EUROCC - National Competence Centres in the framework of EuroHPC

EuroHPC-04-2019: HPC Competence Centres

Best Practices for GDPR, Data Interoperability and Cybersecurity

Version 1.0

This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 951732. The JU receives support from the European Union's Horizon 2020 research and innovation programme and Germany, Bulgaria, Austria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, Greece, Hungary, Ireland, Italy, Lithuania, Latvia, Poland, Portugal, Romania, Slovenia, Spain, Sweden, United Kingdom, France, Netherlands, Belgium, Luxembourg, Slovakia, Norway, Switzerland, Turkey, Republic of North Macedonia, Iceland, Montenegro

Title	Best Practices for Cybersecurity, GDPR and Data Interoperability	
Author(s)	Prof. Vladimir Dimitrov	Sofia University
	Vicky Konstantinopoulou	GRNET
	Prof. Maria Nisheva	Sofia University
Approved by	Project Management Team	USTUTT, HLRS
Dissemination Level	Public	

List of abbreviations

BCR	Binding Corporate Rules
CAPEC	Common Attack Pattern Enumeration and Classification
CISA	Cybersecurity and Infrastructure Security Agency
CNAs	CVE Numbering Authorities
CPE	Common Platform Enumeration
CVE	Common Vulnerabilities and Exposures
CVSS	Common Vulnerability Scoring System
CWE	Common Weakness Enumeration
DHS	U.S. Department of Homeland Security
DIF	Directory Interchange Format
DMP	Data Management Plan
DPIA	Data Protection Impact Assessment
EAD	Encoded Archival Description
EEA	European Economic Area
EU	European Union
FIPP	Fair Information Practice Principles
GDPR	General Data Protection Regulation
LOM	Learning Objects Metadata
IoT	Internet of Things
ISMS	Information Security Management System
ITS	Internationalization Tag Set
METS	Metadata Encoding and Transmission Standard
MITRE	The MITRE Corporation
NIF	Natural Language Processing Interchange Format
NIST	National Institute of Standards and Technology
NVD	National Vulnerability Database
ONIX	Online Information Exchange OWASP Open Web Application Security Project
PIA	Protection Impact Assessment
SANS	Escal Institute of Advanced Technologies
SCC	Standard Contractual Clauses
TEI	Text Encoding Initiative

W3C	World Wide Web Consortium
ZAP	Zed Attack Proxy

Table of Contents

List of figures

List of tables

1 Introduction

Data management is a vast topic including many diverse measures to manage data as **efficiently** and **securely** as possible and at the same time in **compliance** with the law. Depending on the use case, different aspects are more or less important. Within this project, three subtopics are identified that are important covering three aspects of data management:

1. Legal compliance: General Data Protection Regulation

In general, all processed data should be legally compliant. The EuroCC project (especially the management tasks) requires the handling of personal data, e. g. when it comes to training events, workshops or websites etc. The second chapter on t legal compliance introduces the GDPR and subsequent best practices. In addition, it may be the case that personal data is processed in other parts of the project, so the explanation of basic principles is also intended to raise awareness. Since not all project partners are located in an EU-country, reference is made to the procedure for cooperating with non-EU-countries.

2. Efficiency: data interoperability

The subtopic of data interoperability is generally important for the aspect of efficiency, since data must be transferred in order to use HPC. When it comes to cooperation between countries, it makes sense in the EuroCC project to think about data interoperability (technical, syntactic and semantic interoperability must be given), but it can also be a first step towards European-wide standards.

As data interoperability is one of the FAIR principles a quick overview is given on that topic. The FAIR principles for data management provide methodologies and guidelines to improve the findability, accessibility, interoperability, and reusability of digital resources.

In this chapter, the focus is laid on semantic interoperability as the highest and most complete level of data interoperability as well as on the most popular instruments for achieving semantic interoperability – metadata standards and ontologies. Examples of good practices and free software tools supporting semantic interoperability are given at the end of the chapter.

3. Security: Cybersecurity

Data interoperability is one of the main pillars in the EuroCC project. This is only possible with an adequate level of supported cybersecurity of the data. As today's world is becoming more digitized and more interconnected, it leads to the emergence of new unknown threats, which evolve and grow with the Internet proliferation. Therefore, the cybersecurity subtopic is addressed in the third chapter.

Cybersecurity is a complex task. It is not only an issue of some areas in the organization; the approaches to cybersecurity must be systematic managerial and technical. At the same time, security must scope the so-called "extended enterprise" – clients, employees, partners, auditors and regulators. Extended enterprise relationships are built on trust between the parties. Both security management and security enforcement are also based on trust; where trust means trust among partners and their relationships and technology-supported trust. Information is the primary target that is why the international series of standards ISO/IEC 2700K are focused on information security management systems. These management standards will be briefly discussed.

2 General Data Protection Regulation (GDPR)

2.1 Introduction

The General Data Protection Regulation (hereinafter referred to as "GDPR") is a European law on the protection of natural persons with regard to the processing of personal data and on the free movement of such data[1] and went into effect on May 25, 2018. This is not the first time that the European Council has passed a privacy law; the main difference from the previous legislations is that the GDPR tried to "harmonize" the data protection laws of the 28 members of the European Union in order to remove the obstacles to flows of personal data within the Union[2]. Furthermore, GDPR is currently one of the toughest privacy and security laws that affects countries around the world and does not remain secluded within the EU. Organizations are now responsible for the personal data they collect, store and use and should apply technical and organizational measures in order to protect and secure this type of data.

Along with the organizations, come the projects which could be developed by more than one stakeholder. EuroCC as a project has 33 National Competence Centres (NCC) which are the focal points of contact for HPC and related technologies in their country. According to the GDPR guidelines, the NCCs should have a proper data management plan for EuroCC in place. They should be aware of what types of data will be collected, processed or generated. The same principle applies to EuroCC as a H2020 project. For these reasons, the members of the EuroCC consortium developed the Data Management Plan (DMP). The purpose of this deliverable is to describe the types of data that the project will generate or collect, the standards that will be used, and how this data will be exploited, shared and reused, as well as curated and preserved[3]. Furthermore, the DMP is a living document, it will evolve as the project develops: if there are new findings of data generated and processed, then those are to be added to the DMP, aligning with the protection and privacy standards. In the following sections, there will be a brief presentation of the GDPR legal terms, principles and data subjects' rights, as well as a best practice guide for EuroCC.[4]

Here in after are some of the most important definitions of GDPR:

- **'Personal data' –** any information relating to an identified or identifiable natural person.
- **'Data subject' –** an identifiable natural person who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.
- **'Processing' –** any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use,

[1] https://eur-lex.europa.eu/eli/reg/2016/679/oj

[2] https://gdpr-info.eu/recitals/no-10/

[3] Data Management Plan is a separate document that is updated continuously. The main objective of a DMP is to make research data FAIR, i.e. **F**indable, **A**ccessible, **I**nteroperable and **R**eusable, in accordance with the EC's guideline on Fair Data Management in Horizon 2020. Therefore it is a common deliverable for all NCC's.

[4] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

- **'Profiling'** – any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.
- **'Pseudonymisation'** – the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.
- **'Controller'** – the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data;
- **'Processor' –** a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.
- **'Third party'**– a natural or legal person, public authority, agency or body other than the data subject, the controller, the processor and the persons who, under the direct authority of the controller or processor, are authorised to process personal data.
- **'Consent'** – the consent of the data subject is freely, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.
- **'Personal data breach'** – a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed.
- **'Genetic data'** – personal data relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question.
- **'Biometric data'** – means personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data.
- **'Data concerning health'** – personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status.

2.2 Data Protection Principles

GDPR sets its own set of principles for processing personal data. It must be noted that GDPR is not the first tool that tried to set a key of principles. The Organisation for Economic Co-operation and Development in 1980 published guidelines for the Protection of Privacy and Transborder Flows of Personal Data,[5] known as Fair Information Practice Principles (FIPP's) and in 2016 the FAIR principles were published by a consortium of scientists and organizations.[6] The main difference of these documents is that GDPR principles are legally binding, while the others are merely suggestions on fair use of personal data.

In accordance with the article 5 of the GDPR, the processing of personal data must adhere to these principles[7] [8]:

1. **Lawfulness, fairness and transparency** — processing must be lawful, fair, and transparent to the data subject.
2. **Purpose limitation** — data must be processed for the legitimate purposes, specified explicitly to the data subject at the time of collection.
3. **Data minimization** —only as much data should be collected and processed as is absolutely necessary for the purposes specified.
4. **Accuracy** —personal data must be kept accurate and up to date.
5. **Storage limitation** —personally identifying data may only be stored for as long as is necessary for the specified purpose.
6. **Integrity and confidentiality** — processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g. by using encryption).
7. **Accountability** — the data controller is responsible for being able to demonstrate GDPR compliance with all of these principles.

2.2.1 Accountability

The GDPR introduces accountability as a new principle for the data protection rules in Europe. The difference is that the controller shall be responsible for, and be able to demonstrate compliance with the previous principles. This means that the organizations should put in place technical and organizational measures, in order to be able to demonstrate their compliance. A number of those measures but not limited to, could be:

- **Records of processing activity[9]** (article 30, GDPR) – the records contain:
    - the name and contact details of the controller and, where applicable, the joint controller, the controller's representative and the data protection officer.
    - the purposes of the processing.

[5] https://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

[6] https://www.go-fair.org/fair-principles/

[7] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

[8] https://gdpr.eu/what-is-gdpr/

[9] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

  - a description of the categories of data subjects and of the categories of personal data.
  - the categories of recipients to whom the personal data have been or will be disclosed including recipients in third countries or international organisations.
  - where applicable, transfers of personal data to a third country or an international organisation, including the identification of that third country or international organisation and, in the case of transfers, documentation of suitable safeguards.
  - where possible, the envisaged time limits for the deletion of the different categories of data.
  - where possible, a general description of the technical and organisational security measures.

- The appointment of a Data Protection Officer (DPO), if the organization falls under this obligation[10]. A DPO ensures that his/her organisation as a whole, processes the personal data of its data subjects in compliance with the applicable data protection rules. For EuroCC, the Working Package Leader acts as a Data Manager, but that is not to be confused with the DPO. The Data Manager acts on the behalf of EuroCC project and works with the DPO in order to implement the organizational and technical measures for the protection of data.
- Privacy and Security Policy – those two policies contain everything that an organization should implement under the legal and security framework and holds the procedures, registrations and the steps of the risk assessment. More on the Privacy Policy see chapters 2.8 and 2.10.
- Data Protection Impact Assessment – a process in which organizations can identify and minimise the data protection risks of a project. EuroCC, in general does not fall under the obligation to conduct a DPIA, but as the DMP evolves, the consortium will see this through if necessary. For more information on the DPIA see chapter 2.7.
- Privacy statement – is a public document, usually placed on the website of the service which the organization provides and informs the data subjects about the processing of their personal data. For more information see chapter 2.9.

2.3 Lawfulness of processing

Under GDPR[11], organizations should only process data if one of the following conditions is met:

- The data subject has given consent to the processing of his or her personal data for one or more specific purposes.
- Processing is necessary to fulfil a contract to which the data subject is a party or to implement steps at the request of the data subject prior to entering into a contract.
- Processing is necessary for compliance with a legal obligation to which the controller is subject.

[10] See the obligations of appointing a DPO to Art. 43 of GDPR: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R1725&from=EN p. 38

[11] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

- Processing is necessary in order to protect the vital interests of the data subject or of another natural person.
- Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official, authority vested in the controller,
- Processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject, which require protection of personal data, in particular where the data subject is a child.

2.3.1 Consent

Consent is one of the six lawful ways mentioned above, to process personal data. Consent, may seem to be the easiest way to obtain personal data but the requirements for it, are a bit challenging.

Consent, according to GDPR, must be a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her[12].

So, GDPR, establishes four requirements:

- Freely given,
- Specific,
- Informed and
- Unambiguous.

The first requirement concerns the free will of the data subject to provide its consent by making a free choice. If consent is bundled up as a non-negotiable part of terms and conditions, it is presumed not to have been freely given[13]. This also mean that pre-ticked boxes are not considered as an act of free choice.

Furthermore, "specific" means that the data subject should be informed about the purposes for which its data will be processed. For each of the purposed, the data subject should be able to choose whether or not to consent[14].

For the data subjects to be adequately informed, the controller should provide accessible information regarding the identity of the controller, as well as, the purposes of the processing, the type of data to be collected (personal, health, genetic, etc), the way to withdraw consent, the use of the data for automated decision-making and the possible risks of data transfers. For more information regarding this please see the 2.9 Privacy Statement example. In order for the consent to be unambiguous, the data subject should have taken a voluntary action to consent. It is up to the organizations to design mechanisms in ways that

[12] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

[13] The Article 29 Working Party has issued guidelines on consent under GDPR. For the requirement of the "freely given" provides the following example: A mobile app for photo editing asks its users to have their GPS localisation activated for the use of its services. The app also tells its users it will use the collected data for behavioural advertising purposes. Neither geo-localisation or online behavioural advertising are necessary for the provision of the photo editing service and go beyond the delivery of the core service provided. Since users cannot use the app without consenting to these purposes, the consent cannot be considered as being freely given." Page 6, https://ec.europa.eu/newsroom/article29/items/623051/en

[14] Recital 43 GDPR states that separate consent for different processing operations will be needed wherever appropriate. Granular consent options should be provided to allow data subjects to consent separately to separate purposes. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

are clear to the data subjects, GDPR does not define the tools, only the requirements, therefore the action by which consent is given should be distinguished from other actions[15].

