SPEAR

Secure and PrivatE smArt gRid

(Grant Agreement No 787011)

D1.4 Data Management Plan

Date: 2018-12-18

Version 1.0

Published by the SPEAR Consortium

Dissemination Level: Public

Co-funded by the European Union's Horizon 2020 Framework Programme for Research and Innovation under Grant Agreement No 787011

Document Control Page

Document Details

Document Version:	1.0
Document Owner:	G.E. Pukhov Institute for Modeling in Energy Engineering of the National Academy of Sciences of Ukraine (PIMEE)
Contributors:	All partners
Work Package:	WP 1 – Project Management and Coordination
Task:	Task 1.4 – Data Management
Deliverable Type:	Report
Document Status:	Final
Dissemination Level:	Public

Document History

Version	Author(s)	Date	Summary of changes made
0.1	Igor Kotsiuba (PIMEE), Oleksii Vasyliev (PIMEE), Dr. Dimosthenis Ioannidis (CERTH), Odysseas Nikolis (CERTH)	2018-10-03	Initial Draft
0.2	Igor Kotsiuba (PIMEE), Oleksii Vasyliev (PIMEE), Dr. Dimosthenis Ioannidis (CERTH), Odysseas Nikolis (CERTH)	2018-10-04	Definition of Data Management methodology in SPEAR project.
0.3	Oleksii Vasyliev (PIMEE)	2018-10-25	Deliverable version uploaded for Quality Check
0.4	Igor Kotsiuba (PIMEE)	2018-10-26	Quality Check
0.5	Igor Kotsiuba (PIMEE), Oleksii Vasyliev (PIMEE), Inna Skarga-Bandurova (PIMEE)	2018-11-07	Implemented suggestions of Odysseas Nikolis (CERTH) and Dr. Dimosthenis Ioannidis
0.6	Igor Kotsiuba (PIMEE), Inna Skarga-Bandurova (PIMEE)	2018-11-14	Editing
0.9	Michail Angelopoulos (PPC): Antonios Sarigiannidis (SH)	2018-12-14	Final improvements
1.0	Igor Kotsiuba (PIMEE), Oleksii Vasyliev (PIMEE), Inna Skarga-Bandurova (PIMEE)	2018-12-18	Final version submitted to the European Commission

Internal Review History

Reviewed by	Date	Summary of Comments
Michail Angelopoulos (PPC)	13 December 2018	Accepted with reservation. Minor typos and formats should be corrected
Antonios Sarigiannidis (SH)	10 December 2018	Accepted with reservation. Some typos should be corrected. Some figures need a revision. Some acronyms should be defined

Legal Notice

The information in this document is subject to change without notice.

The Members of the SPEAR Consortium make no warranty of any kind with regard to this document, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The Members of the SPEAR Consortium shall not be held liable for errors contained herein or direct, indirect, special, incidental or consequential damages in connection with the furnishing, performance, or use of this material.

Possible inaccuracies of information are under the responsibility of the project. This report reflects solely the views of its authors. The European Commission is not liable for any use that may be made of the information contained therein.

Table of Contents

Acronyms

Acronym	Explanation
APT	Advanced Persistent Threat
BDAC	Big Data Analytics Component
BSU	Belarus State Economic University
CERIF	Common European Research Information Format
CERTH	Centre for Research and Technology Hellas CERTH
CIN	Critical INfrastructures
DMP	Data Management Plan
DoA	Description of Action
DOI	Digital Object Identifier
DoS	Denial of Service
DDoS	Distributed DoS
EC	European Commission
ED	European Dynamics
ENI	ENEL IBERIA S.R.L
IPR	Intellectual Property Rights
LUH	Gottfried Wilhelm Leibniz Universität Hannover
MiTM	Man in The Middle
OAI-ORE	Open Archives Initiative Object Reuse and Exchange
0INF	0 INFINITY Limited
PIMEE	G.E. Pukhov Institute for Modeling in Energy Engineering of the National Academy of Sciences of Ukraine
PM	Project Manager
PPC	Public Power Corporation S.A.
SCHN	Schneider Electric France SAS
SH	Sidroco Holdings Limited
SURREY	University of Surrey
TEC	Fundacion Tecnalia Research & Innovation
TUS	Technical University of Sofia
WP	Work Package

1. Executive Summary

The D1.4 Data Management Plan (DMP) is a framework, that describes how to work with the data and datasets that will be generated during project's lifecycle, including access rights management, storage, backups, data ownership and principles of collaboration within research teams, industrial partners and public bodies. The DMP includes information about data types, formats of generated/collected data, and specifies methods for data gathering, processing, sharing, and archiving. The plan also documents some data management activities associated with the SPEAR project. A list the various types of data that SPEAR consortium expect to collect and create is also represented.

The project will collect the following types of data: network traffic, operating system shell commands, keystrokes, communications and syslogs collected from the devices in smart grid, sensors, gateways, etc.; quantitative data related to day-to-day activity (event data produced after processing collected raw data); and cyber attacks and threats data for information sharing through an anonymous channel/repository. Particularly, data will be obtained from direct observation, industrial enterprises, field instruments, experiments, and compilations of data from other studies.

