D1.7
Data Management Plan (v2)

DOCUMENT	D1.7	**WORKPACKAGE**	WP1
DELIVERABLE STATE	ToC/DRAFT/ REVIEWED/ RELEASE CANDIDATE /**FINAL**	**PROGRAMME IDENTIFIER**	H2020-SU-DS-2020
REVISION	V1.0	**GRANT AGREEMENT ID**	101020560
DELIVERY DATE	31/03/2023	**PROJECT START DATE**	01/10/2021
DISSEMINATION LEVEL	**PU** / CO / EU-RES / EU–CON / EU-SEC	**DURATION**	3 YEARS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 101020560

DISCLAIMER

ACKNOWLEDGEMENT

This document is a deliverable of CyberSEAS project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 101020560.

The opinions expressed in this document reflect only the author's view and in no way reflect the European Commission's opinions. The European Commission is not responsible for any use that may be made of the information it contains.

PROJECT ACRONYM	CyberSEAS
PROJECT TITLE	Cyber Securing Energy dAta Services
CALL ID	H2020-SU-DS-2020
CALL NAME	Digital Security (H2020-SU-DS-2018-2019-2020)
TOPIC	SU-DS04-2018-2020 Cybersecurity in the Electrical Power and Energy System (EPES): an armour against cyber and privacy attacks and data breaches
TYPE OF ACTION	Innovation Action
COORDINATOR	ENGINEERING – INGEGNERIA INFORMATICA SPA (ENG)
PRINCIPAL CONTRACTORS	CONSORZIO INTERUNIVERSITARIO NAZIONALE PER L'INFORMATICA (CINI), AIRBUS CYBERSECURITY GMBH (ACS), FRAUNHOFER GESELLSCHAFT ZUR FOERDERUNG DER ANGEWANDTEN FORSCHUNG E.V. (FRAUNHOFER), GUARDTIME OU (GT), IKERLAN S. COOP (IKE), INFORMATIKA INFORMACIJSKE STORITVE IN INZENIRING DD (INF), INSTITUT ZA KORPORATIVNE VARNOSTNE STUDIJE LJUBLJANA (ICS), RHEINISCH-WESTFAELISCHE TECHNISCHE HOCHSCHULE AACHEN (RWTH), SOFTWARE IMAGINATION & VISION SRL (SIMAVI), SOFTWARE QUALITY SYSTEMS SA (SQS), STAM SRL (STAM), SYNELIXIS LYSEIS PLIROFORIKIS AUTOMATISMOU & TILEPIKOINONION ANONIMI ETAIRIA (SYN), WINGS ICT SOLUTIONS INFORMATION & COMMUNICATION TECHNOLOGIES IKE (WIN), ZIV APLICACIONES Y TECNOLOGIA SL (ZIV), COMUNE DI BERCHIDDA (BER), COMUNE DI BENETUTTI (BEN), ELES DOO SISTEMSKI OPERATER PRENOSNEGA ELEKTROENERGETSKEGA OMREZJA (ELES), PETROL SLOVENSKA ENERGETSKA DRUZBA DD LJUBLJANA (PET), AKADEMSKA RAZISKOVALNA MREZA SLOVENIJE (ARN), HRVATSKI OPERATOR PRIJENOSNOG SUSTAVA DOO (HOPS), ENERIM OY (ENERIM), ELEKTRILEVI OU (ELV), COMPANIA NATIONALA DE TRANSPORT ALENERGIEI ELECTRICE TRANSELECTRICA SA (TEL), CENTRUL ROMAN AL ENERGIEI (CRE), TIMELEX (TLX).
WORKPACKAGE	WP1
DELIVERABLE TYPE	R Document, report DEM Demonstrator, pilot, prototype DEC Websites, patent fillings, videos, etc. OTHER ETHICS Ethics requirement **ORDP Open Research Data Pilot** DATA data sets, microdata, etc.
DISSEMINATION LEVEL	**PU Public** CO Confidential, only for members of the consortium (including the Commission Services) EU-RES Classified Information: RESTREINT UE (Commission Decision 2005/444/EC) EU-CON Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)