Some of the **best practices** in order to obtain consent are:

- Ticking an opt-in box;
- Clicking a link online;
- Selecting from yes/no options;
- Choosing technical settings from preference dashboard settings;
- Responding to an email requesting consent;

The statement of asking the consent is also important. For example, it should read:

- If you would like to receive our newsletter please tick here

And not

- If you don't want to receive our newsletter please tick here

Once the consent is obtained a procedure of keeping a record should also be in place. There must be an effective audit trail of how and when consent was given. The information that is required is:

- Who did consent,
- When they did consent,
- What time did they consent, and
- How did they consent

The same applies to the withdrawal of consent.

2.3.2 Special categories of data

The following special categories of data may not be processed according to GDPR:

- Personal data revealing racial or ethnic origin.
- Political opinions.
- Religious or philosophical beliefs.
- Trade union membership.

[15] The Article 29 Working Party provides two different examples of a distinguished action regarding the consent:

Example 15

Swiping a bar on a screen, waiving in front of a smart camera, turning a smartphone around clockwise, or in a figure eight motion may be options to indicate agreement, as long as clear information is provided, and it is clear that the motion in question signifies agreement to a specific request (e.g. if you swipe this bar to the left, you agree to the use of information X for purpose Y. Repeat the motion to confirm). The controller must be able to demonstrate that consent was obtained this way and data subjects must be able to withdraw consent as easily as it was given.

Example 16

Scrolling down or swiping through a website will not satisfy the requirement of a clear and affirmative action. This is because the alert that continuing to scroll will constitute consent may be difficult to distinguish and/or may be missed when a data subject is quickly scrolling through large amounts of text and such an action is not sufficiently unambiguous. https://ec.europa.eu/newsroom/article29/items/623051/en

- Genetic data and biometric data processed for the purpose of uniquely identifying a natural person.
- Data concerning health.
- Data concerning a natural person's sex life or sexual orientation.

There are, however, limited circumstances set out in article 9 of the GDPR, in which these types of data can be processed the main one is:

- The data subject has given explicit consent to the processing of that personal data for one or more specified purposes.
- Or the processing is necessary for the purposes of carrying out the obligations and exercising specific rights of the controller

Examples of explicit consent:

- The data subject to fill in an online form
- Sent an email with the required data
- Offering Yes and No checkboxes

2.3.3 Consent vs the performance of a contract

The GDPR sets 6 different legal bases for the processing of data and some limited circumstances in which special categories of data can be processed. The controller should be able to decide which of this lawful basis apply, sometimes it can be more than one. In order to do so, the controller should firstly identify the purposes of the processing, there should be a preliminary assessment of goals set and those should be distinguished against business needs. For this, the controller should be able to answer the following questions:

- Is the processing necessary?
- Are there alternative ways to process the data? Less intrusive?
- Is the data subject able to foresee other aspects of the processing?
- Are more than one purposes of processing to the provided service?

There is no simple yes or no answer to these questions, but they can in some ways set the basis for distinguishing between the "necessity of a performance of a contract" or going beyond that and looking for another lawful basis. Two more aspects that should be taken under consideration are the special categories of data and if the data that is processed falls under the e-Privacy Directive[16]. For each of the legal bases, the GDPR principles must be met and the data subject should be adequately informed.

For the **EuroCC project** it should be noted that an NCC shouldn't ask for consent if the process falls under:

- A core service of the project (use instead performance of a contract)

[16] See Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector, in particular Article 5(3).

- Must process personal data by law (this is a legal obligation)
- Or the processing of personal data benefits the NCC in a way that EuroCC users would reasonably expect, with minimal risk taken in consideration and the minimal impact on individuals (this could fall under legitimate interests)

Consent is should be use for a non-essential service, such as:

- Using tracking/advertising cookies. More on this on 3.10 Cookies Policy
- Sending marketing emails or newsletters

In conclusion, the controller should consider whether there is more than one legal basis for the purposes of the processing. Consent must meet certain criteria in order to be adequate with GDPR provisions. Furthermore, the controller is obliged to demonstrate that the data subject has given consent to the processing operation[17]. Also, the withdrawal of consent should be as easy as giving consent in the first place. Additionally, the concept of necessity should be met for the performance of a contract. The processing must be objectively necessary and distinct from business needs.

2.4 Data subject's rights

According to article 12 of the GDPR: "the controller shall take appropriate measures to provide any information relating to processing to the data subject in a concise, transparent, intelligible and easily accessible form, using clear and plain language, in particular for any information addressed specifically to a child.[18]" However, the GDPR does not leave it at that, for the protection of the individual it sets eight specific rights which the data subject can exercise. Namely the rights of access, rectification, erasure, restriction of processing, data portability, the right to be informed, and detailed hereinbelow:

- **Right to be informed** – any information addressed to the data subject must be concise, easily accessible and easy to understand. The controlled should inform the data subject about the identity and the contact details of the controller, the contact details of the data protection officer, the purposes of the processing for which the personal data are intended as well as the legal basis for the processing, the categories of personal data concerned, the recipients or categories of recipients of the personal data, and where applicable, that the controller intends to transfer personal data.
- **Right of Access** – the data subject is entitled to request and obtain confirmation from the controller as to whether or not his/her personal data is being processed and, if so, to exercise the right to access such personal data pursuant to applicable legislation. The data subject may also request a copy of the personal data undergoing processing, as described in the privacy statement, that the controller has in place.
- **Right of Rectification** – the data subject shall have the right to request the controller to rectify any inaccurate personal data concerning him/her. Considering the purposes of the processing, the data subject shall have the right to request that any incomplete personal data

[17] Article 29 Working Party, Guidelines on consent, https://ec.europa.eu/newsroom/article29/items/623051/en

[18] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

be completed, including by means of providing a supplementary statement, in accordance with applicable legislation.

- **Right of Erasure –** the data subject has the right to delete all his / her personal data that have been collected and processed.
- **Right to restriction of processing –** the data subject is entitled to ensure that the controller restricts the processing of his/her data, if any of the conditions laid down by applicable legislation on the protection of personal data, is met.
- **Right to data portability –** the data subject has the right to obtain any personal data concerning him/her, which he/she has provided to the controller in a structured, commonly used and machine-readable format, as well as the right to transmit such data to another controller without any objection from the controller to which the personal data have been provided, in accordance with the provisions of the applicable legislation on personal data.
- **Rights related to automated decision-making including profiling –** the data subject has the right not to be subject to a decision based solely on automated processing.

The controller according to GDPR provisions must answer to the data subject without undue delay and at the latest within one month of receipt of the valid request. Each NCC of the EuroCC project is responsible to answer to any data subject that may request to exercise one of his/her rights. For each of these rights there should be a procedure in place. For example:

First step should be to verify the identity of the requester. This is quite important in order to safely distribute the information. If the wrong person receives the data, the NCC may then be facing a data breach.

The second step is to clarify the data subject’s request. The NCC should ask the data subject to specify the information or processing activities to which their request relates to before responding to the request.

Once the data is collected, the NCC must make sure it doesn’t include personal data of other data subjects since that will probably lead to a data breach.

The data subject should be reminded of their rights and also about their right to object to the processing of their data, and request the rectification of their data, and/or lodge a complaint with a supervising authority.

2.5 Data Protection Agreement

A Data Protection Agreement (DPA) is an agreement between a controller and a processor. It sets the technical and organizational requirements as well as the duration of the processing of personal data. It also defines the limitations of the processing. The DPA is a key element of the GDPR requirement. Article 28 of the GDPR states the sections which the DPA should include:

- Documented instructions from the controller regarding the processing, including the transfer of personal data to a third country or an international organisation.
- The actions that both parties take to ensure that persons authorised to process the personal data have committed themselves to confidentiality or are under an appropriate statutory obligation of confidentiality.

- The security and technical as well as organizational measures required.
- The general information of another processor, if there is one or more.
- The commitment of the processor to assist the controller through appropriate technical and organisational measures, insofar as this is possible, for the fulfilment of the controller's obligation to respond to requests for exercising the data subject's rights.
- The processor is also obliged to:
    - assist the controller in ensuring compliance with the obligations deriving from the GDPR,
    - at the choice of the controller, delete or return all personal data to the controller after the end of the provision of services related to the processing and delete existing copies,
    - provide the controller with all information necessary to demonstrate compliance with the obligations laid down in the GDPR and allow for and contribute to audits, including inspections, conducted by the controller or another auditor mandated by the controller[19].

It should be noted that, in relation to the parties involved, a DPA could also be between controllers in terms of a Joint Controllers Agreement or between a processor and a sub-processor. A DPA is not a best practice, but a legal requirement set by the provisions of the GDPR. A DPA is a different part of a contract or a consortium agreement, has a different legal entity and, as mentioned above, is part of the controller's demonstration of accountability[20].

2.6 Transfers to a third country – Adequacy Decisions-Standard Contractual Clauses

According to the GDPR, personal data can be freely transmitted within the EU and the European Economic Area without restrictions. Outside the EU/EEA, there is no equivalent protection for the transfer of data. The general prohibition of the transfer of data would seclude EU members from business relationships outside the EU/EEA zone. Therefore, GDPR contains a set of rules on the conditions that must be met in order for the transfer to be permitted and for the data to be adequately protected as within the EU/EEA.

For the transfer of data to take place, the following must be present:

- A decision from the European Commission that a certain country outside the EU/EEA ensures an adequate level of protection.
- In the lack of an adequacy decision, the controller must adopt appropriate protection measures, with those either being Binding Corporate Rules (BCR) or Standard Contractual Clauses (SCC).
- Special situations and single cases[21].

[19] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

[20] The GDPR.EU is a website operated by Proton Technologies AG, which is co-funded by Project REP-791727-1 of the Horizon 2020 Framework Programme of the European Union. It provides a DPA template: https://gdpr.eu/data-processing-agreement/. The template is not an official legal document and a further processing is expected.

[21] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN

The way in which the European Commission determines whether a country has an adequate level of protection or not can be found here[22]. For the countries that don't hold an adequacy decision, the European Commission has pre-approved a model of contractual clauses for data transfers outside the EU/EEA[23]. The SCC follow the DPA or the Joint Controllers Agreement.

2.7 Data Protection Impact Assessment

As mentioned above, the controller should be able to demonstrate its compliance with the GDPR. This means that, depending on the processing that is taking place, the controller should be able to determine whether this processing imposes high risks to the rights and freedom of the data subjects. Taking into account the nature, scope, context and purpose of the processing, the controller should conduct an assessment of the impact of the envisaged processing operations on the protection of personal data, as stated in article 35 of the GDPR.

A Data Protection Impact Assessment (DPIA) according to GDPR should take place if one of the following conditions is met:

- A systematic and extensive evaluation of personal aspects relating to natural persons, which is based on automated processing, including profiling, and which is based on decisions that produce legal effects concerning the natural person or similarly significantly affect the natural person.
- Large scale processing of special categories of data or of personal data in connection with criminal convictions and offences.
- Systematic monitoring of a publicly accessible area on a large scale.

DPIA has some similarities with the privacy impact assessment (PIA) but the main differences are:

- DPIA is mandatory under GDPR.
- PIA is conducted to identify and mitigate organizational privacy risks.
- DPIA is to identify and mitigate risks associated with the processing of personal data.
- PIA is to achieve privacy by design in order to protect data efficiently.
- DPIA is to demonstrate that the risk has been mitigated and the processing is aligned with the requirements of the GDPR.

The French supervisory authority has published a guide for conducting a DPIA.[24]

2.8 Privacy Policy vs Privacy Statement

The data privacy policy generally includes the purposes and the objectives set by the controller regarding the protection of personal data, as well as the instructions, procedures, rules, roles and responsibilities related to the protection of this data.

[22] For the list of adequacy decisions see this: https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en

[23] For the Standard Contractual clauses see this: https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/standard-contractual-clauses-scc/standard-contractual-clauses-international-transfers_en

[24] For more information about the DPIA see this: https://www.cnil.fr/en/guidelines-dpia

On the other hand, the privacy statement informs the data subject about the processing of their personal data. The data privacy policy is an internal document, whereas the privacy statement is a public one. The privacy statement is needed for the transparency principle while the data privacy policy stands for accountability. There should be a definitions section in the data privacy policy to define the different terms used throughout the document. It should include the legal basis of the processing, the rights of the data subject s and the data principles. In addition, the roles and responsibilities for employees must be included as well. Access should not be available to everyone within the organization, there should be access control according to the role of each employee. Furthermore, within the data protection policy, the controller should include the procedures for handling a data breach, as well as security measures, data retention practices, and data records[25]. The privacy statement informs the data subjects about the purpose/s for processing their data, the categories of data, the legal bases for processing, the recipients of data, their rights, the retention period, and the contact information of the Data Protection Officer[26].

2.9 Privacy Statement example

A Privacy Statement should list the following categories[27]:

Privacy statement regarding the protection of personal data in the context of the EuroCC project

Controller details:

Name of the Controller

Competent Processing Project:

Which department of the NCC is responsible for EuroCC

Controller's Contact Details:

An email address should be provided for data subjects

Processor:

If there is a processor, its name and its privacy statement (a website link is sufficient) should be added

Scope of this Privacy Statement:

NCCs should provide notice of the scope of this privacy statement, for example:

This Privacy Statement sets out the criteria as well as the terms and conditions under which EuroCC collects, processes, uses, stores and transmits the personal data of the project users, how it ensures the confidentiality of such information, including any law and/or regulation implemented or enacted in accordance with Union and national laws on personal data protection and electronic privacy, as well as any law and/or regulation amending, replacing, issuing or consolidating any of the latter,

[25] For more information see this: https://www.privacypolicies.com/blog/gdpr-data-protection-policy/

[26] For more information please check: https://security.berkeley.edu/how-write-effective-website-privacy-statement#:~:text=Your%20privacy%20statement%20must%20accurately,you%20must%20inform%20your%20users.

[27] For more information see: https://eurocc-greece.gr/privacy-policy/

including any other applicable Union and national laws on the processing of personal data and privacy, which may exist in accordance with applicable law.

