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

D2.1-The SENTINEL privacy & data protection suite for SMEs/MEs: MVP

This work is part of the SENTINEL project. SENTINEL has received funding from the European Union's Horizon 2020 Work Programme for research and innovation 2018-2020 under grant agreement n°101021659.

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	Work Package 2
Deliverable Title	D2.1 – The SENTINEL privacy & data protection suite for SMEs/MEs: MVP
Version	0.4
Date of Submission	24/05/2022
Main Author(s)/ Editor(s)	Christos Dimou (ITML)
Contributor(s)	Eleni-Maria Kalogeraki (FP), Philippe Valoggia (LIST), Giorgos Tsirantonakis (TSI)
Reviewer(s)	Manolis Falelakis (INTRA), Thomas Oudin (ACS)

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

History			
Version	**Issue Date**	**Status**	**Distribution**
0.1	15/04/2022	Draft	Confidential
0.2	11/05/2022	Draft	Confidential
0.3	23/05/2022	Draft	Confidential
0.4	24/05/2022	Final	Public

Table of Contents

Table of Contents 3
List of Figures 5
List of Tables 5
Abbreviations 6
Executive Summary 8
1 Introduction 9
1.1 Purpose of the Document 9
1.1.1 Scope 9
1.1.2 Contribution to WP2 and project objectives 9
1.1.3 Relation to other WPs and deliverables 10
1.2 Structure of the Document 11
1.3 Intended readership 11
2 The privacy and data protection compliance framework 12
2.1 Overview 12
2.2 GDPR CSA MVP version functionalities 14
2.3 Technical specifications 15
2.4 Next steps: towards GDPR CSA full-featured Version (FFV) version 16
3 The integrated Identity Management System 17
3.1 Overview 17
3.2 MyData model 17
3.3 Technical specifications 19
3.4 Future steps 19
4 Contributed cybersecurity components 21
4.1 Overview 21
4.2 Functionalities 21
4.2.1 Vendor Management 21
4.2.2 Threat Intelligence 22
4.2.3 Vulnerability Management 22
4.2.4 Simulation Environment 23
4.2.5 Technical Specifications 23
4.3 Future Steps 24

5 Continuous management and integration of open-source technology offerings and solutions 25
5.1 List of external Plugins 27
5.1.1 Wazuh 27
5.1.2 OWASP Zap 28
5.1.3 SonarQube 28
5.1.4 VirusTotal Mobile 29
5.1.5 Clonezilla 29
6 Conclusions and future steps 30
References 31

List of Figures

Figure 1. Data Protection Dashboard based on GDPR CSA Results 13
Figure 2. GDPR CSA Docker Image 15

List of Tables

Table 1. GDPR Compliance Level - Processes 12
Table 2. Technical Capabilities 25
Table 3. Organizational Capabilities 25
Table 4. List of External Plugins 26
Table 5. List of External Trainings 26

Abbreviations

Abbreviation	Explanation
API	Application Programming Interface
CAPEC	Common Attack Pattern Enumeration and Classification
CPE	Common Platform Enumeration
CPU	Central Processing Unit
CRM	Customer Relationship Management
CSA	Compliance Self-Assessment
CVE	Common Vulnerabilities and Exposures
CVSS	Common Vulnerability Scoring System
CWE	Common Weakness Enumeration
DAPROCADE	Data Protection Capability Determination
DoA	Description of Action
DPLM	Data Protection Lifecycle Management
DPMAN	Data Protection Management
ERP	Enterprise Resource Planning
EU	European Union
FFV	Full-featured Version
GDPR	General Data Protection Regulation
HR	Human Resources
IdMS	Identity Management System
IEC	International Electrotechnical Commission
ISO/IEC	International Standards Organization
IT	Information Technologies
JSON	JavaScript Object Notation
JWT	JSON Web Token
ME	Micro-Enterprise
MITRE	Massachusetts Institute of Technology Research & Engineering
MVCR	Minimum Viable Consent Receipt
MVP	Minimum Viable Product
NIST	National Institute Standards and Technology
NVD	National Vulnerability Database
OIDC	OpenID Connect
OTM	Organizational & Technical Measure
OWASP	Open Web Application Security Project
PA	Processing Activity
PAM	Process Assessment Model
PDML	Product Data Markup Language

PXE	Partial XML (Extensible Markup Language) Envelope
RAM	Random Access Memory
RDBMS	Relational Database Management System
ROPA	Register of Processing Activities
RPT	Requesting Party Token
SA	Self-Assessment
SAML	Security Assertions Markup Language
SME	Small & Medium-Sized Enterprises
SQL	Structured Query Language
SSO	Single Sign On
TIPA	Tudor IT (Information Technology) Process Assessment
UI	User Interface
UMA	User-Managed Access
USB	Universal Serial Bus
URL	Unique Resource Locator
WP	Work Package
XML	Extensible Markup Language
XSS	Cross-Site Scripting

Executive Summary

This deliverable accompanies the Minimum Viable Product (MVP) demonstrator for the SENTINEL privacy and personal data protection technologies. This deliverable has been developed within the scope of *'WP2 – The SENTINEL privacy and personal data protection technologies'*, under Grant Agreement No. 101021659.

As WP2 covers a wide range of data protection aspects technologies, this deliverable presents the work done in each of the Tasks that comprise this work package, relevant to the MVP demonstrator. These Tasks include a) the privacy and data protection compliance framework, b) the integrated Identity Management System (IdMS) (T2.2), c) contributed cybersecurity components (T2.3), and d) continuous management and integration of open-source technology offerings and solutions (T2.4).

The work presented in this document is mainly based on deliverables *'D1.1 – The SENTINEL baseline'* and *'D1.2 – The SENTINEL technical architecture'*, which define in detail the requirements and architecture of the SENTINEL framework, respectively. The specifications, technologies and implementations presented here embark from those grounds, forming a coherent set of technologies that address the overarching challenge of privacy and personal data protection within SENTINEL.

In terms of technical details, this document provides a presentation of the modules, tools and services, including the GDPR Compliance Self-Assessment (CSA) tool, the MVP version of the SENTINEL IdMS, the MITIGATE plugin that offers assessment and protection of the infrastructure of an organization, and a set of selected open-source components that will be offered to the SMEs/MEs for protection.

The presentation for each of the above services, tools and modules contains a brief description of their purpose, role in the context of the MVP and technical details regarding implementation, deployment, and testing. Their integration and function within the context of the defined SENTINEL use cases is further presented in deliverable *'D5.4 – The SENTINEL Minimum Viable Product'*.