The expected data volume will be approximately 150 GB. The document will be updated regularly aimed to improve the data management life cycle for all data generated, collected or processed by the SPEAR project.

2. Introduction

The SPEAR consortium joins the Pilot on Open Research Data project, which is supported by the European Commission through the Horizon2020 program. The SPEAR consortium supports the concept of open science, and shares an optimistic assessment of the prospects of this concept for introducing innovative solutions to the European economy, with the re-use of scientific data on a wider scale. Thus, all data obtained during the implementation of the SPEAR project can be published in open access mode, subject to the additional conditions and principles described in this document below.

2.1 Scope and objectives of the deliverable

The purpose of the Data Management Plan (DMP) deliverable is to provide relevant information concerning the data that will be collected, used, stored, and shared by the partners of the SPEAR project.

The SPEAR project aims at developing an integrated solution of methods, processes, tools and supporting tools for (see Fig. 1):

(a) Timely detection of evolved security attacks such as Threat Advanced Persistent (APT), the Man in the Middle (MiTM) attacks, Denial of Service (DoS) and Distributed DoS (DDoS) attacks using big data source analytics, advanced visual technique for anomaly detection and smart trust security management.

(b) Developing an advanced forensic readiness framework, based on smart honeypot deployment that will collect attack traces and prepare the actionable evidence in court, while also ensuring privacy for the users.

(c) Elaboration and implementation of the anonymous channel for securing smart grid stakeholders during the exchange of sensitive information about cyber-attack incidents and prevent information from leaking.

(d) Performing risk analysis and proposing cyber hygiene procedures, while empowering EU-wide consensus by collaborating with European and global security agencies, standardization organizations, industrial partners and smart grid companies across Europe.

(e) Exploiting the research outcomes to more critical infrastructures (CIN) domains and creating competitive business models for utilizing the implemented security tools in smart grid operators and actors across Europe

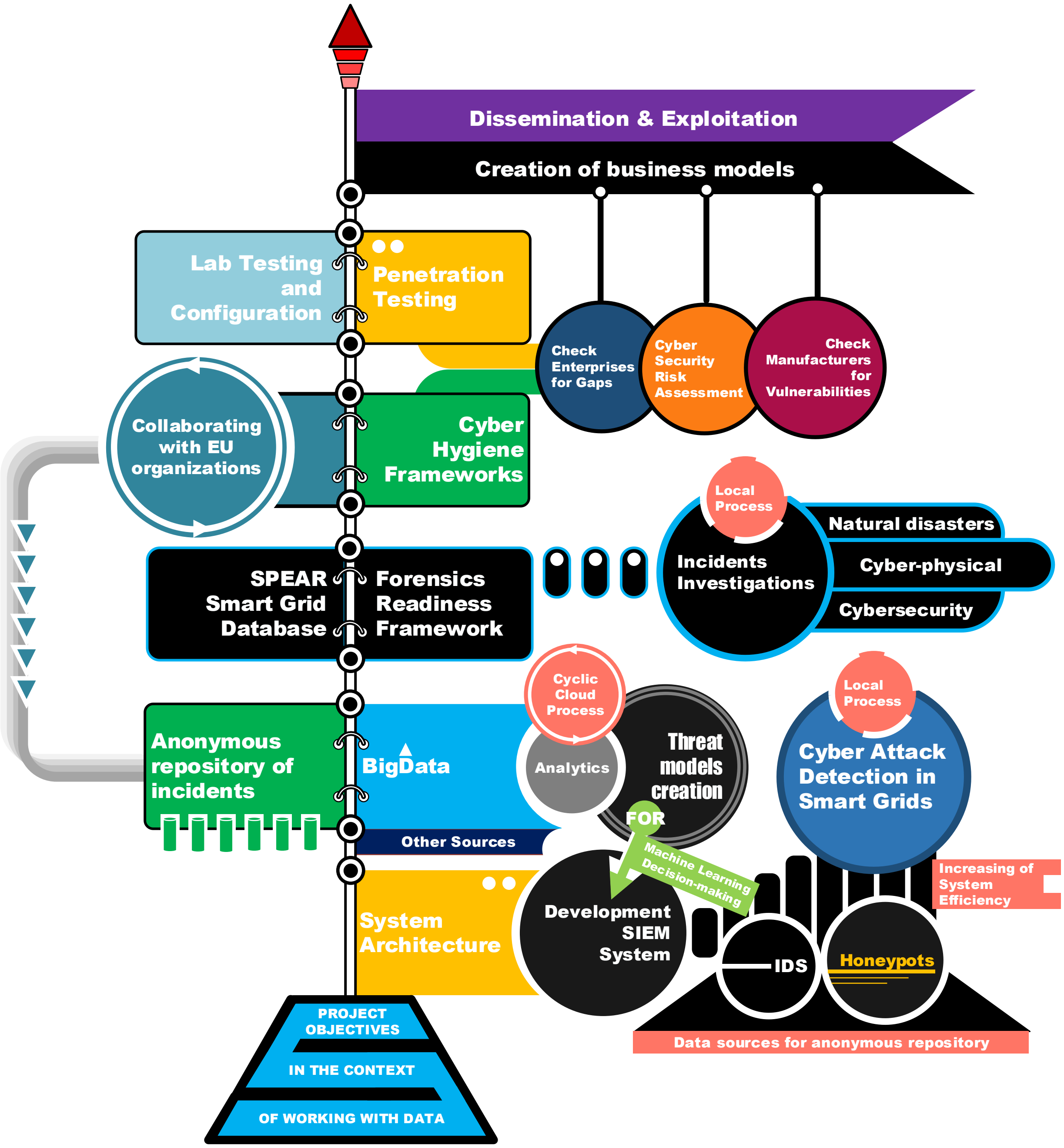
Dissemination & Exploitation
Creation of business models
Lab Testing and Configuration
Penetration Testing
Check Enterprises for Gaps
Cyber Security Risk Assessment
Check Manufacturers for Vulnerabilities
Collaborating with EU organizations
Cyber Hygiene Frameworks
Local Process
Incidents Investigations
Natural disasters
Cyber-physical
Cybersecurity
SPEAR Smart Grid Database
Forensics Readiness Framework
Cyclic Cloud Process
Analytics
Threat models creation
FOR
Local Process
Cyber Attack Detection in Smart Grids
Anonymous repository of incidents
BigData
Other Sources
Machine Learning Decision-making
Increasing of System Efficiency
System Architecture
Development SIEM System
IDS
Honeypots
Data sources for anonymous repository
PROJECT OBJECTIVES
IN THE CONTEXT
OF WORKING WITH DATA