A. Purpose/s for processing the data collected:

The purpose/s for which the project collects and processes the personal data, in detail

B. Categories of personal data processed:

In this section the NCC should list what types of data collects and processes, those could fall under the following categories:

General data (Name, address, date of birth, age, gender, residence, email address, position, area of work, work phone, etc)

Confidence data (National identification numbers, debt,etc)

Sensitive data (Race and ethnic origin Political, religious or philosophical beliefs Trade Union affiliation Genetic and biometric data for the purpose of unique identification Health details Sex life and sexual orientation)

C. Legal bases for processing

As mentioned in 3.3 section Lawfulness of processing in order to legally process any type of personal data one of the 6 legal bases should apply. In this section, NCC should inform the data subject for its legal base.

D. Access to personal data:

Who within the NCC has access to personal data, it could be the project management team, or the security team, or the processor. A simple mention should oblige.

E. Recipients of collected personal data:

Are there any third-party entities, private businesses, natural persons or legal entities, public authorities, agencies or other organizations, where the personal data should be disclosed or transmitted to?

F. Rights of data subject

As mentioned in 3.4 Data subject’s rights, the NCC should inform the data subjects about their rights and how they can exercise those rights. This could be by sending an email or filling in an online form.

G. Personal data retention periods

For how long will the collected data be kept?

H. Privacy and Security of Information:

The processing of personal data by each NCC should be carried out in such a way as to ensure its confidentiality and security. NCC should briefly inform the data subjects about the technical and organizational measures that are taken for data security and protection against accidental or unlawful destruction, accidental loss, alteration, prohibited dissemination or access and any other form of unfair treatment.

I. Contact:

Data subjects should be informed on who to contact to file their complains. This could be the DPO of the NCC or the Data Manager of EuroCC, or a team that will handle this type of requests.

J. Recourse/Complaint

In the event that the request is not satisfied by the NCC, the data subject may at any time address to/ file recourse with the Competent Supervisory Authority.

2.10 Privacy Policy

A privacy policy, as mentioned above should contain all the procedures that the organizations follow in order to protect the data collected and processed. The purpose of a privacy policy is to recognize the risks involved in the processing of personal data carried out by the controller and the implementation of countermeasures to mitigate these risks.The purpose is to apply the rules and techniques in order to satisfy the legal rights of natural persons whose personal data is processed and lastly to comply with the requirements arising from the GDPR. Furthermore, if a data breach arises then the privacy policy holds in place a plan for dealing with this type of incidents, as well as a plan to inform the data subjects and the competent authorities.

For EuroCC the presented document as well as the Data Management Plan constitute the Privacy Policy of the project but do not replace in any way the Privacy policy of each NCC.

2.11 Data Breach Action Plan

If any of the NCCs becomes aware of, or are notified of a data breach which affects the service they provide for EuroCC, they must immediately inform the data manager of EuroCC.

The data manager should be informed for the following:

- The time and date,
- The type of personal data,
- The cause and extent of the breach, and
- The context of the affected data and the breach.

Following this, the data manager along with the security team and the DPO should enable an initial assessment of whether a data breach has or may have occurred and the seriousness of the data breach. There should be record keeping for each data breach in place. Once the incident has been handled, the data manager should consider notifying other NCCs, if affected by the data breach.

2.12 Cookies Policy

A cookies policy provides detailed information to the users about the types of cookies a website uses. A cookies policy should be in place, according to GDPR guidelines, as it is in line with the principle of accountability. There should be a clear distinction between first party and third-party cookies.

- First party cookies are cookies set by the website you're visiting. Only that website can read them. In addition, a website might potentially use external services, which also set their own cookies, known as third-party cookies.
- Persistent cookies are cookies saved on your computer and that are not deleted automatically when you quit your browser, unlike a session cookie, which is deleted when you quit your browser.

There are different types of cookies that target different types of data. The data subject should first be informed about the cookies and their function and whether he/she wishes to consent to the use of cookies. It should be noted that the navigation on a site should not be limited by the use of cookies. One must be able to navigate through it even without providing consent[28].

Any NCC using cookies on their website should consider adding a rather detailed cookie policy and to mention:

- The purposes of the cookies policy
- Which "cookies" are placed on a user's device when using the website
- How long Cookies are stored
- How a user can disable cookies

Example for a cookie policy template:

The cookies used by our website are specified in detail in the table herebelow:

Table 1: Example for a cookie policy template

Cookie type	Cookie provider	Cookie name	Third party cookie	Persistent cookie	Purpose of Cookies
What type of cookie do they use: Session Cookie/Permanent Cookie/1st party cookie/3rd party cookie/Flash cookie/Zombie cookie	The business or entity that owns a cookie. For example, Google Analytics	The name of the cookie.	A third-party cookie is placed on a website by someone other than the owner and collects user data for the third party (social media, live chat pop-ups)	A persistent cookie is stored on a user's device to help remember settings, or sign-on credentials that a user has previously saved. It lasts as long until the user deletes it.	What is the main purpose of the cookie?

[28] For more information please check: https://ec.europa.eu/info/cookies_en

2.13 EuroCC Privacy team

The EuroCC Consortium should initiate the formation of a data privacy team, where involved the members would implement a data privacy framework according to which there are roles with responsibilities for the protection of personal data and should be notified of any privacy incident that occurs and affects the project. The framework will include at least the following roles:

- Data Privacy Manager
- Information Security Officer
- Information Service Officer

The privacy team should have a point of contact that can be notified if privacy issue arise. They could also assist in updating the data management plan and responding to a data subject's requests.

3 Data interoperability

3.1 Introduction

During recent decades, the evolution of information systems and communication technologies, particularly those related to the Internet, has led to the implementation of peer-to-peer communication models among heterogeneous software systems. In this process, the main challenge is to provide **system interoperability** at all levels.

Interoperability in general is the capability of different software systems to communicate. This communication may take various forms such as the transfer, exchange, transformation, mediation, migration or integration of information[29].

System interoperability is provided first of all by **data interoperability,** which is one of the main research areas within the EuroCC project. The term “data interoperability” addresses the ability of systems and services to have clear and shared expectations for the contents and meaning of the data they create, exchange and use.

3.2 General information

3.2.1 The FAIR guiding principles for scientific data management

Data interoperability is the core of the so-called **FAIR principles** for data management which aim to provide methodologies and guidelines for maximizing the usefulness of the data. These principles are based on four aspects of data organization[30]:

- **Findability**: Metadata and data should be easy to find for both humans and software systems. Appropriate metadata are essential for the automatic discovery of useful datasets and services.
- **Accessibility**: Once the user finds the required data, he/she should be provided with the necessary information on how to access it, including information on possible authentication and authorization.
- **Interoperability**: The data usually need to be integrated with other data. In addition, software systems need to interoperate with other systems or workflows for analysis, storage, and processing of data.
- **Reusability**: The ultimate goal of FAIR is to ensure the maximum reuse of data. To achieve this, metadata and data should be well-described so that they can be copied or combined in different ways, as well as used in unforeseen cases.

Data interoperability is the most important result of the **FAIRification process**. It is a feature of information systems which allows data to be easily retrieved, processed, and reused by other systems.

[29] Patel, M. et al. DELOS: A Network of Excellence on Digital Libraries. D5.3.1: Semantic Interoperability in Digital Library Systems. UKOLN, University of Bath, 2005.

[30] Wilkinson, Mark D., et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, Vol. 3 (2016), No. 1, pp. 1-9.

3.2.2 Levels of interoperability

There are several classification schemes of the types of interoperability. Most authors distinguish at least three hierarchical levels of interoperability (see Figure 1):

- Technical interoperability
- Syntactic interoperability
- Semantic interoperability

The availability of **technical interoperability** allows bidirectional data exchange between one software system and another and does not require the ability of both the sending and receiving systems to interpret the data. In general, it is based on the agreement of a common communication protocol between the corresponding software systems.

If two or more information systems are based on a common data model, they can display **syntactic interoperability**. This is an intermediate level of interoperability that defines the structure or format of data exchange (the message format standards). This level of interoperability defines the syntax of the data exchange. It ensures that data exchanges between different information systems can be interpreted at least at the data field level.

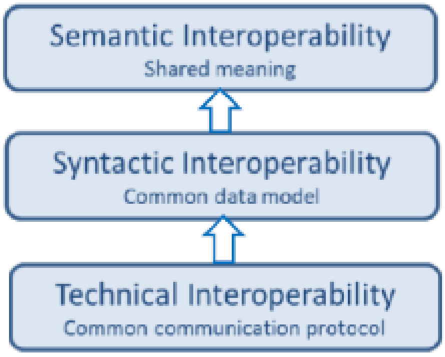

Semantic Interoperability
Shared meaning
Syntactic Interoperability
Common data model
Technical Interoperability
Common communication protocol

Figure 1: Levels of interoperability

Beyond the ability of two or more software systems to exchange information, **semantic interoperability** is the ability to automatically interpret the information exchanged meaningfully and accurately in order to produce useful results as defined by the end users of all systems[31]. Semantic interoperability takes advantage of both the structuring of the data exchange and the codification of the data including proper **vocabulary** so that the receiving software systems can **interpret the data**.

[31] Davies, J., R. Studer, P. Warren (Eds.). Semantic Web Technologies. John Wiley & Sons, 2006.

3.3 Semantic interoperability

3.3.1 Main characteristics of semantic interoperability

Semantic interoperability is characterised by the capability of different software systems to communicate information consistent with the intended meaning of the encoded information (as intended by the creators or maintainers of the systems).

Semantic interoperability involves:

- The processing of the shared information so that it is consistent with the intended meaning
- The encoding of queries and presentation of information so that it conforms with the intended meaning regardless of the source of information

Semantic interoperability issues are relevant to different degrees in each of the following elements of the information lifecycle[32]:

- Creation, modification
- Publication
- Acquisition, selection, storage, system and collection building
- Cataloguing, indexing, knowledge organisation, knowledge representation, modelling
- Integration and linking
- Mediation, personalization, reference, recommendation
- Access, search and discovery
- Annotation, evaluation
- Use, shared application, scholarly communication; reuse
- Maintenance
- Archiving and preservation

Standardization and semantic technologies (primarily ontologies) are indicated as the most effective instruments for providing and maintaining interoperability in information systems.

3.3.2 Standardization

In order to achieve semantic interoperability in a software system environment, **standardization** may comprise widely accepted agreements for the form and meaning of metadata and content schemata as well as for the use of names and the construction of identifiers for concepts and real-world objects.

[32] Patel, M. et al. DELOS: A Network of Excellence on Digital Libraries. D5.3.1: Semantic Interoperability in Digital Library Systems. UKOLN, University of Bath, 2005.

Standardization can direct to the following favourable features of the respective information systems[33]:

- Information can be communicated, transferred, integrated, merged, etc. without transformation
- Information can be kept in a single form
- Information of candidate sources can be enforced to be functionally complete for an envisaged integrated service

Many different metadata schemes are being developed as standards. Some of the most popular **metadata standards** are:

- **Dublin Core** – an interoperable online metadata standard focused on networked resources
- **Encoded Archival Description** (EAD) – a standard for encoding archival finding aids in archival and manuscript repositories
- **Online Information Exchange** (ONIX) – an international standard for representing and communicating book industry product information in electronic form
- **Learning Objects Metadata** (IEEE LOM) – a standard which specifies the syntax and semantics of Learning Object Metadata
- **Text Encoding Initiative** (TEI) – a standard for the representation of texts in digital form
- **Z39.87 Data Dictionary** (NISO MIX) – a technical metadata standard for a set of technical data elements required to manage digital image collections
- **Metadata Encoding and Transmission Standard** (METS) – a XML schema for encoding descriptive, administrative, and structural metadata regarding objects within a digital library
- **Multimedia Content Description Interface** (MPEG-7) – a ISO/IEC standard which specifies a set of descriptors to describe various types of multimedia information
- **Directory Interchange Format** (DIF) – a descriptive and standardized format for exchanging information about scientific datasets

Most of these metadata standards are open but some of them (e.g. LOM) are proprietary.

3.3.3 Ontologies

Semantic interoperability requires that an information system understands both the semantics of the information sent or requested by another system, as well as the semantics of its information sources. In recent years, ontologies have been used as artefacts to represent information semantics.

According to the popular definition of Gruber[34], an **ontology** is an “explicit specification of a conceptualization” which in turn are “the objects, concepts, and other entities that are presumed to exist in some area of interest and the relationships that hold among them”.

[33] Nisheva-Pavlova, M. Providing and Maintaining Interoperability in Digital Library Systems. Proceedings of the Fourth International Conference on Information Systems and Grid Technologies (Sofia, May 28-29, 2010), St. Kliment Ohridski Uiversity Press, 2010, pp. 200-208.

[34] Gruber, T. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal of Human-Computer Studies, Vol. 43 (1995), pp. 907–928.

An ontology is a formal description of domain knowledge as a set of concepts, their properties and the relationships between them. To enable such a description, we need to formally specify components such as classes (concepts), attributes and relations as well as restrictions, individuals (instances or objects), rules and axioms. As a result, ontologies introduce a sharable and reusable knowledge representation that can be flexibly used and extended with new knowledge about the domain. They are the core of the stack of Semantic Web technologies[35] [36] (simply called semantic technologies) and the W3C's vision of Linked Data[37].

Ontologies enable system designers to define the terminology used to represent and share data within a problem domain. As long as the software systems define their data with reference to the same ontology, they can interpret and reason each other's data and collaborate without manually defining any mapping between the systems.

Metadata vocabularies and ontologies are considered as tools for providing semantic context in determining the relevance of resources. The choosing and sharing of appropriate vocabulary elements consistently across information systems is a good basis for semantic interoperability.

