D2.2

Design and Architecture for the ISP Pilot

WP2 – ISP Pilot

C3ISP

Collaborative and Confidential Information Sharing and Analysis for Cyber Protection

Due date of deliverable: 30/09/2017
Actual submission date: 30/09/2017

30/09/2017

Version 13.0

Responsible partner: CNR
Editor: Gianpiero Costantino
E-mail address: gianpiero.costantino@iit.cnr.it

Project co-funded by the European Commission within the Horizon 2020 Framework Programme		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

The C3ISP Project is supported by funding under the Horizon 2020 Framework Program of the European Commission DS 2015-1, GA #700294

Authors: Gianpiero Costantino (CNR), Luca Deri (CNR), Fabio Martinelli (CNR), Maurizio Martinelli (CNR)

Approved by: Charence Wong (3D Repo), Thanh Hai Nguyen and Vincent Herbert (CEA)

Revision History

Version	Date	Name	Partner	Sections Affected / Comments
1.0	06/07/2017	Gianpiero Costantino	CNR	Added ToC
2.0	11/09/2017	Gianpiero Costantino	CNR	Initial writing in System and Pilot architecture sections
3.0	13/09/2017	Gianpiero Costantino	CNR	Added more details regarding the blocks of the ISP Pilot architecture
4.0	14/09/2017	Gianpiero Costantino	CNR	Added security and policy model
5.0	15/09/2017	Gianpiero Costantino, Luca Deri *and* Maurizio Martinelli	CNR	Added Security Scan Software text and detailed analytics for this pilot
6.0	19/09/2017	Gianpiero Costantino	CNR	Added Introduction, Hardware and Software requirements and requirements matrix
7.0	20/09/2017	Gianpiero Costantino	CNR	Polished Section 2,3 and 4
8.0	21/09/2017	Gianpiero Costantino	CNR	Extended Data Model section
9.0	22/09/2017	Gianpiero Costantino	CNR	Extended Data Model section and added Section 4.4
10.0	23/09/2017	Gianpiero Costantino	CNR	Deliverable Ready for internal review
11.0	25/09/2017	Charence Wong	3D Repo	Review and editing of whole document
12.0	26/09/2017	Thanh Hai Nguyen and Vincent Herbert	CEA	Review and editing of whole document
13.0	28/09/2017	Gianpiero Costantino	CNR	Final version

Executive Summary

This deliverable presents the first consolidated architecture of the ISP Pilot with respect to the common architecture presented in D7.2. The ISP Pilot aims at designing and developing a hybrid distributed architecture in which each ISP locally includes a ISI block and a centralised installation of ISI and IAI designed to provide data storage, and usage control of the data and analytics.

In D2.2 are described, in a top-level manner, the blocks related to the ISP Pilot architecture and, in particular, for each block the internal components are described as well as the flow of the requests. The analytics are introduced and are detailed to describe their importance with the ISP Pilot. In addition, the data model of the analytics is presented showing a first version of the data format.

Afterwards, D2.2 introduces the security and DSA model to illustrate the security requirements and policies that Internet Services Providers would like to express with this pilot to protect the privacy of their data.

Finally, this document presents the hardware and software deployment model and recalls the requirements declared in D2.1. They define the minimum requirements to run the software for this pilot and establish which component, designed in the ISP pilot architecture, will cover the use cases.

To conclude, this deliverable can be considered as the starting point for D2.3 that will be delivered at month 26. In particular, D2.2 has defined the main blocks and components of the ISP pilot that in D2.3 will reach a higher maturity level and present the first implementation, test and validation results.

Table of contents

1. Introduction

1.1. Purpose

The purpose of the ISP pilot is to provide analysis on data reports that are given by Internet Service Providers (ISPs) to detect or anticipate cyber-security threats. To support ISPs in discovering security issues, Registro.it offers ISPs a tool to check their status, regarding security vulnerabilities, by providing them security reports. These can be pre-processed using specific operations, called Data Manipulation Operation (DMO), to preserve the privacy during the analysis process. DMO and the analytics can be expressed and protected through policies defined in Data Sharing Agreements (DSAs).

1.2. Scope

The ISP Pilot aims at showing the benefits of this project to the ISPs that, through the Registro.it and C3ISP, can exploit security tools and analytics to mitigate and/or prevent cyber-security threats. This pilot is one of the four pilots available in this project and, in particular, will be designed and developed to follow the common architecture, the analytics, the usage control and the privacy-preserving techniques developed within C3ISP.

The first implementation, test and validations of the ISP Pilot will be released at month 26 and it will show the working integration of the blocks and components designed for this pilot. Conclusive implementation, test and validations are scheduled for month 34, and it will present the complete maturation level of each component as well as the benefits that ISPs will achieve by using the C3ISP Framework.

1.3. Overview

Section 2 provides an overview of this pilot and lists the benefits to the ISPs in using C3ISP functionalities. Section 3 describes the architecture of the ISP pilot merged with the hybrid distributed architecture described in D7.2 [2]. Section 4 takes the architecture blocks presented in the previous section and provide a detailed description of them. Moreover, in Section 4 the analytics for the ISP pilot are introduced. In Section 5, the security model is presented focusing on the more relevant security properties. A first definition of policies through DSA is given Section 6. In Section 7, the hardware and software requirements are presented. Section 8 introduces the matrix that links the requirements presented in D2.1 [1] with the architecture components. Finally, Section 9 concludes this document.

1.4. Definitions and Abbreviations

Term	Meaning
API	Application Program Interface
CLI	Command Line Interface
C3ISP	Collaborative and Confidential Information Sharing and Analysis for Cyber Protection
CVE	Common Vulnerabilities and Exposures
DMO	Data Manipulation Operations
DSA	Data Sharing Agreement
IAI	Information Analytics Infrastructure
IDS	Intrusion Detection System

ISI	Information Sharing Infrastructure
OTP	OpenVAS Transport Protocol
OTP	One-Time Password
LDAP	Lightweight Directory Access Protocol
TLS	Transport Security Layer

2. System Overview

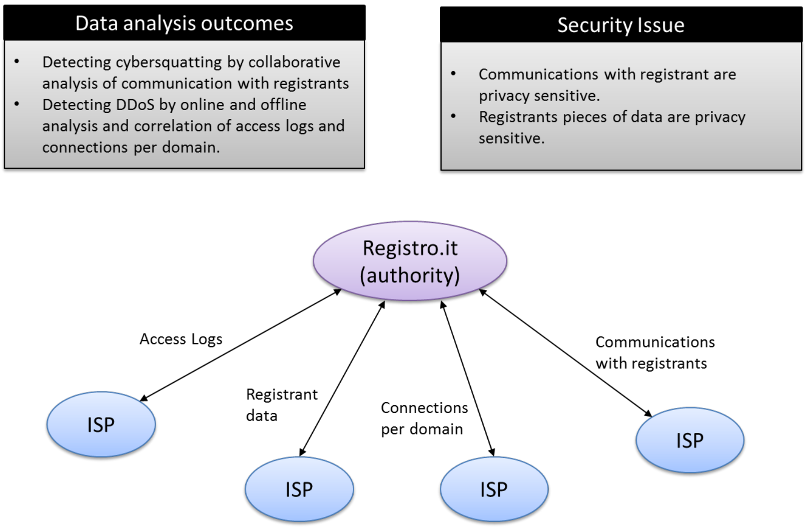
The pilot aims at offering cyber-security solutions to the ISPs by analysing security reports, which can be collected in a collaborative manner. Security reports are produced by Security Scan Software, which Registro.it provides to the ISPs as a remote tool, and the Registrar Local Platform. C3ISP Framework then formats and aggregates the security reports using the CTI format for further analysis. Finally, results of the analytics are provided to the ISPs that can react depending of the kind of results evaluated.

The benefits that the ISPs obtain from this pilot are:

- The ISPs will be able to run security tools to test the soundness of their systems and networks.
- The ISPs will be able to execute the Data Manipulation Operations (DMO) on reports to preserve the privacy by applying techniques like Homomorphic Encryption and anonymization techniques;
- The ISPs will be able to protect the confidentiality of their report using the Data Sharing Agreement (DSA).
- The ISPs will be able to use several analytics functions on the data submitted to C3ISP Framework to discover cyber-security threats.
- The ISPs will be able to share reports highlighting cyber-security issues that can be taken by other ISPs to mitigate or stop cyber-attacks.

3. System Architecture

In Figure 1 the ISP Pilot is illustrated as in its original scenario. The main goal of this pilot is to perform collaborative analysis of data coming from different Internet Service Providers (ISPs), a.k.a. Registrars, to detect cybercrimes and security attacks in real-time. ISPs are public entities and services provided to them, can resolve or prevent cyber-security attacks on incorrect network configurations, suspicious connections and other security issues.



Data analysis outcomes

- Detecting cybersquatting by collaborative analysis of communication with registrants
- Detecting DDoS by online and offline analysis and correlation of access logs and connections per domain.

Security Issue

- Communications with registrant are privacy sensitive.
- Registrants pieces of data are privacy sensitive.

Figure 1: The original schema of the ISP Pilot

The actors involved in the Pilot are the Internet Service Providers (ISPs) and Registro.it (R). They interact with C3ISP Framework to invoke analytics or to share data that can be useful for other ISPs.

