Bridging the security, privacy, and data protection gap for smaller enterprises in Europe

D8.9 The SENTINEL Data Management Plan

Project Information

Grant Agreement Number	**101021659**
Project Acronym	SENTINEL
Project Full Title	Bridging the security, privacy, and data protection gap for smaller enterprises in Europe
Starting Date	1st June 2021
Duration	36 months
Call Identifier	H2020-SU-DS-2020
Topic	H2020-SU-DS-2018-2019-2020 Digital Security and privacy for citizens and Small and Medium Enterprises and Micro Enterprises
Project Website	https://www.sentinel-project.eu/
Project Coordinator	Dr. George Bravos
Organisation	Information Technology for Market Leadership (ITML)
Email	gebravos@itml.gr

Document Information

Work Package	**WP8**
Deliverable Title	The SENTINEL Data Management Plan
Version	4.0
Date of Submission	29/11/2021
Main Editor(s)	Tatiana Trantidou (ITML)
Contributor(s)	ITML, IDIR, STS, CCS, INTRA, AEGIS, LIST, FP
Reviewer(s)	Georgios Tsirantonakis (TSI), Kostas Poulios (STS)

Document Classification							
Draft		**Final**	X	**Confidential**		**Public**	X

History			
Version	**Issue Date**	**Status**	**Distribution**
1.0	01/11/2021	Draft	Confidential
2.0	15/11/2021	Draft	Confidential
3.0	24/11/2021	Draft	Confidential
4.0	29/11/2021	Final	Public

Table of Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Explanation
AAA	Authentication, authorization and accounting
CERTS	Computer Emergency Response Teams
CPU	Central Processing Unit
CS	Cybersecurity
CSIRTS	Computer Incident Response Teams
DMP	Data Management Plan
DOI	Digital Objective Identifier
DPAs	Data Protection Authorities
FAIR	Findability, Accessibility, Interoperability, Reusability
FAQ	Frequently Asked Questions
FVT	Forensics Visualisation Toolkit
GDPR	General Data Protection Regulation
IAM	Identity and Access Management
IdMS	Identity Management System
NVD	National Vulnerability Data
OpenVAS	Open Vulnerability Assessment Scanner
ORDP	Open Research Data Pilot
PDP	Personal Data Collection
RAM	Random Access Memory
RASE	Risk Assessment for Small Enterprises
ROPA	Record Of Processing Activities
SIEM	Security Information and Event Management
SPAP	Security and Privacy Assurance Platform
VM	Virtual Machine

Executive Summary

The SENTINEL Data Management Plan (SENTINEL DMP) describes how research data will be handled by the consortium during and after the end of the project, what data will be collected, processed and/or generated, which methodology and standards will be applied, whether data will be shared/made open, and how data will be curated and preserved. Ethics compliance, privacy, data protection and security of collected/processed/generated data are of the utmost importance for the success of the SENTINEL project. In this context, the SENTINEL DMP provides information on the project data lifecycle, addressing both research data and scientific publications.

Besides addressing the Open Research Data Pilot (ORDP) requirement in terms of implementation of the Findable, Accessible, Interoperable and Reusable (FAIR) principles, this document also includes an ethics compliance dimension. In this sense, it describes the life cycle of any personal data processed for research purposes within the scope of the project. It also enables transparency concerning potential ethics, privacy and data protection risks. Since DMP is a living document, it will be updated to reflect significant changes that may arise in the data sources exploited in the project and on policies and methodologies to be used as the project evolves. Each project partner handling and/or being responsible for data collected, stored or used in SENTINEL will ensure compliance with the strategy outlined in this document. All partners shall refer to this DMP, if questions about the project's data policies and practices arise.

1 Introduction

1.1 Purpose of the Document

Innovative and research projects such as SENTINEL can potentially produce a large set of data and/or process various categories, sometimes including sensitive data. The data processed during the project realisation might relate to numerous objects, as well as natural persons. Also, data might be produced in laboratory-based testing, pilot studies, during various types of observations or by reuse of already existing data.

To satisfy legislative and ethics requirements, the project must be realised in accordance with national, international laws as well as specific sources of regulation applicable on the projects supported by the Horizon 2020 Programme (H2020). For this reason, during the project implementation, the project consortium must ensure that legal and ethical requirements concerning data processing are respected. The latter has been further addressed in Deliverable D9.1 “POPD -requirement No. 1”, which was submitted in M3 (September 2021). Furthermore, they should also ensure that the data deployed by the project stay accessible, namely certain data could be potentially interesting for scientific communities and researchers due to the potential value of the data. Therefore, the project consortium should plan not only how to create and process data properly, but also enable appropriate access to the data.

This document regulates data management by outlining how data collected, processed and/or generated within the SENTINEL project should be handled during and after the project. The document has a form of the project deliverable, and it is titled as ‘SENTINEL Data Management Plan.’ This document is the first version of the DMP. It includes an overview of the data to be produced by the project and the specific rules that should be followed with regard to data processing. The DMP is a living document, and it is expected to be additionally developed as the project progressing.

This DMP is a source of formal self-regulation that outlines how data will be handled during the course of the SENTINEL project and after the project completion. It provides a strategy for managing data generated and collected during the realisation of the project. As a European Union (EU) funded innovation project, SENTINEL includes several phases. During the project phases data are collected, processed and analysed. Afterwards, findings should be published and/or shared (see Figure 1). Also, all published data should be properly preserved and remain available for further use as appropriate. Each phase contains many subphases that involve specific ways of treating data (e.g., validation, transcription, translation, digitalisation, anonymisation, pseudonymisation, creating metadata). Therefore, it is quite necessary to have a plan on how to handle data in all project phases.

Researchers should think about the data that they will process at the very beginning of the research cycle. This requirement creates the need for and explain the purpose of the DMP. The DMP provides an analysis of the main elements of the data management that will be enforced in the SENTINEL project. It covers the complete research data life cycle. Therefore, the DMP provides insights on what data will be generated, how it will be exploited, how it will be made accessible, reusable, curated, and preserved during and after the realisation of this project.

This DMP is the means of proving that the project consortium is aware of the potential ethical implications of the SENTINEL project and confirms its commitment to respect the ethical standards and rules of H2020. Thus, the ethical standards and guidelines of H2020 including those that regulate data processing/protection will be rigorously applied, regardless of the country

in which the research takes place. This deliverable confirms that all project partners will conduct research in accordance with the fundamental principles of research integrity, such as reliability, honesty, respect, and accountability. To ensure the safety and dignity of individuals whose data will be processed within the project, the highest standards of research integrity will be applied.

