Measurements.zip” and “Historical Measurements (Corrected Values).zip” ● for the I&K data with respect to the PV converters and PV plants configuration in pdf format: Documents\COCOON\WP5\I&K Data ● for the Energy Community Topology in Simulink (.slx) to serve as a basis for the USE lab set-up in WP3/T3.2:

	Documents\COCOON\WP3\Task 3.2\Simulink Set Up of Energy Community, file "test_EC_WP3_v4_all_RL_Constant_Q_2023b.slx"
Version Number	V0.1
FAIR Data - Making data openly accessible	
Dataset Openly Accessible	TBD
Methods/Tools for accessing the data	Excel, Python, Matlab
Access Restrictions	TBD
Repository	Not available yet.
FAIR Data - Making interoperable	
Interoperability	To be provided
Standards	To be provided
FAIR Data – Increase data re-use	
Licence	to be determined
Data sharing terms	-
Dataset sharing after the end of the project	Maybe
Preservation and update of the dataset after the project	Yes
Re-use timeframe	-
Allocation of Resources	
Cost for making the dataset FAIR	To be determined at a later stage in the project
Costs for the preservation (including backup) and the update of the dataset	No cost
Data Security	
Security Measures	The dataset does not refer to any actual entity participating in a real industrial environment.
Privacy Measure	-
Ethical Aspects	
Ethical and Legal Aspects	To be provided
Consent form for data sharing and long preservation within the COCOON consortium	To be provided
Consent form for sharing the dataset publicly.	To be provided
Other Issues	No

Table A-7: SEL dataset for DevOps

Dataset Name	SEL dataset for DEvOps
Dataset Summary	
Dataset Description	The SEL dataset includes logs and documentation files (JSON, XML, YAML) related to the management of project data on GitHub and within the DevOps pipeline. This encompasses commit histories, pull request logs, continuous integration/continuous deployment (CI/CD) pipeline logs, and issue tracking data. The logs capture developer activities, code changes, and automated build/test/deployment processes.
Dataset Purpose	To monitor code changes, ensure traceability of development activities, and maintain a historical record of automated build and deployment processes for timely software delivery.
Dataset Type/Format	.json, .yaml, .log, .xml
Re-use of existing data	No.
Dataset Origin	Generated automatically by GitLab and CI/CD tools.
Dataset Collection	Collected using GitLab APIs, CI/CD tool logs, and exported reports
COCOON Work Packages, Tasks and Deliverables	
Relevant Work Packages	WP4: Software Prototype Development
Relevant Tasks	T4.1 – DevOps
Relevant Deliverables	D4.1 COCOON Development Blueprint
Partners Services and Responsibilities	
Partner Owner	SEL
Partners responsible for data collecting	SEL and UCY
Partners responsible for data analysis	SEL and UCY
Partners responsible for data storage	SEL
FAIR Data - Making data findable, including provisions	
Discoverable and Identifiable Data	To be provided
Naming Convention	SEL_Measurements_Dataset_v1
Versioning	v1
Search Keywords	GitLab, DevOps, CI/CD, logs, commit history
Metadata	• Relevant Publication: TBD • Last Update: TBD • DOI: TBD • Data Format *.json, *.yml, *.log, *.xml
FAIR Data - Making data openly accessible	

Dataset Openly Accessible	TBD
Methods/Tools for accessing the data	Jenkins, GitLab, Python
Access Restrictions	TBD
Repository	Not available yet.
FAIR Data - Making interoperable	
Interoperability	To be provided
Standards	To be provided
FAIR Data – Increase data re-use	
Licence	To be determined
Data sharing terms	Please cite the corresponding paper. More specific terms will be provided soon
Dataset sharing after the end of the project	Maybe
Preservation and update of the dataset after the project	Yes
Re-use timeframe	-
Allocation of Resources	
Cost for making the dataset FAIR	To be determined at a later stage in the project
Costs for the preservation (including backup) and the update of the dataset	No cost
Data Security	
Security Measures	The dataset does not refer to any actual entity participating in a real industrial environment.
Privacy Measure	-
Ethical Aspects	
Ethical and Legal Aspects	To be provided
Consent form for data sharing and long preservation within the COCOON consortium	To be provided
Consent form for sharing the dataset publicly.	To be provided
Other Issues	No