	EU-SEC Classified Information: SECRET UE (Commission Decision 2005/444/EC)
DELIVERABLE STATE	**FINAL**
CONTRACTUAL DATE OF DELIVERY	31/03/2023
ACTUAL DATE OF DELIVERY	15/05/2023
DOCUMENT TITLE	Data Management Plan (v2)
AUTHOR(S)	CINI
REVIEWER(S)	Fraunhofer, SQS
ABSTRACT	SEE EXECUTIVE SUMMARY
HISTORY	SEE DOCUMENT HISTORY
KEYWORDS	Data Management, Data Quality

Document History

Version	Date	Contributor(s)	Description
V0.1	10/03/2023	CINI	First draft
V0.2	03/04/2023	ENERIM	First contribution on the Finnish pilot data
V0.3	27/04/2023	ENERIM	Final contribution on the Finnish pilot data
V0.4	07/05/2023	CINI	Prefinal version
V1.0	15/05/2022	CINI	Final version

Table of Contents

List of Figures

List of Tables

N/A

List of Acronyms and Abbreviations

AMQP	Advanced Message Queuing Protocol
AUVP	Advanced Update Validation Platform
BSP	Balancing Service Provider
DACF	Day-Ahead Congestion Forecast
DMP	Data Management Plan
DLAI	Dynamic Line Anti-icing
DLR	Dynamic Line Rating
DoA	Description of Action
DPTR	Dynamic Power Transformer Rating
EC	European Commission
EMS	Energy Management System
EPES	Electrical Power and Energy System
FAIR	Findable, Accessible, Interoperable, and Reusable
ICCP	Inter-Control Center Communications Protocol
IDCF	Intraday Congestion Forecast
IEC	International Electrotechnical Commission
IED	Intelligent Electronic Device
IP	Intellectual Property
MIDA	Machine Integrity, Defense, and Awareness
OT	Operational Technology
RDF	Resource Description Framework
RTU	Remote Terminal Unit
SCADA	Supervisory Control And Data Acquisition
SFTP	Secure File Transfer Protocol
SOAP	Simple Object Access Protocol
TASE	Telecontrol Application Service Element
TSO	Transmission System Operator
XML	Extensible Markup Language
WP	Work Package

1 Executive Summary

Data Management Plans (DMPs) are considered to be a key element to sound data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 Project.

The goal of this document is to provide an update of the initial DMP for the CyberSEAS Project. Specifically, a detailed description of the Finnish pilot data is given as a new contribution with respect to first version of DMP.

This DMP is based on the DMP template provided by the European Commission (EC) and follows the "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020". Furthermore, in accordance with the requirements stated in the description of action covers the following aspects:

- Description of the data to be collected or created during research and solution deployment and piloting, including metadata.
- Standards and methodologies for data collection and management and quality assurance measures.
- Plans for data sharing and access.
- Copyrights and intellectual property of data.
- Data storage and back-up measures.
- Data management roles and responsibilities.

The present document constitutes the second and final version of the DMP and will be submitted within the first eighteen months of the Project.

2 Introduction

2.1 Purpose and scope

Purpose of this DMP is to:

- Create a document, which explains the management of data collected during the Project.
- Support the data management life cycle for all data that will be collected, processed or generated by the Project.
- Provide an analysis of the main elements of the data management policy, which will be used by the Partners regarding all the datasets which will be generated by the Project.
- Provide details and assurances about the preservation of the data collected during the Project, as well as any results derived from the associated research.
- Provide details on how the CyberSEAS consortium plans to address the legal and ethical issues related to data, which will be collected during the Project timeframe.
- Contribution to other Deliverables.

This deliverable is closely related to the following deliverable(s):

- D7.1, D7.2 Report on definition, planning and deployment of pilots' scenarios
- D7.3 Testing and first validation report
- D7.4 Pilot execution and validation report
- D7.5, D7.6 Report on Pilot results and lessons learned
- D9.2 Exploitation Plan (v1)
- D9.3 Exploitation Plan (v2)

The Exploitation Plan will contain identification of Partners' IP assets, including data, and an ownership proposition based on feedback from the consortium and the Project's IP mapping. It will also state the current or the suggested type of protection, and conditions of use for the Project's IP assets, including data.

2.2 Structure of the Document

The rest of the document is structured as follows:

- Section 2 provides a brief description of data sets which will be collected during the CyberSEAS Project, explains the procedures used to collect or create them, as well as standards and methodologies for data collection and management and quality assurance measures.
- Section 3 describes plans for data sharing and access in compliance with the findable, accessible, interoperable, and reusable (FAIR) principles.