According to the level of their generality/specificity, ontologies are classified into three main categories:

- **Upper ontologies** – define basic, domain-independent concepts as well as relationships between them
- **Core ontologies** – define concepts and relationships that are basic in the broad application domain context. They usually capture the semantics of well-accepted domain standards
- **Domain ontologies** – define concepts and relationships used in specific application domains. The concepts defined in domain ontologies often specialise the ones defined in both upper and core ontologies

In general, semantic interoperability depends mainly on the existence of well-defined and widely accepted upper and core ontologies, in which the basic concepts and relationships are defined. Then, the concepts defined in the upper and core ontologies, should be extended by appropriate domain ontologies.

Even when two ontologies describe one and the same domain of interest, they may be based on minor differences such as naming conventions or structures or the ways in which they represent the semantics of information. Therefore, usually a kind of mediation is required between different ontologies in the same domain. **Ontology mediation** is "the process of reconciling differences between heterogeneous ontologies in order to achieve inter-operation between data sources annotated with and applications using these ontologies"[38]. This includes the discovery and specification of **ontology mappings**, as well as the use of these mappings for certain tasks, such as query rewriting and instance transformation.

[35] Berners-Lee, T., J. Hendler, O. Lassila. The Semantic Web. Scientific American, No. 5 (2001), pp. 34-43. https://www.jstor.org/stable/26059207 (accessed 12.03.2022).

[36] MIT-W3C. MIT-W3C DAML program: Final Report. DAML 00-2-0593 Q4CY2005, 2005. https://www.w3.org/2005/12/31-daml-final.html#References (accessed 12.03.2022).

[37] Berners-Lee, T. Linked Data - Design Issues. W3C, 2006. http://www.w3.org/DesignIssues/LinkedData.html (accessed 12.03.2022).

[38] Davies, J., R. Studer, P. Warren (Eds.). Semantic Web Technologies. John Wiley & Sons, 2006.

3.3.4 Semantic annotation and semantic enrichment of data

One of the useful solutions for providing interoperability of software systems is to supply the corresponding datasets and/or individual pieces of data with appropriate semantic annotations. An **annotation** is a form of metadata attached to a particular section of document content or a particular piece of data. **Semantic annotation** is defined as "the action and results of describing (part of) an electronic resource by means of metadata whose meaning is formally specified in an ontology"[39].

A semantic annotation is referent to an ontology. It enriches data with a context that is further linked to the available knowledge about the application domain. Semantic annotation helps to bridge the ambiguity of the natural language when expressing notions and their computational representation in a formal language. By telling a computer how data items are related and how these relations can be evaluated automatically, it becomes possible to process complex filter and search operations and to obtain results that are not explicitly related to the original search queries.

The ontology to support semantic annotation in a web context should address a number of general classes of entities, which use to appear in texts in various domains. Describing these classes together with the most basic relations and attributes means that an upper ontology should be involved. More precisely, a **light-weight upper ontology** (an upper ontology which is poor on axioms and makes no use of complicated logical operators) is what semantic annotations need as a basis.

Semantic annotation is the core of a more general concept called ***semantic enhancement*** or **semantic enrichment** of data that became particularly popular with the entry of so-called big data into research and practice. The term "big data" has emerged to describe this data that is too huge, too unstructured and coming at too great velocity so that it cannot be handled by traditional data management systems. The widespread use of big data has led to a number of new challenges to the methods and tools for data utilization, as for example:

- The **volume of data** to be processed requires an ability to abstract the data in a form that summarizes the situation and is actionable by humans and decision-making software systems
- The **velocity**, i.e. the rapid appearance and change of data, requires the ability to focus on the relevant data and to process it quickly
- Data **veracity** requires the capability of finding anomalies in it and making some types of reasoning based on proper domain knowledge
- Extracting value using data analytics methods on various kinds of data creates the need of ability to extract knowledge from data and integrate it with existing knowledge bases

This kind of challenges need new, flexible approaches to provide a kind of semantic enhancement of data that can be realized with the help of proper ontologies used to incrementally annotate data. Ontologies provide semantic enhancement of data, defining controlled vocabularies for annotations and thus supporting semantic data integration. This type of semantic enhancement of data may be characterized as an" arm extension approach" – it presumes no change of data but "association of each database field with an entire knowledge base". Data should be left as it is but incrementally tagged with terms from a consistent set of ontologies.

[39] Sicilia, M. Metadata, semantics, and ontology: Providing meaning to information resources. International Journal of Metadata Semantics and Ontologies, Vol. 1 (2006), pp. 83-86.

The successful implementation of this approach needs the creation of a shared resource of ontologies to be used for annotation purposes. An effective methodology for dynamic creation, application and extension of ontologies to annotate new sources of streaming data[40] should be necessary as well.

3.4 Examples of good practices and free tools

3.4.1 Apache Stanbol

Apache Stanbol is a cross-domain enhancement framework that combines multiple enhancement engines and domain knowledge resources into pipelines for data (more precisely, text) enhancement.

Stanbol is available as an open source platform intended to extend traditional content management systems with flexible semantic services. It has a pluggable architecture that supports the integration of various text enhancement engines. It provides rich options for integration of various ontologies, which makes it useful in any domain that needs text processing.

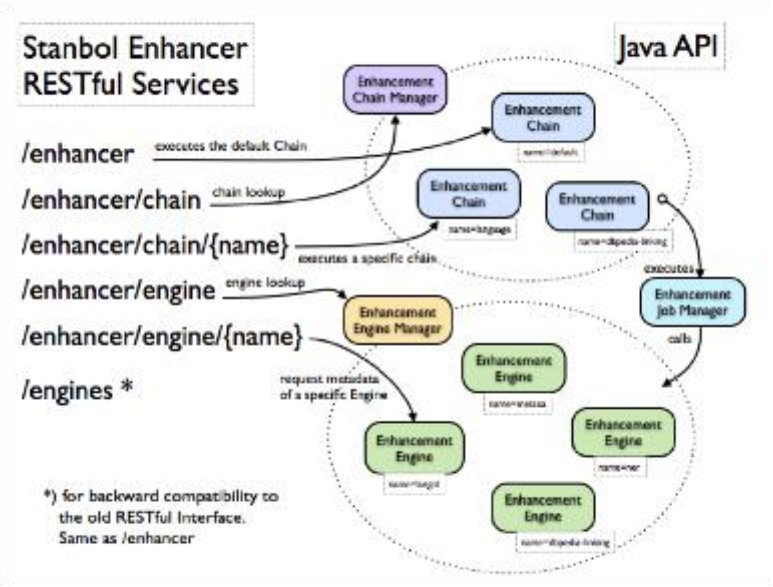


Figure 2: RESTful and Java API provided by the Stanbol Enhancer[41]

Stanbol supports usage with fixed, but configurable chains of enhancement engines to process text data. It provides both a RESTful and a Java API that allows the caller to extract features from the passed text. In more detail the passed text is processed by the available enhancement engines (see Figure 2).

3.4.2 FREME

FREME is a framework for multilingual and semantic enrichment of digital content. It is developed as an open source project that is publicly available at GitHub. It offers a set of reusable services that can be chained to create enhancement pipelines from different components.

[40] Radenski, A. et al. Big Data Techniques, Systems, Applications, and Platforms: Case Studies from Academia. In: Proceedings of the Federated Conference on Computer Science and Information Systems 2016. IEEE, 2016, pp. 883-888.

[41] The Apache Software Foundation. Apache Stanbol – a set of reusable components for semantic content management. https://stanbol.apache.org/ (accessed 12.03.2022).

The FREME project is aimed to provide a set of interfaces for enrichment of digital content (text, video, audio, images, etc.) stored in various formats. It is oriented to a several use cases, particularly[42]:

- Authoring and publishing semantically enriched eBooks
- Enhancing the cross-language sharing and access to open data in specific domains
- FREME-empowered personalized content recommendations

The FRAME framework supports enrichment workflows in various formats. It is oriented to mostly to handling data, oriented to linguistic and natural language processing. Since a large amount language resources are published as a part of the linguistic linked open data (LLOD) cloud, FREME allows processing data available in this cloud as part of its content enrichment workflows. In particular, FREME enrichment workflows support the use of the following formats:

- The Natural Language Processing Interchange Format (NIF)[43] to represent data and enrichment information
- The Internationalization Tag Set (ITS)[44] to represent metadata for improvement of enrichment workflows
- The OntoLex Lemon mode[45] to represent lexicons, including the meaning of terms with respect to ontologies

[42] Sasaki, F., M. Dojchinovski, J. Nehring. Chainable and Extendable Knowledge Integration Web Services. In: Knowledge Graphs and Language Technology, Springer, 2017, pp. 89-101.

[43] Hellmann,S. et al. Integrating NLP using Linked Data. In: International Semantic Web Conference (2013), Springer, 2013, pp. 98-113.

[44] Filip, D. et al. Internationalization Tag Set (ITS) Version 2.0: W3C Recommendation 29 October 2013. https://www.w3.org/TR/its20/ (accessed 12.03.2022).

[45] Cimano, P. et al. Lexicon Model for Ontologies: W3C Community Group Final Report 10 May 2016. https://www.w3.org/2016/05/ontolex/ (accessed 12.03.2022).

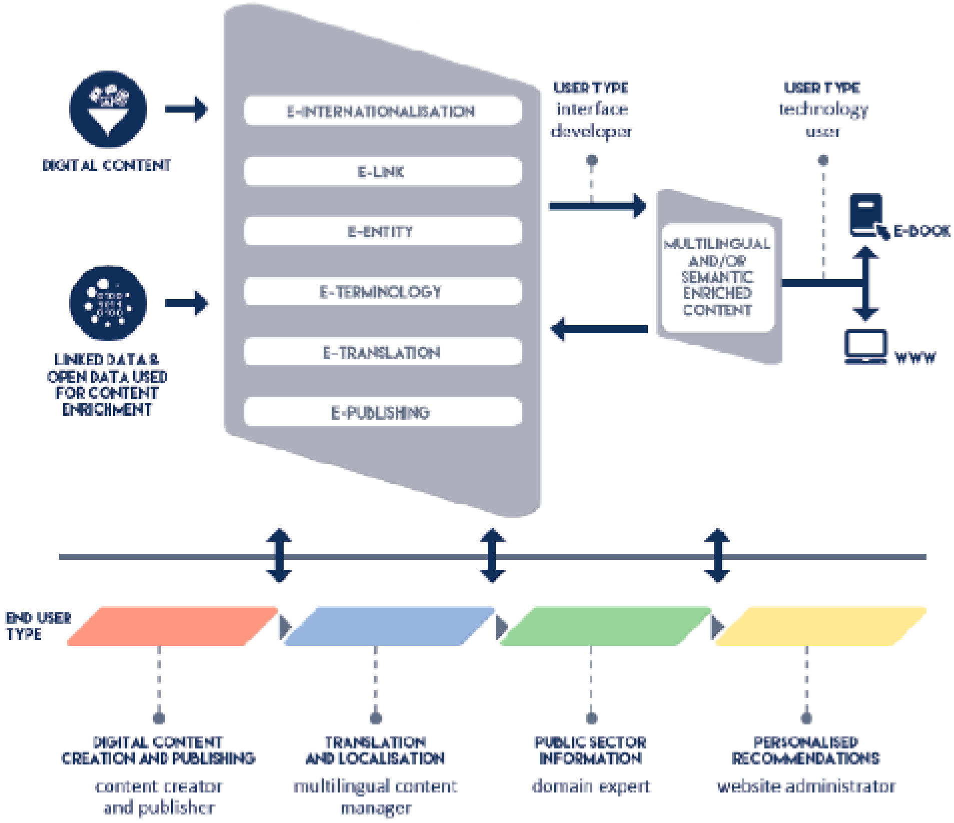

DIGITAL CONTENT
LINKED DATA & OPEN DATA USED FOR CONTENT ENRICHMENT
E-INTERNATIONALISATION
E-LINK
E-ENTITY
E-TERMINOLOGY
E-TRANSLATION
E-PUBLISHING
USER TYPE interface developer
USER TYPE technology user
MULTILINGUAL AND/OR SEMANTIC ENRICHED CONTENT
E-BOOK
WWW
END USER TYPE
DIGITAL CONTENT CREATION AND PUBLISHING
content creator and publisher
TRANSLATION AND LOCALISATION
multilingual content manager
PUBLIC SECTOR INFORMATION
domain expert
PERSONALISED RECOMMENDATIONS
website administrator

Figure 3: Basic structure of the FRAME project[46]

Figure 3 illustrates the basic structure of the FRAME project. The enrichment information is stored within its workflows in two ways[47]: as NIF-encoded information stored in a stand-off manner and inside the enriched content itself.

[46] Sasaki, F., et al. Introducing FREME: Deploying Linguistic Linked Data. CEUR Workshop Proceedings, Vol. 1532 (2015), pp. 59-66.

[47] Sasaki, F., M. Dojchinovski, J. Nehring. Chainable and Extendable Knowledge Integration Web Services. In: Knowledge Graphs and Language Technology, Springer, 2017, pp. 89-101.