Briefly, we recall from D2.1 the description of the ISPs, Registro.it and C3ISP Framework within the ISP Pilot:

- The **Internet Services Provider (ISP)** is in communication both with Registro.it and C3ISP. An ISP interacts with Registro.it to execute security tools to discover vulnerabilities in selected devices, which belong to the ISP-requester. Instead, the ISP interacts with C3ISP to provide security reports, which they include management and distribution policies, and to receive analysis results to prevent cyber-security attacks.
- The **Registro.it** is a relevant actor of the ISP Pilot and its role is to provide security services to the ISPs to find vulnerabilities. An important component of Registro.it is done by the *Security Scan Software*. It provides to the ISPs a set of tools developed for a variety of security purposes.
- The **C3ISP** Framework is the elaboration core of this pilot. Within this pilot, it takes security reports from one or more ISPs, to elaborate them through analytic operations and to send back to ISPs a report of the evaluation. Relevant functionalities of C3ISP Framework regards to the possibility to manage data in different fashion, for instance through data-anonymization or cryptographic techniques, and also to distribute the reports to only ISPs specified through Data Sharing Agreements (DSA).

In Figure 2, the updated scenario for the ISP Pilot is illustrated. In particular, in this top-level view the main components of the architecture are shown. On the left side of the figure, we represent the ISP environment, which is composed by the Internet Service Providers in two different roles: *Producer* and *Consumer*. The ISP interacts with the: *DSA Manager*, the *Registro.it* and also, they can perform local tasks using the *Registrar Local Platform*.

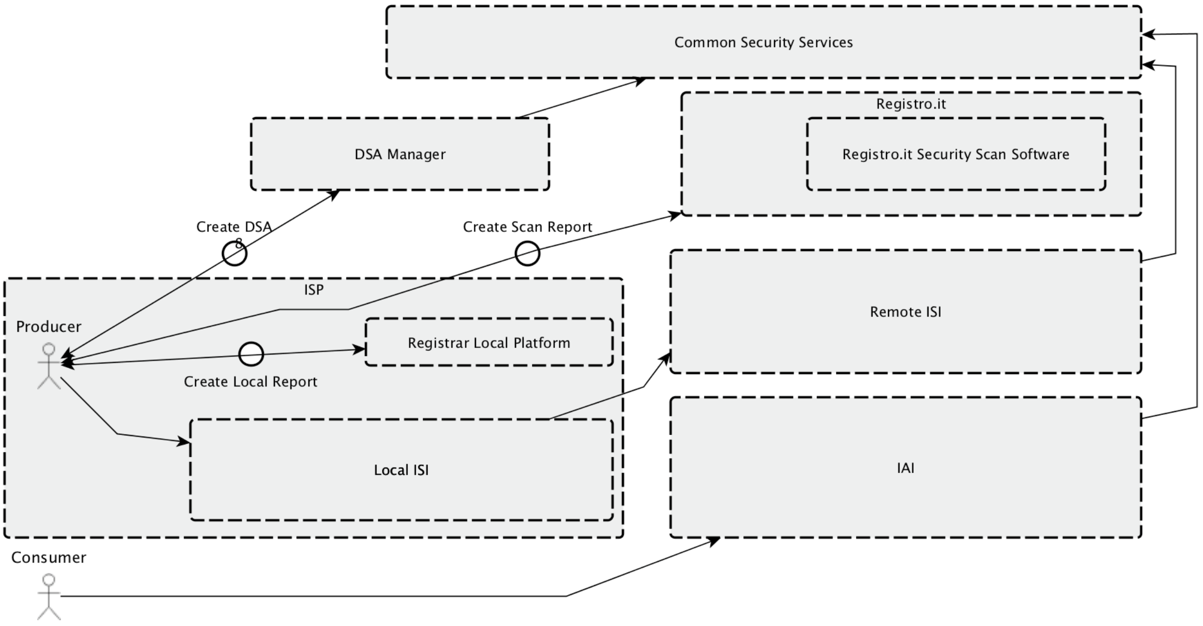


Figure 2: The ISP Pilot – updated scenario

The ISPs interacts with C3ISP, which is composed of the *Information Sharing Infrastructure* (**ISI**) and the *Information Analytic Infrastructure* (**IAI**). In particular, the ISP Pilot scenario is designed to follow the Hybrid Distributed Architecture introduced in deliverable 7.2, see also Figure 4. The hybrid architecture consists of a *Local ISI*, distributed in each ISP plus a *Remote ISI*, centralised and located in the same place of the IAI. The Local ISI prepares data, i.e. report, that are produced by the ISP, which exploits the *Registro.i*t services and the *Registrar Local Platform* security operations. Once the data has been collected, the Local ISI may also perform some pre-processing operations, like data anonymization. When, the Local ISI concludes this pre-processing phase, the ISP may decide to offload the data into the Remote ISI for further analytics or for sharing useful information with other ISPs.

When an ISP invokes analytics (depicted as Consumer in Figure 2), it directly contacts the IAI, which is only available as a remote entity. The IAI is designed to execute analytics and interact with the ISI to retrieve and store the data for the analytics.

In Figure 4, the high-level C3ISP architecture is provided. This is deeply detailed in D7.2, and here we show its integration with the ISP Pilot. The main blocks of the C3ISP architecture are kept and upon these, the ISP Pilot architecture is deployed, see Figure 5. The architecture blocks of C3ISP are described in greater depth in the next section.

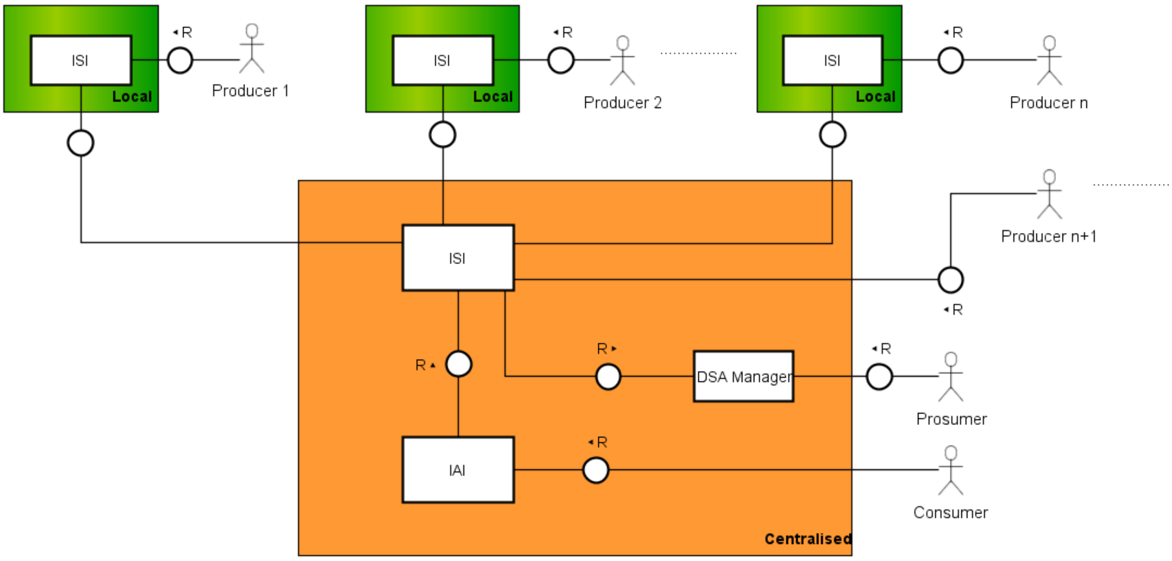


Figure 3: Hybrid deployment model (On-Premises ISI with Centralised ISI and IAI)