Table A-8: SEL dataset for SharePoint

Dataset Name	SEL dataset for SharePoint
Dataset Summary	

Dataset Description	The SEL dataset encompasses data related to document and file management on the SharePoint platform. This includes user access logs, document version histories, and metadata about file uploads and downloads. The logs capture activities such as file creation, modification, deletion, and access permissions.
Dataset Purpose	To enhance collaboration, enable document sharing, track user activities, and maintain a historical record of document versions within the COCOON project.
Dataset Type/Format	.docx, .xlsx, .pptx, .pdf,
Re-use of existing data	No.
Dataset Origin	Uploaded and managed directly by project members using the SharePoint interface.
Dataset Collection	Collected manually by users and stored within SharePoint's folder structure.
COCOON Work Packages, Tasks and Deliverables	
Relevant Work Packages	All work packages
Relevant Tasks	All Tasks
Relevant Deliverables	All Deliverables
Partners Services and Responsibilities	
Partner Owner	SEL
Partners responsible for data collecting	SEL
Partners responsible for data analysis	SEL
Partners responsible for data storage	SEL
FAIR Data - Making data findable, including provisions	
Discoverable and Identifiable Data	To be provided
Naming Convention	SEL_Measurements_Dataset_v1
Versioning	v1
Search Keywords	SharePoint, data management, access, version history
Metadata	● Relevant Publication: TBD ● Last Update: TBD ● DOI: TBD ● Data Format *.docx, *.xlsx, *.pptx, *.pdf,
FAIR Data - Making data openly accessible	
Dataset Openly Accessible	TBD
Methods/Tools for accessing the data	SharePoint
Access Restrictions	TBD
Repository	Not available yet.
FAIR Data - Making interoperable	
Interoperability	To be provided

Standards	To be provided
FAIR Data – Increase data re-use	
Licence	to be determined
Data sharing terms	Please cite the corresponding paper. More specific terms will be provided soon
Dataset sharing after the end of the project	Maybe
Preservation and update of the dataset after the project	Yes
Re-use timeframe	-
Allocation of Resources	
Cost for making the dataset FAIR	To be determined at a later stage in the project
Costs for the preservation (including backup) and the update of the dataset	No cost
Data Security	
Security Measures	MFA required for external users; access controlled and monitored through SIEM; behaviour monitoring tools in place to detect suspicious activity.
Privacy Measure	-
Ethical Aspects	
Ethical and Legal Aspects	To be provided
Consent form for data sharing and long preservation within the COCOON consortium	To be provided
Consent form for sharing the dataset publicly.	To be provided
Other Issues	No

Table A-9: SEL dataset for Firewall Logs

Dataset Name	SEL dataset for Firewall Logs
Dataset Summary	
Dataset Description	The SEL dataset includes log files generated by firewalls in a replicated setup environment. The logs capture network traffic, firewall rule matches, and potential intrusion attempts. This includes data from different layers (network, transport, application) and log entries related to rule bypass attempts by threat models.
Dataset Purpose	To analyze and validate firewall effectiveness against various threat models, understand bypass techniques, and improve CPN for enhanced security.
Dataset Type/Format	.log, .pcap, .txt
Re-use of existing data	No.
Dataset Origin	Generated by network monitoring tools and firewall device in a controlled test environment.
Dataset Collection	Collected using firewall logging tools and network traffic capture software
COCOON Work Packages, Tasks and Deliverables	
Relevant Work Packages	WP6: Secure Regional Electricity Data Operations
Relevant Tasks	T6.1: Pilot Architecture, T6.2: Threat Model Implementation & T6.3: Evaluation
Relevant Deliverables	D6.1: Secure Regional Electricity Data Pilot Design and Configuration, D6.2: Report on COCOON Toolset monitoring, vulnerability assessment and risk profiling evaluation for Secure Regional Electricity Data Pilot & D6.3: Demonstrator final report: Secure Regional Electricity Data
Partners Services and Responsibilities	
Partner Owner	SEL
Partners responsible for data collecting	SEL
Partners responsible for data analysis	SEL and maybe others to be defined
Partners responsible for data storage	SEL
FAIR Data - Making data findable, including provisions	
Discoverable and Identifiable Data	To be provided
Naming Convention	SEL_Measurements_Dataset_v1
Versioning	v1
Search Keywords	firewall, logs
Metadata	• Relevant Publication: TBD • Last Update: TBD • DOI: TB • Data Format *.log, *.pcap, *.txt