4 Cybersecurity

4.1 Introduction

4.1.1 Motivation for cybersecurity violation

Reasons for the emergence of cybersecurity threats are the following:

- Curiosity – script-kiddies, hackers
- Revenge – insiders, using internal information
- Monetary gain – criminals, organized, using complex tools
- Espionage – competitors
- Political activism – hacktivists
- National security – government supported organizations

4.1.2 Factors complicating cybersecurity

The main factors complicating the cybersecurity situation are as follows:

- Data explosion – Big Data era (volume, variety, velocity, and veracity)
- Consumerization of information technologies – the line between personal & working time, devices and data is blurring
- Everything is everywhere – clouds, virtualisation, mobile devices etc.
- Attack sophistication – the mechanics of many attacks with high impact continue to be unresolved

The cybersecurity of IT issues is now a constant management concern. Database servers for financial services are the main targets. Cybersecurity planning and response are complicated by the following:

- Sophisticated attacks – multiple attack vectors, stored zero days etc.
- Cloud computing – public and private clouds, identity and access management, security intelligence, application scanning, end point management, protection for virtualised servers, network intrusion prevention system, data protection
- Mobile computing – IT resources are not only behind the firewalls, in an adverse environment by default; security at the device (enrol, configure, monitor, reconfigure, de-provision), over the network and enterprise (authenticate, encrypt, monitor, control, block), for the mobile applications (develop, test, monitor, protect, update)
- Streaming data, social networking, point solutions – applications and data are distributed on multiple virtual and physical devices
- Regulations and compliance – GDPR, data anonymization

The Security issue is the following complex four-dimensional problem:

- People – employees, consultants, hackers, terrorists, outsourcers, customers, suppliers

- Data – structured, unstructured, semi structured, at rest, in motion
- Applications – system software, web applications, user applications, mobile applications
- Infrastructure – physical and virtual servers, clouds, grids, desktops, laptops, mobile devices, networks

Security is changing as follows:

- Basic, perimeter protection, manual, reactive
- Proficient, on the business operations, automated, proactive
- Optimized, based on security intelligence, automated, proactive

There are many cybersecurity standards, specifications, recommendations etc. with managerial, technical and industry focus. The reasonable question in that situation is "How to start with cybersecurity?"

Our recommendation is to start with the Cybersecurity Framework developed by NIST[48], as described below.

4.2 Cybersecurity Framework

Originally, the Framework has been developed to increase cybersecurity in critical infrastructures like power plants, dams, water supply systems etc. Nevertheless, it happens to be so universal that it can be adapted for different organizations by size and industries. The Framework consists of three main components: **core**, **profiles** and **implementation tiers**.

The core organizes desired cybersecurity outcomes in hierarchy of detailed guidance and controls. The profiles leverage the organization's cybersecurity requirements to the desired Framework outcomes. Profiles can be used to identify opportunities for improving cybersecurity posture by comparing a "Current" profile with a "Target" profile. They are defined in relation to the core outcomes. Implementation tiers are tools for qualitative cybersecurity measurement of risk management practices in the organization.

The key characteristics of the Framework are:

- Common and accessible language, understandable by everyone
- Adaptable to many technologies, lifecycle phases, sectors and uses
- Risk-based, applies to any type of risk management
- Based on international standards, describe desired outcomes
- Living document, defines the entire breadth of cybersecurity
- Guided by many perspectives – private sector, academia, public sector
- Spans both prevention and reaction

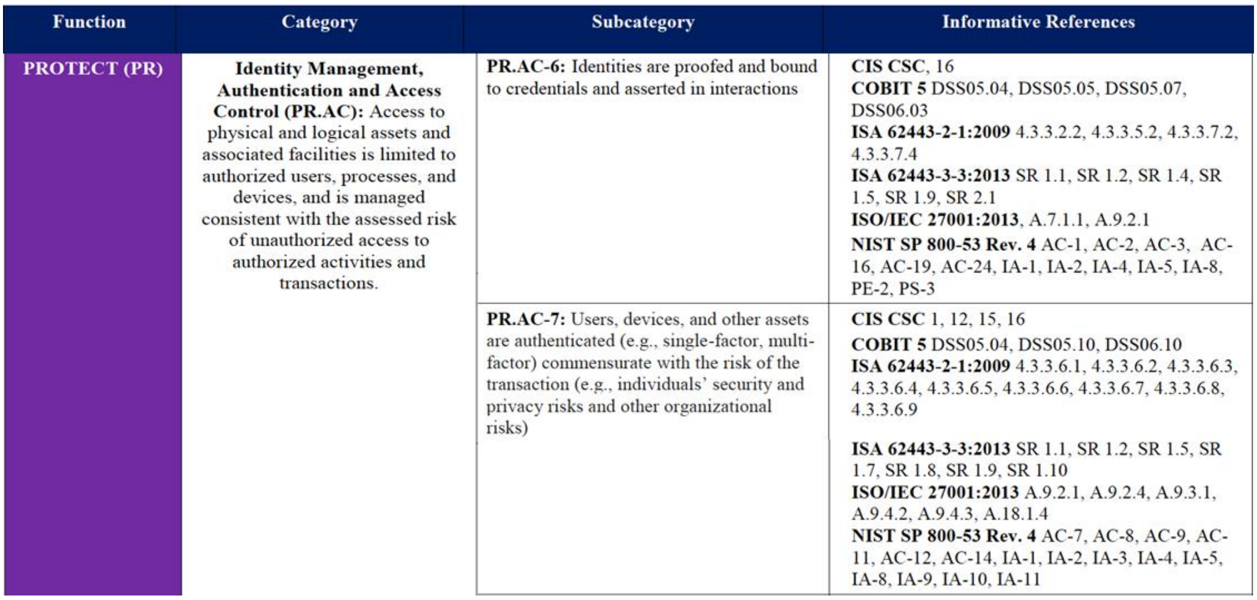
[48] National Institute of Standards and Technology, The Cybersecurity Framework Version 1.1, October 2019, https://www.nist.gov/cyberframework.

The **core** consists of five cybersecurity functions: identify, protect, detect, respond, and recover. In other words, these functions (activities) must be established to support cybersecurity. The functions are detailed in 23 categories, which are further detailed in 108 subcategories. Categories and subcategories are outcomes – guidance and controls supporting the corresponding functions.

Each subcategory is presented via a set of informative references, which refer to standards and specifications applicable to the cybersecurity of the organization under consideration. In the Framework, a recommendation set of informative references is given to each subcategory such as CIS CSC, ISA 62443, COBIT 5, ISO/IEC 27000 series, NIST SP 800 series etc. This set of informative references can be used as a starting point for Framework adaptation – these are the common applicable cybersecurity standards and specification.

The Framework is adaptable to different organizations via these informative references. Functions, categories and subcategories are the same for all organizations, but informative references are different – they depend on the organization industry, size, environment etc. The informative references are specific for each organization. It is possible that some subcategories are not applicable for some organizations.

In Figure 4, an extract from the Framework is given: function – Protect, category – Identity Management, Authentication and Access Control, two subcategories and their informative references.



Function	Category	Subcategory	Informative References
PROTECT (PR)	**Identity Management, Authentication and Access Control (PR.AC):** Access to physical and logical assets and associated facilities is limited to authorized users, processes, and devices, and is managed consistent with the assessed risk of unauthorized access to authorized activities and transactions.	**PR.AC-6:** Identities are proofed and bound to credentials and asserted in interactions	**CIS CSC**, 16 **COBIT 5** DSS05.04, DSS05.05, DSS05.07, DSS06.03 **ISA 62443-2-1:2009** 4.3.3.2.2, 4.3.3.5.2, 4.3.3.7.2, 4.3.3.7.4 **ISA 62443-3-3:2013** SR 1.1, SR 1.2, SR 1.4, SR 1.5, SR 1.9, SR 2.1 **ISO/IEC 27001:2013**, A.7.1.1, A.9.2.1 **NIST SP 800-53 Rev. 4** AC-1, AC-2, AC-3, AC-16, AC-19, AC-24, IA-1, IA-2, IA-4, IA-5, IA-8, PE-2, PS-3
		PR.AC-7: Users, devices, and other assets are authenticated (e.g., single-factor, multi-factor) commensurate with the risk of the transaction (e.g., individuals' security and privacy risks and other organizational risks)	**CIS CSC** 1, 12, 15, 16 **COBIT 5** DSS05.04, DSS05.10, DSS06.10 **ISA 62443-2-1:2009** 4.3.3.6.1, 4.3.3.6.2, 4.3.3.6.3, 4.3.3.6.4, 4.3.3.6.5, 4.3.3.6.6, 4.3.3.6.7, 4.3.3.6.8, 4.3.3.6.9 **ISA 62443-3-3:2013** SR 1.1, SR 1.2, SR 1.5, SR 1.7, SR 1.8, SR 1.9, SR 1.10 **ISO/IEC 27001:2013** A.9.2.1, A.9.2.4, A.9.3.1, A.9.4.2, A.9.4.3, A.18.1.4 **NIST SP 800-53 Rev. 4** AC-7, AC-8, AC-9, AC-11, AC-12, AC-14, IA-1, IA-2, IA-3, IA-4, IA-5, IA-8, IA-9, IA-10, IA-11

Figure 4: Function, category, subcategories and Informative references

There are four **implementation tiers**: partial, risk informed, repeatable, and adaptive. They are described in terms of Risk Management Process, Integrated Risk Management Program, and External Participation. The implementation tiers define how the organization addresses the risks. Each organization must set its desired implementation tier. It is not necessary for all organizations to target adaptive implementation but it is preferable. A higher implementation tier requires higher efforts and investments. Finally, the implementation tier is based on the organization's risk assessment.

4.3 Security Policies

The implementation of the Framework starts with the establishment of policies set by the senior management. The policy influences all decisions in the organization. The activities go around this pillar. SANS Institute has developed 64 policy templates supporting the Framework[49]. They are classified in six categories:

- Application Security
- General
- Server security
- Network security
- Incident handling
- Retired

Every template has: Name, Free Use Disclaimer, Things to Consider, Last Update Status, and is structured in Overview, Purpose, Scope, Policy, Policy Compliance, Related Standards, Policies and Processes, Definitions and Terms, and Revision History.

An example of such a policy template is given in Appendix 6.

SANS site contains useful information focused on the areas:

- Cyber Defence;
- Cloud Security;
- Cyber Security Leadership;
- Digital Forensics;
- Industrial Control Systems;
- Offensive Operations.

SANS Institute offers on cybersecurity:

- Courses;
- Certifications;
- Degree Programs;
- Cyber Ranges.

SANS offers more than 150 free cybersecurity tools.

For more information visit https://www.sans.org.

[49] They are freely available at https://www.sans.org/information-security-policy.

4.4 ISO/IEC 27000 Series

ISO/IEC 27000 series of standards ("ISMS Family of Standards" or "'ISO27K") are devoted to Information Security Management System (ISMS). The full list of ISO/IEC 27000 series of standards is given in Appendix 1. It is process oriented standard at managerial level for developing, supporting and updating ISMS – the core process for supporting cybersecurity in organizations of all types and sizes. The life cycle of ISMS as a business process is given in Figure 5. This family of standards is under permanent update and development. It is a collection of best practices in information security.

ISO/IEC 27000 series of standards are managerial standards. Cybersecurity (and in in particular information security) is not a technical issue – it is a complex issue and therefore managerial issue. These standards are defined in managerial terms.

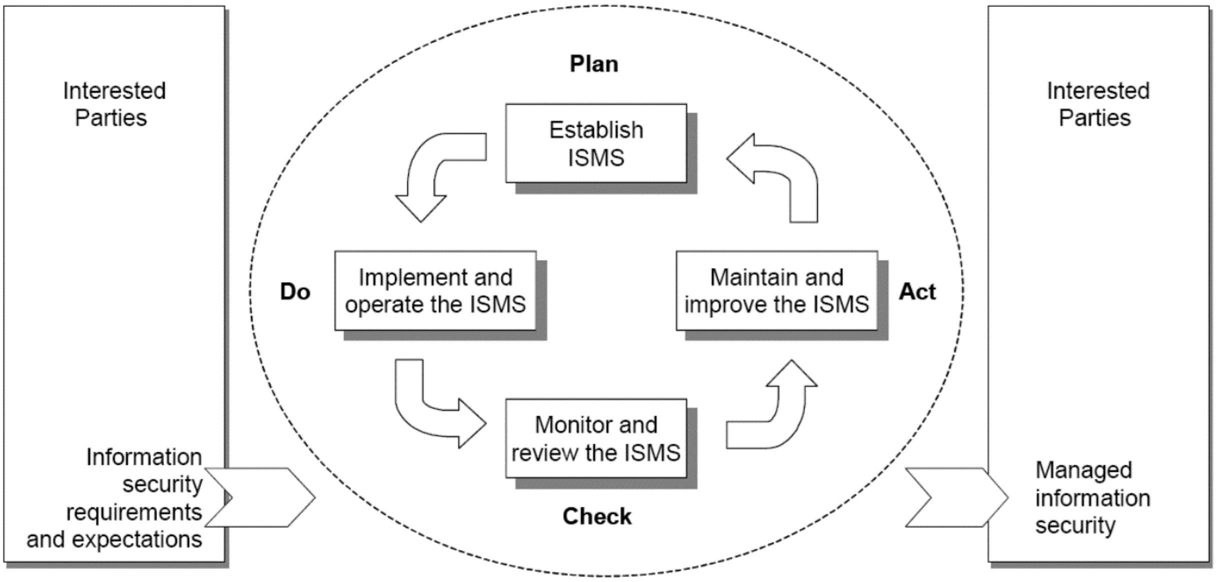

Interested Parties
Information security requirements and expectations
Plan
Establish ISMS
Do
Implement and operate the ISMS
Maintain and improve the ISMS
Act
Monitor and review the ISMS
Check
Interested Parties
Managed information security

Figure 5: ISMS Life Cycle by ISO 27001

Organizations can certify their ISMS on ISO/IEC 27001. This standard has the following clauses:

- Clause 4 Context of the organization
- Clause 5 Leadership
- Clause 6 Planning
- Clause 7 Support
- Clause 8 Operation
- Clause 9 Performance evaluation
- Clause 10 Improvement

Clause 4 requires the organization to determine its external and internal security issues. This process must account the needs and requirements of all interested parties. Based on the collected assessment, the ISMS scope has to be defined.

Clause 5 defines the top management leadership and commitment in respect to ISMS. In this activity, the top management must establish a policy on information security. Following this policy, organization roles, responsibilities and authorities relevant to ISMS must be established and communicated.

Clause 6 requires actions to address the risks and opportunities based on an information security risk assessment and treatment, and a definition of information security objectives as well as a planning to achieve them.