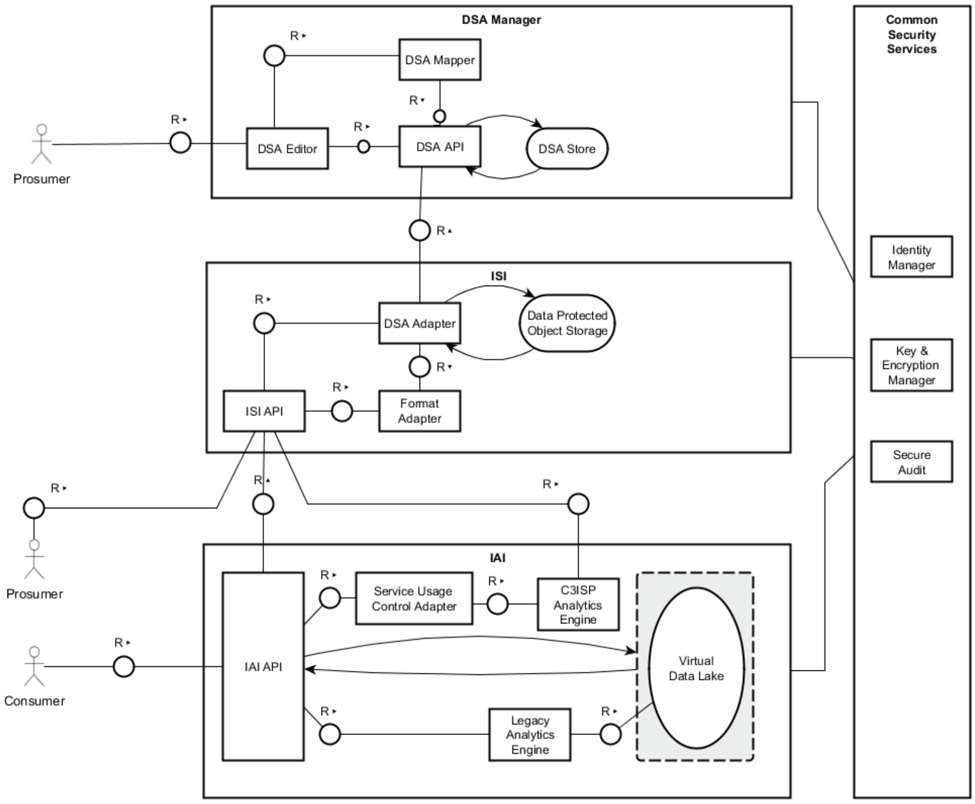


Figure 4: High-level C3ISP architecture

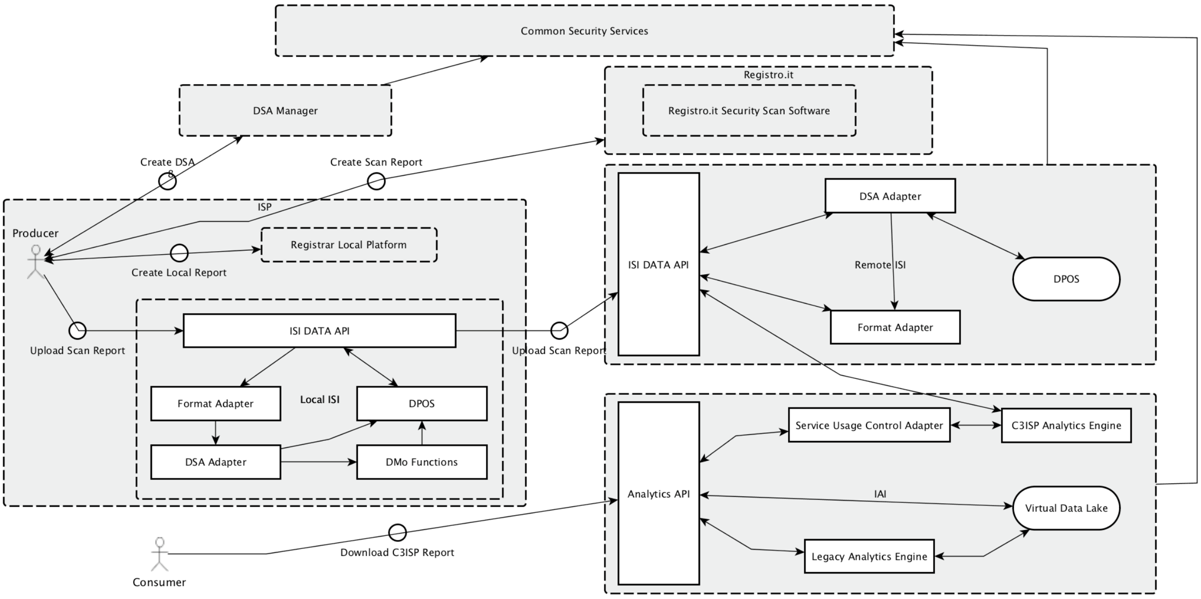


Figure 5: ISP Pilot architecture

4. ISP Pilot Architecture

In this section, we provide an integration of the ISP Pilot components, which involves the Internet Service Providers and Registro.it, with the C3ISP architecture defined in D7.2.

In Figure 5, we illustrate the architecture of the ISP Pilot that uses the Hybrid deployment model. As explained in D7.2, the Hybrid model requires the presence of an ISI block into each ISP, while a remote ISI and IAI are centrally deployed.

Then in this section, we dive down into a greater level of detail to describe better each component involved in the architecture. Moreover, we illustrate the flow that the producer and consumer will perform within this pilot. We introduce the Data Model aiming at showing how the major data and system entities are stored, processed and organised. Finally, we list the analytics that the ISPs will call to exploit the benefits of C3ISP and the sharing of information.

4.1. Block design

4.1.1. ISP report, DSA Manager and Registro.it

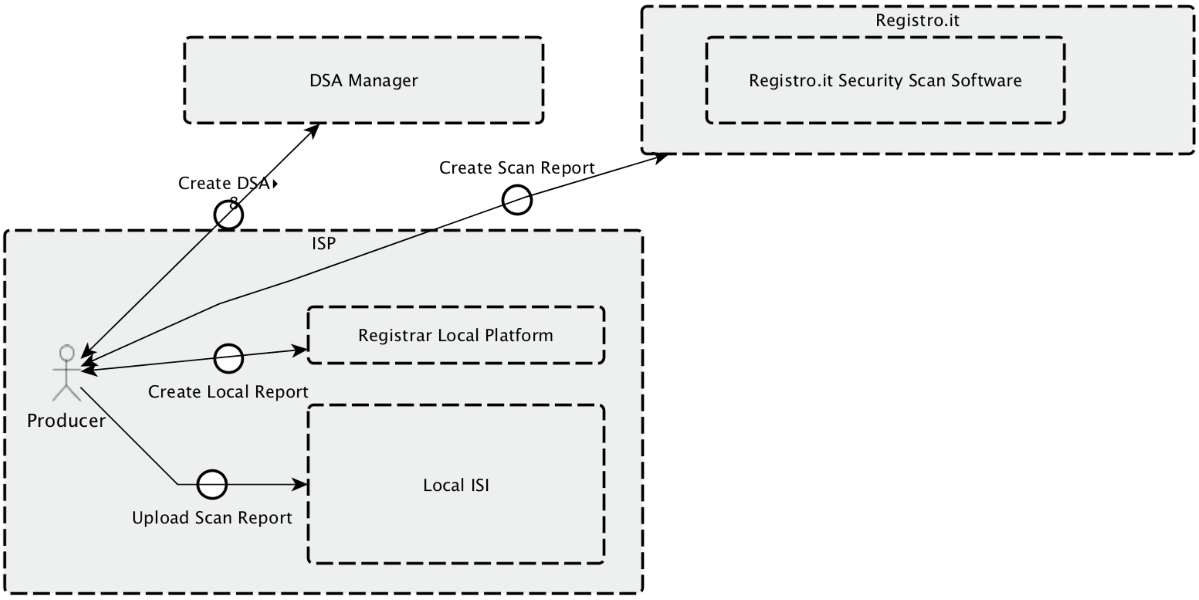


Figure 6: ISP side

Figure 6 focuses on the ISP side and the operations that ISPs can perform before sending the data to the Local ISI. For example, an ISP, acting as *Producer*, may first test the security configuration of its servers by invoking functionalities from Security Scan Software hosted by Registro.it or by the Registrar Local Platform. These operations are both executed following the use cases Create Local Report **(ISP-UC01)** and Create Scan Service **(ISP-UC02)** defined in D2.1.

Once the Producer has obtained the report, it attaches a DSA on it depending on the kind of policies it decides to apply. Policies refer to the data manipulation that are done through the Data Manipulation Operations (DMO), the analytics and the distribution strategy of the results. Policies are expressed through *Authorisation*, *Obligation* and *Prohibitions* and these are explained in D7.2 and D8.1 [3] together with the Data Sharing Agreement (DSA).

So far, the Producer has generated a data-report, which contains the result of a security check, and a DSA file in which the policies expressed. In the next step, the Producer is ready to move the files generated to the remote ISI. Since, this pilot follows a hybrid deployment model, the

Upload Scan Report (**ISP-UC-07**) is a local operation referred to the Local ISI. The Producer sends the report and the DSA files to the Local ISI through the API designed to receive information from the Producer, for instance the *Create* API.

4.1.1.1. Registro.it

The Registro.it interacts with ISPs to provide them security functionalities through the Security Scan Software. In particular, this component exposes to the ISPs functionalities that can be used to discover security vulnerabilities.

In the following, the main functionalities of the Security Scan Software are presented:

4.1.1.2. Functionalities of the Security Scan Software

The core component of the architecture is OpenVAS, an open source tool available at http://www.openvas.org whose architecture is depicted below.