1 Introduction

1.1 Purpose of the Document

1.1.1 Scope

The purpose of this deliverable is to accompany the delivered, functional version of SENTINEL's set of privacy and personal data protection technologies, by providing a description of the relevant services, tools, and modules for the SENTINEL's MVP release. Within the context of SENTINEL, the MVP is an early release that serves as a proof-of-concept for the project's main objectives, as it offers a functional demonstration that is minimum but complete, in terms of end-to-end integration and delivery of value to the end-user.

As MVP overall architecture, integration and use case topics are described in deliverable *'D5.4 – The SENTINEL Minimum Viable Product'*, this deliverable serves as a reference document to the services and modules that only concern this deliverable, namely: a) the GDPR CSA module, b) SENTINEL's IdMS module, c) MITIGATE, and d) a set of open-source cybersecurity plugins.

For each of the above, this document provides an overview, a description of the purpose and role within the context of the MVP, as well as technical details that are useful for the reader to understand the responsibilities, inner workings and offered services of each of the above modules.

1.1.2 Contribution to WP2 and project objectives

This deliverable has been composed within the context of *'WP2 – The SENTINEL privacy and personal data protection technologies'* and constitutes the first major output for this work package. It addresses four of the five Tasks defined in the Description of Action (DoA) that correspond to this work packages objectives. The work presented in this deliverable addresses these objectives as explained below:

Objective 1. *SENTINEL's unified privacy and personal data protection compliance self-assessment framework for GDPR compliance*

In the context of the MVP, the first version of GDPR CSA is delivered, a module that performs an analysis of Processing Activities (PAs) to determine whether personal data are handled in accordance with data protection requirements. GDPR CSA provides SMEs with: a) GDPR Compliance Level of PAs they are responsible for, and PAs they carry out on behalf of another company, and b) a list of recommendations to improve PA's GDPR Compliance Level. This work is described in Section 2 of this document.

Objective 2. *SENTINEL's integrated Identity Management System, based on the decentralised MyData model for human-centric personal data management for SMEs/MEs, enabling a unified European Personal Data Space*

A first version of the SENTINEL's IdMS is delivered. This is a first approach to IdMS ambitious goal to implement, among others, personal data portability and transparent vendor switching. To this end, the SENTINEL MVP showcases a centralized identity management system with Single Sign-On capabilities for SENTINEL end-users. This work will be expanded to apply to the end-

users of each SME/ME in future versions of SENTINEL and is described in Section 3 of this document.

Objective 3. *A curated collection of self-serving, state-of-the-art security- and privacy-enhancing modules, both open-source and contributed by consortium partners, which will be selected to meet individual specific needs of participants*

This objective is addressed in two ways in the context of the MVP. In Section 4, MITIGATE is presented as SENTINEL's a cybersecurity component, used by SMEs and MEs through a simulation environment to investigate their level of security within their organisation, identify and analyse the evolving threat landscape and thereby raise their cybersecurity awareness and strengthen their preparedness towards malware activity. MITIGATE actively contributes to the assessment phase of SENTINEL. In Section 5, a list of open-source data privacy and protection cybersecurity components is presented. Although the presented tools are not part of the delivered MVP, this section will serve as the blueprint for the incorporation of these tools in the full-featured version of the SENTINEL framework.

1.1.3 Relation to other WPs and deliverables

This deliverable expands on the foundational work conducted within *'WP1 – The SENTINEL baseline: Setting the Methodological Scene'*. More specifically, deliverables *'D1.1 – The SENTINEL baseline'* and *'D1.2 – The SENTINEL technical architecture'* define the requirements and refined architecture for the SENTINEL framework, respectively. Within the context of the same work package, deliverable *'D1.3 – The SENTINEL experimentation protocol'* specifies the pilot use cases that, although not directly related to the Core context, serve as an end-goal to the MVP implementation and inform technical decisions on Core module implementations.

This deliverable is also tightly coupled with *'WP5 - SENTINEL continuous integration and system validation'* and more specifically task *'T5.2 – Continuous integration towards the realisation of a complete system'.* Within the activities of that task, all results described in the current document have been integrated in an allocated infrastructure and operate in the context of predefined use cases to deliver the desired services to the end-user. The integration activities, interaction with other SENTINEL contexts and modules, along with the end-user benefits are detailed in deliverable *'D5.4 – The SENTINEL Minimum Viable Product'*.

There is a relationship between the work presented in this deliverable and other technical work packages. More specifically:

- For '*WP3 -The SENTINEL digital Core'*, the GDPR CSA provides the assessment as input to the Core context's Recommendation Engine for the generation of policy drafts. Additionally, the metadata and description of the cybersecurity offerings and open-source solutions are stored in the Common Repository which is an integral part of the Core context.
- For *'WP4 – The SENTINEL services'*, the CyberRange offering developed within the context of task *'T4.1: The SENTINEL Observatory'* is integrated with the SENTINEL IdMS for authentication and authorization purposes. There is also a relationship between the GDPR CSA plugin and task *'T4.2: Data protection Impact assessment and assurance'* where similar assessment tools are grouped in order to make their assessments available as services to the rest of the SENTINEL framework.

Finally, this deliverable will serve as a basis for upcoming deliverables *'D2.2 - The SENTINEL privacy & data protection suite for SMEs/MEs: Full-featured version'* (due M18) and *'D2.3 - The SENTINEL privacy & data protection suite for SMEs/MEs: Final product'* (due M30).

1.2 Structure of the Document

The structure of this document is as follows:

- *Section 2* presents the first version of the GDPR CSA module that performs an analysis of Processing Activities (PAs) to determine whether personal data are handled in accordance with data protection requirements.
- *Section 3* presents SENTINEL's IdMS first version for the MVP that provides authentication, authorization, and Single Sign-On services to SENTINEL end-users.
- *Section 4* presents MITIGATE, a SENTINEL offering for assessment and cybersecurity protection module that facilitates the investigation of level of security within an organisation, identification and analysis of threats and raise an organisation's cybersecurity awareness
- *Section 5* lists selected open-source plugins that contribute to the protection of an SME/ME's infrastructure, which will be delivered in the upcoming full-featured version of SENTINEL.
- *Section 6* summarizes this deliverable with conclusions and future steps

1.3 Intended readership

Deliverable *'D2.1 – The SENTINEL privacy and personal data protection technologies: MVP'* is a public document that accompanies the public demonstrator for the SENTINEL's MVP release. The content found in this document aims to help all stakeholders and potential users of the framework understand the purpose, role and technical details of the services and modules that are grouped under the concepts of privacy and personal data protection. Additionally, this document will serve as a guide for upcoming full-featured releases of the SENTINEL framework that will expand on the MVP in terms of use cases, SENTINEL offerings, technologies and offered services.