Clause 7 is on the support of ISMS, which requires:

- Resources
- Competence (personal)
- Awareness
- Communication
- Documented information (crating and updating; control of documented information)

Clause 8 relates to the operation of ISMS:

- Operational planning and control
- Information security risk assessment
- Information security risk treatment

Clause 9 is on performance evaluation of ISMS and consists of:

- Monitoring, measurement, analyses and evaluation
- Internal audit
- Management review

Clause 10 is on ISMS improvement, which can be achieved with corrective actions in case of nonconformity and with continual improvement.

ISO/IEC 27001 defines requirements for ISMS in terms of 14 sets of controls. Controls are cybersecurity checkpoints implemented as organizational or technical solutions. These sets are as follows:

- Information security policies (2 controls)
- Organization of information security (7 controls)
- Human resource security (6 controls)
- Asset management (10 controls)
- Access control (14 controls)
- Cryptography (2 controls)
- Physical and environmental security (15 controls)
- Operations security (14 controls)

- Communications security (7 controls)
- System acquisition, development and maintenance (13 controls)
- Supplier relationships (5 controls)
- Information security incident management (7 controls)
- Information security aspects of business continuity management (4 controls)
- Compliance (8 controls)

ISO/IEC 27002 is a description of information security controls, including those mentioned in the ISO/IEC 27001 Appendix A. ISO/IEC 27006, ISO/IEC 27007, and ISO/IEC 27008 define the audit and certification processes. The other standards are informative ones – guidelines and recommendations for establishing and maintenance of ISMS. They are focused on different information security activities and areas. For more details, check the above listed standards. They are not freely distributed.

4.5 Databases of Vulnerabilities, Weaknesses and Attacks

4.5.1 CVE (Common Vulnerabilities and Exposures)

CVE is a current list of vulnerabilities and exposures sponsored by the U.S. Department of Homeland Security (DHS), Cybersecurity and Infrastructure Security Agency (CISA) and The MITRE Corporation (MITRE)[50]. The list can be used by cybersecurity experts to investigate the hardware and software products available in the organization.

The list is managed by the CVE program partners with community members worldwide. Its aim is to extend the CVE content and expand its usage. The CVE program partners are CVE Numbering Authorities (CNAs) and Roots. They represent the world community interested in cybersecurity.

The partners can register new vulnerabilities and exposures, the latter being investigated by the community and finally accepted or rejected by voting. The procedure is defined at the site. This means that the site contains not only accepted vulnerabilities but also ones that are under investigation. An example of a vulnerability description from the new site is given in Appendix 2.

Every vulnerability has a unique identifier (CVE-ID) in the database with a short description and source references. All other information is administrative.

4.5.2 NVD (National Vulnerability Database)

NVD is an extraction from CVE supported by NIST[51]. However, NVD is not simply an extraction of accepted vulnerabilities from CVE – it contains additional information elaborated by NIST. An example of CVE in NVD is given in Appendix 3.

CVE in NVD contains:

- **CVE-ID**, same as in CVE list
- **Current Description**

[50] This database is currently available at https://cve.mitre.org. In 2022 the site can found at address https://www.cve.org.

[51] National Institute for Standards and Technology of United States, available at address https://nvd.nist.gov.

- **Analyses Description**
- **Severity** is calculated using the Common Vulnerability Scoring System (CVSS) Version 3.x and CVSS Version 2.0 developed by First SIG. Detailed information about the last organization and particularly on CVSS[52]. NIST is using only publicly available information to calculate the CVSS score. CVSS vector contains details on the score in a coded format
- **References to Advisories, Solutions, and Tools** is an extended version of references in CVE.
- **Weakness Enumeration** contains references to weaknesses – the types of vulnerabilities of which that vulnerability belongs.
- **Known Affected Software Configurations** – every vulnerability is linked to concrete versions of software/hardware in fixed environment configurations. These vulnerable configurations are listed in the section using CPE 2.3 and CPE 2.2. CPE is part of the NVD and is available at the NVD site. The CPE database is maintained by NIST.
- **Change History**

For more details, please refer to above-mentioned site of NVD.

4.5.3 CWE (Common Weakness Enumeration)

CWE is a community-developed list of software and hardware weakness types. It is sponsored by the U.S. Department of Homeland Security, Cybersecurity and Infrastructure Security Agency and managed by the Homeland Security Systems Engineering and Development Institute, which is operated by The MITRE Corporation[53].

The difference between a weakness and a vulnerability is that the vulnerability is a weakness that can be exploited by an attack. Some weaknesses exist in the software/hardware but the attack vector is not available to them, therefore they are not vulnerabilities. CWEs are types of exploitable weaknesses.

CWEs are organized in views, i.e. different perspectives. There are three main views: software development, hardware design, and research concepts. There are more views than the main ones.

Views can be structured in categories, pillars, classes, bases and variants. Following are some definitions of these terms from the CWE website[54]:

- "Category – a CWE entry that contains a set of other entries that share a common characteristic."
- "Pillar – a weakness that is the most abstract type of weakness and represents a theme for all class/base/variant weaknesses related to it. A pillar is different from a category as a pillar is still technically a type of weakness that describes a mistake, while a category represents a common characteristic used to group related things."
- "Class – a weakness that is described in a very abstract fashion, typically independent of any specific language or technology. More specific than a pillar weakness, but more general than

[52] CVSS is available at address https://www.first.org.

[53] CWE is freely available at address https://cwe.mitre.org.

[54] MITRE Corporation, CWE, CWE Glossary, https://cwe.mitre.org/documents/glossary/index.html

a base weakness. Class level weaknesses typically describe issues in terms of 1 or 2 of the following dimensions: behaviour, property, and resource."

- "Base – a weakness that is still mostly independent of a resource or technology, but with sufficient details to provide specific methods for detection and prevention. Base level weaknesses typically describe issues in terms of 2 or 3 of the following dimensions: behaviour, property, technology, language, and resource."
- "Variant – a weakness that is linked to a certain type of product, typically involving a specific language or technology. More specific than a base weakness. Variant level weaknesses typically describe issues in terms of 3 to 5 of the following dimensions: behaviour, property, technology, language, and resource."

An example of a weakness is given in Appendix 4.

CWEs have:

- CWE ID
- Abstraction – view, category, pillar, class, base and variant
- Structure – simple or compound, the latter being rare and can be either chain or composite. Chains are ordered.
- Description
- Extended description
- Relationships – position in the views
- Modes of introduction – in the phases of the life cycle
- Applicable platforms – languages and technologies
- Common consequences – scope and impact
- Demonstrative examples – code
- Observed examples – references to CVEs of that weakness type
- Potential mitigations – very important part organized by the life cycle phases
- Memberships – in CWE categories
- References – external references
- Content history – structured administrative information

For more details, please refer to the CWE website.

4.5.4 CAPEC (Common Attack Pattern Enumeration and Classification)

CAPEC is sponsored by the U.S. Department of Homeland Security, Cybersecurity and Infrastructure Security Agency and managed by the Homeland Security Systems Engineering and Development Institute, which is operated by The MITRE Corporation[55].

[55] CAPEC is freely available at address http://capec.mitre.org.

Attacks can be structured in views, categories, meta attacks, standard attack patterns, and detailed attack patterns. They are defined in CAPEC[56] as:

- "A view in CAPEC represents a perspective with which one might look at the collection of attack patterns defined within CAPEC. There are three different types of views: graphs, explicit slices, and implicit slices."
- "A category in CAPEC is a collection of attack patterns based on some common characteristic. More specifically, it is an aggregation of attack patterns based on effect/intent (as opposed to actions or mechanisms, such an aggregation would be a meta attack pattern). An aggregation based on effect/intent is not an actionable attack and as such is not a pattern of attack behaviour. Rather, it is a grouping of patterns based on some common criteria."
- "A meta level attack pattern in CAPEC is a decidedly abstract characterization of a specific methodology or technique used in an attack. A meta attack pattern is often void of a specific technology or implementation and is meant to provide an understanding of a high-level approach. A meta level attack pattern is a generalization of related group of standard level attack patterns. Meta level attack patterns are particularly useful for architecture and design level threat modelling exercises."
- "A standard level attack pattern in CAPEC is focused on a specific methodology or technique used in an attack. It is often seen as a singular piece of a fully executed attack. A standard attack pattern is meant to provide sufficient details to understand the specific technique and how it attempts to accomplish a desired goal. A standard level attack pattern is a specific type of a more abstract meta level attack pattern."
- "A detailed level attack pattern in CAPEC provides a low level of detail, typically leveraging a specific technique and targeting a specific technology, and expresses a complete execution flow. Detailed attack patterns are more specific than meta attack patterns and standard attack patterns and often require a specific protection mechanism to mitigate actual attacks. A detailed level attack pattern often will leverage a number of different standard level attack patterns chained together to accomplish a goal."

There are two main views: by mechanisms of attack, and by domains of attack. There are other views in CAPEC. An example of attack pattern is given in Appendix 5. It is structured in:

- Name
- Attack pattern ID
- Description
- Typical severity
- Relationships – position in views
- Prerequisites – requirements for successful attack occurrence
- Resources required – required resources for successful attack maintenance
- Related weaknesses – references to CWEs for attack success

[56] MITRE Corporation, CAPEC, Glossary, http://capec.mitre.org/about/glossary.html

- Content history – administrative information

For more details, please refer to above-mentioned website.

4.6 OWASP (Open Web Application Security Project)

OWASP is a non-profit foundation that works to improve the security of software[57]. The OWASP foundation is the source for developers and technologists to secure the web:

- Open source tools and free resources
- Community and networking via hundred worldwide chapters
- Educational and training conferences

The most popular OWASP project is OWASP Top 10. It is an awareness bulletin about the most critical security risks to web applications. Target audience of the bulletin are web application developers and security experts. OWASP Top 10 is regularly supported in CWE as a view. For example, the last view is CWE-1344: Weaknesses in OWASP Top 10 (2021).

OWASP is strictly scoped on the web applications. That is why CWE supports a similar bulletin, but for all weaknesses. The last CWE bulletin is 2021 CWE Top 25 Most Dangerous Software Weaknesses[58]. Some of the other well-known OWASP projects are presented below.

OWASP Dependency-Track is an intelligent component analysis platform that allows organizations to identify and reduce risks in the software supply chain. The system uses a unique approach by leveraging the capabilities of continuous software bill of materials. The system is distributed as Docker containers. OWASP Dependency-Track is integrated with multiple sources of vulnerability intelligence including National Vulnerability Database, GitHub Advisories, Sonatype OSS Index, and VulnDB from Risk Based Security[59].

OWASP Juice Shop is a sophisticated insecure web application. It is used in computer security trainings, awareness demos, computer security and as investigation subject for security tools. Juice Shop encompasses vulnerabilities from the entire OWASP Top 10 along with many other security flaws found in real-world applications. This web application supports a learning process based on the gamification. You can install a free copy of OWASP Juice Shop or you can access some free pre-installed copies on the web[60].

OWASP Mobile Security Testing Guide[61] is a manual for mobile app security testing and reverse engineering for iOS and Android. It is a standard scope:

- Mobile platform
- Security testing in the mobile app development lifecycle
- Basic static and dynamic security testing

[57] OWASP address is https://owasp.org.

[58] 2021 CWE Top 25 Most Dangerous Software Weaknesses is available at address https://cwe.mitre.org/top25/archive/2021/2021_cwe_top25.html.

[59] For more information on OWASP Dependency-Track, refer to https://dependencytrack.org.

[60] As example access OWASP Juice Shop, https://juice-shop.herokuapp.com

[61] The guide is available in several translations and formats athttps://github.com/OWASP/owasp-masvs/releases.

- Mobile app reverse engineering and tampering
- Assessing software protections
- Detailed test cases that meet the requirements in the Mobile Application Security Verification Standard

OWASP ModSecurity Core Rule Set is a set of generic attack detection rules for use with ModSecurity or compatible web application firewalls[62]. ModSecurity will be supported until July 1, 2024. The maintenance of the ModSecurity code is given to the open-source community[63]. ModSecurity 2.x works with Apache 2.0.x or higher, NGINX, and Microsoft IIS. The core rule set provides protection against many common attack categories, including:

- SQL Injection (SQLi)
- Cross Site Scripting (XSS)
- Local File Inclusion (LFI)
- Remote File Inclusion (RFI)
- PHP Code Injection
- Java Code Injection
- HTTPoxy
- Shellshock
- Unix/Windows Shell Injection
- Session Fixation
- Scripting/Scanner/Bot Detection
- Metadata/Error Leakages

The project includes tutorials on:

- Installing ModSecurity
- Including the OWASP ModSecurity Core Rule Set
- Handling False Positives with the OWASP ModSecurity Core Rule Set

ModSecurity Handbook, 2nd edition is available but not free.

OWASP Software Assurance Maturity Model is an effective and measurable way to analyse and improve secure development lifecycles[64]. It is technology and process agnostic.

[62] The official site of the project OWASP ModSecurity Core Rule Set is https://coreruleset.org.

[63] ModSecurity download address is https://github.com/SpiderLabs/ModSecurity.

[64] OWASP Software Assurance Maturity Model site is https://owaspsamm.org/model.

OWASP Security Knowledge Framework is an open source web application that explains secure coding principles in multiple programming languages[65]. It contains manageable software development projects with checklists and labs to practice security verification.

OWASP Web Security Testing Guide is a cybersecurity testing resource for web application developers and security professionals[66].

OWASP Zed Attack Proxy (ZAP) is a Web application scanner[67].

There are many more OWASP projects. For more information, please refer to website.

4.7 Remarks

First, let us clarify the hacker's motivation for attacking HPC/HPDA/AI infrastructure (hereafter simply "infrastructure"). Scientific data is not usually data that the hacker can benefit from, like financial data. In addition, these data are public and freely available. There are some sensitive scientific data like the results of medical research and investigation. These data must be anonymized using procedures that apply to personal data. Anonymization procedures must save the data value for scientific research. This includes the data relationships.