Figure 1 - SPEAR aims diagram

2.2 Structure of the deliverable

The report is structured in 5 chapters:

Chapter 1: Executive summary, including the purpose and the context of this deliverable.

Chapter 2: Introduction concerning the scope of this deliverable.

Chapter 3: An overview of general principles for participation in the pilot on open research data, IPR management and security as well as data protection, ethics and security in SPEAR project.

Chapter 4: An overview of the data management framework along with the specification of the dataset format, the dataset description methods, definition of standards and metadata, approaches and policies for data sharing, archiving and presentation. Datasets list for SPEAR new components is also enclosed.

Chapter 5: Description of datasets from SPEAR partners.

Chapter 6: Conclusions

2.3 Relation to other activities in the project

The following diagram illustrates the relationship between the seven main activities of the SPEAR project.

1. Project Management and Coordination
2. Use Case Preparation
3. Cyber Attack Detection
4. Forensic Readiness
5. EU-Wide Consensus
6. Integration and Development
7. Dissemination and Exploitation

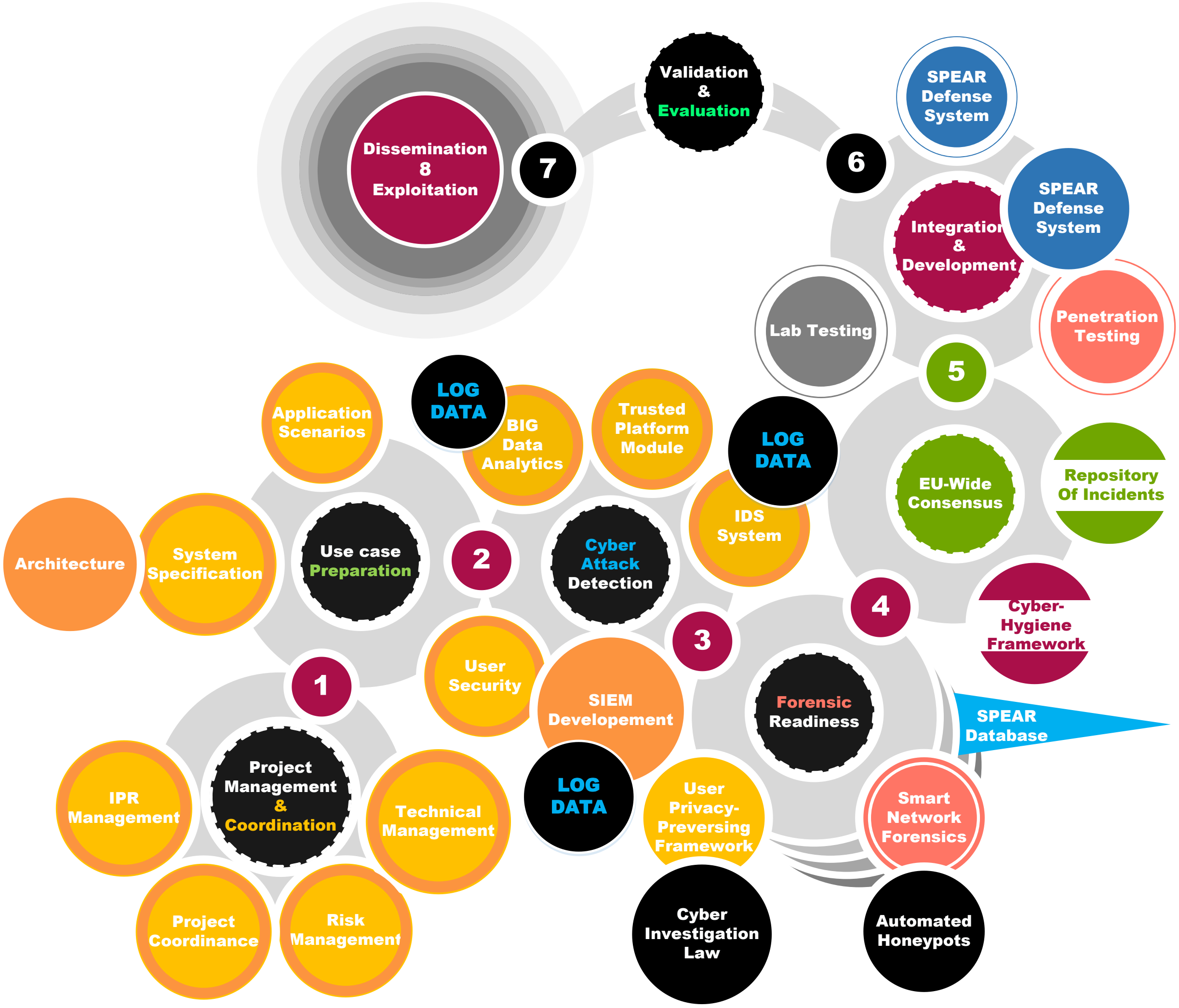

Figure 2 - The main activities of the SPEAR project

3. General Principles

SPEAR project stands for data openness and sharing; hence we are committed to making all data collected during the project to the best of and immediately available for use within the limits of personal privacy and commercial confidentiality following the Fair Data Principles.

3.1 Participation in the Pilot on Open Research Data

3.1.1 Data Availability

All the project data will be publicly available. However, different access levels for different types of data will be allocated. For security reasons, sensitive data such as personal data regulated by data protection rules, will be obscured. Recordings and notes from meetings and workshops as well as survey results will be anonymized. All anonymized data will be available in open-access mode. Technical details of the attacks, from the anonymous repository of smart grid incidents, will be available for everyone. The types of data and rules will be specified in the following sections.

3.1.2 Open Access to Scientific Publications

All Scientific publications will be open, unless there are special requirements or constraints will force to non-open publications.