2 The privacy and data protection compliance framework

2.1 Overview

GDPR Compliance Self-Assessment

The GDPR Compliance Self-Assessment (CSA) module performs an analysis of Processing Activities (PAs) to determine whether personal data are handled in accordance with data protection requirements. GDPR CSA provides SMEs with:

- GDPR Compliance Level of PAs they are responsible for, and PAs they carry out on behalf of another company.
- A list of recommendations to improve PA's GDPR Compliance Level.

GDPR Compliance Level is expressed as capability level of a set of six processes. Each of these processes are related to one specific aspect of data protection requirements as illustrated in the table below:

Table 1. GDPR Compliance Level - Processes

GDPR Process	RECORD	DPLM (Data Protection Lifecycle Management)	RIGHTS	CONSENT	DPMAN (Data Protection MANagement)	BREACH
GDPR requirements addressed	It focuses on compliance with requirements related to PA description	It covers all requirements to meet when handling personal data from collection to data disposal	It describes requirements related to the effectiveness of data subject's rights	It is about requirements where data subject give consent to the collection of personal data	It refers to organizational capability to ensure compliance with GDPR over time.	It allows to verify company's ability to comply with breach notification obligations
Legal references	Art. 30	Art. 5(1) – 6 Art. 9 – 11 Art. 32	Art. 12 – 23	Art. 7 – 8	Art. 5(2), Art. 24 – 29, Art. 35 – 39	Art. 33 – 34

Assessment is performed according to the ISO/IEC 33000 Family Standard on process assessment. Then, GDPR processes are scored to determine their capability level and compare it to the expected target level. However, a simpler version of the scoring is also given to ease the understanding for users that are not familiar with process assessment approach (i.e., GDPR Compliance Level). Assessment results are illustrated by a compliance level expressed on:

- A scale [0;2], the lower the better.
- A qualitative scale [compliant; partially compliant, not compliant]
- A colour scale [green; orange, red] (see Figure 1)

GDPR CSA results can be used by SME to:

- Demonstrate accountability according to Art. 5(2) of GDPR.
- Monitor GDPR Compliance Level. As a monitoring tool, GDPR Compliance Self-Assessment is an OTMs allowing to comply with GDPR.

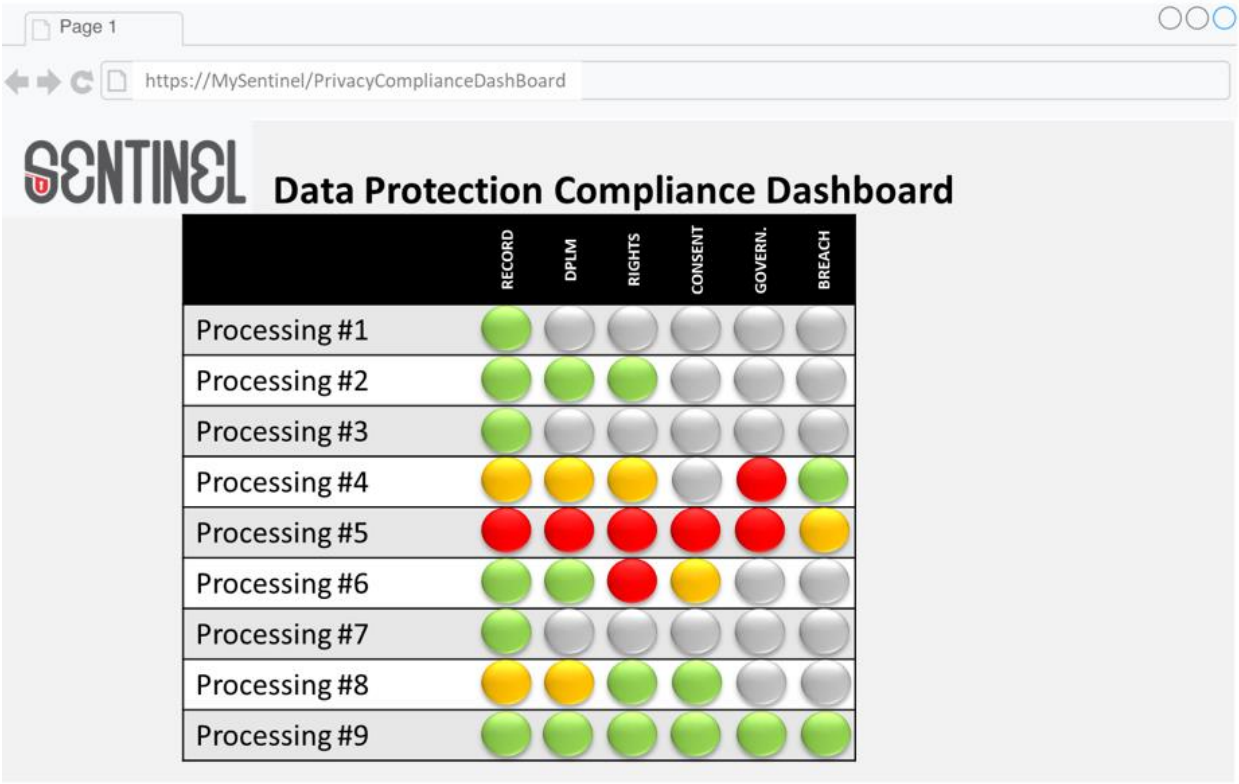

Figure 1. Data Protection Dashboard based on GDPR CSA Results

Digitalizing a process based GDPR compliance assessment

The GDPR Compliance Self-Assessment is the result of the digitalization of a process-based compliance assessment approach developed by LIST (i.e., Data Protection Capability Determination – DAPROCADE). DAPROCADE is based on the combination of an ISO/IEC 330xx compliant process assessment method (i.e., TIPA), with an ISO/IEC 33004 compliant Process Assessment Model (i.e., GDPR PAM). While TIPA specifies how to perform process assessment, GDPR PAM structure GDPR requirements in terms of process elements. Both are used by assessors to identify and to analyse the extent to which implemented Organisational and Technical Measures (OTMs) meet data protection requirements.

Digitalization of DAPROCADE is structured into three sub-tasks.