The data must be briefly classified in sensitive and non-sensitive, but it is possible to apply a more detailed classification scheme. The information classification procedure must be established in the organization following above-mentioned recommendations in the standards. Sensitive data must be appropriately protected "in rest" and "in motion". The main hacker motivation to attack the infrastructure is to take control over it for further attacks on infrastructures that contain valuable information for them.

Usually, government-supported organizations have no suitable infrastructures to conduct some kinds of attacks like DDoS etc. In that case, they try to take unsanctioned control on some commercial or academic infrastructure. There are some examples in the past when this happened even when there were contracts for cooperation between the government and the infrastructure owner. The security of the new technologies is usually very weak at best. One such example is the popular tool for Big Data processing Hadoop in its early versions. There are examples when such infrastructures have been used by infiltrated agents – in these cases; the stored data has been processed illegally. Older technologies like grids and clouds are more secure – learning the lessons from the past.

In our case, hackers must have very advanced skills of the new technologies and enough resources to attack these infrastructures. Therefore, they are a small community that can even be calculated. Usually, security breaches in our kind of infrastructures have been conducted by internal personnel, who sometimes happen to be infiltrated agents. The approach to cybersecurity must be systematic; it is not enough just to rely on the technologies. Every organization must have some ISMS as recommended by ISO/IEC 2700K. There is no need for this system to be certified, but it is important that it is continuously improved because of the dynamic nature of cybersecurity threads.

The framework of the ISMS can be based on the NIST Cybersecurity Framework, which can be adapted for different kind of organizations in different industries using the applicable industry

[65] OWASP Security Knowledge Framework site is https://www.securityknowledgeframework.org.

[66] OWASP Web Security Testing Guide is available at https://owasp.org/www-project-web-security-testing-guide.

[67] OWASP Zed Attack Proxy site is https://www.zaproxy.org.

security standards. Overall, the infrastructure can be protected using open source tools and resources from sources like the ones mentioned above.

Cybersecurity encompasses the following areas:

- Critical infrastructure security
- Application security
- Network security
- Cloud security
- Internet of Things (IoT) security

Cybersecurity planning means planning in all applicable areas from the list above. They are specific in threads, defence, detection, recovery and remedy. Cybersecurity domains are:

- Access control
- Telecommunication and network security
- Information security governance and risk management
- Software development security
- Cryptography
- Security architecture and design
- Operations security
- Business continuity and disaster recovery planning
- Legal, regulations, investigations, and compliance
- Physical (environment) security
- Security credentials

Investigating applicable areas means investigating in these cybersecurity domains. Categories of cybersecurity tools are:

- Network security monitoring tools
- Encryption tools
- Web vulnerability scanning tools
- Network defence wireless tools
- Packet sniffers
- Antivirus software
- Firewall
- PKI services
- Managed detection services

- Penetration testing

Appropriate tools have to be used as cybersecurity controls. Cybersecurity defence can be outsourced but it is extensive.

5 Summary

Cybersecurity is a complex task that must be supported at all system and organizational levels with different aspects. The most important topic in cybersecurity is the information security. Special topic of information security is the application of the GDPR.

GDPR sets a number of restrictions and procedures that the controllers should comply with in order to protect the processing of data. A well detailed data protection policy, along with the records of processing, should provide the controller with all the necessary information to build its privacy program. The compliance with GDPR is more than avoiding hefty fines, a well-managed data lifecycle management is the key to a successful provision of products.

Interoperability of data and metadata plays a key role in facilitating effective data exchange between heterogeneous software systems. It provides many advantages of the respective organizations such as:

- Adaptability,
- Better productivity,
- Data unity,
- Improved data protection,
- Lower costs.

The development of the stack of Semantic Web technologies and the continuous growth of the Linked Open Data Cloud have enabled developers towards cross-domain enhancement frameworks, which integrate multiple enhancement engines and domain knowledge resources into pipelines for data enhancement and thus provide opportunities to achieve semantic interoperability of data.

Appendix 1. ISO/IEC 27000 Series of Standards.

1. ISO/IEC 27000 — Information security management systems — Overview and vocabulary
2. ISO/IEC 27001 — Information technology — Security Techniques — Information security management systems — Requirements.
3. ISO/IEC 27002 — Code of practice for information security controls
4. ISO/IEC 27003 — Information security management system implementation guidance
5. ISO/IEC 27004 — Information security management — Monitoring, measurement, analysis and evaluation
6. ISO/IEC 27005 — Information security risk management
7. ISO/IEC 27006 — Requirements for bodies providing audit and certification of information security management systems
8. ISO/IEC 27007 — Guidelines for information security management systems auditing
9. ISO/IEC TR 27008 — Guidance for auditors on ISMS controls
10. ISO/IEC 27009 — Essentially an internal document for the committee developing sector/industry-specific variants or implementation guidelines for the ISO27K standards
11. ISO/IEC 27010 — Information security management for inter-sector and inter-organizational communications
12. ISO/IEC 27011 — Information security management guidelines for telecommunications organizations based on ISO/IEC 27002
13. ISO/IEC 27013 — Guideline on the integrated implementation of ISO/IEC 27001 and ISO/IEC 20000-1
14. ISO/IEC 27014 — Information security governance.
15. ISO/IEC TR 27015 — Information security management guidelines for financial services (withdrawn)
16. ISO/IEC TR 27016 — Information security economics
17. ISO/IEC 27017 — Code of practice for information security controls based on ISO/IEC 27002 for cloud services
18. ISO/IEC 27018 — Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors
19. ISO/IEC 27019 — Information security for process control in the energy industry
20. ISO/IEC 27021 — Competence requirements for information security management systems professionals
21. ISO/IEC TS 27022 — Guidance on information security management system processes (under development)
22. ISO/IEC TR 27023 — Mapping the revised editions of ISO/IEC 27001 and ISO/IEC 27002

23. ISO/IEC 27031 — Guidelines for information and communication technology readiness for business continuity
24. ISO/IEC 27032 — Guideline for cybersecurity
25. ISO/IEC 27033 [es] — IT network security
26. ISO/IEC 27033-1 — Network security – Part 1: Overview and concepts
27. ISO/IEC 27033-2 — Network security – Part 2: Guidelines for the design and implementation of network security
28. ISO/IEC 27033-3 — Network security – Part 3: Reference networking scenarios — Threats, design techniques and control issues
29. ISO/IEC 27033-4 — Network security – Part 4: Securing communications between networks using security gateways
30. ISO/IEC 27033-5 — Network security – Part 5: Securing communications across networks using Virtual Private Networks (VPNs)
31. ISO/IEC 27033-6 — Network security – Part 6: Securing wireless IP network access
32. ISO/IEC 27034-1 — Application security – Part 1: Guideline for application security
33. ISO/IEC 27034-2 — Application security – Part 2: Organization normative framework
34. ISO/IEC 27034-3 — Application security – Part 3: Application security management process
35. ISO/IEC 27034-4 — Application security – Part 4: Validation and verification (under development)
36. ISO/IEC 27034-5 — Application security – Part 5: Protocols and application security controls data structure
37. ISO/IEC 27034-5-1 — Application security — Part 5-1: Protocols and application security controls data structure, XML schemas
38. ISO/IEC 27034-6 — Application security – Part 6: Case studies
39. ISO/IEC 27034-7 — Application security – Part 7: Assurance prediction framework
40. ISO/IEC 27035-1 — Information security incident management – Part 1: Principles of incident management
41. ISO/IEC 27035-2 — Information security incident management – Part 2: Guidelines to plan and prepare for incident response
42. ISO/IEC 27035-3 — Information security incident management – Part 3: Guidelines for ICT incident response operations
43. ISO/IEC 27035-4 — Information security incident management – Part 4: Coordination (under development)
44. ISO/IEC 27036-1 — Information security for supplier relationships – Part 1: Overview and concepts
45. ISO/IEC 27036-2 — Information security for supplier relationships – Part 2: Requirements

46. ISO/IEC 27036-3 — Information security for supplier relationships – Part 3: Guidelines for information and communication technology supply chain security
47. ISO/IEC 27036-4 — Information security for supplier relationships – Part 4: Guidelines for security of cloud services
48. ISO/IEC 27037 — Guidelines for identification, collection, acquisition and preservation of digital evidence
49. ISO/IEC 27038 — Specification for Digital redaction on Digital Documents
50. ISO/IEC 27039 — Intrusion prevention
51. ISO/IEC 27040 — Storage security
52. ISO/IEC 27041 — Investigation assurance
53. ISO/IEC 27042 — Analyzing digital evidence
54. ISO/IEC 27043 — Incident investigation
55. ISO/IEC 27050-1 — Electronic discovery — Part 1: Overview and concepts
56. ISO/IEC 27050-2 — Electronic discovery — Part 2: Guidance for governance and management of electronic discovery
57. ISO/IEC 27050-3 — Electronic discovery — Part 3: Code of practice for electronic discovery
58. ISO/IEC TS 27110 — Information technology, cybersecurity and privacy protection — Cybersecurity framework development guidelines
59. ISO/IEC 27701 — Information technology — Security Techniques — Information security management systems — Privacy Information Management System (PIMS).
60. ISO 27799 — Information security management in health using ISO/IEC 27002

Appendix 2. CVE-2021-42950.

Table 2: CVE-2021-42950 – new vulnerability under investigation

CVE-ID	
CVE-2021-42950	Learn more at National Vulnerability Database (NVD) • CVSS Severity Rating • Fix Information • Vulnerable Software Versions • SCAP Mappings • CPE Information
Description	
Remote Code Execution (RCE) vulnerability exists in Zepl Notebooks all previous versions before October 25 2021. Users can register for an account and are allocated a set number of credits to try the product. Once users authenticate, they can proceed to create a new organization by which additional users can be added for various collaboration abilities, which allows malicious user to create new Zepl Notebooks with various languages, contexts, and deployment scenarios. Upon creating a new notebook with specially crafted malicious code, a user can then launch remote code execution.	
References	
Note: References are provided for the convenience of the reader to help distinguish between vulnerabilities. The list is not intended to be complete.	
• MISC:http://zepl.com • MISC:https://seclists.org/fulldisclosure/2022/Feb/31	
Assigning CNA	
MITRE Corporation	
Date Record Created	
20211025	Disclaimer: The record creation date may reflect when the CVE ID was allocated or reserved, and does not necessarily indicate when this vulnerability was discovered, shared with the affected vendor, publicly disclosed, or updated in CVE.
Phase (Legacy)	
Assigned (20211025)	
Votes (Legacy)	
Comments (Legacy)	
Proposed (Legacy)	
N/A	

Appendix 3. CVE-2020-0601.

Table 3: CVE-2020-0601 in NVD

CVE-2020-0601 Detail

MODIFIED

This vulnerability has been modified since it was last analyzed by the NVD. It is awaiting reanalysis which may result in further changes to the information provided.

Current Description

A spoofing vulnerability exists in the way Windows CryptoAPI (Crypt32.dll) validates Elliptic Curve Cryptography (ECC) certificates.An attacker could exploit the vulnerability by using a spoofed code-signing certificate to sign a malicious executable, making it appear the file was from a trusted, legitimate source, aka 'Windows CryptoAPI Spoofing Vulnerability'.

Analysis Description

A spoofing vulnerability exists in the way Windows CryptoAPI (Crypt32.dll) validates Elliptic Curve Cryptography (ECC) certificates.An attacker could exploit the vulnerability by using a spoofed code-signing certificate to sign a malicious executable, making it appear the file was from a trusted, legitimate source, aka 'Windows CryptoAPI Spoofing Vulnerability'.

Severity

CVSS Version 3.x CVSS Version 2.0

CVSS 3.x Severity and Metrics:

NIST: NVD

Base Score: **8.1 HIGH**

Vector: CVSS:3.1/AV:N/AC:L/PR:N/UI:R/S:U/C:H/I:H/A:N

NVD Analysts use publicly available information to associate vector strings and CVSS scores. We also display any CVSS information provided within the CVE List from the CNA.

Note: NVD Analysts have published a CVSS score for this CVE based on publicly available information at the time of analysis. The CNA has not provided a score within the CVE List.

References to Advisories, Solutions, and Tools

By selecting these links, you will be leaving NIST webspace. We have provided these links to other web sites because they may have information that would be of interest to you. No inferences should be drawn on account of other sites being referenced, or not, from this page. There may be other web sites that are more appropriate for your purpose. NIST does not necessarily endorse the views expressed, or concur with the facts presented on these sites. Further, NIST does not endorse any commercial products that may be mentioned on these sites. Please address comments about this page to nvd@nist.gov.

Hyperlink	Resource
http://packetstormsecurity.com/files/155960/CurveBall-Microsoft-Windows-CryptoAPI-Spoofing-Proof-Of-Concept.html	
http://packetstormsecurity.com/files/155961/CurveBall-Microsoft-Windows-CryptoAPI-Spoofing-Proof-Of-Concept.html	
https://portal.msrc.microsoft.com/en-US/security-guidance/advisory/CVE-2020-0601	Patch Vendor Advisory

Weakness Enumeration

CWE-ID	CWE Name	Source
CWE-295	Improper Certificate Validation	NIST

Known Affected Software Configurations Switch to CPE 2.2

Configuration 1 (hide)

cpe:2.3:o:microsoft:windows_10:-:*:*:*:*:*:*:* Show Matching CPE(s)
cpe:2.3:o:microsoft:windows_10:1607:*:*:*:*:*:*:* Show Matching CPE(s)
cpe:2.3:o:microsoft:windows_10:1709:*:*:*:*:*:*:* Show Matching CPE(s)
cpe:2.3:o:microsoft:windows_10:1803:*:*:*:*:*:*:* Show Matching CPE(s)
cpe:2.3:o:microsoft:windows_10:1809:*:*:*:*:*:*:* Show Matching CPE(s)
cpe:2.3:o:microsoft:windows_10:1903:*:*:*:*:*:*:* Show Matching CPE(s)
cpe:2.3:o:microsoft:windows_10:1909:*:*:*:*:*:*:* Show Matching CPE(s)
cpe:2.3:o:microsoft:windows_server_2016:-:*:*:*:*:*:*:* Show Matching CPE(s)
cpe:2.3:o:microsoft:windows_server_2016:1803:*:*:*:*:*:*:* Show Matching CPE(s)
cpe:2.3:o:microsoft:windows_server_2016:1903:*:*:*:*:*:*:* Show Matching CPE(s)
cpe:2.3:o:microsoft:windows_server_2016:1909:*:*:*:*:*:*:* Show Matching CPE(s)
cpe:2.3:o:microsoft:windows_server_2019:-:*:*:*:*:*:*:* Show Matching CPE(s)

Denotes Vulnerable Software
Are we missing a CPE here? Please let us know.