3.1.3 Open Access to Research Data

To meet open access policy and be accessible to the research and professional community research data will be uploaded and stored on the Zenodo, EC publications and data repository. Research data archiving and availability will be guaranteed by the Zenodo digital repository.

3.2 IPR management and security

The SPEAR consortium consists of industrial partners form both private and public sector, all of them preserving intellectual property rights on their technology, technical solutions and data. Given this, the SPEAR consortium will pay particular attention to the protection of data, and will consult with the concerned parties prior to data publication.

IPR data management will be conducted within SPEAR PM. The Collection and /or process of personal data are managed by the Data Protection Officer.

Within the project a number of data models will be created to support the various SPEAR modules, e.g. for the Visual-based IDS. Of course, these models will be also populated during the execution of the pilots in SPEAR end-users Infrastructures. If necessary, anonymized data (except the data models that do not have any privacy concern) will be exported. In addition, the DMP is accommodated with a part in the SPEAR website, where the data models / datasets are uploaded (public versions). This website will be created by CERTH (M12).

3.3 Data Protection, Ethics and Security

No data will be collected or processed prior the finalization of the respective deliverables and the relevant Consent Forms.

4. Data Management Framework

SPEAR will develop a data management framework for deliverables which are part of the project and will be shared in the publicly accessible repository Confluence. This repository will provide to the public, for each dataset that will become publicly available, a description of the dataset along with a link to a download section. The portal will be updated each time a new dataset has been provided by research teams and partners, collected and is ready of public distribution.

To reach out industrial partners and smart grid companies across Europe, the anonymous repository of incidents and threats will be developed and anonymous channel for exchanging sensitive information about cyber-attack incidents will be launched.

Data lifecycle related to work packages (WP) of SPEAR project is represented in fig. 3.