Task 2.1.1. GDPR Compliance Self-Assessment Model: It aims at making assessor knowledge explicit, especially knowledge related to:

- Information that is required to perform compliance assessment (contribution to SENTINEL's data model),
- Rules to determine data protection requirement to consider,

- Expected OTMs to meet data protection requirements,
- The extent to which implemented OTMs fulfil data protection requirements.

Task 2.1.2. GDPR CSA module development: This second task aims at coding and testing the GDPR CSA module.

Task 2.1.3. GDPR CSA Interfaces: Last task is dedicated to both integration of the GDPR CSA module to the SENTINEL platform and specification of user interface (UI) to collect additional information to perform GDPR CSA.

Adapt GDPR Compliance Self-Assessment to SMEs

DAPROCADE is not ideally suitable to SMEs, because it requires substantial resources (i.e., time spent to provide data about PAs) that SMEs do not usually possess. Then, digitalization aims at reducing the number of data that is required to perform compliance assessment without depreciating the quality of assessment results. By limiting the time spent by SMEs to feed the GDPR Compliance Self-Assessment systems, the dropout rate is expected to be reduced, and then adoption by SMEs to be increased. To this end, the assessment scope is determined by the risk that is likely to result to the rights and freedom of data subject when handling personal data. Hence, evidence that is required to demonstrate accountability depends on the risk represented by the personal data processing activity. In that sense, while the record of processing activities [Art. 30] is a sufficient accountability evidence for PAs with a low risk to privacy, additional evidence is required for PAs that are likely to give raise to higher privacy risk.

2.2 GDPR CSA MVP version functionalities

Launch GDPR CSA

User launches GDPR CSA for one PA or for all PAs recorded in Register of Processing Activities (ROPA). By doing this, the SENTINEL platform sends to the GDPR CSA module a set of data specified in the API and coming from SENTINEL's databases (i.e., SME Profile and ROPA)

Determine GDPR Compliance Level of RECORD

GDPR CSA module performs an analysis of data transmitted to determine Compliance Level of RECORD. Results provided are related to compliance with obligation to document PA [Art. 30].

Provide recommendations to improve GDPR Compliance Level

Based on the assessment results, the module establishes a list of recommendations to improve RECORD compliance level (i.e., improving comprehensiveness and update of PA description).

Determine GDPR Assessment Scope

The module also determines which processes have to be inspected with regard to Privacy Risk Level of PA. If the assessment scope is not limited to RECORD, then the module sends to the user a set of question to answer via the Questionnaire Engine. It should be noted that the latest will be implemented in Full-featured Version (FFV) of the module only.

2.3 Technical specifications

Data Consumed

The module will rely on three main data sources:

1. The organisational data (from the SME profile database),
2. The processing activities' data (from the ROPA database),
3. Additional data collected from the user through specific questions using the questionnaire engine.

The module will only need the organisations' ID from the shared data and will call the corresponding APIs to retrieve organisational data and processing activities data and to ask questions to the user.

Data generated

The module will generate a GDPR self-assessment profile for the six processes (DPMAN, BREACH, Record, PDLM, RIGHTS, CONSENT), including for each process: (i) an assessment score (integer on a [-3;3] scale, the lower the better, a negative number will mean that the SME over performs with regards to expectations); (ii) a set of recommendations to improve the score (array of text for the MVP, later on links to recommended OTM will be proposed); (iii) a process capability level (integer on a [0;3] scale to comply with ISO/IEC 33000 process assessment principles); (iv) a process target level (integer on a [0;2] scale to comply with ISO/IEC 33000 process assessment principles).

NB: the assessment score corresponds to the difference between the target level and the capability level.

API

The connection between SENTINEL's platform and GDPR CSA module is ensured via an application programming interface (API). Instead of just deploying the code, the GDPR CSA module environment is deployed as well. An docker image is then used to create, run and deploy the application in the container. As illustrated in Figure 2, Docker image contains application code ("assessment rules"), libraries and dependencies ("GDPR self-assessment"), and instructions related to data preparation ("JSON processing").

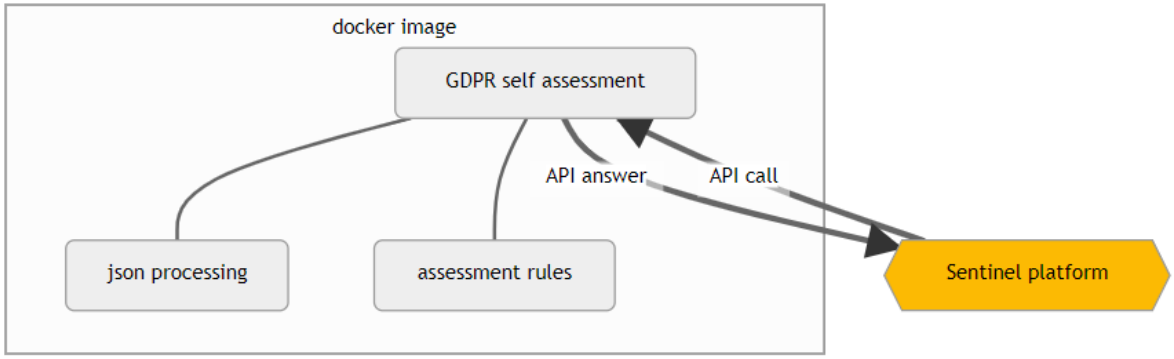

Figure 2. GDPR CSA Docker Image

2.4 Next steps: towards GDPR CSA full-featured Version (FFV) version

Implementation of GDPR CSA full functionalities

Next version of GDPR CSA module will cover all processes of GDPR Process Assessment Model. It supposes then the continuation of current three sub-tasks introduced above. Task 2.1.3 appears to be the most challenging due to the Questionnaire Engine implementation.

Linking up GDPR CSA OTMs evidence with SENTINEL's OTMs database

A first set of OTMs provided by ENISA is structured in SENTINEL's OTMs database. The GDPR CSA module aims at collecting some of them to determine whether they are appropriate regarding PA privacy risk. With GDPR CSA FFV such two databases should be better linking up to:

- Avoid double data entry,
- Associate GDPR CSA recommendations to SENTINEL's OTMs database

Definition of standard PAs

Standard PAs are typical personal processing activities an SME is likely to be responsible for. For instance, SMEs with more than 10 employees should be interested in specific Human Resources (HR) personal data processing activity. Such standard PAs will also be used for testing the GDPR CSA module.