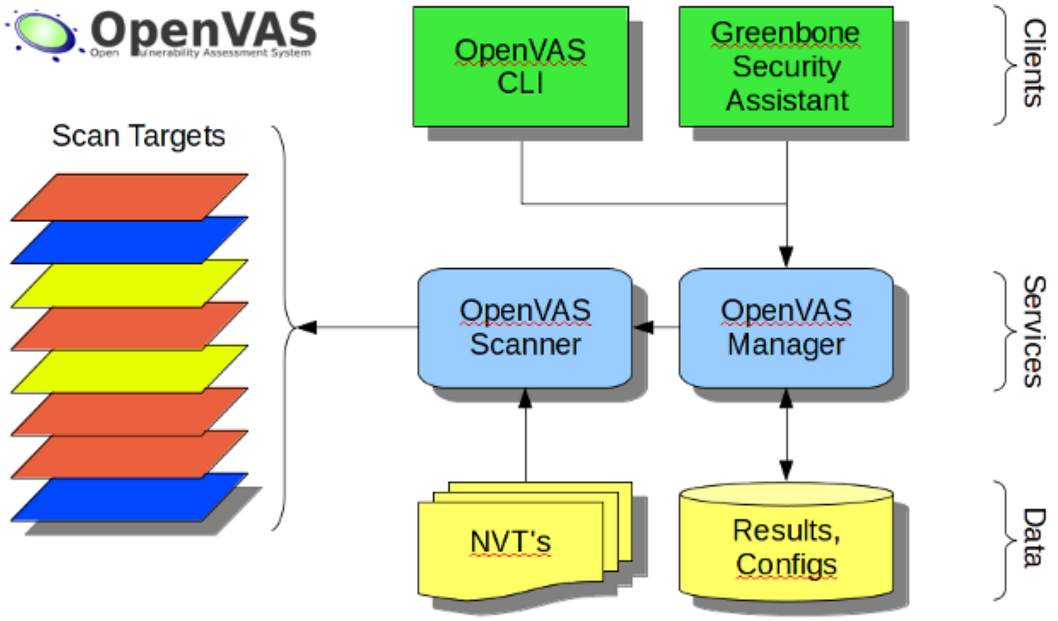


Figure 7: OpenVAS architecture

OpenVAS is a flexible software framework that implements vulnerability scans and management. Licensed under the GNU GPL, it is a fork of a popular tool named Nessus and thus it is used daily by many individuals. Thanks to this, it benefits from a daily feed of new Network Vulnerability Tests that guarantee it to always be an up-to-date tool able to support the most recent CVE (Common Vulnerability Exposures).

The OpenVAS engine can be scripted and configured both using a CLI (Command Line Interface) and a HTTP-based protocol named OTP (OpenVAS Transport Protocol). ISPs scans can thus be scheduled and configured, as well scan results can be retrieved using the same interface. In addition to CLI and OTP, it is possible to access the system using the web interface depicted below that allows to configure scans, and visualise results.

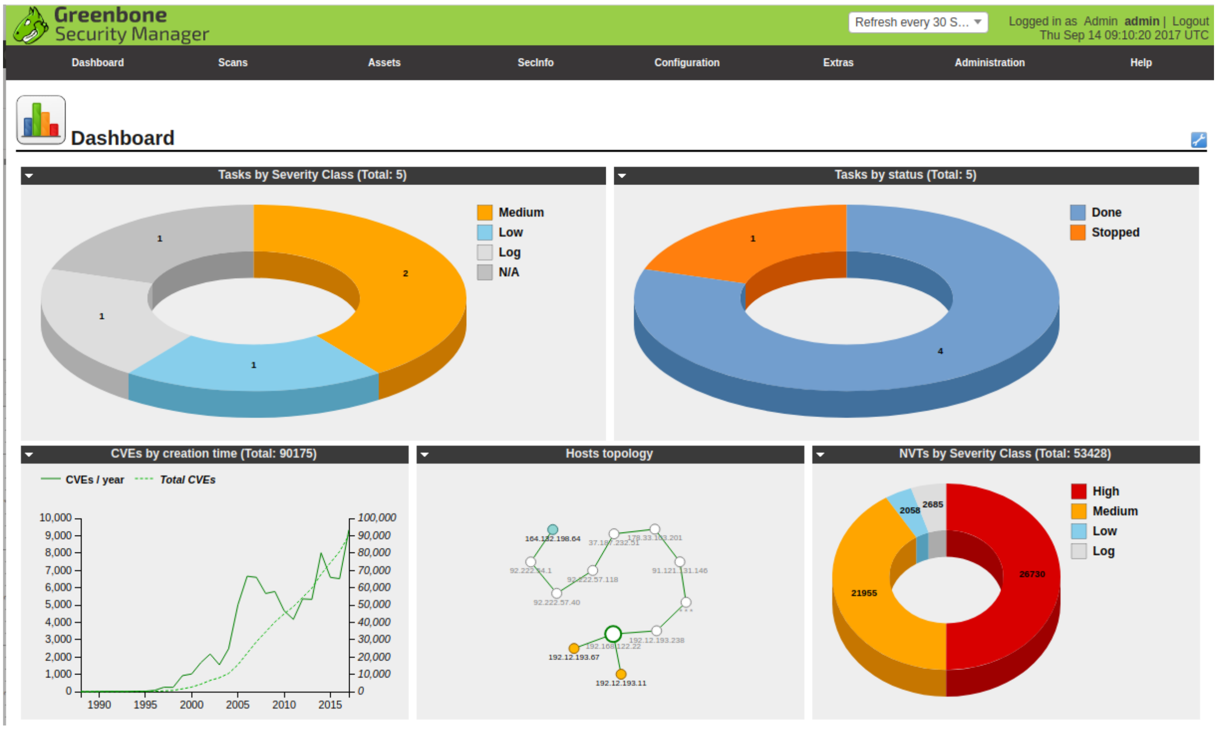
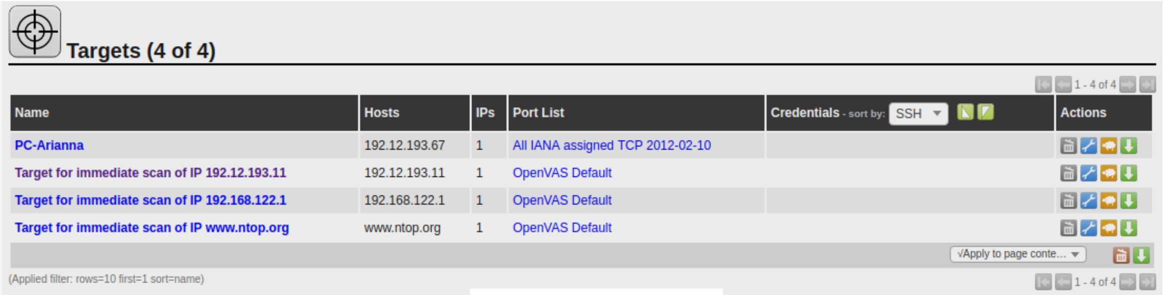


Figure 8: OpenVAS interface

Run a new scan

When a new scan is configured, it is possible to specify:

- The scan targets, i.e. the IP addresses of the hosts that will be scanned.



Targets (4 of 4)

1 - 4 of 4

Name	Hosts	IPs	Port List	Credentials - sort by: SSH	Actions
PC-Arianna	192.12.193.67	1	All IANA assigned TCP 2012-02-10		
Target for immediate scan of IP 192.12.193.11	192.12.193.11	1	OpenVAS Default		
Target for immediate scan of IP 192.168.122.1	192.168.122.1	1	OpenVAS Default		
Target for immediate scan of IP www.ntop.org	www.ntop.org	1	OpenVAS Default		

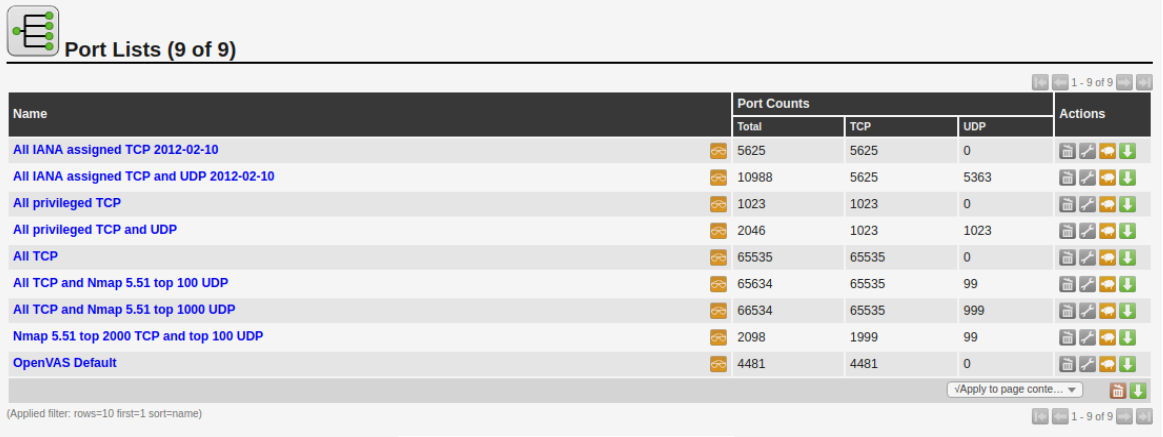
√Apply to page conte...