- Section 4 deals with allocation of resources, data management roles and responsibilities.
- Section 5 presents ethical issues, confidentiality and intellectual property of data.
- Section 6 contains conclusions and further plans for the updating of the DMP.

3 Data summary

3.1 Research data

The notion of "research data" refers to "information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion or calculation" [1]. Research data covers a broad range of types of information, and digital data can be structured and stored in a variety of file formats. Examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. In the context of the CyberSEAS project, these will relate principally to EPES infrastructure. As such, the research data will consist mainly of data describing the EPES infrastructure, its usage, data processed via EPES systems, cybersecurity risks, security measures (including solutions piloted and developed in the context of CyberSEAS), and data on experiences during piloting activities (both qualitative evaluations and quantitative statistics). Additionally, as with most other EU projects, collected data will also relate to consortium activities in general, including project communications, events, surveys, and questionnaires addressed towards the consortium as well potentially external stakeholders.

We note that properly managing data (and records) does not necessarily equate to sharing or publishing that data. Some kinds of data may not be sharable due to the nature of the records themselves, or to ethical and privacy concerns. This refers to, for example:

- Preliminary analyses
- Drafts of scientific papers
- Plans for future research
- Peer reviews
- Communications with colleagues

Moreover, in CyberSEAS some deliverables (and their underlying data) are legally qualified as Classified Information - RESTREINT UE (under Commission Decision 2015/444/EC). As a result, sharing data in such deliverables is not legally possible.

Research data which cannot be shared may also include trade secrets, commercial information, data subject to intellectual property rights claims that impede sharing, or materials necessary to be held confidential by a researcher until they are published or similar information, which is protected under law.

3.2 Collection purposes

It may be summarized that the research data is collected and processed during the Project for the following purposes and in relation to the following Project objectives:

- Identifying, analysing and countering cyber risks related to the high impact attacks against EPES
- Protecting consumers against (personal) data breaches and cyber attacks
- Increasing the security of the Energy Common Data Space

In this regard, CyberSEAS will develop and validate:

- Real-time cyber security monitoring measures

- Continuous cyber risk assessment support
- Confidentiality preserving federated Machine Learning (ML)
- Mechanisms for improved use of cyber threat intelligence
- Techniques for augmented threat detection

3.3 Methodology of work

Given the breadth of the project, and the significant differences between the consortium partners in their activities and infrastructure, it is not feasible at this early stage of the project to identify all relevant data assets at a more granular level. However, a methodology has been established to ensure that the availability of data can be tracked, and that any legal constraints on further use and data sharing can be identified.

Within this methodology, the specific data sets for the Project need to be identified and described with the contribution of all Project Partners, since they are ultimately the only parties that have an accurate overview of data availability and suitability, and the principal stewards of the confidentiality of that data. For this reason, all Project Partners will be asked to describe the specific data sets that will be processed during the Project. Accordingly, a table with the following questions will be circulated very six months to be filled by the WP and Task leaders, further complimented with input from other Partners.

1. OUTPUT DATA (NEW DATA)
    a) What new data will you gather or produce in this task and how?
    b) What is the purpose of the collection/generation of data in relation to the task?
    c) What is the expected size of dataset?
    d) In what manner and format will the data be collected and stored?
    e) What transformations will the data undergo?
    f) What metadata will be created and used? Are there any standards applicable?
    g) Will the data contain or be personal data, subject to intellectual property rights claims, a trade secret, commercially sensitive, or otherwise confidential? Will there be any planned terms and conditions of its use (including within the CyberSEAS project or outside of it)?
2. INPUT DATA (RE-USE OF EXISTING DATA)
    a) What existing data will you re-use for this task and for what purpose?
    b) What is the source of existing data?
    c) What transformations will the data undergo?
    d) What is the expected size of existing data set?
    e) In what manner and format will the data be kept and used?
    f) What metadata will be created and used? Are there any standards applicable?
    g) Is the data personal data, subject to intellectual property rights claims, a trade secret, commercially sensitive or otherwise confidential? Are there any applicable terms and conditions of its use? Has the lawfulness of its further use in the project

been appropriately assessed, taking into account the ethics and lawfulness policies of the CyberSEAS project[1]?