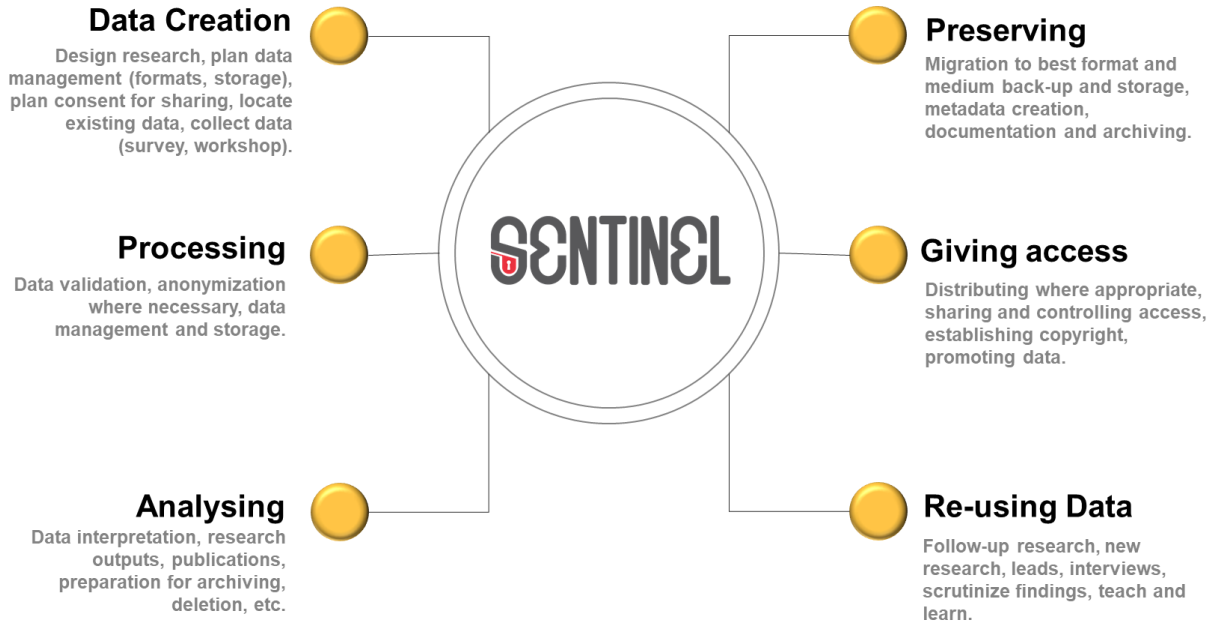

Figure 1. Key elements of the SENTINEL Data Management Plan

1.2 Structure of the Document

This DMP is composed of several sections. After the introduction and brief presentation of SENTINEL activities and the DMP from conceptual perspective, section 2 includes an indicative list with the data that will be collected/created within this project. This DMP is based (inter alia) on the Guidelines on FAIR Data Management in H2020 (European Commission, 2016). Therefore, section 3 on FAIR principles provides a set of criteria that have to be addressed to ensure data to be Findable, Accessible, Interoperable and Re-usable (FAIR). Section 4 contributes to the ethics compliance framework of the SENTINEL project. The composition of this DMP goes beyond the FAIR principles and additionally strengthens ethics compliance within the SENTINEL project. Section 5 briefly explains how available resources will be allocated to properly cover the needs of the development and execution of the DMP. Section 6 provides an overview of data security needs. Finally, the DMP is rounded up in Section 7 by presenting the conclusions.

1.3 Intended readership

The DMP encompasses ethical, legal, and managerial aspects of data processing and data protection. The DMP outlines technical and organisational measures that should be applied to secure data, as well as regulatory and ethical principles applied to protect data. Taking into account that ethics is considered as a benchmark for regulating data protection, the DMP (and this deliverable) should be perceived not only as a complementary but supplementary source of regulation. This deliverable is sectorial and concerns the processing and generation of data in all research activities. It sets up the concrete measures that should enable appropriate processing

and protection of data (in case these are sensitive data). Therefore, its application should prevail in all project activities that include data processing, i.e., WP2 – 'The SENTINEL privacy and personal data protection technologies', WP3 – 'The SENTINEL digital core, and WP4 – 'The SENTINEL services.' Considering that the project's pilots will be based on the processing of substantial data, the DMP is critically important for the proper realisation of project pilots. In addition, the DMP is equally valuable for dissemination purposes with regard to scientific publications and their open-access plans. Therefore, this DMP is intended for a broad audience that transvers all project partners.

1.4 Definitions

In this section, the definitions relevant to the DMP are introduced.

- **Data Management Plan (DMP):** A written formal document that outlines how data will be handled during the course of the project and after the project completion.
- **Personal Data:** Any information that relates to an identified or identifiable living individual.
- **Sensitive Data:** Any classified information that must be protected and is inaccessible to outside parties, unless specifically granted permission.
- **Digital Object Identifier (DOI):** Digital Object Identifier is a unique permanent identifier for a published digital object, standardised by the International Organisation for Standardisation (ISO).
- **FAIR:** Research data that is Findable, Accessible, Interoperable and Re-usable. These principles aim to provide a framework to ensure that research data can be accessed and re-used effectively.
- **Metadata:** Metadata is a set of data that provides context or additional information about other data. The main types of metadata include descriptive metadata, structural metadata, and administrative metadata.
- **Open Access (OA):** Open Access refers to the unrestricted access to research results including scientific peer-reviewed publications and research data. There are two complementary mechanisms for achieving OA to research. Green OA means that the authors will publish the accepted manuscript in an online repository. Gold OA means that the publication is directly available free of charge from the publisher and any related costs referred to as Article Processing Charges (APCs) shall be covered by the authors.
- **Open Research Data:** Research data needed to validate the results of the publications that are openly available in digital form for access and re-use by anyone for any purpose.
- **Open Research Data Pilot (ORDP):** The Open Research Data Pilot aims to improve and maximise the accessibility and re-usability of research data generated by H2020 projects without violating the privacy of sensitive data.
- **Repository:** A digital repository is an archive for storing and managing digital copies.
- **NextCloud:** NextCloud is an open-source software versioning and revision control system that tracks changes made to files, folders, and directories. It eases data recovery and provides history of changes.
- **Zenodo[1]:** Zenodo is a research data archive/online repository created by OpenAIRE and CERN (Conseil Européen pour la Recherche Nucléaire) for sharing research results in a wide variety of formats for all fields of science.

[1] https://zenodo.org

2 Data Summary

In accordance with the SENTINEL framework architecture (presented as part of the Deliverable D1.2 “The SENTINEL technical architecture” – submitted in M6), SENTINEL comprises of a number of ***modules*** (i.e., pieces of software that form part of a SENTINEL context, such as the self-assessment engine, the recommendation engine) and ***contexts*** (i.e., a collection of modules that operate under a common setting, such as the MySentinel front-end, the Self-assessment, the Digital Core, the Observatory, etc). Therefore, the core of SENTINEL comprises a) the MySentinel (front-end) dashboard, b) the DPIA self-assessment, c) the Digital Core and d) the Observatory.

Complementary to these, there are pieces of software (which may store data) that are connected with the SENTINEL platform, but do not form part of the core architecture. These are named as ***plugins*** and can be optionally integrated into the system, as they implement a specific interface; i) Security Infusion, ii) MITIGATE, iii) CyberRange, iv) Security and privacy assessment platform – SPAP, v) GDPR self-assessment, vi) Forensics Visualisation Toolkit, vii) other external plugins. It is worth mentioning that each (external) plugin has its own data management policy, which is independent of the SENTINEL core, and thus, it will not be analysed as part of the Data Management Plan. When the user of the SENTINEL framework is directed to an external plugin, they will be presented with a clear disclaimer message pointing to the plugins' privacy policy.

There are three ***repositories*** in SENTINEL that provide both storage services for important elements used within SENTINEL, as well as functionalities of accessing and updated the stored information; a) the *Policies repository* that collects individualised, validated policy elements that can be used for policy drafting and policy reuse, b) the *Plugins repository* that contains links to SENTINEL plugins, together with list of capabilities offered by each plugin and list of configurations that are necessary for their execution, c) the *Training repository* that is enriched with external training content made available to SMEs/MEs. Additionally, there is a database where the SME profiles will be stored, as filled-in by their representatives, as well as information generated by the SENTINEL platform i.e., their RASE score, and suggestions for plugins and policy drafts, as well as their state of policy compliance and the steps it has undertaken.

An overview of the data that will be collected and generated within the SENTINEL core are summarized in Table 1, while data collected and generated by the external plugins are provided in Table 2. Based on the discrimination of the above framework components, we consider that data collection within SENTINEL’s will be performed from SME participants in two phases: (a) through the self-assessment context during the SME profiling phase to elicit the initial cybersecurity (CS) and Personal Data Protection (PDP) requirements with a view to providing a minimum viable set of RASE score components; and (b) during the SME’s participation in a number of assessment pipelines in the form of plugins.

The data listed below are divided into two categories regarding their accessibility level, namely Confidential and Open Access (OA). Due to the nature of the project, the majority of produced data will be confidential and will only be shared within the consortium. OA is mostly applicable for the *Observatory context*, which constitutes SENTINEL’s connection point to the outside world; the Observatory essentially collects information (cyberthreat and data privacy information) from external knowledge sources (expert-based or digitally available) and makes them available directly to the end users (after anonymisation) or to other modules and tools within SENTINEL.