(Applied filter: rows=10 first=1 sort=name)

1 - 4 of 4

Figure 9: Scan targets

- The list of TCP/UDP ports to scan:



Port Lists (9 of 9)

1 - 9 of 9

Name	Port Counts			Actions
	Total	TCP	UDP	
All IANA assigned TCP 2012-02-10	5625	5625	0	
All IANA assigned TCP and UDP 2012-02-10	10988	5625	5363	
All privileged TCP	1023	1023	0	
All privileged TCP and UDP	2046	1023	1023	
All TCP	65535	65535	0	
All TCP and Nmap 5.51 top 100 UDP	65634	65535	99	
All TCP and Nmap 5.51 top 1000 UDP	66534	65535	999	
Nmap 5.51 top 2000 TCP and top 100 UDP	2098	1999	99	
OpenVAS Default	4481	4481	0	

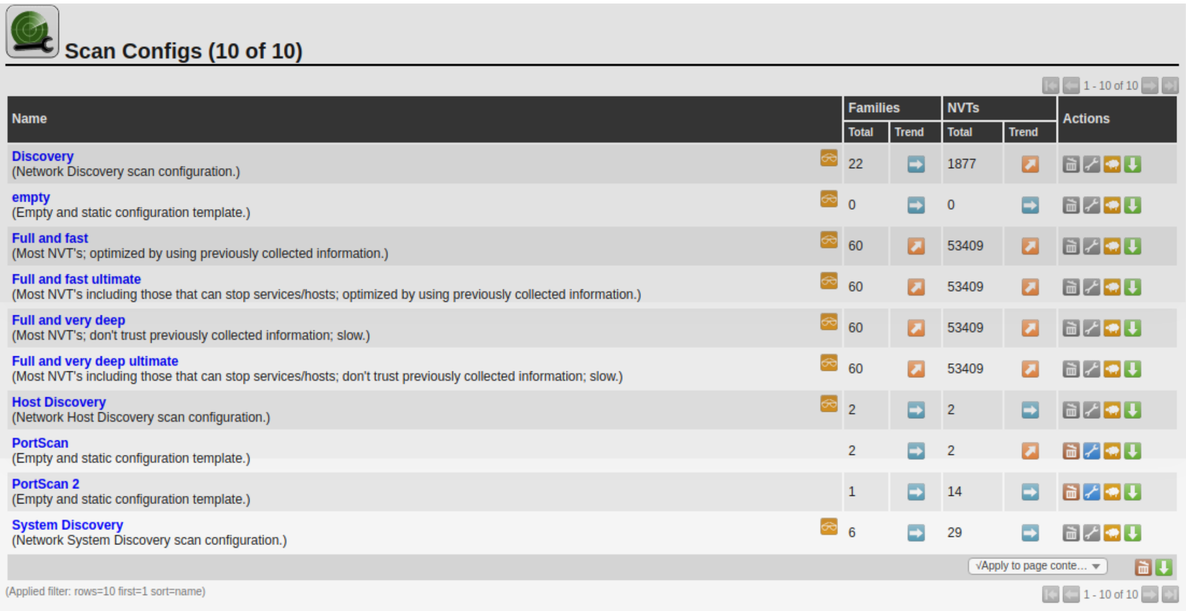
√Apply to page conte...

(Applied filter: rows=10 first=1 sort=name)

1 - 9 of 9

Figure 11: TCP/UDP port list

- The scans types and list of vulnerabilities that will be tested.



Scan Configs (10 of 10)

1 - 10 of 10

Name	Families		NVTs		Actions
	Total	Trend	Total	Trend	
Discovery (Network Discovery scan configuration.)	22	→	1877	↗	
empty (Empty and static configuration template.)	0	→	0	→	
Full and fast (Most NVT's; optimized by using previously collected information.)	60	↗	53409	↗	
Full and fast ultimate (Most NVT's including those that can stop services/hosts; optimized by using previously collected information.)	60	↗	53409	↗	
Full and very deep (Most NVT's; don't trust previously collected information; slow.)	60	↗	53409	↗	
Full and very deep ultimate (Most NVT's including those that can stop services/hosts; don't trust previously collected information; slow.)	60	↗	53409	↗	
Host Discovery (Network Host Discovery scan configuration.)	2	→	2	→	
PortScan (Empty and static configuration template.)	2	→	2	↗	
PortScan 2 (Empty and static configuration template.)	1	→	14	→	
System Discovery (Network System Discovery scan configuration.)	6	→	29	→	

√Apply to page conte...

(Applied filter: rows=10 first=1 sort=name)

1 - 10 of 10

Figure 10:List of Scan Configurations

When invoked, OpenVAS will be instrumented to perform periodic scans of the selected ISP. Based on the discovered target operating system, scans will be adapted automatically and relevant CVEs will be enabled.

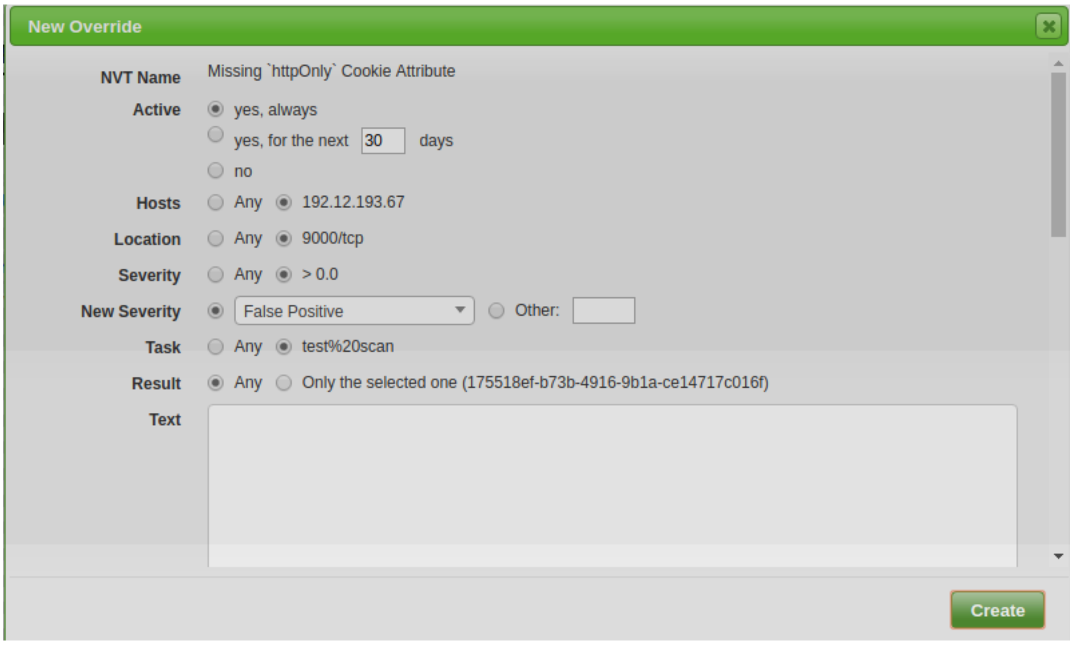
Reports can be exported in various formats including

- Anonymous XML
- ARF (Abuse Reporting Format)
- CPE (Common Platform Enumeration)
- CSV
- HTML
- ITG (Intergraph)
- LaTeX
- NBE (Nessus Report File)

- Topology SVG
- Text
- Verinice ISM and ITG (vernice.com formats)
- XML

Change state of a Report

OpenVAS supports the change state report as shown in Figure 12



NVT Name: Missing `httpOnly` Cookie Attribute

Active: ● yes, always ○ yes, for the next 30 days ○ no

Hosts: ○ Any ● 192.12.193.67

Location: ○ Any ● 9000/tcp

Severity: ○ Any ● > 0.0

New Severity: ● False Positive ○ Other:

Task: ○ Any ● test%20scan

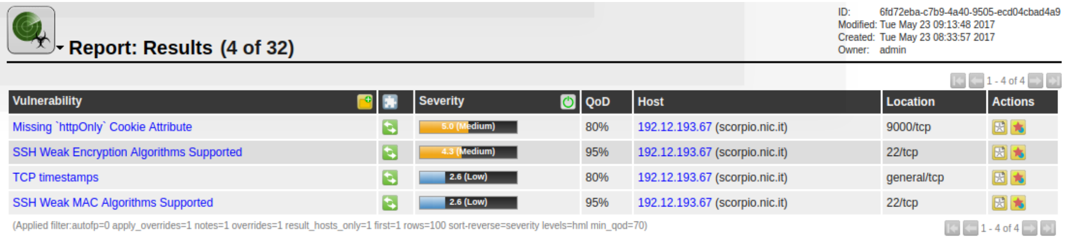
Result: ● Any ○ Only the selected one (175518ef-b73b-4916-9b1a-ce14717c016f)

Text:

Figure 12: OpenVAS change state report

Download State Report

Reports can be accessed and downloaded through the web GUI or through OTP. Based on the results some scans can be disabled (e.g. false positives) to tune future periodic scan. The C3ISP Framework can fetch reports as they are available after a scan is completed.



ID: 6fd72eba-c7b9-4a40-9505-ecd04cbad4a9
Modified: Tue May 23 09:13:48 2017
Created: Tue May 23 08:33:57 2017
Owner: admin

Vulnerability		Severity	QoD	Host	Location	Actions
Missing `httpOnly` Cookie Attribute		5.0 (Medium)	80%	192.12.193.67 (scorpio.nic.it)	9000/tcp	
SSH Weak Encryption Algorithms Supported		4.3 (Medium)	95%	192.12.193.67 (scorpio.nic.it)	22/tcp	
TCP timestamps		2.6 (Low)	80%	192.12.193.67 (scorpio.nic.it)	general/tcp	
SSH Weak MAC Algorithms Supported		2.6 (Low)	95%	192.12.193.67 (scorpio.nic.it)	22/tcp	

(Applied filter:autofp=0 apply_overrides=1 notes=1 overrides=1 result_hosts_only=1 first=1 rows=100 sort-reverse=severity levels=hml min_qod=70)

Figure 13: OpenVAS report download

4.1.2. Registrar Local Platform

Each ISP will have a separate login/password to the OpenVAS platform to guarantee that scan information as well results will be kept private. ISPs credentials will be managed by the Identity Manager, and ISPs will be able to configure the list of targets to scan as well the type of vulnerabilities to test. ISPs can schedule periodic scans (e.g. once a day on target X) as well fetch results and push them to C3ISP Framework.