3 The integrated Identity Management System

3.1 Overview

The overall objective for SENTINEL's Identity Management System (IdMS) is to provide key integrations in the form of plug-in modules for many commercial and open-source applications (such as Cloud ERP (Enterprise Resource Planning), eCommerce, web apps, sales automation / CRM, marketing automation etc.) used by SMEs/MEs, including the capability of revoking or fine-tuning access granularity. An additional objective is to enable a unified "Single European Data Space", facilitating standardisation and governance for data portability, as well as compatibility with the innovative "MyData" paradigm (see Section 3.2 below).

In the context of the MVP, a first approach to reaching the above objectives has been realized. First, a study of the MyData model has been conducted in order to understand in depth the relevant concepts and design the sought solution. Then, a concrete implementation of authentication and Single Sign-on (SSO) is delivered, which is based on the open-source solution Keycloak. With this approach, we explore the potentials of an IdMS solution that is applied first to the SENTINEL users themselves, before expanding and generalizing the solution to the personal data that SMEs/MEs manage for the end-users of their services.

3.2 MyData model

MyData[1] is a human centred approach in personal data management that combines industry needs to data with digital human rights. MyData is both an alternative vision and guiding technical principles for how we, as individuals, can have more control over the data trails we leave behind us in our everyday actions. The core idea is that everyone should have an easy way to see where personal data goes, specify who can use it, and alter these decisions over time.

The benefits of the MyData model and principles include the following:

- It offers a simple EU-wide self-enrolment process.
- It helps securing:
  - the Right of Access - individuals can access their data at any time
  - the Right of Rectification - individuals can alter their data at any time
  - the Right to Information - individuals can check who has accessed their data and when
  - the Right to Data Portability – individuals can provide entities with explicit consent to access their data, backed by a complete GDPR-compliant consent management system
  - the Right to Remove - revoke access to data (or part of the data)
  - the Right to be Forgotten - delete their account altogether, revoking access to all involved entities.

The main concepts of the MyData model are:

[1] https://mydata.org/

PERSON: An individual that manages the use of their own personal data, for their own purposes, and maintains relationships with other individuals, services, or organizations.

DATA SOURCE: A data source collects and processes personal data which the other roles (including Persons) may wish to access and use.

DATA USING SERVICE: A data using service can be authorized to fetch and use personal data from one or more data sources.

PERSONAL DATA OPERATOR: A Personal Data Operator enables individuals to securely access, manage and use their personal data, as well as to control the flow of personal data with, and between, data sources and data using services. Individuals can be their own operator. In other cases, operators are not using the information itself, but enabling connectivity and secure sharing of data between the other roles in the ecosystem.

In the context of SENTINEL, the offered IdMS service should adhere to and make use of these concepts to provide full compatibility with the MyData model.

Based on the MyData principles, the current MyData Operator model [1] is being built upon the **User-Managed Access (UMA)** standard and the Consent Receipt Specification (developed as the **minimum viable consent receipt - MVCR**) realizing core parts of the MyData authentication mechanism and the MyData APIs:

- **UMA** lets individuals control authorizations to share their data and to manage how their data are shared between online services. UMA is a profile of OAuth 2.0 (control access to web APIs) and it shares features with OpenID Connect (federated Single-Sign-On). It brings together two essential elements to the authorization workflow: asynchronous consent and centralized consent management. The UMA specification allows individuals to authorize access to protected resources and defines a means for a client, representing a requesting party, to use a permission ticket to request and gain access to the resource.

- **MVCR** defines a common format for provisioning consent receipts. The record of consent is human-readable and can be represented as standard JSON. This specification defines the requirements for the creation of a consent record and the provision of a receipt. It includes requirements for links to existing privacy notices & policies, as well as a description of what information has been or will be collected, the purposes for that collection, as well as relevant information about how that information will be used or disclosed. An organization can use the standards to self-assert that they are providing notice and getting implied consent in compliance with their policies and applicable regulations. An individual can save the receipt to a personal data store and self-assess if the receipt is compliant with the policies and practices of the organization.

In the MyData model, UMA provides the means for a Person to grant/revoke partial/full access to their personal data either asynchronously or on demand by requesting consent, and the MVCR provides a standard to assist in provisioning, storing, and transferring consent receipts between Persons, Data Sources and Data Using Services.

3.3 Technical specifications

SENTINEL MVP's approach to IdMS and the MyData operator described above is based on Keycloak[2], an open-source Identity and Access Management solution based on standard protocols and provides support for OpenID Connect, OAuth 2.0, and SAML 2.0. Keycloak is a UMA 2.0 compliant authorization server that provides most UMA capabilities. Users authenticate with Keycloak rather than individual applications. This means that your applications don't have to deal with login forms, authenticating users, and storing users. Once logged-in to Keycloak, users don't have to login again to access a different application. Authentication and Authorization in keycloak can be realized either through OIDC or SAML. In both cases client applications are provided with signed JWTs or XML assertions that can be used to get user information or access user resources. Authorization is further enhanced by Keycloak's Authorization Services, which are built upon OAuth 2.0 and UMA offering more control over privacy, part-to-party authorization, and resource sharing.

- **Authentication:** To authenticate a user for an application, after a successful login, the application will receive either a set of tokens (OIDC) or an XML assertion (SAML) that contains information about the user (such as username, email, and other profile information) and access information (like user role mappings) that the application can use to determine what resources the user is allowed to access on the application.

- **Authorization:** An application that wants to gain access to remote services asks Keycloak to obtain an access token (OIDC) or a SAML assertion it can use on other remote services on behalf of the user. In OIDC Keycloak authenticates the user then asks the user for consent to grant access to the application requesting it. The application then receives a signed access token. The client can make invocations on remote services using this token. The service verifies the signature, then decides based on access information within the ticket whether to process the request. In SAML the application asks Keycloak to obtain a SAML assertion it can use to invoke on other remote services on behalf of the user.

- **Authorization Services:** In **UMA**, the authorization process starts when a client tries to access a protected resource. A resource server expects a Requesting Party Token (RPT) in the request. When a client requests a resource at the resource server without a RPT, the server redirects the client to a Keycloak server with a permission ticket in order to obtain an RPT. The permission ticket represents the permissions being requested (e.g., resources and scopes) as well as any other information associated with the request. Only resource servers are allowed to create those tokens. Now that the client has a permission ticket and the location of a Keycloak server, the client can send an authorization request. If the Keycloak assessment process results in issuance of permissions, it issues the RPT, which the client can use on the resource server.

3.4 Future steps

For the full-featured version of SENTINEL (M18), IdMS should implement and provide a first version for the full range of functionalities. The main goal to be achieved is to apply the implemented services to the personal data of the actual end-users for the pilot SMEs/MEs. The principal challenges to address include full compliance with the MyData model and the MyData

[2] https://www.keycloak.org/

operator paradigm, “one-click” integrations with existing SME/ME solutions, secure storage, and communication, enforce data regulations, making sure no breaches are possible, and that the SME will not duplicate sensitive data on their premises.