In some cases, the consortium will (pseudo)anonymise generated data to produce OA data that can be provided free of charge, publicly shared and will be included in the ORDP. However, the data generated from the external plugins are to remain confidential. As the project evolves, this list may require modifications (addition or removal of data) in the updated versions of the DMP.

Table 1. SENTINEL data collection and generation summary

Case #1	Generic SME profiling, assessment and policy recommendations
Responsible	IDIR
Purpose and relation to the objectives of the project	To perform SME profiling in order to elicit the initial cybersecurity (CS) and Personal Data Protection (PDP) requirements with a view to providing a minimum viable set of RASE score components. In some cases, participant SMEs will participate in one or more of assessment pipelines as dictated by their goals and requirements. The outcome of each assessment (assessment results) will weigh in on the corresponding RASE score components as necessary.
Data Types	**PDP Profiling (4 phases)** ***a. Phase 1: SME profile and structure*** i. size (quantitative – predefined employee count(s)) ii. sector/segment and operating environment (structured) iii. structure – PDP departments or persons (structured) ***b. Phase 2: Personal data processing (PDP) operations; Steps below should identify*** i. individual PDP operations (data processor roles) (structured text - list) ii. types of data processed (structured text - list) iii. the purpose of processing (text) iv. the processing operations (structured text - list) v. the means of processing (structured text - list) vi. where processing takes place (structured text - list) vii. the data subjects (structured text - list) viii. the recipients of the processed data (text) ix. the volume of PDP (structured quantitative – predefined templates) x. the timespan of PDP (structured date – time range) xi. the potential involvement of special / sensitive data subjects (1. Boolean; 2. structured text – list) xii. the identifiability, observability and linkability of data (structured text – list - qualitative) xiii. the criticality of the PDP operation in terms of impact (low/medium/high/very high) of a potential CS breach, evaluated as the loss of some or all of {confidentiality, integrity and availability} of this data. (structured text – list - qualitative) xiv. the likelihood of a breach taking place, related to specific security threats (structured text – list - qualitative) xv. the overall risk level for the assessed PDP operation (structured text – list - qualitative) ***c. Phase 3: the SME's self-declared CS & PDP goals and constraints*** i. Self-declared CS gaps (structured text - list) ii. Self-declared PDP gaps (structured text - list) iii. Budget constraints, declaration of range, based on suggestions (quantitative – predefined annual budget range(s))

	iv. HR constraints (ICT staff CS expertise and training requirements), (structured text – lists 3-4 categories) ***d. Phase 4: SME patterns and templates*** i. SME CS and PDP assets inventory analysis (to be defined) ii. Mapping of the SME structure, status and goals to generic predefined templates (to be defined) iii. Adjustment of specific attributes and weights on the template to inform the requirements elicitation process (to be defined)
Data formats	The data will be collected through web forms within the integrated 'self-assessment portal' web app. No file attachments are foreseen as necessary. Most of the data will be collected as plain text, logs, numeric (various), boolean and date formats, including ranges, and stored in structured data types and arrays. The SME participants will also enable the linking between data entities and mapping during phase 4 of the profiling phase. Data will be subject to multiple validation criteria both at the collection and storage phase and when used for inference or to inform RASE score components.
Re-use of existing data	Internal reuse of the data is foreseen within SENTINEL to some degree. The data will not be reused as-is, but they will inform the RASE score components. The RASE score is a persistent composite data structure which a) contains all the evaluated results of both the SME profiling and the SME assessments in a way which is machine-readable by the recommendation engine and b) is attached to the SME during its interaction with every SENTINEL context.
Data origin	All data related to the SENTINEL self-assessment context are self-declarations of the SME participant and are generated at the time of input.
Expected size of data	Estimated 10-500 KB
Data utility	The data are strictly for internal use only (RASE score generation) and will not be accessible to external parties. The consortium is currently examining the potential of anonymization to use these data as a basis for research (e.g., anonymous statistical reports).
Privacy principles	The consortium is currently researching the potential for data pseudonymization/anonymization to use as a basis for openness for research.
Accessibility	Data provided and generated in the self-assessment context are not to be externally shared; they will remain confidential for consortium use only.

Case #2	The SENTINEL Observatory (data reuse and exchange)
Responsible	ITML
Purpose and relation to the objectives of the project	The Observatory constitutes SENTINEL's knowledge hub, making anonymized, CS and PDP-related data available to both SENTINEL end-users and external entities and data sources, for example Open security data sharing platforms, regulators, CERTs, CSIRTs, DPAs etc. There is wide range of data stored in the Observatory knowledge base that are either collected from selected, reputable external sources or produced by

	SENTINEL intelligent services, e.g., reusable policy elements produced by the Policy Drafting module. The Observatory shares these data with end-users via a set of digital collaboration tools, FAQs, forums etc. Regarding external sources, the bidirectional exchange of strictly anonymized data is realized using SENTINEL's dedicated modules that are constantly updated and validated by security experts.
Data Types	Data stored and exchanged are of a variety of formats, including unstructured text and documents, images and videos.
Data formats	All popular formats for the above data types: html, txt, pdf, doc, png, jpg, mp4, avi, etc.
Re-use of existing data	The Observatory Knowledge Base will be constantly updated by existing data from external sources. Additionally, it facilitates data reuse in the case of reusing identified patterns in policy drafting (Data reuse policy module).
Data origin	Observatory data originate either from reputable external sources or from SENTINEL intelligent services (the Recommendation Engine, the Policy drafting, Data reuse policy, Incident reporting, Notification Aggregator)
Expected size of data	Expected moderate data volumes, comprising mainly text descriptions, documents, images, informative videos or tutorials. Although at this point, a precise estimate is not possible, we expect that this should be less than ~250GB.
Data utility	Observatory data will be used for knowledge sharing and information exchange for SENTINEL participants and the external entities interested in accessing constantly updated security and privacy knowledge.
Privacy principles	As data are interchanged with external entities, it is an absolute requirement for all data to be anonymized before stored in the Observatory Knowledge Base.
Accessibility	All data offered by the Observatory can be accessed by all SMEs/MEs, end-users and the above-mentioned categories of external entities and sources.

Table 2. SENTINEL plugin data collection and generation summary.

Case #3	Security Infusion
Responsible	ITML
Purpose and relation to the objectives of the project	Security Infusion is a Security Information and Event Management (SIEM) component that participates in SENTINEL as a plugin. The capabilities that Security Infusion offer include: • Detection of vulnerabilities on a company's digital infrastructure • Detection of cyber threats towards a company's digital operations • Detection of suspicious file modifications and data leaks; • Monitoring of the operation of devices and reporting of anomalous behaviour

	• Managing the infrastructure from a central UI dashboard In the context of SENTINEL, Security Infusion can be recommended to an SME/ME, after the latter has conducted a self-assessment in order to address security shortcomings found in the above list. Additionally, Security Infusion can run independently and periodically on a company's infrastructure for purposes of reporting suspicious events and vulnerabilities to the end-user via the SENTINEL's Notifications Aggregator.
Data Types	Structured textual information for detected events and vulnerabilities, and documents for produced reports.
Data formats	JSON for structured textual information and PDF/HTML for documents
Re-use of existing data	There is no direct reuse of existing data. However, Security Infusion can analyse large amounts of anonymous historical data of the company where it is executed. This analysis facilitates more accurate detections of security related events.
Data origin	Security Infusion produces all data that are then sent to the end-user. The input data that Security Infusion uses are Operating System logs, monitoring information (e.g., CPU/RAM utilisation) and network activity statistics.
Expected size of data	Security Infusion produces data constantly during its operation. For a single organisation, it is expected to produce ~1MB per hour of operation.
Data utility	The data are strictly for **internal use only.** Events and incidents that are identified from application to SMEs infrastructure are used to update external sources.
Privacy principles	Security Infusion **does not collect, process or store any personal data**.
Accessibility	Data produced by Security Infusion are directly accessed only by a representative of the SME/ME that uses this plugin via password authentication mechanism.