4.1.1. The Local ISI

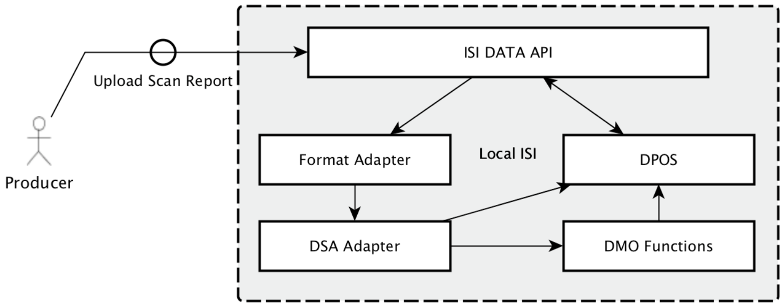


Figure 14: The Local ISI blocks

When the Producer has generated the report and the DSA, it uploads the data to the Local ISI using a particular API. Figure 14 shows the main blocks of the Local ISI. Once, the data reaches the Local ISI, these are formatted accordingly to the common formatting method established within C3ISP Framework. In fact, the report and the DSA, which at this step can be seen as a *raw* format, will be modified by the *Format Adapter* to get a structure that follows the Cyber Threat Information (CTI) format (see D7.2 for more details).

The report and the DSA structured with the CTI format are taken as input by the *DSA Adapter* that evaluates the policies written in the DSA and checks whether the report must be pre-processed through, for instance the Homomorphic Encryption or data anonymization. In case the policies express the need to manipulate the data, the DMO Functions block is invoked, otherwise the CTI bundle is stored in the *Data Protected Object Storage* (DPOS).

4.1.2. The centralised ISI and IAI

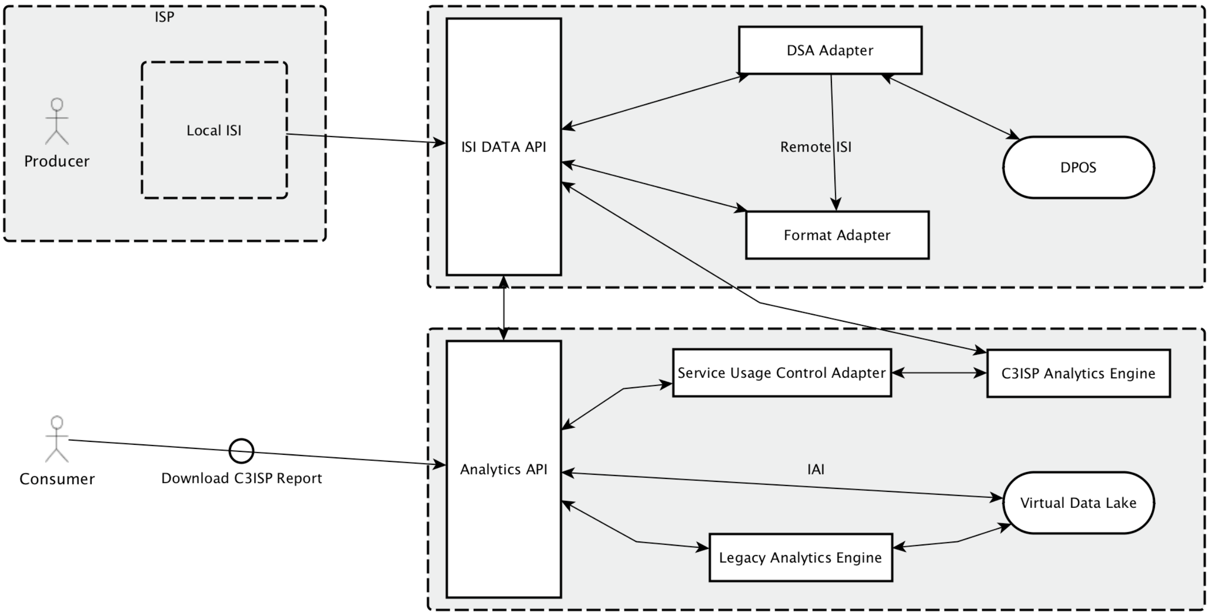


Figure 15: The remote ISI and IAI as centralised blocks

The Local ISI contains all reports that the ISP has produced through the Security Scan Software and the Registrar Local Platform. When an ISP wants its report to be analysed, it moves the CTI bundle to the *Remote ISI.* The move operation is performed through the ISI Data API localised in the Remote ISI. When the bundle reaches the ISI block, it is first evaluated by the *DSA Adapter* and then stored in the *DPOS.*

The IAI sub-system is invoked by the Consumer, which can be either the same ISP that stored the data for further analysis or another ISP that wants to download the result of a previous analysis. In both cases, the consumer first interacts with the Analytics API to perform the desired operations, e.g., running a data analytics or collecting the result of one or more analysis.

Once the desired API is invoked the control goes to the *Usage Adapter* that verifies if the consumer is entitled to execute the desired operation. In case of positive outcome, the C3ISP Analytics Engine is invoked to run the analytic, otherwise the action is denied.

A similar approach is run when a consumer wants to retrieve results of a previous analytic. In this case, the usage adapter always checks the correctness of the request by evaluating the DSA contained in the bundle with the result of the analytic.

Another relevant task of the IAI is managed by the *Legacy Analytics Engine*. It is in charge of providing those legacy engines, such as the Visual Analytics tool, as well as provisioning of its result. The legacy analytics engine uses the *Virtual Data Lake* (VDL) to allow the legacy engine accessing the data that has been shared through the ISI and processed by other C3ISP components such as C3ISP analytics engine, or data manipulation operation modules according to the DSA rules for a particular consumer.

4.2. Analytic and collaborative functions

In the following sections, we provide a list of analytics and other collaborative functions that will be part of the ISPs pilot to validate the benefits of using C3ISP.

4.2.1. [CF-01] Sharing a set of Malicious IPs

On the Internet there are several companies that collect daily lists of malicious or compromised IPs. Such lists are available from security companies as well as non-profit organisations.

In addition to these lists, ISPs are subjected to daily attacks, scans or malicious access attempts. Internet services such as DNS, e-mail or HTTP produce security logs that can be used to enrich the malicious IP set.

Using Unix text-manipulation tools it is possible for ISPs to extract such IPs, aggregate them, and transmit this list to C3ISP through HTTP, in order to create a list of malicious IPs that could be made available to all ISPs in order to enhance their security.

4.2.1. [CF-02] Sharing of Vulnerabilities

OpenVAS is part of the Security Scan Software and it allows ISPs vulnerabilities to be discovered. So, ISPs could periodically deliver scan reports to C3ISP to periodically compile a list of detected vulnerabilities found on ISPs. In order to preserve each ISP's privacy, the list should not contain any ISP information or should be pre-processed using the data manipulation operations.

4.2.2. [CF-03] DGA attack

Domain Generation Algorithms (DGA) are used by malware to generate a list of meeting points where malware apps can meet and implement command and conquer, to decide attacks and spread threats. As names are generated using time-based algorithms and the host name does not matter for humans, such names do not make any sense for a person. Example:

- www.fgd2iwya7vinfutj5wq5we.com
- www.qbtxzhetq4s2f.com
- www.fgd2iwya7vinfutj5wq5we.net
- axwscwsslmiagfah.domain.com.

DGA attempts can be found using Intrusion Detection Systems (IDS), such as Bro.org, or developing simple tools that try to see if the requested host name looks like a DGA. This can be achieved by keeping a list of impossible bigrams and thus discovering these DGA. More information can be found at:

http://www3.nd.edu/~busiforc/handouts/cryptography/Letter%20Frequencies.html

In this way, analysing DNS server logs and applying the above algorithm it is possible for ISPs to spot potential DGAs and report it in real-time to C3ISP that can re-distribute this information to other ISPs.

4.2.3. [AF-01] Analysing authentication Log

Most Internet services such as email, HTTP and SSH produce log files containing access attempts. For instance, SSH feeds auth.log and HTTP error.log files.

In addition to what described in the previous section, ISPs could deliver (in real-time) to C3ISP IPs the log structure in the CTI format to be then analysed to discover malicious attempts to the services.

4.2.4. [AF-02] Connections to malicious hosts

In addition to malicious attempts, it would be interesting also to monitor in real-time accesses from malicious IPs to ISP servers. This can be implemented by enabling on ISPs routers and switches protocols, such as NetFlow[1] and sFlow[2], that are designed to provide monitoring data in standard format.

ISPs could collect the generated flows and report them to C3ISP Framework to find patterns of malicious IPs. Even in this case, a producer may require that its personal information will be treated anonymously by applying homomorphic encryption or data anonymization techniques.

4.3. Data Model

Section 4.2 has introduced the analytics and the collaborative functions that the ISPs will be able to execute exploiting the Registrar Local Platform, the Security Scan Software and the IAI sub-system of C3ISP Framework. This section introduces the data model that defines a first structure of data that ISPs will prepare before submitting them to the ISI sub-system. In particular, this section will provide a first overview of the data model considering each analytic and functionality discussed in Section 4.2.