3. SOFTWARE TOOLS & STORAGE
    a) With what IT tools will the data in this task be processed?
    b) Where will the data be stored and backed-up? Does the data remain within the EU?
    c) How will the data be secured?
4. DATA SHARING
    a) Will the data be shared during the Project (internally within the consortium, or externally)?

        a1) If yes, how and with whom? Have appropriate safeguards been implemented, and which?
    b) Will the data be shared after the Project (internally within the consortium, or externally)?

        b1) If yes, how and with whom? Have appropriate safeguards been implemented, and which?
5. PRIVACY/ETHICS
    a) Will the data – either new or existing - include any personal data?
    b) If yes, will the data subject be informed about data processing activities, and be able to consent to it? If not, is there an alternative legal basis available?
    c) Will personal data be anonymized/pseudonymized?
    d) Is there any ethics approval or prior authorisation required?

The DMP questionnaire is intended as a living document and more detail will be added as the Project progresses. All Partners commit to continuously keep track of the specific data sets processed under the tasks they are involved in and to report them internally by updating the DMP questionnaire for each task of the Project. Partners will be periodically reminded to update their responses.

3.4 CyberSEAS pilots data

This section provides preliminary information about the data that will be used in the Estonian and Slovenian-Croatian pilots. Data from the other CyberSEAS pilots will be described in the next version of Data Management Plan D1.7.

[1] Notably as set out in D10.2 - POPD - Requirement No.2, Section 5 - Further processing of previously collected personal data.

3.5 Estonian pilot

3.5.1 Advanced Update Validation Platform (AUVP)

The aim of the AUVP is to ensure the integrity of the operational technology (OT) device configuration through firmware control and signing.

Currently, critical systems of the Estonian energy grid are supported by managed services providers. It is their responsibility to update the whole platform. The OT devices in substations are rarely updated. As a rule, the OT device updates come from manufacturers that may not be well secured.

The AUVP is a control mechanism where only centrally controlled and signed firmware can be applied to OT devices. The use of other versions is inherently prohibited in the central management system. The controlling is achieved through applying Machine Integrity, Defense, and Awareness (MIDA) that provides real time situational awareness and continuous monitoring of all assets deployed.

Data sources are described in the following:

- IED (Protection Relays)
  - Real-time substation automation data
  - System and audit logs
- Communication devices (Gateways)
  - Real-time substation automation data
  - System and audit logs
- RTUs (Remote Terminal Units)
  - Real-time substation automation data
  - System and audit logs
- Smart Meters
  - System and audit logs
- Advanced Update Validation Platform
  - System and audit logs
  - Manufacturers' firmware update database
- KSI blockchain
  - Trust token - digital confirmation data

The formats are defined by Syslog and IEC 60870-5-104 standards.

3.6 Slovenian – Croatian pilot

3.6.1 Dynamic Rating System - SUMO

SUMO is an indirect dynamic thermal rating system (DRS) comprising of the dynamic line rating (DLR), the dynamic power transformer rating (DPTR) and the Dynamic Line Anti-icing (DLAI) subsystems. SUMO is a fully software-enabled and meteorological model-based system with the possibility to also integrate data from meteorological weather stations.

SUMO supports real-time and short-term forecast operations, calculations of transmission capacities for up to two days ahead (IDCF, DACF, D-2), and allows for mitigation of N and

N-1 overloading operational situations. It also features an inverse DLR algorithm for icing prevention and alarms for extreme weather conditions along the power lines.

It is a modular IT system featuring an integration bus with programming APIs and templates for the integration of additional third-party modules.

Data sources and sinks are described in the following:

- Dynamic Rating System – Service Bus
  - SOAP Web Services - More than 25 Web services with more than 100 methods
    - Weather data – real-time and forecasts, extreme weather alarms, DTR/DLR result (ampacity, conductor temperature, max. time of overloading), power line currents, power line loading forecasts, structural data
- Weather stations
  - Proprietary protocol
    - Weather data: ambient temperature, wind speed and direction, solar irradiation, humidity, air pressure.
- SCADA/EMS
  - ICCP (TASE.2) protocol
    - Power lines currents, DTR/DLR results (ampacity, conductor temperature, max. time of overloading)

3.6.2 Balancing Services Platform

The Balancing Services Platform allows Balancing Service Providers (BSPs) to provide following ancillary services to the TSO:

- FCR (frequency containment reserve),
- aFRR (automatic frequency restoration reserve),
- mFFR (manual frequency restoration reserve).