Change History
2 change records found show changes

Appendix 4. CWE-942.

Table 4: CWE-942 invariant

Weakness ID: 942

Status: Incomplete

Abstraction: Variant
Structure: Simple

Description

The software uses a cross-domain policy file that includes domains that should not be trusted.

Extended Description

A cross-domain policy file ("crossdomain.xml" in Flash and "clientaccesspolicy.xml" in Silverlight) defines a list of domains from which a server is allowed to make cross-domain requests. When making a cross-domain request, the Flash or Silverlight client will first look for the policy file on the target server. If it is found, and the domain hosting the application is explicitly allowed to make requests, the request is made.

Therefore, if a cross-domain policy file includes domains that should not be trusted, such as when using wildcards, then the application could be attacked by these untrusted domains.

An overly permissive policy file allows many of the same attacks seen in Cross-Site Scripting (CWE-79). Once the user has executed a malicious Flash or Silverlight application, they are vulnerable to a variety of attacks. The attacker could transfer private information, such as cookies that may include session information, from the victim's machine to the attacker. The attacker could send malicious requests to a web site on behalf of the victim, which could be especially dangerous to the site if the victim has administrator privileges to manage that site.

In many cases, the attack can be launched without the victim even being aware of it.

Relationships

Relevant to the view "Research Concepts" (CWE-1000)

Nature	Type	ID	Name
ChildOf	B	183	Permissive List of Allowed Inputs
ChildOf	P	284	Improper Access Control
CanPrecede	C	668	Exposure of Resource to Wrong Sphere

Relevant to the view "Architectural Concepts" (CWE-1008)

Modes Of Introduction

Phase	Note
Implementation	
Architecture and Design	COMMISSION: This weakness refers to an incorrect design related to an architectural security tactic.

Applicable Platforms

Languages

Class: Language-Independent *(Undetermined Prevalence)*

Technologies

Class: Web Based *(Undetermined Prevalence)*

Common Consequences

Scope	Impact	Likelihood
Confidentiality Integrity Availability Access Control	**Technical Impact:** *Execute Unauthorized Code or Commands; Bypass Protection Mechanism; Read Application Data; Varies by Context* An attacker may be able to bypass the web browser's same-origin policy. An attacker can exploit the weakness to manipulate or steal cookies, create requests that can be mistaken for those of a valid user, compromise confidential information, or execute malicious code on the end user systems for a variety of nefarious purposes. Other damaging attacks include the disclosure of end user files, installation of Trojan horse programs, redirecting the user to some other page or site, running ActiveX controls (under Microsoft Internet Explorer) from sites that a user perceives as trustworthy, and modifying presentation of content.	

Demonstrative Examples

Example 1

These cross-domain policy files mean to allow Flash and Silverlight applications hosted on other domains to access its data:

Flash crossdomain.xml :

(bad code)

Example Language: **XML**

```
<cross-domain-policy xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation="http://www.adobe.com/xml/schemas/PolicyFile.xsd">
<allow-access-from domain="*.example.com"/>
<allow-access-from domain="*"/>
</cross-domain-policy>
```

Silverlight clientaccesspolicy.xml :

(bad code)

Example Language: **XML**

```
<?xml version="1.0" encoding="utf-8"?>
<access-policy>
<cross-domain-access>
<policy>
<allow-from http-request-headers="SOAPAction">
<domain uri="*"/>
</allow-from>
<grant-to>
<resource path="/" include-subpaths="true"/>
</grant-to>
</policy>
</cross-domain-access>
</access-policy>
```

These entries are far too permissive, allowing any Flash or Silverlight application to send requests. A malicious application hosted on any other web site will be able to send requests on behalf of any user tricked into executing it.

Observed Examples

Reference	Description
CVE-2012-2292	Product has a Silverlight cross-domain policy that does not restrict access to another application, which allows remote attackers to bypass the Same Origin Policy.
CVE-2014-2049	The default Flash Cross Domain policies in a product allows remote attackers to access user files.
CVE-2007-6243	Chain: Adobe Flash Player does not sufficiently restrict the interpretation and usage of cross-domain policy files, which makes it easier for remote attackers to conduct cross-domain and cross-site scripting (XSS) attacks.
CVE-2008-4822	Chain: Adobe Flash Player and earlier does not properly interpret policy files, which allows remote attackers to bypass a non-root domain policy.
CVE-2010-3636	Chain: Adobe Flash Player does not properly handle unspecified encodings during the parsing of a cross-domain policy file, which allows remote web servers to bypass intended access restrictions via unknown vectors.

Potential Mitigations

Phase: Architecture and Design

Strategy: Attack Surface Reduction

Avoid using wildcards in the cross-domain policy file. Any domain matching the wildcard expression will be implicitly trusted, and can perform two-way interaction with the target server.

Phases: Architecture and Design; Operation

Strategy: Environment Hardening

For Flash, modify crossdomain.xml to use meta-policy options such as 'master-only' or 'none' to reduce the possibility of an attacker planting extraneous cross-domain policy files on a server.

Phases: Architecture and Design; Operation

Strategy: Attack Surface Reduction

For Flash, modify crossdomain.xml to use meta-policy options such as 'master-only' or 'none' to reduce the possibility of an attacker planting extraneous cross-domain policy files on a server.

Memberships

Nature	Type	ID	Name
MemberOf	C	1349	OWASP Top Ten 2021 Category A05:2021 - Security Misconfiguration

References

[REF-943] Apurva Udaykumar. "Setting a crossdomain.xml file for HTTP streaming". Adobe. 2012-11-19. <http://www.adobe.com/devnet/adobe-media-server/articles/cross-domain-xml-for-streaming.html>.

[REF-944] Adobe. "Cross-domain policy for Flash movies". Adobe. <http://kb2.adobe.com/cps/142/tn_14213.html>.

[REF-945] Microsoft Corporation. "HTTP Communication and Security with Silverlight". <http://msdn.microsoft.com/en-us/library/cc838250.aspx>.

[REF-946] Microsoft Corporation. "Network Security Access Restrictions in Silverlight". <http://msdn.microsoft.com/en-us/library/cc645032.aspx>.

[REF-947] Dongseok Jang, Aishwarya Venkataraman, G. Michael Sawka and Hovav Shacham. "Analyzing the Crossdomain Policies of Flash Applications". 2011-05. <http://cseweb.ucsd.edu/~hovav/dist/crossdomain.pdf>.

Content History

Submissions

Submission Date	Submitter	Organization
2014-06-05	CWE Content Team	MITRE
	Created by MITRE with input from members of the CWE-Research mailing list.	

Modifications

Previous Entry Names

Appendix 5. CAPEC-145.

Table 5: CAPEC-145 – detailed attack pattern

CAPEC-145: Checksum Spoofing

Attack Pattern ID: 145

Status: Draft

Abstraction: Detailed

Description

An adversary spoofs a checksum message for the purpose of making a payload appear to have a valid corresponding checksum. Checksums are used to verify message integrity. They consist of some value based on the value of the message they are protecting. Hash codes are a common checksum mechanism. Both the sender and recipient are able to compute the checksum based on the contents of the message. If the message contents change between the sender and recipient, the sender and recipient will compute different checksum values. Since the sender's checksum value is transmitted with the message, the recipient would know that a modification occurred. In checksum spoofing an adversary modifies the message body and then modifies the corresponding checksum so that the recipient's checksum calculation will match the checksum (created by the adversary) in the message. This would prevent the recipient from realizing that a change occurred.

Typical Severity

Medium

Relationships

Nature	Type	ID	Name
ChildOf	M	148	Content Spoofing

View Name	Top Level Categories
Domains of Attack	Software
Mechanisms of Attack	Engage in Deceptive Interactions

Prerequisites

The adversary must be able to intercept a message from the sender (keeping the recipient from getting it), modify it, and send the modified message to the recipient.

The sender and recipient must use a checksum to protect the integrity of their message and transmit this checksum in a manner where the adversary can intercept and modify it.

The checksum value must be computable using information known to the adversary. A cryptographic checksum, which uses a key known only to the sender and recipient, would thwart this attack.

Resources Required

The adversary must have a utility that can intercept and modify messages between the sender and recipient.

Related Weaknesses

CWE-ID	Weakness Name
354	Improper Validation of Integrity Check Value

Content History

Appendix 6. SANS Policy Templates

An example of policy template follows below.

Acquisition Assessment Policy

Free Use Disclaimer: This policy was created by or for the SANS Institute for the Internet community. All or parts of this policy can be freely used for your organization. There is no prior approval required. If you would like to contribute a new policy or updated version of this policy, please send email to policy-resources@sans.org.

Things to Consider: Please consult the Things to Consider FAQ for additional guidelines and suggestions for personalizing the SANS policies for your organization.

Last Update Status: Updated and converted to new format June 2014

1. Overview

The process of integrating a newly acquired company can have a drastic impact on the security poster of either the parent company or the child company. The network and security infrastructure of both entities may vary greatly and the workforce of the new company may have a drastically different culture and tolerance to openness. The goal of the security acquisition assessment and integration process should include:

- Assess company's security landscape, posture, and policies
- Protect both <Company Name> and the acquired company from increased security risks
- Educate acquired company about <Company Name> policies and standard
- Adopt and implement <Company Name> Security Policies and Standards
- Integrate acquired company
- Continuous monitoring and auditing of the acquisition

2. Purpose

The purpose of this policy is to establish Infosec responsibilities regarding corporate acquisitions, and define the minimum security requirements of an Infosec acquisition assessment.

3. Scope

This policy applies to all companies acquired by <Company Name> and pertains to all systems, networks, laboratories, test equipment, hardware, software and firmware, owned and/or operated by the acquired company.

4. Policy

4.1 General

Acquisition assessments are conducted to ensure that a company being acquired by <Company Name> does not pose a security risk to corporate networks, internal systems, and/or confidential/sensitive information. The Infosec Team will provide personnel to serve as active members of the acquisition team throughout the entire acquisition process. The Infosec role is to detect and evaluate information security risk, develop a remediation plan with the affected parties for the identified risk, and work with the acquisitions team to implement solutions for any identified security risks, prior to allowing connectivity to <Company Name>'s networks. Below are the minimum requirements that the acquired company must meet before being connected to the <Company Name> network.

4.2 Requirements

4.2.1 Hosts

4.2.1.1 All hosts (servers, desktops, laptops) will be replaced or re-imaged with a <Company Name> standard image or will be required to adopt the minimum standards for end user devices.

4.2.1.2 Business critical production servers that cannot be replaced or re-imaged must be audited and a waiver granted by InfoSec.

4.2.1.3 All PC based hosts will require <Company Name> approved virus protection before the network connection.

4.2.2 Networks

4.2.2.1 All network devices will be replaced or re-imaged with a <Company Name> standard image.

4.2.2.2 Wireless network access points will be configured to the <Company Name> standard.

4.2.3 Internet

4.2.3.1 All Internet connections will be terminated.

4.2.3.2 When justified by business requirements, air-gapped Internet connections require InfoSec review and approval.

4.2.4 Remote Access

4.2.4.1 All remote access connections will be terminated.

4.2.4.2 Remote access to the production network will be provided by <Company Name>.

4.2.5 Labs

4.2.5.1 Lab equipment must be physically separated and secured from non-lab areas.

4.2.5.2 The lab network must be separated from the corporate production network with a firewall between the two networks.

4.2.5.3 Any direct network connections (including analog lines, ISDN lines, T1, etc.) to external customers, partners, etc., must be reviewed and approved by the Lab Security Group (LabSec).

4.2.5.4 All acquired labs must meet with LabSec lab policy, or be granted a waiver by LabSec.

4.2.5.5 In the event the acquired networks and computer systems being connected to the corporate network fail to meet these requirements, the <Company Name> Chief Information Officer (CIO) must acknowledge and approve of the risk to <Company Name>'s networks

5. Policy Compliance

5.1 Compliance Measurement

The Infosec team will verify compliance to this policy through various methods, including but not limited to, business tool reports, internal and external audits, and feedback to the policy owner.

5.1 Exceptions

Any exception to the policy must be approved by the Infosec team in advance.

5.1 Non-Compliance

An employee found to have violated this policy may be subject to disciplinary action, up to and including termination of employment.

6. Related Standards, Policies and Processes

None.

7. Definitions and Terms

The following definition and terms can be found in the SANS Glossary located at:

https://www.sans.org/security-resources/glossary-of-terms/

- Business Critical Production Server

8. Revision History

Date of Change	Responsible	Summary of Change
June 2014	SANS Policy Team	Updated and converted to new format.

Among these policies SANS site contains useful information focused on the areas:

- Cyber Defence;
- Cloud Security;
- Cyber Security Leadership;
- Digital Forensics;
- Industrial Control Systems;
- Offensive Operations.

SANS Institute offers on cybersecurity:

- Courses;
- Certifications;
- Degree Programs;
- Cyber Ranges.

SANS offers more than 150 free cybersecurity tools.

For more information visit https://www.sans.org.