4 Contributed cybersecurity components

4.1 Overview

MITIGATE ([2], [3]) is the cybersecurity component of SENTINEL provided by Focal Point (FP) which can be used by SMEs and MEs through a simulation environment to investigate their level of security within their organisation, identify and analyse the evolving threat landscape and, thereby, raise their cybersecurity awareness and strengthen their preparedness towards malware activity. It contributes to the assessment phase of SENTINEL.

MITIGATE is a standards-based risk management tool providing a collaborative, evidence-driven risk assessment approach, which delves into the technical specificities and security particularities of an organisation's infrastructure, analyses assets' interdependencies, detects all cyber threats and assets' vulnerabilities and calculates all cyber risks related to the underlined infrastructure, including potential cascading effects. It promotes collaboration among business entities in assessing and exploring their risks allowing them to manage their cybersecurity in a holistic and cost-effective manner.

The SENTINEL cybersecurity component provides a bundle of automated processes and routines, which enables SMEs/MEs to explore threat propagation and experiment on attack scenarios through this simulation environment. In this respect, it can be utilized by enterprises as a guide to undertake proper decisions to alleviate cyber threats and thus enhance their level of security and ensure data protection. Within this report, the functionalities of the cybersecurity component to identify weaknesses and threats of the organisation's IT infrastructure are described. Specifically, it supports the following main functionalities:

- the Vendor Management, which allows the Security Expert to select specific products of vendors from a dropdown list that reflects the organisation's assets.
- the Threat Intelligence, which is responsible for delivering threat and vulnerability related information identified on the selected products.
- the Simulation Environment, which provides graph analytics and reports on the security related information identified on the selected products. In particular, it allows attack scenarios development for each of the underlined products selected by the Security Expert.

4.2 Functionalities

In this section, the functionalities of the cybersecurity component are presented which may facilitate SMEs/MEs to explore security-related information concerning detected threats and vulnerabilities recognized upon selected products of vendors, which correspond to the organization's assets.

4.2.1 Vendor Management

Cyber assets which are utilised in daily operations to support the organisation's services engage vendor and product characteristics. In MITIGATE, assets vendor and product details are

synchronised with the asset's "Common Platform Enumeration" (CPE)[3] catalogue of National Institute of Standards and Technology (NIST) which is fetched in json format. The catalogue is parsed for the embedded vendor names and products along with their CPE-id, name, version, and edition, which are then extracted and assigned with a unique id. Moreover, these details are inserted automatically and afterwards they are enumerated. Dropdown lists of vendors and products are generated, which the Security Expert of the organisation can use to develop attack scenarios and review the delivered security-related information.

4.2.2 Threat Intelligence

Threat Intelligence provides all security-related information concerning threats and vulnerabilities of vendors' products. This information stems from vulnerability and threat open repositories utilizing open Intelligence and Big-Data Analytics to provide near real-time notifications on such security details. Thereby, vulnerabilities and threats are mapped with the corresponding products. In particular, threat Intelligence encompasses the following processes:

4.2.3 Vulnerability Management

The current process identifies all known vulnerabilities identified on the selected products (cf. Section 4.2.1) which reflect the assets of the organisation's IT infrastructure using the open online repository "Common Vulnerabilities and Exposures" (CVE) of MITRE[4]. The process is enabled by the CPE (see Section 4.2.1) and CVE connection already catalogued in the "National Vulnerability Database" (NVD) of NIST[5]. Moreover, a Vulnerabilities Synchronization and Management process realizes the following events:

- Initiate a call to NIST's API to download the latest version of NVD.
- Compare the fetched files with existing entries.
- Alter existing entries and add new vulnerabilities to the local database.

Based on the selected product version (see Section 4.2.1) and the vulnerability records, that are replicated in the persistency engine from the NVD open source, the exact vulnerabilities related to the declared asset is automatically inherited. Afterwards, a vulnerability analysis is provided following the "Common Vulnerability Scoring System" (CVSS) vector upon which the vulnerability attributes are determined along with the CVSS severity score illustrated per product vulnerability.

- **Threat Management**

Threat management allows the Security experts to be aware of the threat landscape the underlined organisation's IT infrastructure may be exposed to. In this vein, the current functionality provides an up-to-date catalogue of known threats, which raises the security awareness of the Security Expert of the organisation. The current functionality utilizes a combination of the NIST NVD, the "Common Weakness Enumeration" (CWE) of MITRE[6] and the "Common Attack Pattern Enumeration and Classification (CAPEC) of MITRE[7]. In particular, for each vulnerability entry a related CWE id is enumerated through NVD, whereas in CWE for each entry the related CAPEC

[3] https://nvd.nist.gov/products/cpe
[4] https://www.cve.org/
[5] https://nvd.nist.gov/
[6] https://cwe.mitre.org/
[7] https://capec.mitre.org/

ids are enumerated. According to these relationships threats are automatically extracted in the context of either CWE or CAPEC entries. A local instance of all three databases is created and threats are synchronised with the corresponding vulnerabilities identified on the selected vendors' products.

4.2.4 Simulation Environment

The simulation environment offers the user interface, where the Security Expert may review the information which is produced from the previous functionalities by setting experiments on possible attack scenarios. An attack scenario is considered a relation (triplet) of a vendor's product, vulnerability and threat. The current functionality initiates an evaluation process that builds a hybrid model of the information catalogued through the previous processes which delivers the attack scenarios. The Security Expert may develop and explore different possible attack scenarios upon selecting various vendors' products from the respective lists delivered from the Vendor Management (Section 4.2.1) that rely on the organisation's assets. The selected vendors' products are automatically linked to available vulnerabilities and threats or attack-types that are relevant, which are derived from the respective vulnerabilities list and threat catalogue of the open sources, described in "Threat Intelligence" (Section 4.2.2). The vendors' products along with the linked information (i.e., vulnerabilities, threats/attack types) are intuitively visualised using a graph visualization modality. The specific component offers this visualization functionality, where graphical analytics facilitates the Security Expert to better comprehend the generated results and alleviate some of the analysis burden from the assessor's point of view.

To this aim, the Security Expert may review information concerning:

- the affected vendor's product
- the corresponding identified vulnerability
- the threat that can impact the respective vulnerability

Reports containing lists of products associated with the corresponding vulnerabilities and threats are generated.