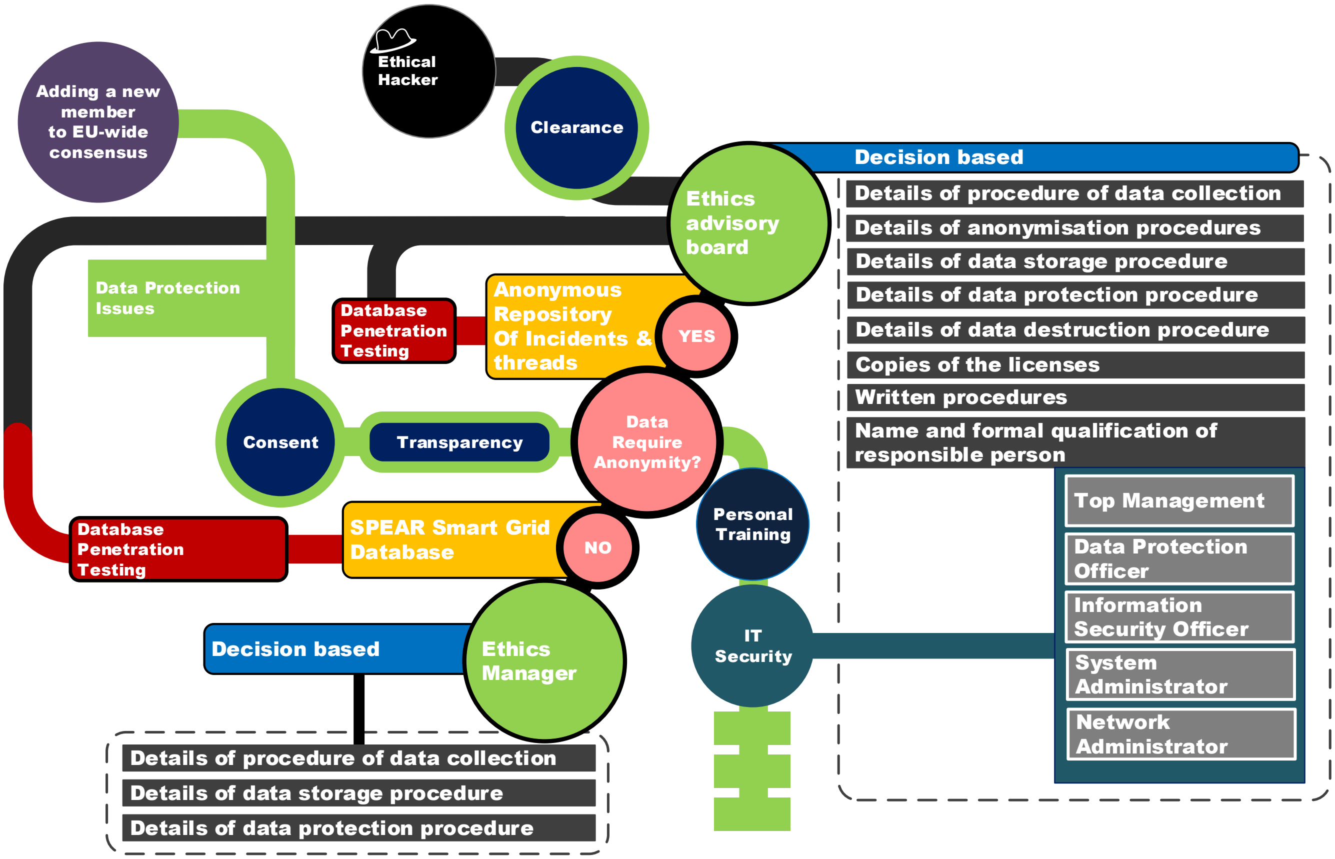
Adding a new member to EU-wide consensus
Ethical Hacker
Clearance
Decision based
Ethics advisory board
Details of procedure of data collection
Details of anonymisation procedures
Details of data storage procedure
Details of data protection procedure
Details of data destruction procedure
Copies of the licenses
Written procedures
Name and formal qualification of responsible person
Data Protection Issues
Database Penetration Testing
Anonymous Repository Of Incidents & threads
YES
Consent
Transparency
Data Require Anonymity?
Personal Training
Top Management
Data Protection Officer
Information Security Officer
System Administrator
Network Administrator
Database Penetration Testing
SPEAR Smart Grid Database
NO
IT Security
Decision based
Ethics Manager
Details of procedure of data collection
Details of data storage procedure
Details of data protection procedure

Figure 3 - Project Data lifecycle

4.1 Format of datasets

For each dataset the following characteristics will be specified:

Table 1 - Format of Datasets

X PARTNER Name_New Component/Existing Tool Name	
Dataset Information	
Dataset / Name	*<Mention an indicative reference name for your produced dataset>*
Dataset Description	*<Mention the produced datasets with a brief description and if they contain future sub-datasets>*
Dataset Source	*<From which device and how the dataset will be collected. Mention also the position of installation>*
Beneficiaries services and responsibilities	
Beneficiary owner of the component	*<Partner Name>*
Beneficiaries in charge of the data collection (if different)	*<Partner Name>*
Beneficiaries in charge of the data analysis (if different)	*<Partner Name>*

Beneficiaries in charge of the data storage (if different)	*<Partner Name>*
WPs and tasks	*<e.g. WP3, T3.4>*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*<Provide the status of the metadata, if they are defined and their content>*
Standards, Format, Estimated volume of data	*<Mention the data format if it is available, the potential data volume and refer also to the standards concerning the communication and the data transfer>*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*<Purpose of the data collection/generation and its relation to the objectives of the project>*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*<Access for partners & access for the public (open access>, refer to the data management portal if available and to dissemination acitivities>*
Data sharing, re-use and distribution (How?)	*<Provide if available the data sharing policies, the requirements for data sharing, how the data will be shared and who will decide for sharing>*
Embargo periods (if any)	
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*<Who will be the owner of the collected information, define the adherence to partner policies and mention any potential limitations>*

4.2 Description of methods for dataset description

The datasets will be generated by the project research team as well as industrial partners.

All incident-related data will be entered manually and will be stored in one anonymous repository.

Folders will be organized in a hierarchical structure.

Files will be supported with identification and number of version by using such structure: project name, dataset name, ID, place and date.

Keywords will be added by using the thesaurus.