Case #4	CyberRange
Purpose and relation to the objectives of the project	The purpose of this plugin is the creation of scenarios for simulation and training purposes with the aim to increase the SME's cybersecurity awareness and build staff capacity. The creation of the simulation and training scenarios is based on the infrastructure of the SME for network topology creation and the collection of a template (Virtual Machine – VM or Docker).
Data Types	Text, VM
Data formats	Json, ovf Text (report)
Re-use of existing data	Existing data can be reused, such as the VM templates and topology. Participant SME can share the VM and topology, if there agree, by creating a new template.
Data origin	VM templates are created within Airbus CyberSecurity. Information on scenarios and infrastructure is collected from the participating SMEs.

Expected size of data	GB for VMs, MB for reports
Data utility	**The data are strictly for internal use only**. Reports will be sent to the participant SME.
Privacy principles	CyberRange **does not collect, process or store any personal data**. Data (templates for VM, scenarios) are limited to the purposes of the project and are saved in the CyberRange repository. A backup system is in place.
Accessibility	The data will be saved in the CyberRange platform and will be accessible via password authentication mechanism.

Case #5	Forensics Visualisation Toolkit (FVT)
Responsible	AEGIS
Purpose and relation to the objectives of the project	FVT offers a complete toolset for IT security data collection, processing and visualisation that analuses monitored components (i.e., computers, network devices, etc) and provides a comprehensive overview of the system operation.
Data Types	Data consumed will be structured, quantitative, and a combination of text and numeric type regarding the metric's needs.
Data formats	Data collected and generated will be of JSON format.
Re-use of existing data	Data will not be reused. The consortium will consider the anonymisation of reported incidents to make them openly available in the SENTINEL Observatory.
Data origin	For the purpose of SENTINEL, fabricated synthetic data provided by public open-source online data repositories will be used, if required, for testing purposes and only within the context of the SENTINEL pilot cases.
Expected size of data	The expected size of data transmitted at any time will not exceed 8KB/second. And the total stored data size is not expected to reach more than a few hundred MB.
Data utility	The data are strictly for **internal use only**.
Privacy principles	AEGIS **does not collect, process or store any personal data.**
Accessibility	Results will be strictly for internal use within the consortium. Results of testing and outcomes will be locally stored by the use case providers.

Case #6	Security and Privacy Assessment Platform (SPAP)
Responsible	STS
Purpose and relation to the objectives of the project	To identify and describe the processes within the participant SME, its personnel, the systems software, hardware, physical and data assets, the threats corresponding to these assets and the sequence of events that leads

	to the manifestation of these threats, the security properties that must be maintained for each asset, the vulnerabilities that compromise the security properties and the security controls that mitigate the exploitation of the vulnerabilities.
Data Types	Numeric, text, logs
Data formats	OpenVAS, NVD (National Vulnerability Database) and other custom formats
Re-use of existing data	The consortium will consider anonymising the produced reports for making them available for reuse via the SENTINEL Observatory.
Data origin	Data (IT assets) are provided by the participant SMEs.
Expected size of data	10-15Kb per day per assessment per SME
Data utility	The data are strictly for **internal use only**, as they will expose the SME's security posture. They will thus remain confidential for use within the consortium. The consortium will consider producing anonymous statistical reports for making them available in the SENTINEL Observatory.
Privacy principles	SPAP **does not collect, process or store any personal data.** SPAP will be used with the consent of the participating SME.
Accessibility	Data will remain confidential for use strictly within the consortium. Anonymised data could be shared via the SENTINEL Observatory.

Case #7	MITIGATE for policy drafting
Responsible	FP
Purpose and relation to the objectives of the project	The collection and generation of data will support the amelioration of CS and PDP posture of SMEs and MEs
Data Types	Qualitative and quantitative data
Data formats	Doc, pdf
Re-use of existing data	The consortium will consider anonymising the produced policy reports for making them available for reuse via the SENTINEL Observatory.
Data origin	Data are provided by the participant SMEs/MEs
Expected size of data	N/A
Data utility	The data are strictly for **internal use only.**
Privacy principles	MITIGATE **does not collect, process or store any personal data.** The data that will be collected will be proportional to the purpose of the project since they will be related only to the risk management consistent process that will be followed throughout the project. Data will not be further shared/transferred.
Accessibility	Datasets to be collected will not be widely shared due to potential contractual reasons.

Case #8	GDPR Compliance Assessment
Responsible	LIST
Purpose and relation to the objectives of the project	To establish and maintain a Record of Processing Activities (ROPA) for the participant SME, each record providing a description of personal data processing activities. GDPR Compliance Assessment is based on the collection and processing of data from ROPA to determine the data protection capabilities level.
Data Types	Structured data
Data formats	XLS, CSV or/and JSON
Re-use of existing data	Data re-use is based on re-using the processing activity descriptions (i.e., Records of Processing Activities)
Data origin	Records of processing activities are provided by the online self-assessment performed by the end users (SMEs/MEs)
Expected size of data	Not yet known precisely
Data utility	The data are strictly for **internal use only.** Data produced will be useful for the participant SME (CEO, DPO, Product owner, etc).
Privacy principles	GDPR Compliance Assessment **does not collect, process or store any personal data.**
Accessibility	Access of data is restricted to the participant SME and the SENTINEL project members via the SENTINEL platform, where access control rules will apply. SMEs can make data accessible by providing the description of processing activity available in ROPA hosted by the SENTINEL platform.

3 The FAIR requirements

The project incorporates the FAIR data principles (Wilkinson et al. 2016), building on the *'Guidelines on FAIR Data Management in Horizon 2020'* (European Commission Directorate-General for Research & Innovation, 2016) and will state in detail the type of data and research outputs and how those data will be findable, accessible, interoperable and reusable. The FAIR principles naturally apply to data that will be openly available for public use, which primarily concerns the Observatory context. These data will comply with the provisions described in this section. Data that are considered confidential due to internal regulations and/or legal reasons that data providers ought to comply with will be shared only within the consortium. The SENTINEL data management policy that will be followed for all the available data is summarised below and will be described in the following sections.

3.1 Findable data

Making data findable, including provisions for metadata: Data and metadata collected/generated within the project will be made readable by humans and computer systems (machine-actionability).

- **F1.** (Meta)data should be assigned a globally unique and Persistent Identifier (PID); The use of PID will be used to enhance findability of **both data and metadata** in accordance with the European Open Science Cloud (EOSC) PID Policy (EOSCsecreatariat.eu, 2019), including documentation files (e.g., pdf files, reports, etc).
- **F2.** Data needs to be described with rich metadata.
- **F3.** Metadata should clearly and explicitly include the identifier of the data they describe.
- **F4.** (Meta)data needs to be registered or indexed in a searchable resource.

As a good practice towards findability of metadata by search engines, the consortium will consider the use of OAI-PMH[2] mechanisms. Whenever possible, data will also be made openly available to the community – as part of the ORDP - by deposition in trusted, free-of-charge data-sharing platforms (e.g., the SENTINEL Observatory, Zenodo) and on the project's website (e.g., public deliverables). In case of scientific publications, we will prioritise the data sharing platform of the Publisher, as it will allow associating the data repository with a Digital Object Identifier (e.g., DataCite).