4.2.5 Technical Specifications

The cybersecurity component of SENTINEL is an intelligent mechanism of an optimized solution which communicates with the SENTINEL digital core. It is built on the Java Spring Framework, which is an open source, enterprise-level framework for creating standalone, production-grade applications. It offers dependency injection features providing a list of objects defining their own dependencies which the Spring container later injects into them. This process enables the creation of modular applications consisting of loosely coupled components that are ideal for microservices and distributed network applications as in the SENTINEL case. The PostgreSQL is used for the data store of vendors, products, threat and vulnerabilities catalogues, which is a free and open-source Relational Database Management System (RDBMS) emphasizing extensibility and SQL compliance.

In addition, the MITIGATE cybersecurity component provides near real-time notifications of security related information (vulnerabilities and threats) on products extracted from open online repositories (CPE of NIST; CVE of MITRE; CAPEC of MITRE) through text processing, utilising

open intelligence techniques and Big Data analytics. Such Mining techniques are extremely computationally intensive; thus, the component relies on Apache SPARK big-data framework[8] to achieve linear scalability. Server push notifications are provided to the Security Expert concerning any type of messages that are published in the pub/sub queue to ingest and distribute the relevant data. The consistency of the various 'enumerations' concerning vendor/products, threats and vulnerabilities is managed through semi-automated updates from the open sources. The Graph Visualiser is responsible to render all types of the supported graphs in the simulation environment.

4.3 Future Steps

Future actions regarding the cybersecurity component of SENTINEL may be the deployment of the Asset Inventory functionality. This functionality may allow SMEs/MEs to register their assets, including all IT and networking related devices of the company. The aim of representing the company's IT Infrastructure may be to create a valid representation of the various types of assets and their dependencies with other elements of the company. As a result, the organisation's asset cartography may be developed illustrating information on assets' vendors and products' characteristics, assets interrelations and asset graph visualisations. On this account, assets may be connected with corresponding threats and vulnerabilities depending on their version of the respective vendor's product.

In addition, the Control Management functionality may be deployed to support the security controls implemented on the organisation's assets which may be registered by the Security Expert. Thereafter, the security controls may be mapped to the corresponding vulnerabilities and threats that mitigate.

Furthermore, the "Threat Intelligence" capabilities could be enhanced by giving the opportunity to the Security Expert to define risk appetites.

Over and above these proposed features, calculation of the vulnerabilities exploitability on the identified assets may be produced. Eventually, estimation on threat/vulnerabilities/risk levels may be provided by initiating a risk assessment process which the Security Expert could have the opportunity to conduct to review the recorded results and further expand the cybersecurity awareness. Specifically, the latter could guide the Security Expert and decision makers how to undertake optimal mitigation strategies and thus maintain organisation's security and data protection.

[8] http://spark.apache.org/

5 Continuous management and integration of open-source technology offerings and solutions

In this section we will present the external plugins and trainings which, along with SENTINEL's technology offerings, will help complete the SMEs security capabilities. Each external plugin will correspond to a mature open-source tool which will be able to cover one or more security capabilities. The open-source solutions are selected based on their maturity and community support to ensure the long-term maintainability of these solutions, their ease of integration and the number of capabilities they can cover. The trainings will be a variety of articles and online presentations and will be selected by their ability to cover capabilities and explain a concept to audiences of different levels of technical background. The recommendation engine will parse this information and make their recommendation to SMEs depending on their requirement analysis.

The Tables below list organizational and technical capabilities, which have been derived from requirements laid out originally in D1.1 and carried forward to D1.2 (Appendix A). A more detailed version of categories / OTMs / Generic Policy description) can be found in D3.1 and specifically at the Policy Drafting module.

Table 2. Technical Capabilities

T1	tec_auth_acl	Authentication and Access control
T2	tec_logging_monitoring	Logging, monitoring and alerting
T3	tec_server_database	Server and database security
T4	tec_endpoint_workstations	Endpoint security (workstations)
T5	tec_endpoint_mobile	Endpoint security (mobile devices)
T6	tec_network	Network security
T7	tec_backup	Backup policy
T8	tec_app	Application lifecycle security
T9	tec_disposal	Data disposal
T10	tec_physical	Physical security

Table 3. Organizational Capabilities

O1	org_policy_drafting_enforcing	Defining and enforcing a policy
O2	org_assigning_roles	Assigning roles and responsibilities
O3	org_access_policy	Enforcing an access control policy
O4	org_asset_management	Securely managing assets
O5	org_change_management	Managing change

O6	org_gdpr_management	Managing data processors for the GDPR
O7	org_incident_handling	Handling incidents
O8	org_business_continuity	Managing business continuity
O9	org_hr	Managing human resources
O10	org_awareness	Cybersecurity awareness, education and training

The external plugins and trainings will aim to focus on the derived capabilities not covered by the SENTINEL project offerings. A repository will be maintained which will contain a description for each plugin along with its basic functionalities, the capabilities it can meet and the technical specifications. All this information will be stored in a format readable by the Recommendation Engine in order to parse and make decisions based on the information provided. In the table below we can see a list of open-source plugins and the capabilities they match, currently 6 out of 10 technical capabilities are being covered by external plugins.

Table 4. List of External Plugins

List of External Plugins	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
Wazuh		■	■	■						
OWASP Zap			■					■		
SonarQube			■					■		
VirusTotal Mobile			■		■					
Clonezilla			■				■			

Table 5. List of External Trainings

List of External Trainings	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	Link
Guidelines 07/2020 on the concepts of controller and processor in the GDPR						■					https://edpb.europa.eu/system/files/2021-07/eppb_guidelines_202007_controllerprocessor_final_en.pdf
GDPR data controllers and data processors						■					https://www.gdpreu.org/the-regulation/key-concepts/data-controllers-and-processors/
GDPR Compliance: "Explain Like I'm Five" with Data Privacy Expert						■				■	https://www.youtube.com/watch?v=nG9RJLhDTXc

An introduction to GDPR											https://www.virtual-college.co.uk/courses/compliance/introduction-to-gdpr
Regulatory Spotlight: GDPR and Incident Response (Incident Response Forum Europe 2020)											https://www.youtube.com/watch?v=wY7mL_QMPok
Webinar: Responding to a Data Breach | What you should know!											https://www.youtube.com/watch?v=iryvaQicnKU

A second repository will be created which will contain the information of the external trainings of the SENTINEL platform. These trainings will be a collection of articles, webinars and videos which will help SMEs meet their organization capabilities. Currently 3 out of 10 organization capabilities are being covered by external trainings.