In Section 4.1.1, we wrote that reports are prepared using the Security Scan Software and the Registrar Local Platform. Both entities generate a report, a.k.a. log file, in which the result of the security analysis is provided. However, reports are generated in specific formats that depend on the operation itself. For instance, in the case of an authentication log, this will be structured as a list of IPs that attempted to log in to the system but for some reason have failed. So, the report is provided in a *raw* format that is not compatible with the C3ISP analytics, i.e., the CTI format.

4.3.1. [CF-01] Data Schema

List of malicious or compromised IPs are available from security companies as well as non-profit organisations. ISPs can use public lists of malicious IPs or produce their own to share

[1] https://www.cisco.com/c/en/us/products/ios-nx-os-software/ios-netflow/index.html

[2] http://www.sflow.org

this information with other ISPs. In addition, since attackers often change IPs to perform their attacks, the lists must be constantly updated.

The list in its raw state should have the following parameters:

Table 1: CF-01 Data Schema

Element Name	Element Type	Required	Description
IP	String	Yes	The IP of the device exploited to perform the attack
Status	String	Yes	The status of the IP. It can assume two values: *active* or *inactive*
Data Added	Timestamp	Yes	When the IP has been added in the list
Updated	Timestamp	Yes	Timestamp of when the IP has been checked again

The following table shows an example of the list of malicious IPs:

Table 2: List of malicious IPs - example

IP	Status	Data Added	Updated
X.X.X.X	Active	2017-09-21 07:56:23	2017-09-21 07:56:23
X.X.X.Y	Active	2017-09-20 08:56:23	2017-09-20 10:06:13
X.X.Y.Y	Active	2017-09-20 11:56:23	2017-09-20 11:56:23
…	…	…	…
Z.Z.Z.Z	Active	2017-09-20 19:56:23	2017-09-20 23:00:09

4.3.2. [CF-02] Data Schema

ISPs exploit the Security Scan Software to check the status of their system and verify if they have specific vulnerabilities that can threaten their system. The vulnerabilities found by the Security Scan Software may be shared through C3ISP to other ISPs that may react depending on their current system configurations.

The format of the vulnerability found should be the following:

Table 3: [CF-02] Data Schema

Element Name	Element Type	Required	Description
ISP	String	Yes	The ISP in which the vulnerability was found
Device-IP	String	Yes	The IP of the device that has the vulnerability
Scan Date	Timestamp	Yes	When the scan was done
CVE-ID	String	Yes	The vulnerability found expressed through the Common Vulnerabilities and Exposures (CVE) identification

Description	String	No	The description of the vulnerability

An example of the [CF-02] Data Schema is the following:

Table 4: Vulnerability found - Example

ISP	Device-IP	Scan Date	CVE-ID	Description
Helloworld-ISP	X.X.X.X	2017-09-15 09:56:00	CVE-2017-XXXX	A divide-by-zero error in the library ABC that may lead to a DoS attack when opening a crafted text file

Since Table 4 contains sensitive parameters, such as the *ISP-name* and the *Device-IP* of ISPs that have vulnerabilities, the ISP may require to mask those parameters by using anonymization techniques available through the DMO. This operation is in charge of the Local ISI but the need to mask policies must be expressed through policies within the DSA.

4.3.1. [CF-03] Data Schema

DGA found by ISPs can be shared with other ISPs to make them unreachable in case of malware or other security threats that make use of DGA. The data schema for this function can follow a similar structure of the [CF-01] function. So, the parameters are illustrated in Table 5:

Table 5: [CF-03] Data Schema

Element Name	Element Type	Required	Description
URL	String	Yes	The DGA URL, for instance www.qbtxzhetq4s2f.com
Status	String	Yes	The status of the URL. It can assume two values: *active* or *inactive*
Data Added	Timestamp	Yes	When the URL has been added in the list
Updated	Timestamp	Yes	Timestamp of when the URL has been checked again

The following table shows an example of the list of DGA:

Table 6: List of DGA - example

URL	Status	Data Added	Updated
www.qbtxzhetq4s2f.com	Active	2017-09-21 07:56:23	2017-09-21 07:56:23
www.qbtxz23q4s2f.com	Active	2017-09-20 08:56:23	2017-09-20 10:06:13
www.absdasdahh32.org	Active	2017-09-20 11:56:23	2017-09-20 11:56:23
…	…	…	…
www.fgfasgf7rwh19.net	Active	2017-09-20 19:56:23	2017-09-20 23:00:09

4.3.2. [AF-01] Data Schema

ISPs may publicly expose services that can be used by malicious entities to access sensitive contents. Services like SSH, HTTP, or database management system like MySQL. Logs on failed connections to the service are stored by the service itself. So, ISPs may use these logs for analysis by the IAI analytics.

Logs sent to the Local ISI and in their raw format should contain the following information:

Table 7: [AF-01] Data Schema

Element Name	Element Type	Required	Description
Service-Name	String	Yes	The kind of service, for instance SSH, HTTP
IP	String	Yes	The IP of the device that attempted the connection and it failed
Port	String	No	Port of the connection attempt
Account	String	No	Username used to attempt the connection
Date	Timestamp	Yes	When the connection was done

In the following, an example of a service log:

Table 8: Log structure - example

Service-Name	IP	Port	Account	Date
SSH	X.X.X.X	42570	root	2017-09-15 09:56:00
SSH	X.X.X.X	42570	root	2017-09-15 09:56:01
SSH	X.X.X.X	42570	root	2017-09-15 09:56:02
SSH	X.X.X.X	42570	root	2017-09-15 09:56:03
SSH	Y.Y.Y.Y	42571	root	2017-09-15 09:57:00

4.3.1. [AF-02] Data Schema

ISPs can use monitoring tools to sniff the traffic of their network, collect these pieces of information and send them to C3ISP for analysis of patterns or devices connected to malicious entities. Data collected from the network can be formatted using the following parameters:

Table 9: [AF-02] Data Schema

Element Name	Element Type	Required	Description
IP-src	String	Yes	The IP of the device that started the connection, e.g., a server of a ISP
IP-dest	String	Yes	The IP of the remote device, for instance it may be malicious entity
Port	String	Yes	Port of the connection attempt
Service-name	String	No	The kind of service (*if known*)
Date	Timestamp	Yes	When the connection was detected

The above parameters can be expressed in the format:

Table 10: Traffic connection - Example

IP-src	IP-dest	Port	Service-name	Date
X.X.X.Y	X.X.X.X	42570	SFTP	2017-09-15 09:56:00
X.X.X.Y	X.X.X.Z	17892		2017-09-15 09:56:01
X.X.Y.Y	X.X.Z.Z	12631	SSH	2017-09-15 09:56:02
X.Y.Y.Y	X.Z.Z.Z	22222		2017-09-15 09:56:03

4.4. Analytics and collaborative functions vs Data Manipulation Operation

Reports generated by the Security Scan Software or the Registrar Local Platform may contain sensible information that should not be made public. To this purpose, the DSAs help the ISPs to obligate C3ISP Framework to treat some data in a privacy-preserving fashion. In particular, the Data Manipulation Operation (DMO) can be exploited by ISPs to protect or anonymise the content of their own data before sending them to the ISI sub-system, and analysed through the IAI sub-system.

In the following table, we built a matrix in which we compare the analytics and collaborative functions and we show the willing of an ISP to use one of the available DMO operation. The DMOs are detailed in D8.1 and here we consider a first match analytics vs DMO.

Table 11: Analytics and collaboration functions Vs DMO

	No DMO	Homomorphic Encryption	Anonymization
[CF-01]	X		
[CF-02]	X		X
[CF-03]	X		
[AF-01]	X	X	
[AF-02]	X	X	

Table 11 provides a first definition of which DMO may be applied on the different analytics and collaboration functions. In particular, for the case of [CF-02] the anonymization is the only alternative to “No DMO”. However, the whole anonymization of the report could hide pieces of information and also could make them not relevant anymore. So, a partial anonymization of the content could solve the issue, for instance by making anonymous the ISP and Device-IP parameters. This DMO may be termed as *filtering* and will cover those ISPs interested in removing parts of the data reports that may be considered as sensitive.

5. Security Model

The ISP pilot architecture has been designed to follow the C3ISP hybrid distributed architecture. The reason behind this choice is due to the fact that ISPs may need to locally manipulate some reports before sending them to the remote ISI. In fact, reports will contain data that are related to security aspects, such as weak security configuration, brute-force attacks, intrusions and others. So, the ISPs may prefer to not share the complete details of this information with other entities and for this reason the Local ISI needs to apply the data manipulation operations.

Moving to the security properties, the pilot should guarantee that the most important security property, such as *confidentiality*, data *integrity*, *authentication* and *authorisation*, will be covered.

- Confidentiality: is the property that guarantees that messages transmitted from a source to a destination is not accessed by unauthorised entities.
- Integrity: is the property that allows a recipient to verify whether the message has been altered during its transmission;
- Authentication: is the property that allows a recipient to verify whether the message is sent by a legitimate sender;
- Authorisation: is the property that verifies that an entity is not able to perform the requested action if it is not allowed.

5.1. Confidentially

It mainly involves the communications from ISPs to the DSA Manager, Registro.it and the centralised ISI and IAI. These communications can be protected with the use of a standard and well-known protocol like the Transport Layer Security (TLS)[3].