As mentioned above, the **Observatory** constitutes SENTINEL's connection point to the outside world. Its main goal is to share all meaningful findings of the platform with the external actors (i.e., participant SMEs/MEs), while making good use of the constantly updated security-related knowledge that exists in a wide range of external sources. The Observatory will provide three key functions: (a) a centralised threat intelligence *Knowledge Base* (KB), which aggregates information from external sources or provided by other SENTINEL modules about recently identified data and privacy breaches, attack vectors, forensic intelligence and cyberthreats signatures and related data as well as RASE scoring information for the SENTINEL ecosystem of SMEs/MEs, and produces anonymised information to be shared with external actors and sources; (b) the *Observatory Information Exchange*, an open API platform to exchange threat intelligence with SME/ME associations, such as open source incident response platforms, CERTs, CSIRTs and DPAs, in coordination with the Digital Core's Incident Reporting, Handling

[2] https://www.openarchives.org/pmh/ (last accessed: 8.09.2021)

and Sharing and (c) the *Data Reuse policy module*, an intelligent module that coordinates the reuse of policy elements when drafting new security and privacy policies for participants, through exchanging data with the Digital Core's Policy Drafting module.

Access to the contents of this KB by end users is achieved via the MySentinel front-end module in various ways, such as a searchable knowledge base, a structured FAQ and collaboration tools. The KB exchanges information with two other modules; (b) the Observatory information exchange, an open API platform that transfers information between external open security data sharing platform and the KB.

The observatory will incorporate global benchmarks identified early in the project (included in D1.3 "The SENTINEL experimentation protocol"), while it will also provide the means to report basic recommendations and best practices to mitigate the challenges, problems and vulnerabilities identified. The SENTINEL Observatory will empower both human and machine collaboration (API-enabled) collaboration on the exchange of critical data and knowledge.

Zenodo, is a well-known data repository, developed by OpenAIRE and CERN aiming to allow researchers to store both publications and data. It gives also the opportunity of linking a given publication and a used dataset in order to produce the results the use of persistent identifiers and citations. Zenodo can easily share datasets in various sizes and formats, provide flexible licensing, and access and re-use of research data. Zenodo was developed to facilitate the FAIR principles. The consortium will examine the potential of (pseudo)anonymizing data (e.g., anonymous statistical reports, anonymized policy recommendations) to use them as a basis for openness for research. Partners will examine whether the (pseudo)anonymized data can be documented and uploaded to Zenodo with their related metadata. Knowing the impact of Zenodo in the research community, SENTINEL is willing to use Zenodo as a dissemination means to foster the impact of the SENTINEL Observatory. The users in Zenodo will be re-directed to SENTINEL Observatory in case they require more data and/or documentation. The gain of this approach is twofold; SENTINEL will make open an amount of data to the research community via Zenodo, while the SENTINEL Observatory will reach a broader audience with the help of Zenodo maximising its impact.

When data will be uploaded in an online data repository, a Digital Objective Identifier (DOI**)** will be assigned to achieve effective and persistent citation. The DOI will be further used for all the related publications so that readers will be able to link them with the underlying datasets. SENTINEL will use a **standardised naming convention** for all the project files and folders at data repositories that will be constructed using the following characteristics:

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

Search keywords describing the dataset or content of the data will be provided when a dataset is uploaded to a repository aiming at optimising possibilities of re-use.

3.2 Accessible data

Making data openly accessible: A user needs to know how the data can be accessed once

discovered.

- **A1.** (Meta)data should be retrievable by their identifier using an open, free, and universally implementable standardised communications protocol that should allow for an authentication and authorisation procedure, where necessary;
- **A2.** Metadata needs to be accessible, even when the data are no longer available.

As mentioned above, accessibility of data will be limited to white papers, recommendation reports, templates and incident reports that are all anonymized and contain no SME data. Policy recommendations and incident reports – after anonymization – will become available in the SENTINEL Observatory KB. Data provided and generated by the other SENTINEL contexts and modules (e.g., the Self-assessment context) or plugins (Security Infusion, SPAP, MITIGATE, CyberRange, FVT, etc) will remain confidential (shared within the consortium) and will not be externally shared, since they would expose the security posture of the SME to which they were applied.

As per Article 29.2 of the Model Grant Agreement (GA) under H2020, the SENTINEL consortium will ensure Open access (OA) to all peer-reviewed scientific publications relating to its results. Figure 2 below provides a general overview of the decision process that will be followed when publishing research results. The decision on whether to publish through OA will have to account for the potential necessity of data protection. The consortium will follow ‘green’ OA in the publications. In cases, where the timely OA dissemination is not possible by following the ‘green access’ model, SENTINEL will opt for ‘gold’ OA. In any case, the Consortium will carefully examine before making public any research output/information the aspects of potential conflicts against Intellectual Property Rights (IPR) protection issues.

According to Article 26 of the SENTINEL GA, research data that are generated during the action of the project, are owned by the beneficiary that generates them. The repositories that will be used for the project data have been decided by the consortium and are Zenodo and SENTINEL Observatory. They will be used for any generated open research data during the lifetime of the project. Confidential data, on the other hand, will not become public.

Research data needed for validation of results presented in scientific publications will be uploaded to Zenodo and/or SENTINEL Observatory as soon as possible. In case an embargo period should be applied before the publication of the results, data will be deposited in the project’s repository (NextCloud). OA articles will be deposited in the selected repository within 6 months of publication. Partners will also consider using ResearchGate as a repository for articles. All public deliverables provided on the project’s website will be maintained for several years after the project, and links to published journal articles will also be included. Partners that have institutional repositories (e.g., LIST) can make published articles openly available.

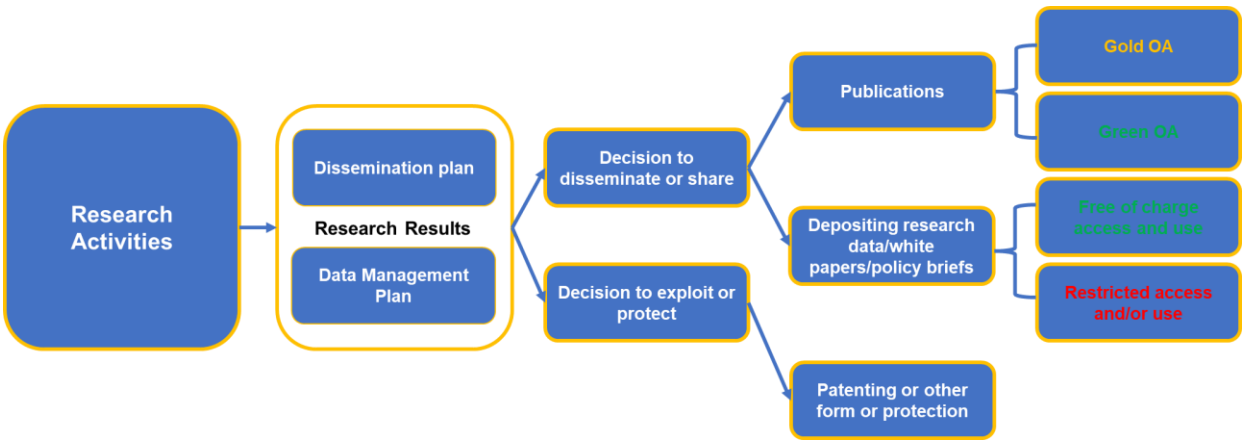

Figure 2. Open access to scientific publication and research data in the context of dissemination & exploitation (European Commission, 2017)

3.3 Interoperable data

Making data interoperable: Data should be able to be integrated with other data and interoperated with applications or workflows for storage, analysis, etc.

- **I1.** (Meta)data needs to use a formal, accessible, shared and broadly applicable language for knowledge representation;
- **I2.** (Meta)data should use vocabularies that follow FAIR principles;
- **I3.** (Meta)data should include qualified references to other (meta)data.

The SENTINEL consortium uses a variety of file formats. In order that documents are accessible and usable by all partners and can be shared with the public, partners will use common file formats as much as possible (for example: .docx, .xslx, .pdf). Such an approach is also required to facilitate partners in their everyday use of the project's shared drive, which will support software that can use common file formats.