Data sources and sinks are described in the following:

- EccoSP platform (TSO)
  - AMQP protocol between client and server/platform
    - Business data – bids for balancing service
- ICCP (TASE.2) gateway (TSO)
    - Operational data for balancing services
- Virtual power plant platform VE.TER (BSP)
  - Web Services, AMQP and file exchange integration possibilities with the EccoSP client for business data exchange
  - ICCP (TASE.2) protocol for operational data exchange
- Distributed generation and/or loads controllers/gateways
  - Various industrial protocols
  - Operational data for activation/deactivation, monitoring and control

3.6.3 Eles (TSO) – Hops (TSO) Virtual Cross-border Control Center

Aim of the Virtual Cross-border Control Center is voltage control and loss optimisation in both transmission systems.

Data sources and sinks are described in the following::

- SCADA/EMS (source)
- Voltage control and loss optimisation application (sink)

Data are network model of Slovenian Transmission network and of Croatian Transmission network.

Data Format is Common Grid Model Exchange Specification - IEC technical specification (TS 61970-600-1, TS 61970-600-2) based on the IEC CIM (Common Information Model).

Serialization is done in RDF XML.

The used protocol is SFTP.

3.6.4 Threat Sharing intelligence platform - MISP

MISP Threat Sharing (MISP) is an open-source threat intelligence platform. The project develops utilities and documentation for more effective threat intelligence, by sharing indicators of compromise.

The MISP core format is a simple JSON format used by MISP and other tools to exchange events and attributes. The JSON schema 2.4 is described on the MISP core software and many sample files are available in the OSINT feed. The MISP format is described as Internet-Draft in misp-rfc (https://github.com/MISP/MISP-rfc).

Data sources and sinks are described in the following:

- MISP exchange - JSON format
- MISP freetext import - text format
- MISP IDS export - Snort/Suricata single line text format

3.7 Finnish pilot

3.7.1 Enerim CIS platform

Enerim CIS platform is a software module developed by Enerim company to help energy companies automate their customer information and invoicing. The platform is a flexible and modular tool for all multi-utility company functions. It comprises of customers contracting data and electricity consumption/production data.

The CIS paltform provides variety of processes and services including but not limited to customer management, product management, contract management, consumption location management, invoicing, reporting and archiving for energy suppliers and distribution system operators. The platform increases speed of cashflow by doing calculations

in real-time. The real-time calculation capability serves real-time data and dashboards which lead to full visibility and transparency to business operations.

The platform exchanges data with some other platforms including but not limited to online platform and customer relationship management (CRM) platform. The platform also indirectly exchanges data with national datahub for electricity retail services. The data exchange is through information exchange system (IXS) platform which is developed and operated by Enerim company.

The CIS platform has a modular architecture which enables microservices/functional modules being updated individually without interrupting the service. The platform has easy to browse interface for users and open APIs for easy integartion with third-party modules.

Data sources and sinks are listed in the following:

- Datahub (through Information Exchange System (IXS) which has been developed by Enerim)
    - Protocol: HTTPS for sending data to IXS and TCP for receiving data from IXS
    - Data: consumers electricity consumption data and contracting information
    - Data format and standards: Finnish Datahub format
- Head-end systems
    - Through API
    - Data: consumers electricity consumption data
    - Data format and standards: file based integration which is usually customer specific
- Third-party platforms
    - Through API
    - Data: consumers electricity consumption data and contracting infromation
    - Data format and standards: file based integration which is usually customer specific
- User interface
    - Protocol: https
    - Data: product information
    - Data format and standards: web protocol and IIS server

4 Findable, Accessible, Interoperable and Re-usable data (FAIR data)

4.1 Making data findable, including provisions for metadata

4.1.1 General principles

The CyberSEAS Project attaches great importance to making its research data findable, discoverable and identifiable. The DMP defines what documentation and metadata will accompany the data.

"Metadata" is structured information describing the characteristics of a resource. For example, the dates associated with a dataset or the title and author of a book. Metadata supports discovery, re-use and long-term preservation of resources. Metadata needs to vary across scientific fields, but typically cover the following:

- Descriptive metadata, such as title, abstract, author, and keywords;
- Administrative metadata which are used to provide information to help manage a source, such as when and how it was created, file type and other technical information, and who can access it;
- Archive terms and access policies.