4.3 Standards and metadata

For common project data the following standards and metadata will be applied:

Table 2 - Standards and Metadata

Purpose	***Standard***	***Link***
Recording information about research activity	CERIF (Common European Research Information Format)	http://rd-alliance.github.io/metadata-directory/standards/cerif.html
Data exchanging	Data Package	http://rd-alliance.github.io/metadata-directory/standards/cerif.html

Data citation and retrieval purposes	DataCite Matadata Schema	http://rd-alliance.github.io/metadata-directory/standards/datacite-metadata-schema.html
Data authoring, deposit, exchange, visualization, reuse, and preservation	OAI-ORE (Open Archives Initiative Object Reuse and Exchange)	http://rd-alliance.github.io/metadata-directory/standards/oai-ore-open-archives-initiative-object-reuse-and-exchange.html
Data registration	DOI (Digital Object Identifier)	https://fairsharing.org/biodbcore-001020/

4.4 Data sharing

All research data will be shared in the publicly accessible repository Confluence using descriptive metadata as it provided by this repository. To perform identification and access to citation all research data will be supported by DOIs.

For all other cases, in accordance with project policy, credentials are needed in order to obtain information from the repository.

Table 3 - Data Types and Repositories for Storage and Sharing Data

Data types	***Users***	***Repository***	***Type of Repository***	***Link***	***Access***
Research data, e.g. statistics, visualization analytics, measurements, survey results, results of experiments available in digital form	*University researchers*	*University of Reading Research Data Archive*	*External*	http://www.reading.ac.uk/reas-RDArchive.aspx	*Open*
Publications	*All*	*Zodeno*	External	https://zenodo.org/	*Open*
Project documentation	*SPEAR Partners*	*Confluence*	External	*https://space.uowm.gr/confluence*	
Security related data, e.g. Network traffic data and syslogs, operating system shell commands, Abnormal network traffic dataset, database records that tracks the	*SPEAR Partners*	*Anonymus repository,* *SPEAR webcloud*	Internal		*Closed*

changes in reputation and trust of home nodes over time, Cyber attacks and threats data					

4.5 Archiving and preservation (including storage and backup)

In accordance with EC FAIR (Findable, Accessible, Interoperable, and Re-usable) Policy and Horizon 2020 Data Management Guidance, SPEAR project data will be archived and preserved in open formats. For this reason, the data will remain re-usable until the repository withdraws the data or goes out of business.

All project-related data will be stored in *Confluence* repository.

4.6 Datasets List

Table 4 - Datasets List for SPEAR New Components

SPEAR New Component Name	***Sub-compo-nents Name***	***Related Task***	***Partner***	***SPEAR Pilot***	***Produced Datasets***
SPEAR - SIEM	***OSSIM SIEM SIEM Basis (Data collector)***	***T 3.1***	***TEC***	*UC1- The Hydro Power Plant Scenario UC2- The Substation Scenario UC3- The combined IAN and HAN scenario UC4- The Smart Home Scenario*	**OSSIM is an open-source SIEM, https://www.alienvault.com/products/ossim** *Network traffic data and syslogs from the devices in Smart grid scenarios. Event data produced after processing collected raw data (Network traffic data and syslogs)*
SPEAR - SIEM	***BDAC***	***T 3.2***	***SURREY UOWM CERTH***	*ALL*	*Normal and abnormal network traffic dataset, including different types of modern attacks, application layer attacks and several network traffic features.*
SPEAR - SIEM	***Visual-based IDS***	***T 3.3***	***CERTH***	*ALL*	*Visualization of multiple attributes of network traffic as well as common attributes among the records, the features extracted from the data, the (dis-)similarities among them and the combination of multiple types of features in clusters.*

SPEAR - SIEM	***GTM***	***T 3.4***	***SURREY CERTH***	*ALL*	*A set of database records that tracks the change in reputation and trust of home nodes over time.* *A set of database records that tracks the change in reputation and trust of nodes over time.*
SPEAR - FRF	***AMI HONEYPOTS***	***T 4.3***	***TEC***	*UC2- The Substation Scenario*	*Network traffic data, operating system shell commands, keystrokes, communications and syslogs.*
SPEAR - FRF	***PIA frame-work***	***T 4.4***	***ED***		
SPEAR - FRF	***Forensic Database Services***	***T 4.5***	***ED***		
SPEAR - CHF	***SPEAR-RI***	***T 5.1***	***TEC***		*Cyber attacks and threats data*

5. Description of Datasets

The SPEAR data management repository will enable project partners and research teams to manage and distribute their public datasets through a common cloud infrastructure in secure and efficient manner. The datasets on repository will provide a holistic list of data resources, generic and easy to handle datasets, and ability to move to industrial datasets. Datasets are to be identifiable, with allowance to segregate access rights and with accessible backups.