With regard to data, technology providers will mostly use proprietary software in order to carry out their tasks in SENTINEL and so this will not be interoperable. The consortium will attempt to increase the interoperability of the provided data – where possible, by leveraging commonly used vocabularies for the metadata within the data will be used. Interoperability can be ensured by using the same standards for data and metadata capture/creation. Metadata contents, formats and schema may be updated as the project progresses.

Information generated with respect to incidents (by the Incident Broker) will be transformed in order to comply with external APIs of CERTs, CSIRTs and DPAs. The same applies for generated anonymised policies to ensure compatibility with external bodies and open security data sharing platforms.

3.4 Reusable data

Increase data re-use (through clarifying licences)*:* Metadata and data should be well-described to optimise the reuse of data.

- **R1.** (Meta)data needs to be richly described with a plurality of accurate and relevant attributes.

- **R1.1.** (Meta)data should be released with a clear and accessible data usage license.
- **R1.2.** (Meta)data should be associated with detailed provenance.
- **R1.3.** (Meta)data should meet domain-relevant community standards.

Data reusability will be achieved via two modules of the SENTINEL Observatory context; (i) the Data Reuse policy module and (ii) the Incident broker.

The Data Reuse policy module coordinates policy reuse elements when drafting new security and privacy policies for SME participants, exchanging data with the Policy drafting module (from the Digital Core). It provides the means for SMEs/MEs to report recommendations and best practices in a concise human readable way. Additionally, the module uploads anonymised policy recommendations, white papers, incident reports, *etc* to the SENTINEL Observatory repository for reuse for similar cases in the future. The Data reuse policy module identifies and shares patterns of reusable policy elements, assisting the creation of policy drafts, as well as updating the Observatory Knowledge Base with useful policy information.

The Incident broker module facilitates data transfer from the Incident Reporting module (of the Digital Core) to external actors, such as regulators, CERTs, CSIRTs and DPAs. It receives security-related incidents reported either automatically or manually by the end users, and subsequently anonymises them and forwards them to external actors.

According to the SENTINEL time plan, the aforementioned data will become available to third parties after M18. However, the consortium will examine the possibility to release these data as soon as they are generated. The data will remain public for 1 year after the completion of the project in the SENTINEL Observatory. Any restrictions that may arise during the course of the project, such as setting embargo periods or restrictions from editors of scientific journals and organisers of conferences, will be examined case by case.

The owner of the data is the beneficiary that generates it. The owner will be responsible for maintaining the data after the completion of the project. To protect the ownership of the publicly provided data, Creative-Commons-Licences (CC) will be used. CC are machine-readable licenses. The consortium will discuss and decide per case which license will be more suitable.

In terms of software licenses for algorithms, components and modules to be utilised, the consortium will consider the business-friendly Apache Software License (ASL) license, which allows the redistribution of the program's source code in any form (compiled binary or plain text). ASL is used by a wide array of Open-Source Software projects (e.g., Apache web server), and the Lesser General Public License (LGPL) software licenses.

4 DMP and Ethics Compliance

This section highlights potential ethical issues with regard to data processing. The chapter explains how they should be managed in accordance with applicable regulatory frameworks, ethical standards, and other requirements.

This section contributes to the ethics compliance framework of the SENTINEL project and to its risk management component (data protection risk management aspects). The composition of this DMP goes beyond the FAIR principles and additionally strengthen ethics compliance within the SENTINEL project. This chapter provides brief presentation of Ethics Appraisal Scheme relevant for this project as well as an overview of the legal grounds for data processing and the applicable standards, principles, and guidelines for personal data processing/protection.

4.1 SENTINEL within the EC's Ethics Appraisal Scheme

The ethics requirements that were deemed relevant for the project by the ethics panel that evaluated the proposal concerned protection of personal data (Table 3). These were properly addressed as part of WP9 under the responsibility of the Project Coordinator (PC) (ITML). The corresponding deliverable D9.1 "POPD – Requirement No. 1" was submitted in M4, Sept. 2021).

Table 3. SENTINEL Post-Grant POPD Ethics Requirements

Req. No.	Description	Action taken in fulfilment
#1	The beneficiary must submit an explicit confirmation that they have lawful basis for the data processing of previously collected data and that the appropriate technical and organisational measures are in place to safeguard the rights of the data subjects. The beneficiary must provide justification for the processing of sensitive personal data and an explanation why synthetic data cannot be used instead. In addition, the beneficiary must check if special derogations pertaining to the rights of data subjects or the processing of genetic, biometric and/or health data have been established under the national legislation of the country where the research takes place and submit a declaration of compliance with respective national legal framework(s).	Deliverable D9.1: POPD – Requirement No. 1 submitted

4.2 The legal grounds for data processing

The self-assessment phase collects data from a participant SME which are sensitive in the sense that a potential data breach might expose the SME's PDP vulnerabilities and open them up for CS attacks. SENTINEL **will only collect the minimum amount of data necessary** during both the profiling phase and the other assessment phases in order to calculate the necessary RASE score components.

It should be noted that **no special categories of personal data** (e.g., genetic data, biometric data, health data) will be collected or processed. Although the participant SMEs might be processing sensitive data of these types, **SENTINEL does not ask for the data itself**, only inquires whether such data may be processed by the SME in order to assess the related risk.

SENTINEL does not collect, process or store personal data, since SME participants can create accounts and access the platform anonymously, when allowed. SENTINEL does collect data that pertain to the CS and PDP structures and practices of the SME and will obtain the necessary consent for the processing of such data. Therefore, the primary legal ground for such

lawful processing will be the participant SME's informed consent as it is regulated by Art. 6(1)(a) of the GDPR. SMEs whose data will be processed will have the right to withdraw consent at any time without any negative consequences as well as to object processing of data related to them. Data processing could rely on the legitimate interest (Art 6(f) GDPR), where the context of data processing allows that. Data processing based on this legal ground should ensure lawfulness only if the interests of data controllers (or third parties) are not overridden by interests of rights and freedom of data subjects. It is worth mentioning that data provided and generated in the self-assessment context and through the distinct external plugins will not be externally shared and will only be used internally in SENTINEL.

The legal grounds for the lawful processing of data laid down by the GDPR are summarised in Figure 3.

Figure 3.Legal grounds for the processing of data

4.3 Applicable standards, principles and guidelines

All SENTINEL partners will comply with ethical principles and principles laid down by applicable international, EU and national laws relevant to protection of personal data (in particular the GDPR (EU) 2016/679). These laws set up principles for data protection, appropriate legal grounds for data processing, standards for data sharing and transferring as well as measures for securing data (see Figure 4).

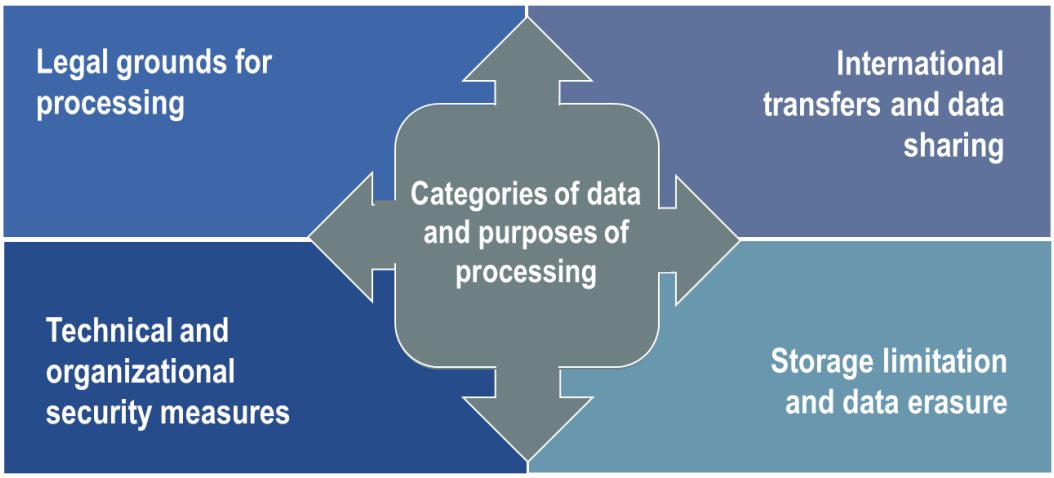

Figure 4. Elements for data classification

In order to satisfy the requirements of transparency principle, all information and consent forms will be created in an understandable manner and in the languages that participant SMEs understand. The provided information will describe the aims, methods and implications of the research, as well as the nature of the participation and any benefits or potential risks that might occur. Also, the way of collection and protection of data, its destruction and possible reuse of data will be outlined.