A metadata record consists of a set of predefined elements that define specific attributes of a resource. Each element can have one or more values; for example, a dataset may have multiple creators or more keywords may be added to a particular image to enable its finding. Documenting data enables other researchers to discover the data. Metadata about the nature of the files is also critical to the proper management of digital resources over time.

4.1.2 Implementation in CyberSEAS

As for the documents produced within the Project, including reports, following the Consortium Agreement and the D.1.1 CyberSEAS Project Guidelines, as well as the procedures agreed in D1.2 – Project Quality Plan (v1), they are subject to the following principles and measures to facilitate FAIR data policies:

- For non-Classified Information (under Commission Decision 2015/444/EC), a structured repository of Project documents, including restricted information has been developed. For Project-internal data sharing, such as the sharing of working documents, reports and deliverables, the Project uses Microsoft Teams with restricted access to efficiently manage the Project information amongst the Project Partners and to enable the preservation of Project data and appropriate versioning of the documents. A dedicated section of the project website will be used for data sharing with external parties.
- For Classified Information (under Commission Decision 2015/444/EC), a parallel track has been established, with project participants that handle such data being required

to use approved cryptographic tools, and respecting the relevant legislation and procedures both at the national and EU level. Classified Information is not shared via the standard communication tools, and is not added to the repository.

- All Project documentation needs to conform to specific templates., made available in Microsoft Teams.
- Recommended document naming convention has been developed. The naming convention for all documents to be produced within the Project is provided in Section 3.3 of the D.1.1 CyberSEAS Project Guidelines.
- It is prescribed to use versioning property when modifying a document uploaded in the Project document repository or when producing different versions of code.
- Every document circulated to other Partners in the consortium should include a version number and date.
- When multiple contributors need to work on a document, it is recommended to use online documents that allow synchronous co-editing.

The research data which will be published should contain include the reference period, Project funding information (e.g., EU logo and information about the Grant Agreement and the action/program that funds the Project, official Project name and Project ID), release policy including dissemination rules, information about the collection of the data such as the data source, geographic coverage of the data, language, and file format.

4.2 Making data openly accessible

Materials generated under the CyberSEAS project will be disseminated in accordance with the Consortium Agreement. The project deliverables that are marked as 'PU' (public) in the Description of Action will be made openly available via the project website, and can be further shared through related platforms such as Zenodo, OpenAIRE, etc, in accordance with the Grant Agreement and the Horizon 2020 Open Access Guide.

Certain data fall outside the scope of the open access strategy. These obviously include any deliverables and underlying data qualified as Classified Information: RESTREINT. It also includes confidential data, such as different types of data that can be used to identify individuals or that are of a commercially sensitive in nature, or that reveal exploitable security weaknesses.

As a consequence, non-anonymised personal data of research participants, project partners or other stakeholders (including household energy consumption data at the individual level), raw qualitative research data from interviews, focus groups and workshops, draft reports, unfinished work, personal notes, plans for future research, preliminary analyses, peer reviews, and communication outside of a test setting, fall outside of the scope of the open access strategy.

Therefore, any data in PU deliverables will be anonymised. This implies that (1) evidences as a data set are entirely out of scope; (2) household level energy consumption patterns will only be reported on at the aggregate, statistical level, except when piloting can be done through fictitious test data; and (3) survey outcomes will similarly be only reported on at the aggregate, statistical level. It is envisaged that such data will therefore not be traceable to individual users (persons or companies), nor to individual administrations or authorities when this would be reasonably likely to impair their functioning. Original (non-aggregate and thus identifiable) information will not be made openly accessible, although source information

will be retained by the CyberSEAS partners for as long as legally permissible under the Consortium Agreement and/or as required under applicable law.

The open research data will be made available with the lowest technical threshold possible, i.e. without any prior requirement of identification or authentication. Nonetheless, in order to protect the identity of research participants and in order to encourage participants to speak freely and truthfully, all reporting and communication relating to research participants will be shared only in a pseudonymised or anonymised manner.

Original (non-anonymised or non-pseudonymised data) will be stored in order to allow identification and traceability for research validations and follow-up, but such storage will be organised separately from the research data and in adherence to state-of-the-art confidentiality and security standards (including encryption, access logs and seals). If the document itself cannot be made secure with a password, it will be stored in an encrypted container with password protection.