5.2. Data integrity

It is necessary to verify that data will be not altered during a communication or by other unauthorised entities. In this pilot, ISPs must be able to protect their content from being altered. This will be achieved with the data bundle introduced in the D7.2 that allows ISPs to apply an integrity tag to be then verified later by the recipient. A solution for this issue could be the Message Authentication Code (MAC) or the Hash-based Message Authentication Code (HMAC)[4].

5.3. Authentication

It is managed within C3ISP with the Identity Manager and this verifies that the ISPs are properly authenticated before interacting with components like IAI, ISI, the Registro.it and others. The Identity Manager is described in D7.2 and its working should consider protocol like the Lightweight Directory Access Protocol (LDAP)[5].

[3] https://tools.ietf.org/html/rfc5246

[4] https://www.ietf.org/rfc/rfc2104.txt

[5] https://tools.ietf.org/html/rfc4511

5.4. Authorisation

It is another important security property that is foreseen in C3ISP Framework and it is managed through the DSA Adapter and the Service Usage Control Adapter available in the ISI and IAI. In particular, the DSA Adapter will enforce the execution of some operation on the data is requested, e.g., read. In particular, since the DSA consists of a Usage Control policy, the DSA Adapter retrieves the attributes required for the evaluation from other components of the architecture, evaluates the authorisations and conditions to decide whether the access can be performed or not, and performs the resulting obligations, which can even change the data itself before being released to the requestor. Instead, the Service Usage Control Adapter enforces policies that are strictly related to the service to be used. For instance, a policy monitored by the Service Usage Control Adapter may control that a specific service cannot be invoked more than n-times in parallel. This could be the case of homomorphic encryption analysis that are heavy to be computed.

6. DSA Model

In this section, we present a first definition of policies that may be expressed by the ISPs to protect their data once they are moved to the centralised infrastructure of ISI and IAI.

The policy will be defined through the DSA Authoring available as a tool of the DSA Manager. More specifically, an operator of the ISP should be able to write policies that are in a sort of human-readable format and easy to be expressed. In particular, an ontology will help the operator to express policy that will be written in the Data Sharing Agreement and will be used to control the access to data.

6.1. Authorization and Prohibition policy

As declared in D7.2 policies are encoded in a Controlled Natural Language (CNL) [1] and they express rules about authorisations, obligations and prohibitions. An example of policy written in CNL could be:

if Subject_ID(Subject,ISP_X) and hasType(Data, LOG_X) AND hasActionID(Action, Analytics_X) , then Subject CAN Action Data

This means that ISP_X, which can also be the owner of the data, has the authorisation to execute ANALYTICS_X on the data, i.e., LOG_X. The above policy is written within the DSA and, in particular, in the authorisation body of the DSA. A similar policy could be expressed to extend access to the data to another consumer, which for instance could be ISP_Y. In this case, the policy will be:

if (Subject_ID(Subject,ISP_X) Or Subject_ID(Subject,ISP_Y)) and hasType(Data, LOG_X) AND hasActionID(Action, Analytics_X) , then Subject CAN Action Data

Another policy can also be written to allow or deny the access to the result of an analysis. The following policy labelled has prohibition, says that the ISP_X cannot access the data result identified with the *object_id: 1234.*

if Subject_ID(Subject,ISP_X) and hasId(Data, 1234), then Subject CANNOT access Data

All policies defined into a DSA by a Producer are stored in a bundle together with the data generated with the report. The whole bundle is first kept on the Local ISI and then it may be moved to the remote ISI to be stored into the DPOS. Each action, e.g., read, on the data is blocked and enforced by the DSA Adapter. In particular, it takes the action and enforces it with the policies written in the DSA. If a policy permits the action then it is allowed, otherwise it is denied.

6.2. Obligation policy

Another important aspect of the DSA is related to the Obligation body, here the ISP can express policy on actions that must be performed. This can be the case of a pre-processing on the data, such as encrypting data with homomorphic friendly cryptosystem like Kreyvium algorithm described in D7.2 or data anonymization. So, once the report is sent to the Local ISI, it is processed first by the Format Adapter, to give the correct format to the bundle, and then the DSA Adapter evaluates the obligations written in the DSA. If one or more policies refers to the Data Manipulation Operation, then the DSA Adapter will trigger the DMO Functions component to pre-process the data. The transcryption operation is an example of DMO pre-processing on the data bundle that may be requested by an ISP to protect the privacy of its data.

if Subject_ID(Subject,ISP_X) and hasId(Data, 1234) AND hasActionID(Action, Analytics_X) , then Subject MUST Action Data

7. Deployment Model

7.1. Hardware Requirements

7.1.1. Hardware Requirements for the ISP

In the following, the hardware requirements to achieve a proper and efficient running of the Registrar Local Platform and Local ISI within the ISP premise:

- *Processors*: 2 Intel/AMD 64-bit (4cores, if provided as Virtual Core)
- *Minimum RAM*: 8 GB
- *Hard Disk*: 200 GB

7.1.2. Hardware Requirements for the Registro.it

In the following, the hardware requirements to achieve a proper a proper and efficient running of the Security Scan Software within the Registro.it premise:

- *Processors*: 2 Intel/AMD 64-bit (4cores, if provided as Virtual Core)
- *Minimum RAM*: 8 GB
- *Hard Disk*: 200 GB

7.2. Software Requirements

7.2.1. Software Requirements for the ISP

ISPs will host the Registrar Local Platform and Local ISI to execute the local scripts and to pre-process the report collected from the Security Scan Software. So, it will require the presence of Ubuntu Desktop LTS[6] 16.04 to run the local scripts and the Local ISI. Moreover, depending on how the DPOS will be developed, additional software, such as Apache Hadoop Distributed File System (HDFS) [5] or a DBMS like *MongoDB [6] may be required.*

7.2.2. Software Requirements for the Registro.it

Similar to the requirements of the ISP, the Registro.it will host other local script and the OpenVAS software to create reports. This will require Ubuntu Desktop LTS 16.04.

Registro.it will provide its functionalities through the Security Scan Software that will be invoked through APIs that will be developed with the SOAP[7] or REST[8] technology. The first one exposes a set of methods that can be invoked remotely by clients, while REST allows developers to define a set of resources that clients can request using the HTTP/HTTPs protocol.

The output of a REST API can be them formatted using different encoding, such as XML or JSON. These in the case of the ISP pilot will report the result of the security analysis done.

[6] https://www.ubuntu.com/desktop

[7] https://www.w3.org/TR/soap/

[8] https://www.ibm.com/developerworks/library/ws-restful/index.html

8. Requirements Matrix

In Table 12, we recall the requirements introduced in D2.1 and we show which component satisfies the requirements.

Use Case	Description	Component
ISP-UC-01	It refers to the possibility of an Operator, who work for the ISP, to create a local security report.	Registrar Local Platform
ISP-UC-02	It refers to the possibility to create a new security scan depending on the services requested.	Security Scan Software
ISP-UC-03	It refers to the design and development of the Security Scan Software to identify security threats.	Security Scan Software
ISP-UC-04	It refers to the action of downloading report by the ISP.	Security Scan Software
ISP-UC-05	It refers to the action of opening a report after a scan.	Security Scan Software
ISP-UC-06	It refers to the action of changing the state of a report.	Security Scan Software
ISP-UC-07	It refers to the possibility to offload a security report to C3ISP for further analysis.	Local ISI

Table 12: Requirements matrix

9. Conclusion and Future Work

This document has presented the design and architecture for the ISP Pilot. In particular, the designed architecture follows the hybrid distributed architecture presented in the D7.2 that consists of a ISI block installed locally within each ISP and a centralised installation of ISI and IAI is designed to provide data storage, usage control of the data and analytics.

In addition, the D2.2 has defined a first set of analytics and collaborative functions that the ISPs will be able to run to discover cyber-security issues and to provide collaborative information that can useful for the ISPs. Then, the security and DSA model has been presented as well as the deployment model and the requirements matrix, which recalls the use case requirements defined in D2.1.

This deliverable, released at month 12, can be considered a base for the blocks and components that will reach a first implementation, test and validation phase planned at month 26.

10. References

[1] G. Costantino, L. Deri, F. Martinelli, M. Martinelli, *Requirements for the ISP Pilot*. C3ISP Deliverable 2.1

[2] C. Gambardella, M. Manea , T. Nguyen, V. Herbert, I. Herwono, R. de Lemos, D. Chadwick, F. Di Cerbo, P. Mori, A. Saracino, G. Costantino, I. Matteucci, J. Dobos, *First version of C3ISP Architecture,* C3ISP Deliverable 7.2

[3] P. Mori, C. Gambardella, A. Sajjad, V. Herbert, F. Di Cerbo, A. Saracino, I. Matteucci, M. Manea, G. Costantino, *Components Requirements,* C3ISP Deliverable 8.1

[4] I. Matteucci, M. Petrocchi, and M. L. Sbodio. CNL4DSA: a Controlled Natural Language for Data Sharing Agreements. In SAC: Privacy on the Web Track. ACM, 2010. 21, 23, 52

[5] Hadoop Distributed File System (HDFS) URL: https://hadoop.apache.org/docs/r1.2.1/hdfs_user_guide.html

[6] MongoDB official web site: https://www.mongodb.com/