Special care will be given to ensure that participant SMEs have fully understood relevant information and do not feel pressured or forced to give consent. Participants will be provided with a consent form to read and sign. The signed forms containing their consents will be kept on file for inspection. Considering that the participation will be entirely voluntary, participants will be able to reject participations or to withdraw participation without any consequences. They would be able to opt-out from any further communications and have their data deleted from all project's records.

It is intended that most project activities will not contain any personal data or reference to personal data. Therefore, the amount of data collected and processed in this project are limited to the ones necessary for the profiling phase and the other assessment phases (ultimately enabling the calculation of the RASE score), and thus will be minimised. Also, SENTINEL will not process or store personal data; only policy recommendations and anonymized incident reports will be extrapolated and stored (storage limitation). The project partners will not hold more information than what is properly needed to fulfil specific processing purpose. All stored data will be secured and protected by appropriate technical and organisational measures, including anonymisation or pseudonymisation of data.

5 Allocation of resources

According to Article 6 and Article 6.2.D.3 of the GA, costs related to OA to research data in H2020 are eligible for reimbursement during the duration of the project. For this project, costs for making data FAIR are mainly related to personnel costs and include preparation of data by each project partner for publishing, updating and maintaining data, data hosting and backup, data sharing, data security etc. These costs will be covered by the project funds. Project beneficiaries will be responsible for applying for reimbursement of costs related to making data accessible to others beyond the consortium.

The OA data that will be uploaded in the SENTINEL Observatory will be preserved for 1 year. These costs will be covered by the infrastructure provider (INTRA) and will be free of charge. Copyright licensing with Creative Commons is also free of charge.

Another category includes the fees for the publication of scientific articles containing project's research data in "Gold" OA journals. These costs are eligible and might be shared among multiple authors involved in the publication or will be examined case by case.

The costs related to the long-term preservation of data after the end of the project are difficult to be estimated precisely in this version of the DMP and the final decision has not been taken yet.

Dr Tatiana Trantidou (ITML) is responsible for the data management within the SENTINEL project. The responsibilities include the creation of the DMP, the necessary updates, and the control of appropriate storage, management, and sharing. Moreover, each partner is responsible for implementing and respecting the policies of the SENTINEL DMP.

6 Data security

All partners of the SENTINEL consortium will implement good practices regarding data security. This will ensure prevention of data breach including unauthorized and/or unlawful disclosure, use, alteration and modification of data. In order to appropriately protect data, all project partners must meet requirements imposed by relevant obligations (including GDPR) regarding data security:

- Article 32 (GDPR) on security of processing;
- Article 35 (GDPR) on data protection - Impact Assessment;
- Article 5(1)(c) (GDPR) on the principle of Data Minimisation;
- Article 25 (GDPR) on Data Protection by Design;
- Article 33 (GDPR) on Data Breach Notification;
- Article 5(1)(f) (GDPR) on Integrity and Confidentiality.

All data collected and generated by the various SENTINEL contexts, modules and plugins will be (if applicable) safely stored at the responsible partners' (pilot partners or technology providers) local private repositories. These repositories should be secure and non-accessible to the public. Responsible partners should follow appropriate and GDPR-compliant procedures for recovery, secure storage and transfer of data. The partners are also encouraged to use appropriate repositories and carry out regular back-ups for any work conducted for the purpose of the project realisation. Selection of appropriate repositories should include due diligence of the company/entity that offers repository service.

All security measures should be in-depth considered with caution. The responsibility of the service provider, configuration settings, functionalities of the service should be taken into account. All partners may store local copies of research data on their institutional servers and or business cloud-based servers with appropriate access controls, encryption and/or password protection. The main responsibility regarding security of data collected and processed during the project realization, as well as after its completeness lies with the owners/managers of the repositories where these data are stored.

After the end of the project, open data (e.g., policy recommendations, white papers, anonymized incident reports, etc) will be archived and preserved in the SENTINEL Observatory repository. Regarding confidential data that will be deposited to SENTINEL Observatory repository, they will be retained for 1 year, whereas data that will be kept at Zenodo will be retained for 5 years. However, the final decisions regarding long-term preservation and curation will be take/confirmed in due time.

As the minimum threshold for ensuring data security each partner should:

- ensure that data processed for the purpose of project realization are stored at institutional serves and are regularly backed up;
- ensure devices and data are safely and securely stored, and access control measures are defined at the user level (e.g., encryption is implemented, password protection is applied, and there is restriction regarding access privileges);
- support good security practices by protecting their own devices and installing and updating anti-malware software, anti-virus software and enabling firewalls;
- ensure appropriate security and confidentiality of data, including for preventing unauthorised access to or use of such data and the equipment used for the processing (this is relevant when/if personal data is processed);

- where necessary, the controller or processor of sensitive data evaluate the risks inherent regarding processing and implement measures necessary to mitigate risks;
- apply appropriate technical and organisational measures to ensure security and confidentiality will be applied; when assessing appropriateness of the measures, the state-of-the-art, the costs of implementation in relation to the risks and the nature of the personal data to be protected will be considered.

SENTINEL will implement technical measures to ensure data confidentiality, integrity and availability for information processed within the self-assessment context. The primary technical measure employed is encryption for data in transit (SSL, HTTPS and end-to-end encrypted data communications at every stage) and data at rest (encryption at the data layer using strong cryptographic primitives). The consortium is also researching methods for the pseudonymisation and anonymisation of data for potential further data reuse for research or other purposes.

The self-assessment context will be covered by SENTINEL's strong integrated IAM (Identity and Access Management) and AAA (Authentication, authorization and accounting) techniques, including robust enforced authentication, authorisation and access management for every object or entity in the web apps.

SME participants may elect to either download or permanently delete their account data, which includes self-assessment data. If they opt for permanent deletion, secure erasure will be employed to eradicate the data beyond restoring from both their primary storage and all secondary sites or backups.

It should be noted that data provided and generated in the self-assessment context are not to be externally shared. Data that will be shared within the consortium will be stored in the SENTINEL Observatory hosted by the integration leader (INTRA). This repository is intended to be a centralized password-protected repository over secure network connection. Regular backup strategy and data recovery procedures will be followed. All partners will have access to the repository. INTRA will capitalise on its CI/CD tools and environment, to be hosted in the cloud, for easy access by all other technical partners. The CI/CD tools and environment consist of:

- GitLab for source control, acting as code repository and allowing code versioning
- Jenkins for automated build and testing
- Docker for containerisation of services and components
- Eclipse IDE (as one exemplary case) for software development and debugging

All data storage and processing that is performed on INTRA services is executed within INTRA controlled premise. Access to all services is based on contracts where identity of each person has to be confirmed – specifically no anonymous usage is allowed. All INTRA personnel is trained and aware of GDPR and legal obligations regarding data privacy.

In addition, INTRA has ISO/IEC27001:2005 certification for information security management procedures, and there are procedures for keeping all infrastructure managed by INTRA safe, patched and up to date in terms of security.

Conclusion

Data processing within a research project such as SENTINEL generates various challenges. Data sharing in the open domain can be greatly beneficial to society in terms of building knowledge and facilitating reuse of data for further research. However, inappropriate disclosure of data brings many risks. Therefore, the consortium should balance openness and protection of data. As stated in the Guidelines on FAIR Data Management, data should be 'as open as possible and as close as necessary'. Therefore, the SENTINEL project must be realised in accordance with all relevant national and international sources that regulate data processing.