4.3 Making data interoperable

The Partners will focus on harmonising the input data from the different Data Providers and in this way make it available and integrated, as well as interoperable for the purposes of subsequent research challenges within the project, i.e. allowing re-use of this data by the researchers within the consortium, although they are datasets coming from different origins.

From a practical perspective, standard and commonly used file formats will be agreed in advance and used wherever available.

4.4 Increase data re-use (through clarifying licenses)

Specific details on future use of the data will be provided in the sustainability and exploitation work of CyberSEAS (see notably Exploitation Plan (D9.2 and D9.3)). In general, the project consortium aims to apply open licensing through common, standardised and widely known license models, such as e.g. a 'CC-BY-SA 3.0' license, as a general rule to all research data in order to facilitate the widest re-use as possible, or e.g. the EUPL for software releases.

There will be no restriction on the use of such open data by third parties after the end of the project. The open data will remain reusable (labelled accordingly with the applicable licenses) without time duration.

5 Allocation of resources

5.1 Roles in data management

The main coordinating roles in data management are provided for in the Consortium Agreement and Grant Agreement as follows:

1. Ethics and Legal Manager (TLX): is responsible for ensuring that an appropriate data management plan is developed and used to protect the privacy of data and address all other data management aspects.
2. The Innovation Manager (CRE): is responsible for managing the knowledge produced during the Project lifecycle; manages execution of the overall exploitation plan of the Project and supports the Partners in setting up their individual business plans, in order to exploit the Project results.
3. WP8 Leader (ICS): is responsible for raising public awareness and ensuring wide communication of the Project results and will also be responsible for the coordination of the scientific dissemination, clustering and standardization activities.

Furthermore, the WP/Task leaders will be expected to provide first level of data management within the scope of their role and ensure that the data of their WP/Task are treated according to the agreed Project principles and processes.

5.2 Resources

CyberSEAS researchers will follow the Open Access Green method in the case of conferences, workshop contributions, and journal publications. In this case the published article or the final peer-reviewed manuscript is archived by the researcher in an online scientific repository before, after or alongside its publication. The authors must ensure open access to the publication within a maximum of six months. The Open Access Infrastructure for Research in Europe will be explored to determine which repository to choose (http://www.openaire.eu). Moreover, CyberSEAS will exploit other support infrastructures provided by the EC towards data preservation, e.g. the Horizon Results Platform.

As the Project progresses and the different types of results are produced, it will be possible to provide further details on approaches for long-term preservation and accessibility of each type of dataset beyond the end of the funding period.

6 Ethical aspects and intellectual property rights

6.1 Ethical issues

The CyberSEAS Partners have committed to comply with the ethical principles as set out in Article 34 of the Grant Agreement, which, among other, states that all activities must be carried out in compliance with:

- Ethical principles (including the highest standards of research integrity)
- Applicable international, EU and national law.

The ethical aspects of the Project will be assessed and monitored under WP10, which sets out the ethics requirements that the Project must comply with. More specifically:

- D10.1 describes the procedures and criteria that will be used to identify/recruit research participants. Principally, this is a statement of principles and safeguards in relation to the onboarding of pilot participants during the project. Additionally, it comprises templates of the informed consent forms and information sheets relating to personal data processing, and current details on incidental findings policy (including identification, assessment and management of incidental findings.
- D10.2 confirms that a Data Protection Officer (DPO) has been appointed and that the contact details of the DPO are made available to all data subjects involved in the research. This deliverable also contains a description of the anonymisation/pseudonymisation techniques that will be implemented, as also referenced in section 4.2 above. Finally, it describes the assurances related to the lawfulness of any further processing of previously collected personal data.

Additionally, the Project partners confirm to respect the EU and national law requirements on privacy and data protection and to adhere to the research ethics standards applicable to Horizon 2020 research. In accordance with the data minimization, data retention and purpose limitation principle, personal data will not be collected beyond the scope of the processing objectives and will not be stored for longer than necessary.