FAIR Data - Making data openly accessible	
Dataset Openly Accessible	TBD
Methods/Tools for accessing the data	Wireshark, Tshark, Python, Excel
Access Restrictions	TBD
Repository	Not available yet.
FAIR Data - Making interoperable	
Interoperability	To be provided
Standards	To be provided
FAIR Data – Increase data re-use	
License	to be determined
Data sharing terms	Please cite the corresponding paper. More specific terms will be provided soon
Dataset sharing after the end of the project	Maybe
Preservation and update of the dataset after the project	Yes
Re-use timeframe	-
Allocation of Resources	
Cost for making the dataset FAIR	To be determined at a later stage in the project
Costs for the preservation (including backup) and the update of the dataset	No cost
Data Security	
Security Measures	The dataset does not refer to any actual entity participating in a real industrial environment.
Privacy Measure	-
Ethical Aspects	
Ethical and Legal Aspects	To be provided
Consent form for data sharing and long preservation within the COCOON consortium	To be provided
Consent form for sharing the dataset publicly.	To be provided
Other Issues	No

Table A-10: USE dataset of the cyber physical power system laboratory

Dataset Name	**USE dataset of the cyber physical power system laboratory**
Dataset Summary	
Dataset Description	This USE dataset comprises all the data regarding the testing of the COCOON solutions in the cyber physical power system laboratory consisting of: a) Static information about the network topology and its corresponding parameters. b) Real-time measurements of relevant electrical magnitudes c) OT traffic captures. All this information will be particularized for each case study tested in the laboratory. The type of file within the dataset depends on the information to be shared, particularly: a) Static information about the network topology and its corresponding parameters: .txt files and .png files (one-line diagrams). b) Real-time measurements of relevant electrical magnitudes: csv files. c) OT traffic captures: .pcap files.
Dataset Purpose	The purpose of the dataset is to collect the performance in the laboratory environment of the COCOON solutions to prevent cyberattacks based on false data injection (FDI).
Dataset Type/Format	.csv, .txt, .png, .pcap
Re-use of existing data	No
Dataset Origin	Laboratory testing
Dataset Collection	The dataset is collected depending on the type of information: a) Static information about the network topology and its corresponding parameters: manually. b) Real-time measurements of relevant electrical magnitudes: commercial SCADA available at the laboratory which collects the relevant electrical magnitudes from the injection nodes. c) OT traffic captures: wireshark.
Data size	~1 GB
COCOON Work Packages, Tasks and Deliverables	
Relevant Work Packages	WP3
Relevant Tasks	T3.1 and T3.2
Relevant Deliverables	D3.2
Partners Services and Responsibilities	
Partner Owner	USE
Partners responsible for data collecting	USE
Partners responsible for data analysis	USE, ING, CUE, AUTH, ING, HEDNO

Partners responsible for data storage	USE
FAIR Data - Making data findable, including provisions	
Discoverable and Identifiable Data	To be provided
Naming Convention	USE_measurements_EC_test#_dataset_v1 USE_measurements_PV_test#_dataset_v1
Versioning	v1
Search Keywords	Electrical measurements, PV power plant, energy community, false data injection
Metadata	Relevant Publication: TBD Last Update: TBD DOI: TBD Data Format *.csv, *.txt, *.pcap, *.png
FAIR Data - Making data openly accessible	
Dataset Openly Accessible	TBD
Methods/Tools for accessing the data	Excel, Python, text editor
Access Restrictions	TBD
Repository	Not available yet.
FAIR Data - Making interoperable	
Interoperability	To be provided
Standards	To be provided
FAIR Data – Increase data re-use	
Licence	TBD
Data sharing terms	-
Dataset sharing after the end of the project	TBD
Preservation and update of the dataset after the project	TBD
Re-use timeframe	-
Allocation of Resources	
Cost for making the dataset FAIR	To be determined at a later stage in the project
Costs for the preservation (including backup) and the update of the dataset	No cost
Data Security	
Security Measures	The dataset does not refer to any actual entity participating in a real industrial environment.
Privacy Measure	-
Ethical Aspects	
Ethical and Legal Aspects	To be provided
Consent form for data sharing and long preservation within the COCOON consortium	To be provided
Consent form for sharing the dataset publicly.	To be provided
Other Issues	No

7 References

[1] https://zenodo.org/communities/

[2] Zenodo, “ZENODO-GeneralPolicies”, [Available online]: https://about.zenodo.org/policies/ (last Accessed on 28/08/2024).

[3] European Commission, “” Guidelines for Researchers: How to comply with HORIZON Europe mandate for publications”, Available [Online]: https://www.openaire.eu/how-to-comply-with-horizon-europe-mandate-for-publications (last Accessed on 28/08/2024)

[4] https://www.cyber-cocoon.eu/

[5] https://gdpr.eu/what-is-gdpr/