This deliverable brings one more source of regulation relevant to data processing. This DMP is a sort of self-regulation composed of not only general standards and principles, but also implementable measures that all project partners must adhere to. Enshrined obligations and measures are grounded in the legal and ethical deriving from not only national and international legislations, but also H2020 rules.

This DMP regulates the data management during the project lifecycle and it is relevant for all data processed and generated within the project. This DMP will build on appropriate resources that should ensure a certain level of data quality. The general contribution of the DMP might be seen as twofold. It ensures having findable, accessible, interoperable, and reusable data. Also, it is an important source for building trusted data.

References

EOSCsecreatariat.eu (2019). *PID Policies.* In Eosc Symposium 2019. [Online]. Available: *https://www.eoscsecretariat.eu/eosc-symposium2019/pid-policies* [Accessed 8.09.2021]

European Commission Directorate-General for Research & Innovation (2016). *H2020 Programme - Guidelines on FAIR Data Management in Horizon 2020*. [Online]. Available: *https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf* [Accessed 15.11.2021]

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J. (2016). *The FAIR Guiding Principles for scientific data management and stewardship*. Scientific data, Vol. 3, No. 1, pp.1-9.

Annex 1 – DMP Questionnaire

Partner Name	*Please Specify the short name of your organisation*
Corresponding person (email)	*Please Specify the corresponding person*
Case	*Please Specify the use case where these data will be used*
A – Data Summary	
What is the purpose of the data collection/generation and its relation to the objectives of the project?	*Please Specify* *[State the purpose of the data collection/generation]*
What types of data you will generate/collect?	*Please Specify* *[Define the nature of the data, e.g., quantitative, qualitive, numeric, text, audio, structured, unstructured, etc]*
What formats of data you will generate/collect?	*Please Specify* *[Define the format of the data, e.g., txt, xlsx, doc, pdf, etc]*
Will you re-use any existing data and how?	☐ No ☐ Yes (please specify): *[Specify if existing data is being re-used (if any)]*
What is the origin of the data?	*Please Specify the origin of the data*
What is the expected size of the data?	*Please Specify* State the expected size of the data (if known).
To whom might the data be useful?	*Please Specify* Outline the data utility, to whom will it be useful, e.g., research groups, private sector, citizens, etc.
Wil you apply data aggregation, minimization and anonymization methods to the data?	☐ No ☐ Yes (please specify): Define if privacy preservation techniques are required before making your data open.
B - FAIR Data	**Findable, Accessible, Interoperable, Reusable**
B1 - Making data Findable	
Are the data produced and/or used in the project discoverable with metadata?	☐ Yes ☐ No
What metadata will be created?	*Please Specify*
Do metadata standards exist in your discipline?	☐ Yes (please specify): ☐ No (please specify): *Define the metadata standards you use. If there are no standards in your discipline, do you agree to use minimum DataCite metadata standards[3]?*
Will search keywords be provided that optimize possibilities for re-use?	☐ Yes ☐ No *[Outline the approach towards search keyword (e.g., search keywords will be provided when the data will be uploaded to the repository]*
B2 - Making data openly accessible	
Accessibility	☐ Public data ☐ Confidential data *[Will data produced and/or used in the project be made openly available?]*
If certain data cannot be shared (or need to be shared under restrictions), explain why	*Please Specify* *[If some data is kept closed provide rationale for doing so, clearly separating legal and contractual reasons from voluntary restrictions]*
How will the data be made accessible	*Please Specify* *[Specify how the data will be made available, e.g., by deposition in a repository, e.g., SENTINEL Observatory, Zenodo, etc]*
Which repository will be used?	*Please Specify* *Specify which repository will you use, e.g., SENTINEL Observatory, Zenodo]*
What methods or software tools are needed to access the data?	*Please Specify*
Is documentation about the software needed to access the data included?	☐ Yes ☐ No
Is it possible to include the relevant software (e.g., in open-source code)?	☐ Yes ☐ No (please specify):

[3] https://support.datacite.org/docs/schema-40

Where will the data and associated metadata, documentation and code be deposited?	*Please Specify* *Specify which repository will you use, e.g., SENTINEL Observatory, Zenodo, GitHub, etc.]*
If there are restrictions on use, how will access be provided?	*Please Specify*
Is there a need for a data access committee?	☐ Yes (please specify why): ☐ No (please specify why):
Are there well described conditions for access (e.g., a machine-readable license, etc.)?	*Please Specify*
How will the identity of the person accessing the data be ascertained?	*Please Specify*
B3 - Making data interoperable	
File format	
Spreadsheet: ☐ ODS ☐ XLS ☐ CSV **Documentation:** ☐ DOC ☐ PDF ☐ TXT ☐ HTML	**Structured data:** ☐ XML ☐ JSON **Other (please specify):**
What data and metadata vocabularies, standards or methodologies will you follow?	*Please Specify*
B4 - Increase data re-use	
How will the data be licensed to permit the widest re-use possible?	*Please Specify* *[License conditions: Copyright, Creative Commons, Open License, etc. A list of licenses can be found here https://opendefinition.org/licenses/ . Do you agree to use Creative Commons Attribution 4.0?]*
Data owner	*Please Specify* *[List the data owner, the copyright owner, the intellectual property owner]*
When will the data be made available for re-use?	*Please Specify* *[E.g., Immediately, after the end of the project (specify the exact time), along with the publication of main results, etc. If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible]*
Are the data produced and/or used in the project useable by third parties, in particular after the end of the project?	☐ Yes ☐ No (please specify): *[If the re-use of some data is restricted, explain why]*
How long is it intended that the data remains re-usable?	*Please Specify* *[Specify the length of time for which the data will remain re-usable, e.g. 5 years after the conclusion of the project]*
C - Allocation of Resources	
What are the costs for making data FAIR in the project?	*Please Specify*
How will these be covered?	*Please Specify* *[Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions)]*
What are the resources for long term preservation?	*Please Specify* *[Resources: Costs and potential value, who decides and how, what data will be kept and for how long]*
D - Protection of Personal Data & Data Security	
How are the data you process relevant and limited to the purposes of the project? (Please show that what you collect is proportional to the purpose - data minimisation principle)	*Please Specify*
Will you process "special categories of personal data" (racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, health data, sex life or sexual orientation)? If so - why?	*Please Specify*
Will you obtain the data subject consent for processing personal data? If not, please provide a valid legal basis for the processing of personal data (e.g., legitimate interest).	*Please Specify*

Will you process previously collected personal data (secondary use)? If so - please describe the data and justify your right to use data for this project (e.g., consent).	*Please Specify*
Are you going to transfer/share your data? If so - where and how will you transfer your data?	*Please Specify*
Will you process any data that cannot be shared? If so - please explain.	*Please Specify*
Are there any ethical or legal issues that can have an impact on data sharing?	*Please Specify*
What technical and organisational measures will be implemented to safeguard the rights and freedoms of the data subjects / research participants?	*Please Specify*
Where will the data be stored or are already stored?	*Please Specify*
What provisions are in place for data security (including data recovery as well as secure storage and transfer of data including personal data)?	*Please Specify*
What security measures will be implemented to prevent unauthorised access to personal data or the equipment used for processing?	*Please Specify*
Please describe your procedures for erasure and deletion of data, if you have them.	*Please Specify*
Will you safely store data in certified repositories for long term preservation and curation?	*Please Specify*
Please reference your organisational policy and procedures on personal data management (include national/funder/sectorial/departmental procedures, and ethical standards, for processing of personal data, if any).	*Please Specify*
E - Other Issues	
Do you make use of other national/funder/sectorial/departmental procedures for data management?	*Please Specify* *[Indicate if you must adhere to other policies and procedures for data management]*