5.1 Datasets for SPEAR-SIEM

5.1.1 Datasets for OSSIM SIEM

Table 5 - TEC-SIEM Basis (Data Collector)

TEC- *SIEM Basis (Data collector)*	
Dataset Information	
Dataset / Name	*network traffic, syslog and event dataset for BDAC and Visual IDS*
Dataset Description	*The dataset includes network traffic data and syslogs from the devices in Smart grid scenarios, and also event data produced after processing collected raw data (network traffic data and syslogs).*
Dataset Source	• *In: Smart grid systems of the use case scenarios* • *How: Wireshark, Suricata, AlienVault OSSIM, syslog protocol (RFC5424*)
Beneficiaries services and responsibilities	
Beneficiary owner of the component	*TEC*
Beneficiaries in charge of the data collection (if different)	*TEC*

Beneficiaries in charge of the data analysis (if different)	*SURREY, UOWM, CERTH, 0INF, TEC, SH*
Beneficiaries in charge of the data storage (if different)	*TEC*
WPs and tasks	*WP3, T3.1*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*Metadata not yet defined.*
Standards, Format, Estimated volume of data	• *Proprietary format using common data model of SPEAR*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*The dataset will be used for the anomaly detection algorithms of the big data analytics component (T3.2) and visual IDS component (T3.3)*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The datasets will be confidential and only for the members of the consortium.*
Data sharing, re-use and distribution (How?)	*The datasets can be shared to support other WP and tasks as defined in the DoA.*
Embargo periods (if any)	
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Data will be stored in a suitable form (e.g. security mechanisms will be studied since the collected data needs to fulfil forensics requirements) in servers indicated by the pilots or the technology providers.*

5.1.2 Datasets for BDAC

Table 6 - CERTH – Big Data Analytics Component

CERTH-Big Data Analytics Component	
Dataset Information	
Dataset / Name	*Smart Home network traffic dataset for anomaly detection*
Dataset Description	*The dataset includes both normal and abnormal network traffic and several network traffic features to be used for anomaly detection.*
Dataset Source	• *In: Smart devices, gateways and sensors of the smarthouse* • *How: Wireshark, AlienVault OSSIM*
Beneficiaries services and responsibilities	
Beneficiary owner of the component	*SURREY*
Beneficiaries in charge of the data collection (if different)	*CERTH*
Beneficiaries in charge of the data analysis (if different)	*SURREY, CERTH*
Beneficiaries in charge of the data storage (if different)	*CERTH*
WPs and tasks	*WP3, T3.2*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*Metadata not yet defined.*
Standards, Format, Estimated volume of data	• *Proprietary format using common data model*

	of SPEAR - *Data volume In = number of smart devices x time duration of capture x type of network traffic*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*The dataset will be used for the anomaly detection algorithms of the big data analytics component*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The datasets will be confidential and only for the members of the consortium.*
Data sharing, re-use and distribution (How?)	*The datasets can be shared to support other WP and tasks as defined in the DoA.*
Embargo periods (if any)	
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Data will be stored in a suitable form (e.g. encrypted) in servers indicated by the pilots or the technology providers.*

Table 7 - SURREY – Big Data Analytics Component

SURREY - Big Data Analytics Component	
Dataset Information	
Dataset / Name	*network traffic dataset for anomaly detection*
Dataset Description	*The dataset includes both normal and abnormal network traffic and several network traffic features to be used for anomaly detection.*
Dataset Source	- *In: Use case devices, gateways and sensors from the pilots* - *How: Wireshark, AlienVault OSSIM*
Beneficiaries services and responsibilities	
Beneficiary owner of the component	*UOWM*
Beneficiaries in charge of the data collection (if different)	*CERTH*
Beneficiaries in charge of the data analysis (if different)	*UOWM, SURREY, CERTH*
Beneficiaries in charge of the data storage (if different)	*CERTH*
WPs and tasks	*WP3, T3.2*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*Metadata not yet defined.*
Standards, Format, Estimated volume of data	- *Proprietary format using common data model of SPEAR* - *Data volume In = number of devices x time duration of capture x type of network traffic*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*The dataset will be used for the anomaly detection algorithms of the big data analytics component*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The datasets will be confidential and only for the members of the consortium.*
Data sharing, re-use and distribution (How?)	*The datasets can be shared to support other WP and tasks as defined in the DoA.*

Embargo periods (if any)	
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Data will be stored in a suitable form (e.g. encrypted) in servers indicated by the pilots or the technology providers.*

5.1.3 Datasets for Visual-Based IDS

Table 8 - CERTH – Visual-based IDS

CERTH_Visual-based IDS	
Dataset Information	
Dataset / Name	*Smart Home clustered network traffic dataset*
Dataset Description	• *In: Real-time network traffic capture* • *Out: Visualization points and coordinates*
Dataset Source	• *In: Smart devices, sensors, gateways* • *How: Wireshark, AlienVault OSSIM*
Beneficiaries services and responsibilities	
Beneficiary owner of the component	*SH*
Beneficiaries in charge of the data collection (if different)	*CERTH*
Beneficiaries in charge of the data analysis (if different)	*SH, CERTH*
Beneficiaries in charge of the data storage (if different)	*CERTH*
WPs and tasks	*WP3, T3.3*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*Graph coordinates, timestamp*
Standards, Format, Estimated volume of data	• *Proprietary format using common data model of SPEAR* • *Data volume In = number of smart devices x time duration of capture x type of network traffic* • *Data volume Out = number of nodes x graph space dimensions x frequency and amount of communications*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*The datasets will be used for the visual identification of normal/abnormal activities in the network in the pilot sites.*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The datasets will be confidential and only for the members of the consortium.*
Data sharing, re-use and distribution (How?)	*The datasets can be shared to support other WP and tasks as defined in the DoA.*
Embargo periods (if any)	
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Data will be stored in a suitable form (e.g. encrypted) in servers indicated by the pilots or the technology providers.*