5.1 List of external Plugins

Here we will present a more detailed analysis for each open-source component selected. For each external plugin we provide a link to the official distributor, link it to capabilities SMEs will need to cover, a description of the tool and its main advantages and finally technical specifications which can help with the installation and the usage of the tool.

5.1.1 Wazuh

Link: https://wazuh.com/

Capabilities: T4:tec_endpoint_workstations, T3:tec_server_database, T2: Logging - monitoring and alerting

Overview:

Wazuh is an established open-source solution for host-based intrusion detection which can be deployed as an endpoint protection mechanism in a company. It can perform log analysis, file integrity checking, policy monitoring and vulnerability detection among others. Wazuh is ideal for small enterprises, since, once you deploy the server, it is easy to add additional agents for monitoring in a centralized location. In addition, it has multiplatform support and can be installed in Linux, OpenBSD, FreeBSD, MacOSX, Solaris and Windows. Finally, it is easy to integrate with known network intrusion detection systems such as Suricata[9] and Bro[10] and log aggregators such as ELK[11].

Technical Specifications:

[9] https://suricata.io/
[10] https://zeek.org/
[11] https://www.elastic.co/

For the Server installation a machine with a 64-bit Linux operating system is needed such as Ubuntu and Debian. The all-in-one installation is recommended which includes the ELK stack and is easier to setup and can support up to 100 agents. The minimum requirements for this type of deployment are 4 GB of RAM and 2 CPU cores, and the recommended are 16 GB of RAM and 8 CPU cores. For the server installation an official guide can be found here[12]. After setting up the server, the Wazuh agent must be installed in each device to be monitored. Details depending on the operating system of each device can be found here[13].

5.1.2 OWASP Zap

Link: https://www.zaproxy.org/

Capabilities: T8:tec_app, T3:tec_server_database

Overview: Zap is an open-source web application pentesting (penetration testing) tool maintained by the OWASP organization. It acts as a proxy between the browser and the targeted web application and it can be used do conduct vulnerability assessments. It has a variety of automated web application attacks and can detect SQL injections and attacks. For any vulnerabilities it reports, it will suggest mitigations and best practices which can be very valuable for a SME which aims to cover the application lifecycle security capability.

Technical Specifications:

ZAP can be installed in Windows, Linux, and Mac OS/X and requires java8+, alternatively there is a docker release for easier integration. More details can be found here[14].

5.1.3 SonarQube

Link: https://www.sonarqube.org/

Capabilities: T8:tec_app, T3:tec_server_database

Overview: SonarQube in an open-source platform which can perform static code analysis and can help SMEs to improve their code quality by reporting on bad coding habits, bugs and vulnerabilities. It supports more than 25 programming languages including C, C++, Java & Python and is a good way to ensure best practices are used during the development phase of an application lifecycle.

Technical Specifications:

Java (Oracle JRE or OpenJDK) is needed on the machine where SonarQube will run. A small-scale instance requires at least 2GB of ram, the amount of disk space depends on the size of the code you want to scan. The server can only be installed in 64-bit systems. More information on how to set up SonarQube can be found here[15].

[12] https://documentation.wazuh.com/current/installation-guide/open-distro/index.html
[13] https://documentation.wazuh.com/current/installation-guide/wazuh-agent/index.html
[14] https://www.zaproxy.org/getting-started/
[15] https://docs.sonarqube.org/latest/setup/get-started-2-minutes/

5.1.4 VirusTotal Mobile

Link: https://support.virustotal.com/hc/en-us/articles/115002146549-Mobile-Apps

Capabilities: T5:tec_endpoint_mobile, T3:tec_server_database

Overview: VirusTotal is an application which can be used for scanning of files and URLs to detect malicious activity. For each database it uses over 70 antivirus scanners and URL/Domain blacklisting services. It can provide protection to the end user from phishing attacks by preventing the user from visiting blacklisted URLs and the download and use of files which have been flagged as malicious.

Technical Specifications:

For Android devices 4.4+ it can be found and installed from the PlayStore.

5.1.5 Clonezilla

Link: https://clonezilla.org/

Capabilities: T7:tec_backup, T3:tec_server_database

Overview: Clonezilla is an open-source disk cloning application, it can copy the contents of a hard drive to another storage device. Having a backup and recovery solution can be vital for a SME to deal with ransomware and virus attacks which can have devastating consequences to the liveability of a company. Two types of Clonezilla are available, Clonezilla live and Clonezilla SE (Server Edition). Clonezilla Live is suitable for single machine backup and restore and can be a great solution for SMEs with limited infrastructure.

Technical Specifications:

Clonezilla live needs an X86 or x86-64 processor, 196 MB of system memory (RAM) and a Boot device, e.g., CD/DVD Drive, USB port, PXE, or hard drive. First the end user must create a bootable media following this guide[16]. Then he can follow this guide[17] with pictures on how to save a restore a disk image.

[16] https://clonezilla.org/clonezilla-live.php#make
[17] https://clonezilla.org/clonezilla-live-doc.php

6 Conclusions and future steps

In this document, we presented the technical description of the MVP version for the services and modules relevant to privacy and personal data protection. We embarked from the previous work on requirements and framework architecture, and proceeded with the specification of the scope, role and technologies for each of the pieces that offer services for data protection. All information presented here serves as complementary documentation to the functional prototype delivered as part of the SENTINEL MVP.

Especially for the full-featured version of the SENTINEL integrated framework (M18), this document constitutes a natural continuation of the integration efforts that have started with the MVP. The work presented here serves as a proof-of-concept for the potentials of SENTINEL. For the upcoming versions, this work will be expanded to offer the full range of the envisioned services of SENTINEL.

References

[1] Langform J., *et al.* Understanding MyData Operators (white paper). [Online]. Available: https://mydata.org/wp-content/uploads/sites/5/2020/04/Understanding-Mydata-Operators-pages.pdf [Accessed 13 May 2022]

[2] Kalogeraki, E.-M., Papastergiou, S., Mouratidis, H., Polemi N., (2018) "A novel risk assessment methodology for SCADA maritime logistics environments", Applied Sciences, MDPI AG, Switzerland, 8(9): 1477, ISSN: 2076-3417, https://doi.org/10.3390/app8091477

[3] Schauer, S., Polemi, N. & Mouratidis, H. (2019). MITIGATE: A dynamic supply chain cyber risk assessment methodology. Journal of Transportation Security, Vol. 12, pp. 1-35, https://doi.org/10.1007/s12198-018-0197-x