Finally, as is also described in the aforementioned ethics deliverables, compliance with ethics policies and procedures is continuously monitored and evaluated through a four tiered model:

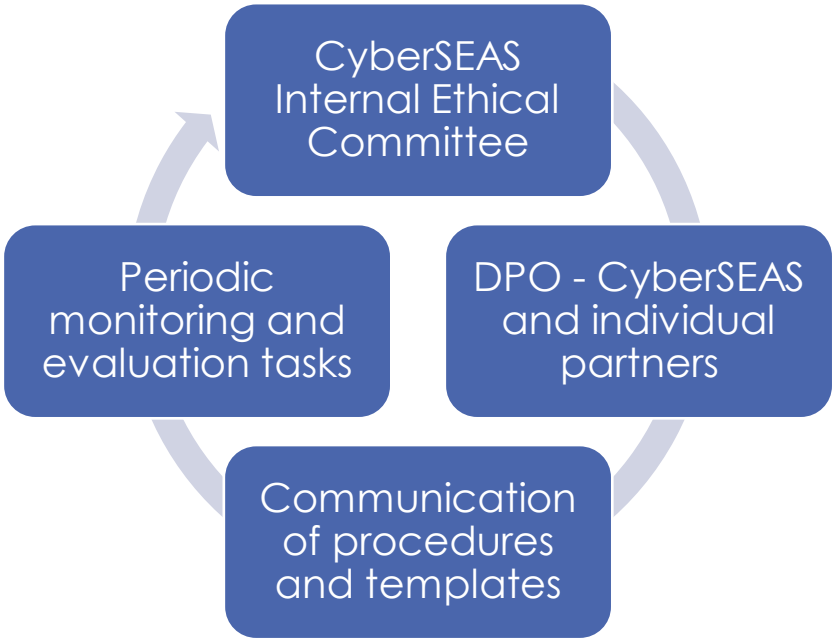

Figure 1 – Monitoring and evaluation structure

- **Establishment of a CyberSEAS Internal Ethical Committee** (IEC), which has the assignment of ensuring clarity and consistency in communicating with CyberSEAS project partners on ethics issues, assessing compliance with ethics policies, and supporting interactions with the users. It has the responsibility for results monitoring, ethical, privacy and data protection issues compliance and assessment of the sensitivity of deliverables before publication.
- **Appointment of Data Protection Officers** (DPOs) in accordance with the GDPR, to oversee data protection compliance. Moreover, a list of DPOs at the partner level is maintained, to facilitate interaction with local end users, and to ensure that there is hands-on involvement at the partner level.
- **Communication of procedures and templates:** the ethics guidance from the WP10 deliverables are actively disseminated and explained towards all CyberSEAS partners, to ensure that they are known and used in practice.
- **Periodic monitoring and evaluation tasks:** CyberSEAS will evaluate to what extent the ethics principles are respected during the project's execution. Beyond the ethics reporting in the periodic activity reports, CyberSEAS has defined specific tasks to conduct data protection impact assessments (T2.5) and to create and monitor SELP (Security, Ethical, Legal and Privacy) requirements (T3.2), which will be used to further detail, monitor and report on ethics compliance, and to take any corrective actions needed.

In this way, CyberSEAS can ensure compliance throughout the project's duration, including with respect to data management.

6.2 Confidentiality

All CyberSEAS Partners must keep any data, documents or other material confidential during the implementation for the Project and for four years after end of the Project in accordance with Article 36 of the Grant Agreement. Further detail on confidentiality can be found in Article 36 of the Grant Agreement.

6.3 IPR

Issues regarding the protection of intellectual property rights (IPRs) and confidential information were addressed in detail within the Consortium Agreement. In particular, the Consortium Agreement regulates the IP-Ownership, Access Rights to Background and Foreground IP (Section 9). Moreover, in accordance with Article 24 of the Grant Agreement, Background was identified for all Partners, if applicable. In the first months of the Project, IPR control spreadsheets have been circulated and existing IP (Background), foreground IP and contributed assets were identified by the Partners. The details will be described in the Exploitation Plan (D9.2 and D9.3).

7 Conclusions

The document presented initial CyberSEAS Data Management Plan. The DMP will be revised and updated during the entire duration of the Project. The DMP will be updated at least by the mid-term and final review to fine-tune it to the data generated and the uses identified by the consortium since not all data or potential uses are clear from the start. New versions of the DMP will be created whenever important changes to the Project occur due to inclusion of new data sets, changes in consortium policies or external factors.

8 References

[1] European Commission, Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, Version 3.2, 21 March 2017,4.