5.1.4 Datasets for GTM

Table 9 - CERTH – GTM

CERTH_GTM	
Dataset Information	
Dataset / Name	*Smart home's nodes reputation over time*
Dataset Description	*A set of database records which capture the change of reputation and trust of smart home's devices, sensors and gateways over time.*
Dataset Source	*Smart devices, sensors, gateways*
Beneficiaries services and responsibilities	
Beneficiary owner of the component	*SURREY*
Beneficiaries in charge of the data collection (if different)	*CERTH*
Beneficiaries in charge of the data analysis (if different)	*SURREY*
Beneficiaries in charge of the data storage (if different)	*SURREY, CERTH*
WPs and tasks	*WP3, T3.4*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*Type of device, timestamp of reputation change*
Standards, Format, Estimated volume of data	• *Proprietary format using common data model of SPEAR*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*The dataset will be used for the validation of GTM component in the smart home scenario.*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The datasets will be confidential and only for the members of the consortium.*
Data sharing, re-use and distribution (How?)	*The datasets can be shared to support other WP and tasks as defined in the DoA.*
Embargo periods (if any)	
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Data will be stored in a suitable form (e.g. encrypted) in servers indicated by the pilots or the technology providers.*

5.2 Datasets for SPEAR-FRF

5.2.1 Datasets for AMI Honeypots

Table 10 - TEC-AMI HONEYPOTS

TEC- ***AMI HONEYPOTS***	
Dataset Information	
Dataset / Name	*System activity*
Dataset Description	*The dataset includes network traffic data, operating system shell commands, keystrokes, communications and syslogs.*

Dataset Source	• *In: UC2- The Substation Scenario* • *How: As a basis open-source honeypots can be used (conpot, CryPLH…)*
Beneficiaries services and responsibilities	
Beneficiary owner of the component	*TEC, SCH*
Beneficiaries in charge of the data collection (if different)	*TEC*
Beneficiaries in charge of the data analysis (if different)	*TEC*
Beneficiaries in charge of the data storage (if different)	*TEC*
WPs and tasks	*WP4, T4.3*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*Metadata not yet defined.*
Standards, Format, Estimated volume of data	• *Proprietary format using common data model of SPEAR* •
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*The dataset will be used for the identification of cyber attacks, collection of intelligence about attack strategies and possible countermeasures needed and also as deception technology against attackers.*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The datasets will be confidential and only for the members of the consortium.*
Data sharing, re-use and distribution (How?)	*The datasets can be shared to support other WP and tasks as defined in the DoA.*
Embargo periods (if any)	
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Data will be stored in a suitable form (e.g. security mechanisms will be studied since the collected data needs to fulfil forensics requirements) in servers indicated by the pilots or the technology providers.*

5.3 Datasets for SPEAR-CHF

5.3.1 Datasets for SPEAR-RI

Table 11 - TEC-AMI HONEYPOTS

TEC- *SPEAR-RI*	
Dataset Information	
Dataset / Name	*Cyber attacks and threats data*
Dataset Description	*The dataset includes Cyber attacks and threats data for information sharing through an anonymous channel/repository.*
Dataset Source	• *In: Smart grid systems of the use case scenarios*

	- *How: to be defined. There are different options: to be filled by a system operator/administrator or automatically by the IDS system and confirmed manually by a system operator/administrator*
Beneficiaries services and responsibilities	
Beneficiary owner of the component	*TEC (UOWM, 8BL – to be defined)*
Beneficiaries in charge of the data collection (if different)	*TEC, UOWM, 8BL*
Beneficiaries in charge of the data analysis (if different)	*TEC, UOWM, 8BL*
Beneficiaries in charge of the data storage (if different)	*TEC, UOWM, 8BL*
WPs and tasks	*WP5, T5.1*
Standards	
Info about metadata (Production and storage dates, places) and documentation?	*Metadata not yet defined.*
Standards, Format, Estimated volume of data	- *Proprietary format using common data model of SPEAR*
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	*The dataset will be used for the threat intelligence information sharing among industrial partners.*
Data access policy / Dissemination level (Confidential, only for members of the Consortium and the Commission Services) / Public	*The datasets will be confidential and only for the members of the consortium.*
Data sharing, re-use and distribution (How?)	*The datasets can be shared to support other WP and tasks as defined in the DoA.*
Embargo periods (if any)	
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	*Data will be stored in a suitable form in servers indicated by the pilots or the technology providers.*

6. Conclusions

Table of Figures

Table of Tables

References

(Metadata Standards) Metadata Standards Directory Working Group http://rd-alliance.github.io/metadata-directory/

(H2020, 2016) H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. Version 3.0 26 July 2016. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

(EC) EC guidance on Data management in H2020 http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm