Proposal No. 830929
Call H2020-SU-ICT-03-2018

Project start: 1 February 2019
Project duration: 42 months

Cyber Security for Europe

D3.11

Definition of Privacy by Design and Privacy Preserving Enablers

Document Identification	
Due date	30 September 2020
Submission date	24 September 2020
Revision	1.0

Related WP	WP3	Dissemination Level	PU
Lead Participant	NEC	Lead Author	Alessandro Sforzin
Contributing Beneficiaries	AIT, ATOS, C3P, CNR, CYBER, NEC, UM, UMA, UMU, UNILU	Related Deliverables	D3.2

Abstract: This document presents D3.11 – "Definition of Privacy by Design and Privacy Preserving Enablers". It is a supplement to D3.2 – "Cross Sectoral Cybersecurity Building Blocks": while D3.2 gives an overview of all CyberSec4Europe's enablers (also known as "assets" within the project), D3.11 focuses on those dealing with privacy issues in today's cybersecurity landscape. For each asset, it describes not only its privacy-preserving properties, but also its position w.r.t. CyberSec4Europe's core research and development work packages, namely: WP3 - "Blueprint Design and Common Research", WP4 - "Research and Development Roadmap", and WP5 - "Demonstration Cases".
Privacy is a topic of paramount importance for the EU economy and society. Therefore, this document aims to give a clear picture of CyberSec4Europe's research and development capabilities by providing an introduction to the concepts of privacy and privacy-by-design, as well as a set of cybersecurity challenges that CyberSec4Europe intends to address by the end of the project.

This document is issued within the CyberSec4Europe project. This project has received funding from the European Union's Horizon 2020 Programme under grant agreement no. 830929.

Executive Summary

This document present D3.11 – "Definition of Privacy by Design and Privacy Preserving Enablers", a product of the joint work of the members of task T3.2 – "Research and Integration on Cybersecurity Enablers and underlying Technologies". As anticipated by its title, this deliverable focuses on privacy, beginning with a short history of what has been done in the past to safeguard the privacy of individuals, and where we stand today.

The discussion begins with the definition of a set of challenges in today's privacy research, presented as three broad categories: "data privacy challenges", "identity privacy challenges", and "legal and development challenges". These are open problems that CyberSec4Europe's enablers want to address over the course of the project's lifetime. An exhaustive state of the art section – preserving the three categories-based structure – provides the reader with an outlook on the privacy research landscape.

The deliverable then shifts its attention to CyberSec4Europe's and its privacy-preserving enablers (also known as "assets"). Before delving into the assets themselves, it presents CyberSec4Europe's privacy-preserving architecture, of which the assets are critical components.

The presentation of the enablers comprises an explanation of the assets' inner working and privacy-preserving functionalities, followed by a three-part discussion of its relationship with CyberSec4Europe's core research and development work packages. Namely, its relation to WP3 lines of research, as well as other assets; its place within WP4 research roadmap; and how WP5 demonstrator could leverage its functionalities.

This document is of interest to anyone looking for an overview of CyberSec4Europe's portfolio of privacy-preserving technologies, as well as the project's plans to address today's privacy research challenges.

Document information

Contributors

Name	Partner
Stephan Krenn	AIT
Thomas Lorünser	AIT
Juan Carlos Pérez Baún	ATOS
Miryam Villegas Jimenez	ATOS
Luís Antunes	C3P
João Resende	C3P
Rolando Martins	C3P
Said Daoudagh	CNR
Eda Marchetti	CNR
Liina Kamm	CYBER
Alisa Pankova	CYBER
Pille Pullonen	CYBER
Alessandro Sforzin	NEC
Claudio Soriente	NEC
Boštjan Kežmah	UM
Marko Kompara	UM
Javier Lopez	UMA
Ruben Rios	UMA
Jorge Bernal Bernabe	UMU
Antonio Skarmeta	UMU
Rafa Torres	UMU
Alireza Esfahani	UNILU
Paulo Esteves-Verissimo	UNILU

Reviewers

Name	Partner
Natalia Kadenko	TUD
Jozef Vyskoc	VAF

History

Version	Date	Authors	Comment
0.1	2020-08-07	AIT, ATOS, C3P, CNR, CYBER, NEC, UM, UMA, UMU, UNILU	1st Draft. Submitted to internal reviewers.
0.2	2020-09-16	AIT, ATOS, C3P, CNR, CYBER, NEC, UM, UMA, UMU, UNILU	Addressed internal reviewers comments.
0.3	2020-09-21	NEC	Addresses internal reviewers 2nd round of comments.
1.0	2020-09-22	GUF	High level review and final check before submission

Table of Contents

List of Figures

List of Tables

List of Acronyms

#	**6LowPan**	IPv6 over Low-Power Wireless Personal Area Network
A	**ABAC**	Attribute-Based Access Control
	AC	Access Control
	ACM	Access Control Mechanism
	ACP	Access Control Policy
	ADLC	Authorization Development Life Cycle
	AED	Advanced Encryption Standard
	API	Application Programming Interface
	Art. 29 WP	Article 29 Data Protection Working Party
B	**BPMN**	Business Process Model and Notation
C	**CA**	Certificate Authority
	CNIL	Commission Nationale de l'Informatique et des Libertés (Data Protection Authority in France)
	CEF	Connecting Europe Facility
D	**DANS**	Data ANonymization Service
	DHE	Diffie-Hellman Ephemeral
	DLT	Distributed Ledger Technology
	DPIA	Data Protection Impact Assessment
	DPO	Data Protection Officer
	DSM	Digital Single Market
E	**EC**	European Commission
	ECDHE	Elliptic Curve Diffie-Hellman Ephemeral

	ECDSA	Elliptic Curve Digital Signature Algorithm
	ECIES	Elliptic Curve Integrated Encryption Scheme
	eID	electronic IDentifier
	eIDAS	electronic IDentification, Authentication and trust Services
	ENISA	European Union Agency for Cybersecurity
	EU	European Union
F	**FEC**	Forward Error Correction
G	**GDPR**	General Data Protection Regulation
	GENERAL_D	Gdpr ENforcEment of peRsonAL Data
H	**HIPAA**	Health Insurance Portability and Accountability Act
	HTTPS	Hyper Text Transfer Protocol Secure
I	**ICO**	Information Commissioner's Office (Data Protection Authority in United Kingdom)
	ID	IDentifier
	IdM	Identity Management
	Intel SGX	Intel Software Guard eXtensions
	IoT	Internet of Things
	IPv6	Internet Protocol version 6
J	**JWT**	JSON Web Token
L	**LoT**	Lab of Things

O	**OTP**	One-Time Password
P	**PDP**	Policy Decision Point
	PKI	Public Key Infrastructure
	PFS	Perfect Forward Secrecy
R	**REST**	Representational State Transfer
	RFID	Radio-Frequency IDentification
S	**SAML**	Security Assertion Markup Language
	SDK	Software Development Kit
	SME	Small and Medium-sized Enterprise
	SPeIDI	Service Provider eIDAS Integrator
	SPOF	Single Point of Failure
	SSI	Self-Sovereign Identity
T	**TLS**	Transport Layer Security
U	**UID**	User IDentifier
X	**XACML**	eXtensible Access Control Markup Language

Introduction

The right to privacy is one of the fundamental rights included in more than a hundred national constitutions. It sets boundaries that protect individuals from external interference. The debate around privacy has gained traction since Snowden's revelations about governmental mass surveillance programs and, more recently, with the advent of artificial intelligence and data mining. As technology puts the privacy of individuals at great risk, the goal of minimizing such risk by creating privacy-friendly technologies has been a central part of the debate in privacy research circles.

In this context, the term "Privacy by Design" broadly refers to the application of data-protection best practices to system design. It is based on the idea that building privacy into a product or a service from the beginning of the design process is preferable to the alternative of adding privacy on top of an already existing system as an afterthought [198]. In contrast, privacy by default designates a situation where the default settings in a product or a service provide the user with the protection against privacy risks by themselves, without the need for any additional configuration or other changes [198].

1.1 The First Framework

The term "privacy by design", together with a first privacy by design framework, was introduced in 1995 by Ann Cavoukian, the Information and Privacy Commissioner of Ontario, Canada. It defined seven foundational principles that helped shaping the literature and legislation on the matter [199]. Below, we briefly review the principles and how they apply to the privacy of individuals:

- *Proactive not Reactive; Preventative not Remedial*: there should be measures to anticipate and prevent privacy-infringing events, rather than recovering as quickly as possible once one such event has happened. This principle ensures that a system includes means to protect privacy from foreseeable risks.
- *Privacy as the Default Setting*: data should be private "by default"', without requiring data owners to explicitly state their will to protect their data. As such, this principle protects the privacy of individuals prior to any acknowledgment or consent. For example, a data collection tool should require users to "opt-in" before harvesting their data, rather than harvesting users' data by default and allow them to "opt-out".
- *Privacy Embedded into Design*: one should integrate privacy into system design rather than adding it "on top". In other words, privacy becomes a basic system service. For example, users' data protection mechanisms should be implemented first, and their impact on the system should be considered at design time.
- *Full Functionality – Positive-Sum, not Zero-Sum*: privacy by design should create benefits for companies and users, allowing both to obtain added value from the system without trade-offs. This principle states that privacy provides added value for users, without being an obstacle to a company's business.
- *End-to-End Security – Full Lifecycle Protection*: data security and privacy must be ensured from data collection to data destruction. No intermediaries or third-parties should have access to the data, and it should be available only if necessary and with limited scope.
- *Visibility and Transparency – Keep it Open*: system components as well as stakeholders must be audited to verify that all other principles have been properly implemented. Transparency ensures that each party complies with its promises and existing regulations, thus providing individuals with guarantees of their privacy being respected.

- *Respect for User Privacy – Keep it User-Centric*: The best way to achieve great results in implementing privacy by design is to create products with end-users in mind. Products should be designed to meet users' needs, and include user-friendly functionalities for them to control and oversee how their data is processed.

Interestingly, the second principle is "Privacy as the Default Setting", shortened to "privacy by default". It mandates to clearly state the purposes for which the data is being processed (*purpose specification*), limitations on what data can be collected (*collection limitation*), minimization of the collected data (*data minimization*), limitations on use, retention, and disclosure of the data (*use, retention, and disclosure limitation*), and the notion that there should be a presumption of privacy, meaning that the default settings should provide the best possible privacy protection for users. These are all issues that still are objects of debate in today's politics and research circles.

Even though this framework was clearly ahead of its time, it was criticized for being vague and difficult to apply. It was, however, used as an inspiration to include the concept of "privacy by design" in the General Data Protection Regulation (GDPR).

1.2 First Steps in EU Legislation and the GDPR

These concepts were initially codified in the European Union RL 95/46/EC data protection directive [200], specifically Recital 46 and Article 17, even though the term "privacy by design" is never mentioned. Later, in 2009, The Article 29 Data Protection Working Party (Art. 29 WP), together with the Working Party on Police and Justice published a long essay [201], destined for the EU Commission, in which they strongly supported the idea of including privacy by design into future frameworks and for it to be binding for data controllers, as well as technology designers and producers. Their text also underlined how existing measures did not ensure privacy in practice, and how regular users could not be expected to take appropriate security actions by themselves. Therefore, it advocated that solutions using personal data should be designed with privacy by default settings, mentioning the term "privacy by default".

The legal framework for the fundamental right to protection of personal data eventually became the General Data Protection Regulation (GDPR; officially Regulation EU 2016/679) [77], which repealed the previously mentioned directive. The regulation embraces privacy by design by its Article 25, which mandates appropriate technical and organizational measures to safeguard data, both at the design phase of the processing and at its operation. Art. 29 WP suggestions mentioned above were clearly taken seriously, and privacy by design and privacy by default were combined into the new term "data protection by design and default".

1.3 Privacy by Design vs Data Protection by Design

Since the introduction of the concept of privacy by design, data protection has found itself at the core of system design, whereas it was previously considered a complementary activity, confined to legal departments. Essentially, privacy by design introduces best practices on data protection in the early stages of system design. Compared to a scenario where data protection is added a posteriori to an existing system, privacy by design provides less overhead, better security, and improved modularity.

The general consensus in the field is that there is no significant difference between privacy by design and data protection by design and default. ENISA does not distinguish between the terms "privacy by design" and "data protection by design" [198]. The difference in wording is inconsequential, as the terms "privacy" and "data protection" are effectively two sides of the same coin – the protection of (personal) data.

The GDPR refers to data protection by default as the set of choices, affecting the degree of data protection, made by a controller regarding any pre-existing configuration or processing option selected in a software application, a computer program, or a device [139]. Namely, it restricts the processing of personal data to

what is necessary for achieving specific purposes, and forces the data controller to clearly state what data it requires beforehand, and to timely inform users of the data's use [202].

Data protection by default is related to the principles of data minimization and purpose limitation. This means that it requires data controllers to use privacy preserving settings as the default settings, prevents them from giving users a false sense of choice and the illusion of control, prevents them from processing additional data unless users gives specific consent, and forces them to give users all the necessary means to exercise their rights. However, to the expert, these important measures fall under the umbrella of privacy by default as well, thus suggesting that there is no significant difference between privacy by design and data protection by design and default. Nevertheless, as applied in [140], the terms "privacy by design" and "privacy by default" would be more suitable when discussing the generic concepts, while the terms "data protection by design" and "data protection by default" are more fitting when addressing concrete requirements, such as those set by the GDPR.

1.4 The Privacy Debate

Today, privacy – especially online privacy – is at the centre of an important debate. On one side, there are governments and corporations that harvest users' data indiscriminately, using national security and "services tailored to your needs" as justification. These practices gained support by leveraging users' psychological state – fear after a terrorist attack, or the comfort of using a recommender system – but as time goes by, they are increasingly perceived as dubious at best, and against human rights at worst. Indeed, the unveiling of government surveillance programs, such as PRISM [241], and industry scandals such as the Cambridge Analytica affair [242] have caused widespread outrage, and renewed demands to limit unsolicited users' data collection.

On the other side of the debate there are users worried about being tracked against their will. Because of their protests, organizations are scrutinized more than ever before, with consequences to their public image if they are found guilty of breaching their users' privacy, or lacking adequate data security protocols. The most recent example thereof is Zoom, a video conferencing tool that gained popularity after the world was forced to stay at home because of the global pandemic [243, 244, 245, 246]. The GDPR itself can be seen as the translation of users' demands into legislation.

Perhaps the biggest obstacle against reaching a satisfying conclusion of this debate is the myth that privacy and usability are mutually exclusive. Security and privacy researchers often scorn users, for they seem incapable of understanding their guidelines for "online well-being": users post their credit card numbers on social networks and use weak passwords. However, the responsibility of creating usable and user-friendly software does not fall on the users.

The strategy of today's software is to shower users with security warnings and pop-ups – often full of technical jargon, thus hard to understand – whenever a security incident happens or is about to happen. The strategy of today's users is to ignore the warnings, close the pop-ups, and carry on. Therefore, it is crucial that researchers and industry leaders collaborate to create software that prevents such incidents, while providing users with the simplest and most direct way to achieve their goals, with actions that were designed to preserve the privacy and security of their data.

Interstate regulations, such as the GDPR, are important steps to further promote best practices and shield users against malicious third-parties, but they are a complement to secure software, not a substitute. Together with increasing efforts to educate users on the steps they can take to secure their data, it could lead to a future in which everyone's online well-being is secured.

1.5 Structure of the Document

This document is a follow-up of D3.2 - "Cross Sectoral Cybersecurity Building Blocks". As the title says, it presents WP3 enablers (also known as "assets") that focus on issues of privacy and privacy by design. These are existing technologies to be integrated with CyberSec4Europe's demonstrators, and technologies under development that will mature thanks to the project's research efforts. For each asset, the discussion is structured as follows:

- An introduction to the asset and its modus operandi;
- An overview of its privacy-preserving capabilities;
- Its relationship with CyberSec4Europe's core research and development work packages, that is, WP3, WP4, and WP5.

The document is structured as follows:

- Section 2 presents a number of challenges in today's privacy research landscape;
- Section 3 reports the state of the art in privacy-preserving technologies, with regard to the challenges presented in Section 2 and the assets that Section 5 will describe in detail;
- Section 4 gives an overview of CyberSec4Europe's privacy-preserving architecture;
- Section 5 describes CyberSec4Europe's privacy-preserving assets.;

Section 6 concludes the document.

2 Challenges in Privacy Research

This section presents a number of challenges in today's privacy reserach landscape. The discussion is structured in three broad categories: *data privacy challenges*, *identity privacy challenges*, and *legal and development challenges*. Each categories lists a number of challenges that the CyberSec4Europe's wants to address with its privacy-preserving assets (Section 5). For easy reference, a short code uniquely identify each challenge:

- DP-XX: refer to *data privacy* challenges;
- IDP-XX: refer to *identity privacy* challenges;
- LDP-XX: refer to *legal and development* challenges.

2.1 Data Privacy Challenges

DP-01 *Performing a Data Protection Impact Assessment in accordance with the General Data Protection Regulation to promote/achieve Privacy by Design*.

Considering all the requirements set forth by the General Data Protection Regulation (GDPR), it is a challenge for businesses to fulfill all of them and ensure compliance. The requirements are sometimes vague or too open and therefore subject to interpretation. One of the most challenging aspects of achieving the GDPR compliance is the creation of a Data Protection Impact Assessment (DPIA). This is a significant problem, especially for small and medium organisations, with cost and personnel restrictions. This is a privacy related challenge because DPIA is one of the cornerstones for achieving Privacy by Design as set by the GDPR.

DP-02 *When using secure multi party computation (MPC) to analyse data, analysts are not able to see the individual data values*.

Often this is cited as a barrier to using MPC as analysts are used to looking at individual values to determine different properties of the dataset. As MPC does not allow access to individual values, data analysts tend to feel that this barrier makes data analysis impossible.

DP-03 *Using secure MPC is so slow that data analysis becomes infeasible*.

As MPC introduces a computational overhead to what is already a computationally strenuous task, the methodology is often disregarded purely on this assumption without considering the properties of the specific datasets and possible hybrid solutions that would still improve data privacy when compared to the status quo.

DP-04 *The current legal status of MPC w.r.t. GDPR makes its use harder to justify*.

With the introduction of the legal definition of anonymisation in GDPR, the status of MPC in the context of law has become unclear. As per GDPR, MPC and homomorphic encryption are not considered anonymisation techniques as it is possible to combine or decrypt the data. Since analysing secret shared or encrypted data makes it impossible for analysts to see indiviual values, and it is also hard to identify data subjects.

DP-05 *Lack of mechanisms for controlling and limiting access to the data collected from numerous and geographically disperse IoT devices*.

The number of personal devices tracking health conditions, users' whereabouts, and so on continue to grow with the development of the IoT and the market of wearable devices. Moreover, these devices typically share this information with third party applications or services without the user being able to decide when, how and to what extend their personal information is shared with others. Therefore, it is paramount to devise tools and mechanisms for allowing the users to keep their personally sensitive information under control.

DP-06 *Common ways of securing data include firewalls, cryptography and using public/private keys, but nothing prevents information from propagating on.*

The hard problem in protecting privacy is preventing private information from leaking through computation. Access control mechanisms do not help with this kind of leak, since they only control information release, not its propagation once released. It is important to enhance the traditional Access Control by providing continuous, data-centric, cross-application and end-to-end control of data flows.

DP-07 *Before sending information to the cloud is important to sanitize the data from known identifiers.*

With the introduction of the legal definition of Personal Identifiable Information (any data that can be used to clearly identify an individual) in GDPR, it is mandatory to use privacy mechanisms of this type of data, especially when the data is stored in public clouds (outside the user's control). For the creation of these mechanisms, it is important to build tools with anonymization techniques to detect and protect personally identifiable information on documents.

DP-08 *When uploading information to the cloud the user partially loses control over the data.*

It is important to create mechanism to store the data accordingly to the user preferences identified. Therefore, it is paramount to enhance tools and mechanisms for allowing the users to keep their personally sensitive information under control in the cloud.

2.2 Identity Privacy Challenges

IDP-01 *Insufficient know-how of the re-identification risks of various pseudunymisation and anonymisation techniques*.

Attackers can leverage the lack of know-how or misuse of pseudonymization and anonymization techniques when personal and sensitive data are altered for analysing. Re-identification risk must be avoided by using the appropriate group of methods and models which assure the utility of the data is not affected but keeping the data privacy protection.

IDP-02 *Unnecessary over-identification and information disclosure due to a lack of awareness and usability drawbacks.*

When authenticating to cloud service providers or other third parties on the Internet, users often disclose far more information than actually needed to these parties. For instance, often proving that one is eligible to access a resource would be sufficient, yet user need to fully identity themselves in order to be granted access. Similarly, proving certain requirements (e.g., age constraints, country of residence, etc.) often leads to the full disclosure of birth date, etc. in contrast to only proving that the requirements are satisfied. We believe that two of the main reasons of such unnecessary over-identification and information disclosure are the following:

Firstly, end-user as well as the service provider are unaware that mature solutions exist and are ready for real-world deployment. Secondly, many of the existing solutions suffer from usability drawbacks, e.g., with regards to compatibility with low-cost IoT devices or user interfaces.

IDP-03 *User's privacy-preservation of transactions in distributed and immutable systems (e.g. blockchains).*

DLTs offer a decentralized, immutable and verifiable structure that can record transactions of digital assets. However, they are subject to different issues, such as transaction linkability, on-chain data privacy, or compliance with privacy regulations (e.g. GDPR). Emerging cryptographic technologies are making it possible to apply privacy preserving approaches to DLTs increasing privacy features. A balanced solution is needed to allow good privacy and compatibility with IoT scenarios.

IDP-04 *Honest but curious Service Providers and Identity Providers that might be tracking users, IdP could also become a potential point of failure (identity/data leakage in the IdP).*

Data has become the most important resource today. Identity and service providers have become entities that are able to track and obtain sensitive information about their users. While a service provider that acts maliciously is a problem, an identity provider that acts that way is a very serious problem, as it is an entity that must be fully trusted. Introducing cryptographic or privacy techniques that limit the ability to obtain information beyond the necessary to identity and service providers is a necessity.

IDP-05 *Lack of ways for provisioning of IoT devices reducing the risks of impersonating attacks and enabling the devices to authenticate.*

Given the proliferation of the IoT devices and their interconnections, it is paramount to manage their identity in a highly scalable way. It should be noted that the devices should not only be identified by their attributes, but also based in their context. IoT device has a Unique IDentifier with metadata. However, depending on the object, the identity will be different. For example, RFID uses UIDs, ZigBee sensors uses an address and 6LowPan objects will have IPv6 address. Hence, it’s difficult to let all these objects communicate with different IDs.

IDP-06 *Lack of control over the private keys on cloud storage.*

If an attacker accesses the private key of an HTTPS server or even a session key, the attacker can compromise even the previous communications. A module that secures the private key behind a secure enclave is necessary, never exposing it to attackers since it never leaves the enclave.

2.3 Legal and Development Challenges

LDP-01 *A lack of well established and easy to understand privacy metrics or privacy analysis tools*.

Mainly, differential privacy is used to quantify privacy (or provide quantifiable protection) however the guarantees it provides are not easy to understand and the necessary parameters are complicated to choose.

LDP-02 *A gap in the skill sets and communication tools between business analysts and privacy engineers.*

This gap makes it harder to discuss and document privacy decisions and the effect privacy enhancing technologies have on the business processes.

LDP-03 *GDPR-based development life cycle.*

The available development life cycles do not completely incorporate the privacy-by-design principles, and proposals targeting the GDPR demands are needed. Therefore a reference GDPR-based development life cycle for the specification, implementation and testing of software systems and applications which takes into account (European) legal requirements is needed.

LDP-04 *Enforcing and demonstrating the privacy principles compliance.*

Considering the peculiarities and the complexity of the currently available systems and applications, specific automatic approaches, facilities and tools for enforcing and demonstrating the privacy principles compliance are a crucial aspect for the successful and lawful privacy-by-design process development.

2.4 Mapping Challenges to Enablers

The tables below provide a bird's eye view of the relationship between the challenges defined above and CyberSec4Europe's enablers. Namely, each table shows which enablers possess functionalities capable of solving one or more challenges in that particular category.

2.4.1 Data Privacy Challenges

Table 1 shows CyberSec4Europe's enablers tackling data privacy challenges.

Name	DP-01	DP-02	DP-03	DP-04	DP-05	DP-06	DP-07	DP-08
GENERAL_D	X				X			
DPIA Template	X							
Edge-Privacy					X			
Sharemind		X	X	X				
pTASC					X	X		
ARGUS			X				X	X
DANS							X	
SPeiDI						X		

Table 1: Mapping of Data Privacy challenges to enablers.

2.4.2 Identity Privacy Challenges

Table 2 shows CyberSec4Europe's enablers tackling identity privacy challenges.

Name	IDP-01	IDP-02	IDP-03	IDP-04	IDP-05	IDP-06
Self-Sovereign Privacy-Preserving Idm in Blockchains		X	X	X		
Mobile p-ABC		X		X		X
pTASC					X	
ARGUS						X
DANS	X			X		
SPeiDI		X		X		

Table 2: Mapping of Identity Privacy challenges to enablers.

2.4.3 Legal and Development Challenges

Table 3 shows CyberSec4Europe's enablers tackling legal and development challenges.

Name	LDP-01	LDP-02	LDP-03	LDP-04
PLEAK	X	X		
GENERAL_D			X	X
DPIA Template				X
DANS	X			
SPeiDI			X	

Table 3: Mapping of Legal and Development challenges to enablers.

3 State of the Art

This section gives an exhaustive overview of the state of the art in privacy research. It maintains the structure of the previous section, that is, the same three broad categories – *data privacy*, *identity privacy*, *legal and development* – here further expanded with additional meaningful subcategories.

The works presented here are closely related to both the challenges presented in Section 2 and the assets that will be presented in Section 5.

3.1 Data Privacy

3.1.1 Data Protection Impact Assessment Templates

Data Protection Impact Assessment (DPIA) is a vital part of data protection by design as defined and mandated by the General Data Protection Regulation (GDPR). Performing a DPIA does not have to be a complicated and time-intensive procedure (depending on the circumstances). Still, the required level of rigour in addressing and analysing the risks to the rights and freedoms of personal data owners can be a daunting task for many organisations, especially smaller organisations. That is why since the GDPR has come into effect, there have been many good practices, recommendation lists, and tools to help organisations implement DPIAs. There are more than a few commercial solutions, but in keeping with the challenge DP-01, we focus on freely available ones because we are primarily trying to provide guidance for small and medium-sized organisations that require a DPIA but are not in a situation where they can afford to spend many resources on it.

The first is the DPIA template [1] by the Information Commissioner's Office (ICO) of the United Kingdom. As a template, it is in a document form the same as the enabler discussed in this deliverable. The ICO template itself provides little information to the reader; however, their web pages provide additional information. The risk assessment documentation is similar to the one used in our enabler, but there is no explanation of the procedure or inclusion of common threats. As such the template is more difficult to use for those that have never before performed a DPIA. The next tool [2] is provided by the European Union Agency for Cybersecurity (ENISA). However, it is only a tool for security risk assessment, and as such, it only covers DPIA partially. It comes with accompanying relevant information and explanations about the process and is as such reasonably easy to use. Risk assessment is performed in a similar way to our enabler. This tool is available online, which is very good for convenience (the final report is generated and can be downloaded by the user), but this does also mean it is less easy to adapt to the user's needs/circumstances. The third tool designed to help perform DPIA is the PIA software [3] provided by the French supervisory authority Commission Nationale de l'Informatique et des Libertés (CNIL). The tool is available in a desktop and online version; however, both have to be installed/deployed by the user. This tool is the most complete of the solutions presented here, but consequently, it is also the most complicated to use. The process itself is not well documented, which might cause problems to those not familiar with the DPIA or similar assessments.

Edinburgh Business School has created a Privacy by Design and Data Protection Impact Assessment (DPIA) Toolkit [4] for use within the Heriot-Watt University. It is a template of questions and guidelines to guide the user through the process - similar to the ENISA template mentioned before, however it covers the entire DPIA process. Considering its functionalities this template is arguably the most user friendly on this list. Family Links Network has published its own Code of Conduct and DPIA template [5] for their work in the Restoring Family Links program. The template is elementary without any guidance outside the possible risks/compliance issues, which can negatively affect the usability of the template. While the last two templates are intended for specific projects/activities, they can be adopted to other domains.

There have also been projects explicitly funded to create tools for the support of the DPIA process. Project funded in the United Kingdom is producing the Digital data protection impact assessment (DPIA) tool [6]. The project has produced an alpha version, and the partners are still developing and refining the tool even after the initial project has ended. Another project, this one funded by the German government [7], has, however, as far as we can tell, not yet produced any results.

3.1.2 Privacy Managers

Since the early days of Ubiquitous Computing privacy risks scaled to a whole new dimension due to the incorporation of numerous sensing devices in the vicinity of the users capable of capturing sensitive information about them. In this context, several researchers started to think about the need for privacy guards, privacy assistants, personal data managers or privacy helpers [8, 9, 10]. In essence, an entity allowing users to express their privacy preferences and controlling access to their personal information. The general consideration in this scenarios was to introduce contextual information in the definition of privacy policies.

Also, at that time there was significant effort in the design and development of privacy managers for managing user data while accessing the web. For example, [11] proposes a system in charge of encrypting data according to the privacy preferences of the user (expressed in APPEL [12]). Once data are encrypted, they are sent to a collector. The collector is in charge of interfacing data consumers: receives encrypted queries and returns encrypted results, which can be decrypted by data consumers informing the Trusted Privacy Manager about the privacy practices (expressed in P3P [13]).

The evolution of the web into the so-called Web 2.0, brought also privacy concerns due to the raise of social networks [14]. Privacy Managers were also developed to prevent users over sharing personal information. The main goal of these systems [15, 16, 17] was to automatically generate access rules to users' profiles in social networks based on users' privacy policies, the sensitivity of the data and the risk of exposing this data to other users.

With the advent of data outsourcing to the cloud, new privacy risks aroused and privacy managers were also developed for this context. A client-side privacy manger with obfuscation capabilities is presented in [18] and extended in [19] to consider also different probable architectures. The obfuscation module is intended to alter data before it is uploaded so that application in the Cloud cannot access it. Moreover, when plaintext data is uploaded to the cloud it should be attached to its privacy policy (i.e., sticky policies [20]). Another privacy guardian for the Cloud is defined in [21]. In this case, the focus is on the privacy policies which are defined using XACML [22] for attribute-based access control.

Although several works have been devoted to the definition of privacy assistants and managers for the Internet of Things [23, 24, 25] (some of them in the scope of Cloud-IoT architectures [26]), very few works have taken advantage of the recent development of Edge Computing to that end. To the best of our knowledge there are only two papers in this respect [27, 28]. The first one is a position paper [27] that suggests to use cloudlets to deploy one privacy manager for each raw sensor stream. The privacy manager would perform privacy policy enforcement as well as data aggregation and obfuscation but being a position paper, the authors only outline the desirable components of such a system. The second one [28] proposes to use policy enforcement on data but it distinguishes when applications are local to the fog node collecting the data from remote applications. In the former case, policy enforcement is done as usual but in the latter case they propose the creation of an execution environment that contains both data and the policies. This execution environment is sent to another fog node closer to the data requester.

3.1.3 Privacy-Preserving Data Analysis Using Secure Multi-party Computation

Computations on encrypted data/secure computing were proposed by Andrew C. Yao [247] in 1986 as a way for parties to jointly process data to learn new information together without having to compromise privacy. This methodology is prevents even the owner of the computer from accessing the data. In addition to the garbled circuit style of secure computing proposed by Yao, secure multi-party computation methods can be based on secret sharing [148], fully homomorphic encryption [248] or trusted execution environments [249]. Combinations of these methods have led to techniques like federated learning that have been deployed by Google [250]. Secure multi-party computation based on secret sharing can perform any computation and performs better than several competing techniques.

A thorough survey of general-purpose compilers for secure multi-party computation (MPC) has been done in [28]. In that paper, the authors consider eleven systems: EMP-toolkit, Obliv-C, ObliVM, TinyGarble, SCALE-MAMBA (formerly SPDZ), Wysteria, Sharemind, PICCO, ABY, Frigate and CBMC-GC. They evaluate these systems from a practical use point of view, basing their evaluation on language expressibility, capabilities of the cryptographic back-end, and accessibility to developers.

The real world applications of secure MPC in a number of use cases are discussed in [29]. The authors highlight a number of these, ranging from securing cryptographic keys to securing an entire database. This paper considers the Jana private data as a service (PDaaS) system, the Sharemind private data analytics system, the MPC applications developed by Partisia and Sepior, the Unbound Tech vHSM (virtual HSM) used for securing cryptographic keys. More specifically, due to the strong security guarantees that MPC provides, the technology has been shown to be usable and even feasible in various areas, such as privacy-preserving statistical analysis [30, 31, 32, 33, 34, 35], financial data analysis [36, 37, 38, 39], genome-wide association studies and personalised medicine [40, 41, 42, 43, 44, 45, 46], and satellite collision detection [47, 48].

3.1.4 Access Control

In literature there are several works that use Access Control (AC) as main means of protecting personal data. Different proposals are currently available and can be divided the following categories:

- **Modeling:** i.e., using Access Control elements and extensions to address specific concepts that can be related to a given law, such as consent and purpose [49], or to investigate the feasibility of translating the articles of the EU data protection directive into access control rules [50]. In this area, an initial proposal for an automatically enforceable policy language is discussed in [51]; whereas, a formal definition of the consent is introduced in [52]. Other proposals focus on matching actual attributes gathered from legal use cases and translating the resulting policies into a given formalism or language [53] in order to comply with GDPR's principle of "data protection by design and by default" [54].
- **Developing:** authors in [55] extended the XACML architecture to support context-aware security policies. [56] focuses on the enforcement of the privacy policies throughout different kind of systems and environment. Other proposals focus on integrating Access Control and business processes for GDPR compliance [57] or providing mechanisms to enforce GDPR compliance during business activities of data management and analysis [58].
- **Verification & Validation**: the GDPR is changing how Personal Data should be processed. Part of the scientific and industrial worlds are replying to this exigencies by modifying the Access Control Mechanisms (ACMs) and the way of writing their policies. However, the current proposals are still manually developed, or at proposal stage, and therefore they are exposed to the risk of encoding data protection vulnerabilities or threats. Indeed, they may require specific testing strategies or validation approaches [59, 60]. However, the software testing world is not moving at the same

speed. Currently, there are very few proposals targeting the GDPR-based Access Control Policies Verification and Validation [61]. The problem is that developing GDPR-based ACPs is per se an error-prone activity; indeed, extracting, translating and encoding the GDPR requirements into enforceable ACPs is a challenging task [53]. The major issue that can result in serious consequences, is that the generated data protection policies are not aligned with the GDPR. This can lead in developing ACP that allows an unauthorized user to access protected personal data (security perspective) and consequently resulting in an unlawful processing (legal perspective). Therefore, having facilities for verifying the compliance of the derived policy with respect to the requirements expressed in the GDPR could help the controller to provide evidence of the correctness of the developed policies. To this purpose [61] proposes an integrated testing framework able to: i) validate the currently available testing strategies by assessing their effectiveness in the context of the GDPR; ii) suggest possible improvements or integrations to the derived test suites in order to avoid specific data protection vulnerabilities.

3.1.5 Genomic Data

The advent of next-generation sequencing (NGS) machines made DNA sequencing cheaper, but also put pressure on the genomic life-cycle, which includes aligning millions of short DNA sequences, called reads, to a reference genome. However, numerous privacy attacks have demonstrated the sensitivity of genomic, and more generally, biological data. In this context, genomic privacy is considered one of the major challenges of the biomedical community [152, 153, 153, 155]. There are three different ways to protect the privacy of genomic data: (i) data deidentification [156, 157]; (ii) data augmentation [158, 159]; and (iii) methods based on cryptographic methods [160]. The data deidentification methods tend to remove or encrypt personal identifiers, such as social security numbers, zip codes, or names, which are initially associated with genomic data. Nevertheless, these methods cannot guarantee sufficient privacy protection and are not able to deal with reidentification problems [161, 162]. Data augmentation methods achieve the goal of privacy protection by generalizing, each record so that it cannot be distinguished from some other shared records [158, 159]. With this kind of methods, the privacy of genomic data can be enforced, at the expense of controlled loss of utility. Cryptology-based methods do not access the original data. They maintain data utility by using privacy-preserving data querying methods that can be applied to genomic sequences [160, 163]. Protection methods based solely on cryptographic algorithms are not sufficient, since encryption mechanisms can be broken in a comparably shorter time than the personal genomic privacy protection requires [164].

It is shown that efficient plaintext alignment methods have been developed, and can be used in parallel in public clouds (either with or without encrypting the data transfers) to study large amounts of data. However, these highly optimized methods, which include CloudBurst [165] and DistMap [166], are not privacy-preserving. Secure alignment algorithms have also been developed, for example using garbled circuits [167] or homomorphic encryption schemes [168], however, they suffer from poor performance. Recently, researchers have been searching for approaches that combine high performance and privacy. Chen et al. [169] proposed a seed-and-extend alignment method, where the seeding step is executed in a public cloud based on keyed-hashes, and the extension step runs in a private cloud. Differently, [170] makes use of Locality Sensitive Hashing (LSH), secure k-mer voting, and a MinHash algorithm, and Maskal [171] relies on a read filter and Intel SGX enclaves.

Ayday et al. [172] proposed to store encrypted reads in a biobank that enables classified people (e.g., data analyst in a hospital) to retrieve a subset of the reads from a biobank to perform genetic tests while keeping the nature of the tests private. In this approach, the biobank masks parts of the reads, for example, those located outside the request range, or those that the patient did not agree to share. Filtering approaches that identify potential genomic variations at the reads or at the nucleotides level have been described [173].

However, contrary to DNA-SeAl, those approaches do not support sensitivity levels. Concurring with our position, of classifying genomic data into sensitivity levels, Dyke et al. [174] proposed a Data Sharing Privacy Test to distinguish degrees of sensitivity for the GA4GH Beacon Project to facilitate data sharing.

3.1.6 Trust

In recent years, many companies migrate their systems to the cloud, especially to achieve agility, flexibility, and scalability. With the use of clouds, companies can obtain a high flexibility for dynamic workloads, allowing them to store new types of data or new business opportunities without significant infrastructure commitments. However, these changes have created limits on the company's exposure to the world. It is essential to analyze the cost of integrating the cloud to solve scalability and flexibility problems, because there are issues regarding data privacy, specially when considering trust, since all services will be exposed on the Internet.

Current cloud systems try to overcome issues regarding the trust in the provider, the limitation of sharing files, or the bottleneck of having a local hybrid cloud.

ARGUS is a novel platform for managing the information privately in the cloud [175]. The main advantage of this system is the possibility of the user choose between encrypt locally or in ARGUS, this implementations also uses system such as DropBox to store the files in a way that users can combine the low cost of public cloud or free tier to encrypt the information.

Charon [176] is a newer version of system such as SCFS [177] or SafeFS [178]. This system focuses on the client performing all the security and reliability computation to split the information among the different cloud providers. Even though ensuring an extra layer of security with BFT, this type of system creates battery limitations on small footprint devices. It is a good solution for server or desktop usage (since power is widely available).

3.1.7 IoT Privacy Middleware

Recent literature [179, 180, 181] highlighted the fact that privacy concerns could be a significant barrier to the IoT's growth, identifying security and privacy as a major IoT research challenge. Some of the missing characteristics are: models for decentralized authentication and trust; data-protection technologies; data ownership; repository data management; access and use rights; integration with or connection to privacy-preserving frameworks; and privacy policies management. There are open research questions that needs to be addressed in order to understand which problems are still unsolved:

Q1. How do we manage the identity of the connected things with a PKI?

   a. How do we manage the identity of the connected things without a PKI?

Q2. How do we allow users to control and manage the consent to share their data?

Q3. How do we present privacy policies and terms to IoT users in a user friendly and understandable manner?

The development of new middleware solutions for the IoT is an active area of research. The following list covers some of academic research start-ups [184] with a critical analysis in the scope of the stated research questions:

1. Microsoft Research's Lab of Things (LoT) is a platform that uses connected devices in homes [182]. However, the LoT assumes that privacy concerns must be manually handled, with the deployer signing an agreement with data owners.

   - Unsupported: Q1, Q2, Q3

2. Xively [184, 185] offers a platform as a service that lets IoT devices connect to the cloud. It does not address any privacy issues other than providing secure data storage. It has a PKI identity management [183] however the responsibility of the privacy management belongs to the person building applications and services on top of the Xively platform.
    - Supported: Q1
    - Unsupported: Q3, Q2
3. Datacoup [186] focuses on social media, and they do not focus on IoT data. Datacoup pays for each user that shares data returning an undisclosed fraction of the profit to users, for a maximum of $10 per month [187]. However, users must trust Datacoup as it sells user data through its own servers.
    - Unsupported: Q1, Q2, Q3
4. Mydex [184, 188] is a British social enterprise that make it easier and safer for individuals to hold, control, and reuse their personal information in effective and secure ways. Mydex is also a personal data sharing platform.
    - Supported: (partially) Q2, Q3
    - Unsupported: Q1
5. Axeda [189, 190] is a complete IoT solution that specializes in connecting and managing devices at low cost and complexity and remotely servicing the machines on behalf of client companies.
    - Unsupported: Q1, Q2, Q3
6. OpenIoT [191] proposes an IoT cloud platform that supports sensing as-a-service, which is a service based on the contextual enrichment of sensor data [184]. However, OpenIoT does not properly address the existing privacy concerns, as it promotes the use of public data sources.
    - Unsupported: Q1, Q2, Q3
7. FIWARE [192, 193, 194, 195] is a platform supports several IoT protocols with a modular architecture where the modules are called “IoT Agents” [196]. Users can configure roles and permissions with the PKI identity management [197].
    - Supported: Q1, (partially) Q2
    - Unsupported: Q3

3.2 Identity Privacy

3.2.1 Personal Data

The amount of data in the world is growing fast. Social networks, online shopping and in general online services are gathering huge amounts of data.

The benefits of a data-driven society are obvious and we all enjoy them however, there is a growing concern about the privacy of our data. Private and public organizations are accumulating personal and sensitive information for their own purposes while individuals are losing control over their personal data without knowing how it is stored and for what purposes it is used. Scandals such as the one involving Facebook [62] highlight the need for better management.

A lesson learned from this scandal is that we must become more aware of how much personal information is available about us online and how it is possible for us to limit this. One approach is of course to simply

not share anything with sites we do not fully trust. However, even trusted sites suffer from breaches that leak personal information [63, 64] and sometimes this information is even collected without our knowledge [65].

From a privacy perspetive, some data anonymization methods attempt to protect personal information for example *k-anonymity*, a common property of anonymized datasets requires that sensitive information ofeach record is indistinguishable from at least $k-1$ other records [66]. Related extensions to *k-anonymity* include *l-diversity*, which ensures the sensitive data is represented by a diverse enough set of possible values [67]. These techniques are not perfect and in fact, it has been shown that it is possible to eliminate anonymization [68].

Legal regulations are emerging in order to protect users and give back the control over their personal data, for example the GDPR [69].

The appearance of DLT technologies (Blockchain) introduces new approaches to address privacy and personal data management issues.

3.2.2 Identity Management

The big data era is undermining the user's privacy in multiple digital scenarios. Large third-parties benefit from the management of their users data, by collecting, analyzing, correlating and controlling massive amounts of personal data. These organizations, and their services, are subject to security breaches and user data misuse, which might compromise users' privacy, even without user-awareness.

Transactions in the blockchain are not immune to these privacy issues. Besides, individuals are given few options to control their personal data and their privacy during their online transactions, encompassing how, when, where, by whom, and which particular personal information is disclosed in each particular transaction. This problem is intensified in blockchain, as the private data included in the ledger is immutable and the user's rights to control and rectify personal information decrease.

In this context, the research community and stakeholders institutions are working to strengthen information privacy, which was highlighted by [70] as a key multilevel concept that has been studied by diverse disciplines. Diverse taxonomies of privacy have been defined in the literature [71, 72]. Furthermore, [73] provided an interdisciplinary review on information privacy, which can be defined as the ability to control information about oneself [74]. In this regard, as analyzed by [75], the notion of Privacy embraces two main areas, Confidentiality and Control.

On one hand, when it comes to Confidentiality, privacy is seen as the protection of personal data against unauthorized accesses, keeping personal data protected, anonymized and therefore private with regard to the general public. In this sense, many different mechanisms can be employed to anonymize the collected information, secure protected information, encrypt data, protect connectivity channels, etc., thereby ensuring integrity, anonymity, unlinkability, communication protection, undetectability and unobservability [76].

On the other hand, privacy refers also to the right given to citizens to Control and manage their personal data at any time, ensuring user self-determination, as defined in the European GDPR [77]. Privacy as Control can be implemented through Privacy Enhancing Technologies (PET), ensuring selective and minimal disclosure of credentials and personal attributes using, for instance, Anonymous Credential Systems [78] such as Idemix [79], which employs ZKPs to reveal the minimal amount of information to the verifier (usually a service provider), even without disclosing the attribute value itself.

In the past, traditional centralized IdM solutions, based on central authorities, set up silos of trust, meaning subjects cannot sign-on across different domains. This kind of IdM system is subject to different problems and threats such as data breaches, identity theft and privacy concerns. The rise of federated IdM models helped to mitigate partially those problems enabling Single Sign-on (SSO).

This kind of server-centric systems enables users to adopt the same identity system across different domains. The user is redirected for authentication and user identity data retrieval to his home identity provider. Some federated IdM initiatives such as Stork [80], have gone a step forward implementing a cross-border and user-centric approach, since users are put in the middle to take control of their personal data.

Several technologies, such as OpenId [81], SAML [82] or Fido [83], can be used as a baseline for implementing this user-centric approach, empowering users to share its identity across different services.

Unlike those traditional approaches, IdM based on self-sovereign identities [84] (SSI) focuses on providing a privacy-respectful solution, enabling users with full control and management of their personal identity data without needing a third-party centralized authority taking over the identity management operations. Thus, citizens are not anymore data subjects, instead, they become the data controller of their own identity. This is, they can determine the purposes, and ways in which personal data is processed, as they manage directly their personal data during their online transactions.

Chritofer Allen described in the detail the self-sovereign identity path [85], and detailed the ten Principles of Self-Sovereign Identity: Existence, Control, Access, Transparency, Persistence, Portability, Interoperability, Consent, Minimalization, Protection.

Diverse investigations in the scope of SSI have been conducted recently embracing a user-centric and privacy-preserving approach for mobiles using anonymous credential systems, verifiable attribute credentials and claims that use ZKPs, thereby giving full control to users to interact directly with the verifier. Examples of those research investigations can be found in the scope of Irma EU project [86], H2020 EU Aries project [87], H2020 EU Olympus project [88] as well as for IoT scenarios [89, 90].

Unlike those proposals, nowadays, SSI has been brought forward, as it is being materialized through blockchain, which facilitates the governance of the SSI system, increasing the performance to Internet scale and enabling the accessibility of identities to everyone. Blockchain enables sovereignty as users can be endowed with means to transfer digital assets, including user decentralized identifiers (DID) [91], DID documents, identity attributes, verifiable claims and proofs of identity [92] (including ZKPs), to anyone privately, without rules in behind, which ultimately increases the global democracy in the world.

In this sense, latest blockchain solutions [84, 93] make use of DLTs, along with user-centric and mobile-centric approaches, and therefore, empowering users to maintain securely protected (in their mobile wallet) the needed crypto credentials. In this scenario, blockchain acts as distributed and reliable identity verifier, providing provenance and verifiability of identities. Thus, the ledger provides a cryptographic root of trust, which facilitates identity management without external authorities. In this sense, [94] has recently described the main SSI concepts on blockchain and the road ahead. A User (holder) might have DIDs and obtain verifiable claims and credentials from the Issuer authority, in a user-centric way, using his smartphone whereby the private-keys are kept securely protected in the wallet. To increase the privacy-preserving capabilities in the SSI model, the user can be empowered with means to present Zero-Knowledge crypto proofs against a Service Provider acting as verifier that checks in the blockchain the attestations and signatures. Despite the features and benefits brought by SSI and blockchain, they are subject to diverse privacy issues and challenges.

3.2.3 eID

eIDAS [95] can be seen as the regulation that allows citizens to have a European national ID document, with the advantages that that has when using it in a wide range of industries and across borders. For example in view to access benefits or services provided by government authorities, banks or other companies, for mobile payments, etc. Apart from online authentication and login, many electronic identity services also give users the option to sign electronic documents with a digital signature.

3.2.4 Anonymization

When personal data are taken and modified in order not to be used to identify a person, a process of data anonymization takes place. The resulting data set, properly anonymized, can be used in a free way: it can be shared or transferred *without* being protected by GDPR or any other regulation.

Sometimes pseudonymization is used to protect personal data, since anonymization processes are so tough. The main difference is that with pseudonymization the personal identifiers are replaced with a random identifier. In this way, one can believe that the personal data are protected and not to be aware of the possible inference attacks. Moreover, pseudonymous data have to be considered as personal data within the GDPR, and the related procedures have to be applied.

However, an anonymization process can make the data useless for the initial goal. This is the problem of the privacy versus utility balance. The more privacy protection to be applied, the more probability that the resulting data mismatch the usability expectations. The relation is shown in Figure 1:

Figure 1: Privacy vs Utility trade-off

Due to the above behavior, a good anonymization process, not revealing identifying data, would generate a data set not useful anymore. It seems appropriate to have functionalities to change certain anonymization parameters and to compare the different results, to find the optimum anonymization process to be applied to a given data set to be used for a certain purpose.

In relation to those anonymization parameters, different anonymization algorithms can be applied apart from a wide variety of privacy models.

For example, k-Anonymity, k-Map, δ-presence, risk-based privacy models, differential privacy and the game-theoretic model are privacy models commonly used for attributes which are going to be transformed. In contrast, l-diversity, t-closeness, β-likeness and δ-disclosure privacy are privacy methods to be used on sensitive attributes. Some models further require particular settings (e.g. a value generalization hierarchy must be specified to be able to use t-closeness with hierarchical ground distance). Some privacy models (e.g. k-map and δ-presence) require a population table.

When an anonymization solution is chosen, two approaches can be taken. The approach called *static* anonymization, means all the data are anonymized before any management of them. It is a traditional approach that shows several drawbacks like the difficulty of select the proper anonymization algorithm to obtain a good anonymized dataset, and how accurate the final data analysis is [206]. Figure 2 shows static anonymization process considerations:

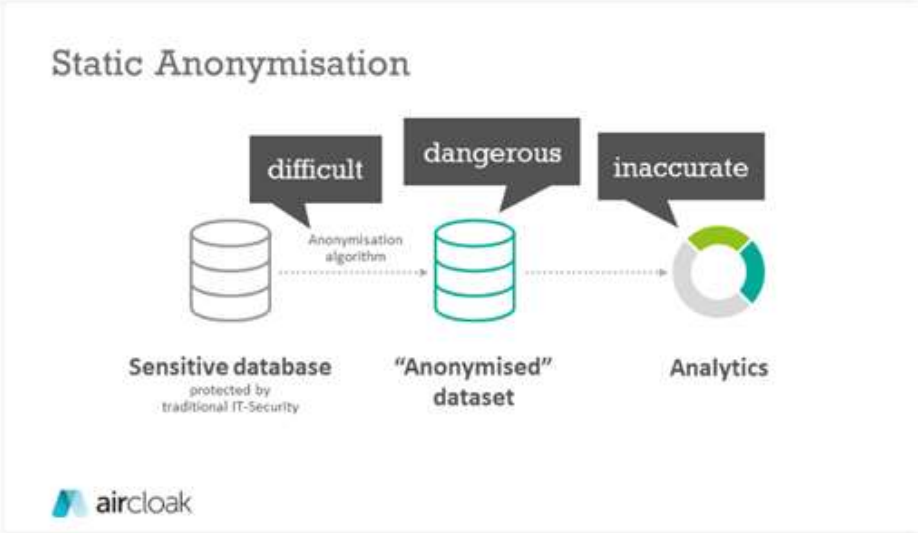

Figure 2: Static anonymization process considerations.

The other type of anonymization is the *dynamic* one, *or interactive* anonymization. In this approach, data anonymization is another part of the data query process. The data analysis is more accurate in the sense of usefulness, but it is more problematic to implement. Notice how sensitive data must be properly protected before the specific anonymizations takes place depending on the query/use of the data.

Figure 3 depicts the interactive anonymization process.

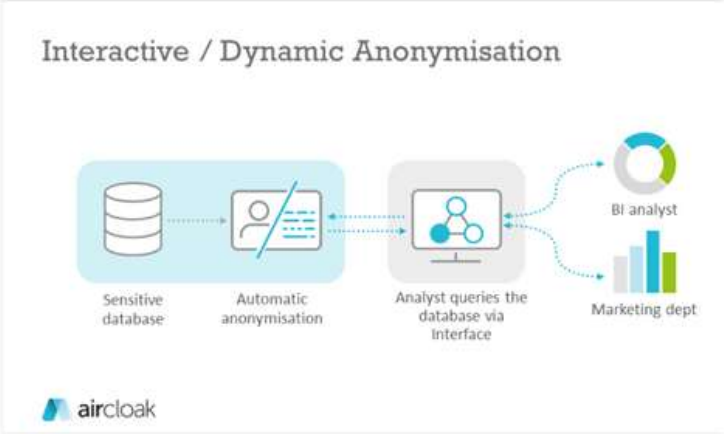

Figure 3: Interactive anonymization process.

Defining what dynamic/interactive anonymization is, seems not to be clear. According to [207], there should be available a machine learning algorithm used in an interactively mode by a data protection expert (or with her help), who would evaluate the resulting dataset by a given algorithm and, would improve the privacy/usefulness rate as much as possible.

On the other hand, the interactivity would be limited to the adjustment of a list of anonymization parameters in the tool. The anonymized dataset would be a static data set to be query afterwards [208].

Given the spectrum of anonymization tools, some experts think the dynamic/interactive anonymization tools assure the privacy in a more optimum level that the static tools. Due to the large parameters to take into account (data usability versus data protection again), the probability of generating good (useful) anonymized data with static anonymization is low.

But a static anonymization process would be enough for applications with well-defined tasks plus certain legal actions. Interactive anonymization seems more suitable in large projects, where sensitive datasets are managed, and query processes are defined with the data protection and privacy requirements.

In the end, the anonymization approach selection is based on the use cases and the importance of the data privacy in those use cases.

Finally, in some cases the anonymization could be not possible. Anonymization can be considered like a *privilege*: a specific dataset cannot be anonymized to be used for a given analytic task. In such case, the data must be protected in a properly way, applying any pseudonymization procedure, and be covered by the corresponding legislation measures.

Some examples of static and dynamic anonymization tools are collected in Table 4 and Table 5.

Static tool	Current version	Main features	Ref.
UTD Anonymization Toolbox	2012	Toolbox with 6 different anonymization methods over 3 different privacy definitions: Datafly, Mondrian, Multidimensional k-Anonymity, Incognito, , ncognito with l-diversity, Incognito with t-closeness, Anatomy	[209]
Cornell Anonymization Toolkit (CAT)	2014	Data generalization, risk analysis, utility evaluation, sensitive record manipulation, visualization and interaction. Output when a satisfactory anonymization result.	[210]
TIAMAT	2009	Interactive comparison of anonymization methods, quasi-identifier change-impact analysis, statistics.	[211]
SECRETA	2013	Supports: • For relational (R) datasets: Incognito, Cluster, Top-down, and Full subtree bottom-up • For transaction (T) attribute data sets: COAT, PCTA, Apriori, LRA and VPA • RMERGEr, TMERGEr, and RTMERGEr when anonymising RT-datasets by combining two algorithms, Two operation modes: evaluation and comparison.	[212]
Open Anonymizer	2009	k-anonymity	[213]

Static tool	Current version	Main features	Ref.
AnonTool	v086-01 (2019)	Medical data anonymization based on k-anonymity	[214]
PPS		K-anonymity algorithms: • PTA; GSC Privacy-preserving data mining algorithms: • Greedy, PSO2DT, sGA2DT, pGA2DT, cpGA2DT, SIF-IDF Privacy-preserving utility mining algorithms: • HHUIF, MSICF, MSU-MAU, MSU-MIU, pGAPPUM	[215]
ARX	v3.8.0 (2019)	Data Anonymization Tool used in DANS provide high scalability and ease of use.	[216]
Amnesia	1.2.0 (2020)	Generalization and suppression. k-anonymity and km-anonymity. Two algorithms for k-anonymity, Incognito and a parallel version of the Flash algorithm.	[217]
μ-ARGUS	5.1.3 (2018)	Process of Disclosure Control: Global Recoding and Local Suppression. Additionally, modification of numerical variables by 'top/bottom coding', 'rounding', 'numerical Micro Aggregation', 'numerical Rank Swapping' or by creating synthetic data. Manually, Global Recodings (via Modify|Global Recode).	[218]
sdcMicro	5.5.1 (2020)	Statistical Disclosure Control Methods for Anonymization of Data and Risk Estimation. Based on mu-Argus plus several new methods. Results are fully reproducible, included GUI, no time-consuming meta-data management is necessary. Necessary a detailed knowledge about SDC when applying the methods on data.	[219]
Anonimatron	1.13 (2020)	Using synonyms. Generates fake email addresses, fake Roman names, and UUID's out of the box. Default anonymizers.	[220]

Table 4: Static anonymization tools.

Dynamic tool	Current version	Main features	Ref.
Google's RAPPOR	2017	Randomized Aggregable Privacy-Preserving Ordinal Response. Differential Privacy. Inferring statistics about populations while preserving the privacy.	[221] [222]
GUPT	2012	Differential Privacy. Model of data sensitivity that degrades privacy of data over time. Different levels of privacy for different user applications, guaranteeing a constant level of privacy and maximizing the utility of each application. Techniques that improve the accuracy of output while achieving the same level of privacy.	[223]
PINQ (and wPINQ)	2009	Private Integrated Queries (PINQ). Language and framework. Differential Privacy. Language Integrated Queries (LINQ)-like API for computing on privacy-sensitive data sets. wPINQ extends PINQ with weighted datasets.	[224] [225]
Harvard Privacy Tools Project: Private Data Sharing Interface	2018	PSI ("a Private data Sharing Interface"), developed to share and explore privacy-sensitive datasets with the strong privacy protections of differential privacy	[226]
FLEX	2018	For SQL queries using elastic sensitivity. Compatible with any existing database, implements differential privacy for real-world SQL queries, with low performance overhead.	[227]
DiffprivR toolbox	2017	General-purpose mechanisms for privatizing statistics, models, and machine learners, within the framework of differential privacy of Dwork. Example of Laplace mechanism for releasing numeric aggregates, and the exponential mechanism. Sensitivity sampler available.	[228]
Aircloak Insights		Non-formal approach. It is a proxy between analysts and the sensitive data. Works with existing SQL and NoSQL databases. To be integrated directly in the existing workflow: the system queries like normal, using SQL or dashboards are intercepted by the tool. Results are returned via the proxy which ensures they are aggregated and fully anonymized.	[229]

Table 5: Dynamic anonymization tools

3.2.5 Authentication

The Service Provider eIDAS Integrator (SPeIDI) enabler has been developed for easing the use of eID, issued by the EU Member States, in the context of the cross-border authentication to online services provided by the private sector. The integration of Service provider with the country eIDAS node is made through the SPeIDI enabler (more detailed information is provided in section 5.14). The cross-border

identification the eIDAS interoperability framework allows contribute to increase the trust using these online services.

Currently the use of national eID schemas for authentication purposes against public online services is mandatory and is widely spread in each country. But In spite of several projects have been launched by EC in different programs, the use of eID on the private sector is still very low.

The EC is boosting the use of eID among SMEs (https://ec.europa.eu/futurium/en/blog/future-steps-towards-digitalisation-beyond-eidas-smes-pilot-programme) during the last years Also, initiatives such as go.eIDAS [230] are engaging the use of eID with trust services (signing, timestamp, etc.).

In the context of private sector some innitiatives have been developed for connecting real operational online services from private sector with the eIDAS network. Examples such as CEF LEPS[1] EU project shows how postal services from Spain and Greece and the Hellenic Exchanges-Athens Stock Exchange (Athex) company from Greece leverage the eIDAS network for user registration process on the online services by using the eID means issued by EU Member States. In the context of this project a mobile app for using eID card supporting NFC technology was developed[2], which facilitates the use of eID means for authentication purposes when citizens get access to online services, improving the user experience and service trustworthiness. Recent innitiative for providing login and Wi-Fi access servces by using the eIDAS network has been conducted. Benefits for users and service providers were found in terms of security against attacks and facilitating the access to services avoiding the user registration in advance [240].

Currently, the EC is trying to integrate new created building blocks, as the Blockchain one, with eIDAS. The project European Self Sovereign Identity Framework (eSSIF) [231] is part of the European blockchain service infrastructure (EBSI), which is supported by the EC for using blockchain technologies by online public services. eSSIF allows the users to obtain verifiable credentials (VCs) which can be used for identification/authentication purposes in a decentralized manner. It means that the user can “create and control their own identity without relying on centralized authorities” [231]. Several projects are funded by the European Union such as the eSSIF-Lab for increasing the uptake of the Self-Sovereign Identities (SSI) on cross-border online transactions [232].

3.3 Legal and Development

3.3.1 Development Lifecycle

In literature there are several works that use Access Control (AC) as main means of protecting personal data. Different proposals are mainly divided into two main categories. The former uses AC to address specific concepts that can be related to a given law, such as consent and purpose. In this area an initial proposal for an automatically enforceable policy language is discussed in [51], whereas, a formal definition of the consent is introduced in [52]. In particular, in [49] the authors have evaluated whether the XACML standard is adequate to express the consent and the purpose as expressed in the GDPR, whereas in [50], the authors investigated the feasibility of translating the articles related to access control of the previous EU data protection directive.

[1] http://www.leps-project.eu/
[2] http://www.leps-project.eu/node/15

However, as for any other software requirement, a fundamental step for guaranteeing the GDPR compliant realization of a given system is that the data protection concepts have to be integrated into overall software life cycle [53, 57, 96]: from gathering of the requirements to deployment and subsequent maintenance of the system. Considering in particular the authorization systems, they are recognized by scientific communities and private companies, as the successful elements for the development of the GDPR compliant solutions [97], [98]. However, most of the available proposals targets just a single aspect of authorization system development and few integrated solutions for guiding their GDPR compliant development through the entire life cycle are provided.

Recently, in the industrial environment, authors in [99], proposed a systematic methodology for the implementation of Attribute-Based Access Control (ABAC) solutions in real contexts, but without taking in consideration any legal framework. Inspired by the principle of Data Protection by-design, some proposals are integrating into a GDPR focused process development life cycle the specification, deployment and testing of adequate fine-grained authorization mechanisms [100]. More precisely, this work provides a unified environment where means and facilities are made available for assisting the different stages of development life cycle such as: modeling access control policies that are by-design compliant with the GDPR; testing those policies by means of state-of-the-art testing tools and monitoring the policies application during the production time so as to suggest possible improvements in case of deviation of the expected behavior.

3.3.2 Privacy-Enhanced Business Process Modelling with Differential Privacy Analysers (PLEAK)

The main area of PLEAK is privacy analysis of business processes and incorporating privacy enhancing technologies into business process models. As far as we know there are no tools that do similar analysis, however there are some other ideas how to approach privacy questions in business processes. The following paragraphs references some of the more closely related work. For a longer exposition regarding working with privacy or security in BPMN please see the related work section of [101].

Accorsi et al [102, 103] perform analyses based on Petri nets similar to PLEAK's Boolean analysis. However, instead of privacy enhancing technologies they divide the objects and subjects of the process to different security levels and analyze data movement between the security levels.

A BPMN extension with privacy awareness is introduced in [104]. It adds annotations about privacy concerns (or goals) to the BPMN model, for example to denote access control, user consent etc. However, it does not propose analysis based on this notation.

In addition, [105, 106] define privacy additions. First, [105] is used to specify identity negotiation and [106] to cover security and privacy notions, including data minimization and fairness.

There are many different works on differential privacy and sensitivity. Especially considering different flavors of sensitivity. Our focus is on computing the sensitivity of SQL queries and finding the noise distribution based on the desired parameters of differential privacy. PLEAK can use local sensitivity and derivative sensitivity.

Adding differential privacy to SQL queries is also considered in [107]. They define elastic sensitivity to estimate the sensitivity of the queries and use this as a basis for differential privacy. They also build a tool called FLEX to add differential privacy to SQL database engines.

In addition to FLEX also PrivateSQL [108] is a differentially private SQL engine that allows the user to specify which elements of the database schemas are private. They also develop their own flavor of differential privacy to better handle multi-relational data. Similar goal is also addressed by Wilson et. al in [109].

For many queries computing the exact sensitivity can not be done well, [110] found a subclass of the queries for which the exact sensitivity can be computed. However, they also establish that in the general case it is not computable.

There are also several works [111, 112, 113, 114] on computing or bounding the sensitivity of SQL queries. These works are either based on abstract syntax tree of the query and tracking the domains of the values or use types to derive the bounds.

On a wider view, Pufferfish [115] is a framework for privacy definitions that can be customized to specific applications. Especially, they explore notions close to PLEAK's guessing advantage.

4 CyberSec4Europe Privacy-Preserving Architecture

The CyberSec4Europe Privacy-Preserving Architecture consist of a number of building blocks which expand over several intertwined domains, including the user domain, the web domain and the IoT domain, as shown in Figure 4 and Figure 5. The building blocks are defined for different purposes which range from the compliance with current legal frameworks such as eIDAS and GDPR to mechanisms related to hardware-based solutions for managing keys and applications securely. Next we give an overview of the different building blocks that are being proposed.

In the Control and Management plane of the CyberSec4Europe architecture, the **Identity and Privacy-preservation Services** plane includes the building blocks considered in the CyberSec4Europe Privacy-Preserving Architecture devoted to enabling privacy-respectful authentication based on the provision of anonymous credential systems and privacy-preserving identity management services, some of which rely on the use of secure distributed ledger technologies such as a Blockchain to provide a self-sovereign identity (SSI) model. The Identity and privacy-preservation Services also includes mechanisms for privacy-preserving computation technologies to reduce information leakage during the computations in the managed domain, thereby verifying that the systems comply with the users' privacy policies. Those privacy-preservation services can be run in the Cloud so that the architecture includes confidentiality-preserving and end-to-end secure sharing of sensitive data in the cloud among stakeholders using, for instance, secret sharing technologies. Besides, the architecture considers the privacy brokerage aiming at enhancing user trust in public cloud storage systems, guaranteeing data confidentiality and improving availability. The Privacy-preserving architecture includes functional building blocks for confidential and privacy-preserving storage that can employ techniques such as secret sharing to anonymize personal information during data analysis processes. Similarly, it also embraces privacy-preserving mechanisms for analyzing data from potentially different stakeholders in a way that gives high authenticity guarantees on the computation’s result, while protecting the confidentiality and privacy of the input data, and ensuring data integrity.

On top of that, the Privacy-Preserving Architecture includes several mechanisms that use Trusted Execution Environments (TEE) for different purposes that range from securely storing and managing secret keys to remote anonymous attestation even in the presence of compromised hardware. The building blocks can be used on the virtualized applications in the Cloud or directly installed in the user domain.

In the **User Domain**, the privacy-preserving architecture encompasses the wallets and TEE needed to maintain securely protected the credentials and manage key material obtained during the issuance and enrollment in diverse identity providers. The user domain is exemplified either with user mobiles, or software for desktop browsers. It contains the client-side software needed to perform authentication against service providers, eIDs-based authentication, and run protocols for proving privacy-Attribute Based credentials and claims (including zero-knowledge proofs).

Therefore, the user domain plays the role of *Recipient* and *Prover* in the privacy-ABC model. To this aim, user domain interacts with diverse online identity services (including IdPs, Attribute providers, PKIs, biometric verifiers, eID verifiers) placed in the *Control and Management Domain* of the CyberSec4Europe architecture. In addition to credentials, the user domain needs to manage the attestations obtained from diverse attributes and identity providers, and short tokens obtained from IdPs (for single sign-on in Service Providers). The user-domain might also include ID-Proofing mechanisms, with client-side biometrics software needed to authenticate in biometric servers as second authentication factor.

Furthermore, the user-domain considers the data anonymization building blocks to share in a privacy-preserving way data in transactions online and between organizations using diverse different privacy models (e.g., the k-anonymity, k-Map, Average risk model, among others). In addition, in the user-domain, the privacy-analyzer allows reducing the attack surface preventing privacy breaches when sensitive personal data are managed.

Decentralized authorization, privacy-preservation and distributed access control are also important features considered in this architecture. In the **Blockchain privacy-preserving SSI Layer**, this is achieved by means of building blocks that are aimed at making blockchain technologies and consensus mechanisms more scalable, efficient, guarantying on-chain transactional privacy. Besides, it includes building blocks for modifying transactions (fine-granular rewriting) already present in the blockchain in a limited and traceable manner, which may be important for legal reasons.

The architecture considers privacy-preservation of identities and personal data in blockchains. To that aim, and following the *identity.foundation* (DIF)[3] standards and specifications, the architecture features the building blocks needed for the creation, resolution, and discovery of decentralized identifiers (DID identifiers[4]) and names in heterogeneous blockchains through resolvers. In addition, the Identity Hubs keep secure, encrypted, privacy-preserving personal data storage and computation of data. Where the resolver services links user's DID's employed in blockchain with Identity Hubs. The blockchain Identity services provide means to create, exchange, and verify crypto credentials and claims in a decentralized identity ecosystem with the User, following a self-sovereign identity management model. Besides, the blockchain identity services might rely on authentication protocols open standards and cryptographic protocols, including DIDs and DID Documents.

Another group of solutions is intended to enable privacy preservation in Cloud computing environments as well as its extension towards the user side with **Edge computing**. The Privacy-Preserving architecture provides building blocks for secure data storage and processing in public clouds. In particular, it considers distributed data storage and privacy-preserving analytics as well as mechanisms for compliance with the provisions of GDPR regarding interoperability and cross-border data transfers.

The Edge is considered in this architecture as a security and privacy enabler especially for the **IoT domain**, where devices are typically extremely resource-constrained and may be subject to compromise or interference. In this respect, the proposed architecture includes data broker for both handling sensitive data according to a set of privacy policies as well as tools for monitoring and sanitizing IoT devices for reducing the attack surface in this domain. Likewise, the privacy-preserving architecture considers the privacy-preserving middleware and software for the IoT domain aimed to ensure secure and authenticated communication channels between IoT devices. The managed domain in the global IoT architecture of Figure 4 can be also instantiated through processes related to **Web domain** (e.g. eCommerce) in the CyberSec4Europe privacy-preserving architecture. In this case the Web domain is comprised of set of

[3] DIF Identity Foundation. https://identity.foundation.org.
[4] Decentralized Identifiers (DIDs) v1.0. W3C. November 2019. https://w3c.github.io/did-core/.

functional components needed for the Service providers to authenticate their users, verify claims and privacy-preserving crypto-proofs (e.g. Zero-knowledge proofs). These service providers play the role of *Verifier* in the privacy-ABC model.

Finally, our privacy architecture also considers the application of security and privacy by design mechanisms by introducing components for GDPR-compliant software development as well as analyzing the information leakage produced by some particular privacy solutions.

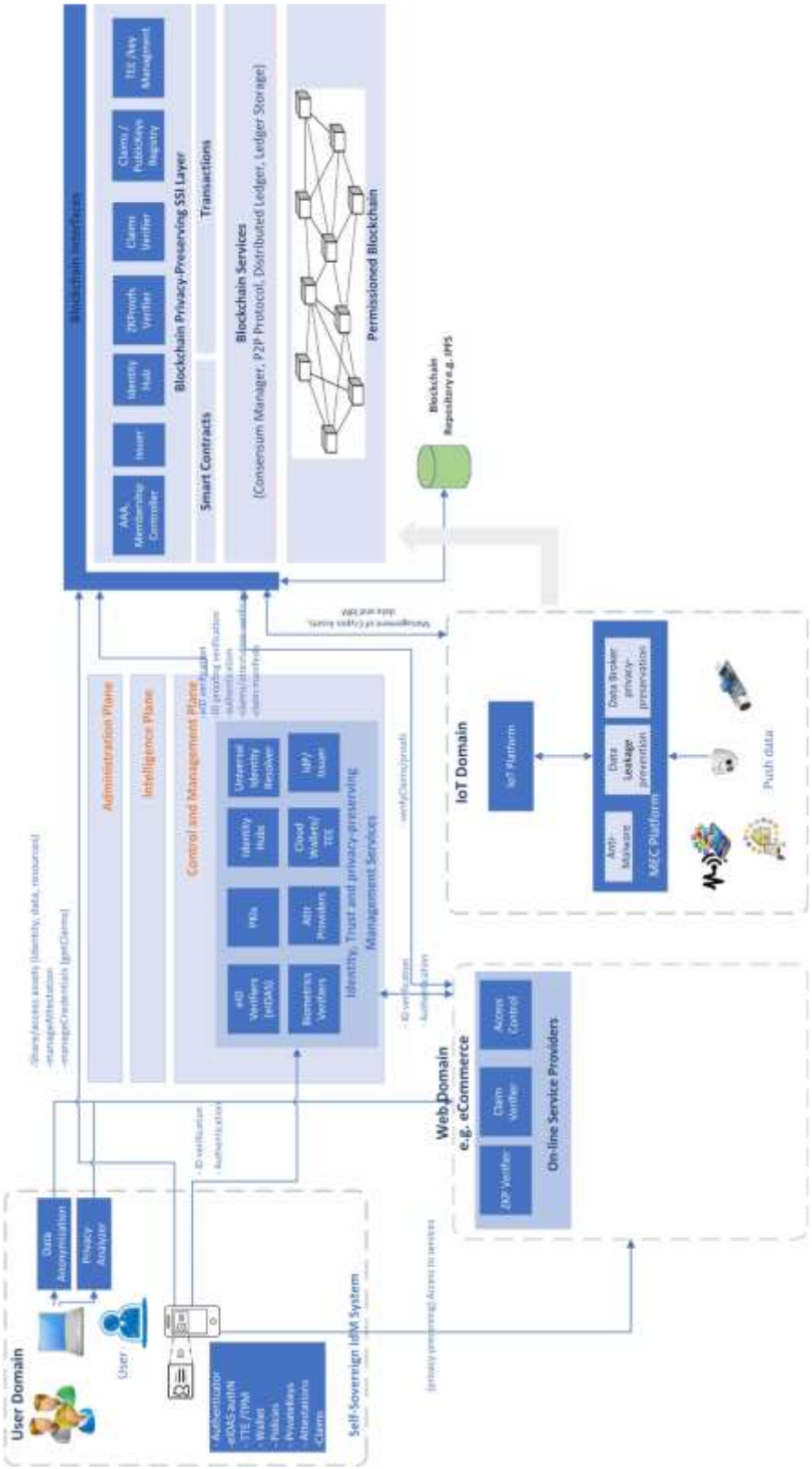

Figure 4: CyberSec4Europe Global Architecture

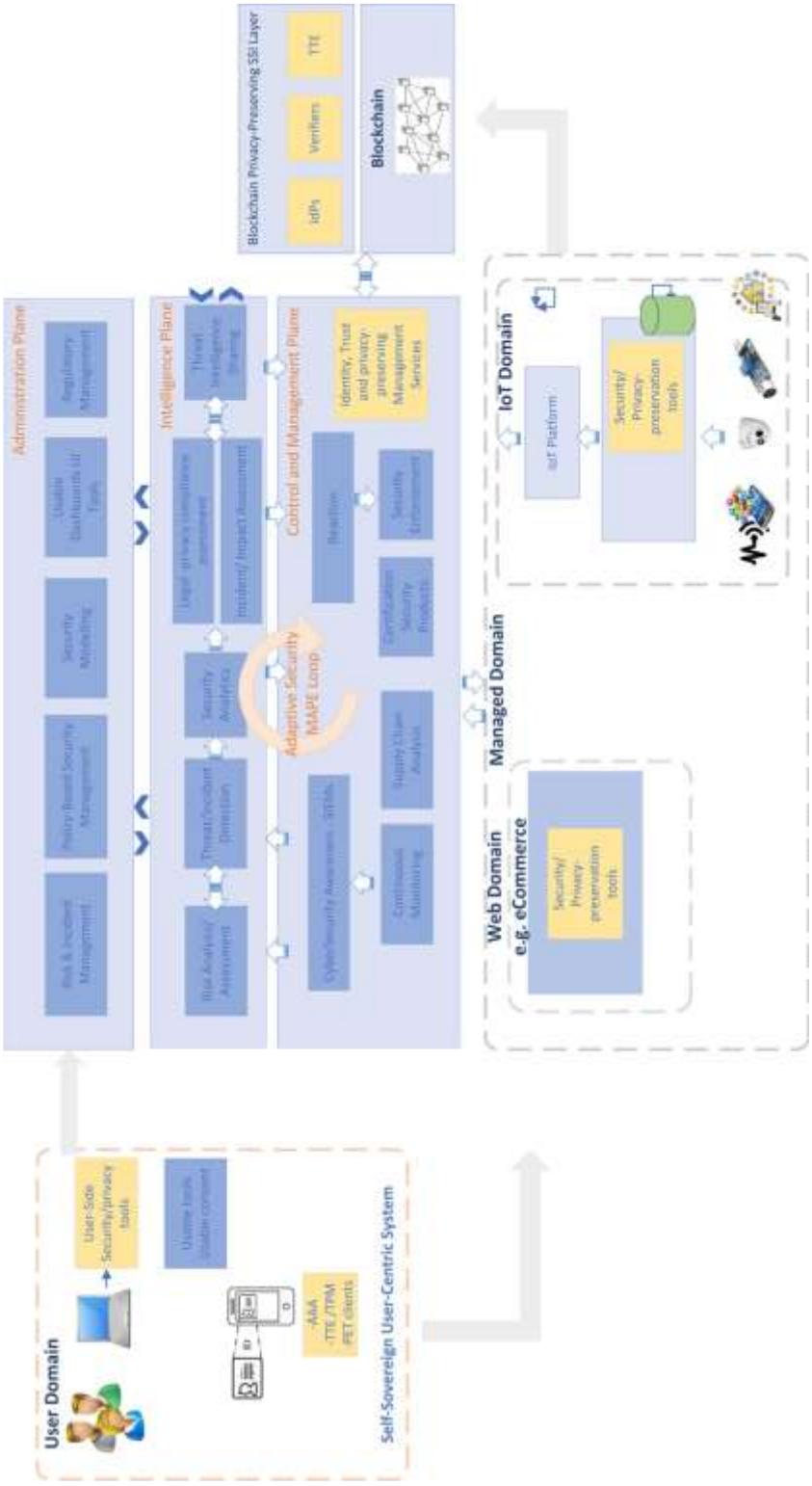

Figure 5: CyberSec4Europe Privacy-Preserving Functional Architecture

5 CyberSec4Europe Privacy-Preserving Assets

This section present CyberSec4Europe's privacy-preserving enablers[5]. The discussion is structured as follows:

- An introduction to the assets and its inner workings;
- An overview of its privacy-preserving functionalities, and how they address the challenges described in Section 2;
- The asset's relationship to CyberSec4Europe's core research and development work packages. Namely:
    - Its correlation with WP3 research and other assets;
    - Its place within WP4 research roadmap;
    - Its possible integration with WP5 demonstrators.

5.1 Self-Sovereign Privacy-Preserving Idm in Blockchains

Emerging privacy-preserving proposals for Blockchain [116, 117, 118, 119], and platforms, such as uPort[6] or Sovrin[7], propose enhanced decentralized ledgers that empower users with mechanisms preserve their privacy in their digital transactions.

The management of user's related information in permissioned blockchains is being characterized by its privacy-preserving nature. With the rise of blockchain, Identity Management (IdM) systems are switching from traditional web-centric approach or identity federation approaches, towards the self-sovereign identity (SSI) paradigm [84] facing privacy-preservation of transactions in those systems (IDP-03). Self-sovereign identities allow citizens to take control of their data in any-time in any online situation reducing unnecesary over-identification drawbacks described in IDP-02. Under this approach, user personal data is no longer kept in raw in third-parties services, neither in Service Providers or Identity Providers, and information regarding transactions and interactions of users in services can be anonymized. It avoids that third-parties can leak personal data, and, in the worst case, become a potential source of other, more important, risks, such as identity-related cybercrimes (e.g. identity-theft).

Identity Management based on Self Sovereign Identities (SSI) focuses on providing a privacy-respectful solution, enabling users with full control and management of their personal identity data without needing a third-party centralized authority taking over the identity management operations, avoiding data leakage from IdPs or SPs problems described in IDP-04. Thus, citizens are not anymore data subjects, instead, they become the data controller of their own identity. This is, they can determine the purposes, and ways in which personal data is processed, as they manage directly their personal data during their online transactions

5.1.1 Privacy-Preserving Properties

A User might have DIDs [91] and obtain verifiable claims and credentials from the Issuer authority, in a user-centric way, using his smartphone whereby the private-keys are kept securely protected in the wallet. To increase the privacy-preserving capabilities in the SSI model, the user can be empowered with means to

[5] Within the project, enablers are usually called "assets". Therefore, this document uses the two terms interchangeably.
[6] https://www.uport.me/
[7] https://sovrin.org/

present Zero-Knowledge crypto proofs against a Service Provider acting as verifier that checks in the blockchain the attestations and signatures. Indeed, SSI systems in permissioned blockchains can be leveraged with additional privacy-preserving capabilities, by using oblivious and distributed privacy-preserving crypto-solutions (using threshold cryptography) [121] where the IdP role is split up into several authorities, so that a single entity is not able to impersonate or trace user behaviors. The use of unconscious technologies in IdM architectures is already being addressed in H2020 projects such as OLYMPUS [88, 120] in which the role of IdP is segmented and distributed across several IdPs for which it is not necessary to have established a full trust relationship. This technology integrates p-ABC systems that allow the generation of user credentials through which it is possible to generate cryptographic proofs , based on PS-MS cryptographic schemes [121], which allow users to prove compliance with certain requests.

The p-ABC credentials described do not have a standard format so this solution integrates with the definition of verifiable credentials[8] and decentralized identifiers[9] described by the W3C improving the possibilities of adoption of the solution by making it compatible with web standards.

With the handling of the credentials and the application of oblivious schemes , this enabler proposes to give a further boost to the trust of the infrastructure by adding DLT support so that there is a Blockchain structure that ensures certain actions. For example, supporting the issuance of credentials, associated identities and providing traceability to the architecture operation. The privacy in this structure is guaranteed, it does not contain any personal information of any kind, only stores hashes that can be auditable to have the certain that an action occurred. The integration with DLT will require the incorporation of some Blockchain platform, for example Hyperledger Indy, so that the credentials and operations that can be derived from its use can be registered in a privacy preserving way and keeping the control of the self-sovereign identity in the users.

[8] https://www.w3.org/TR/vc-imp-guide/
[9] https://www.w3.org/TR/did-core/

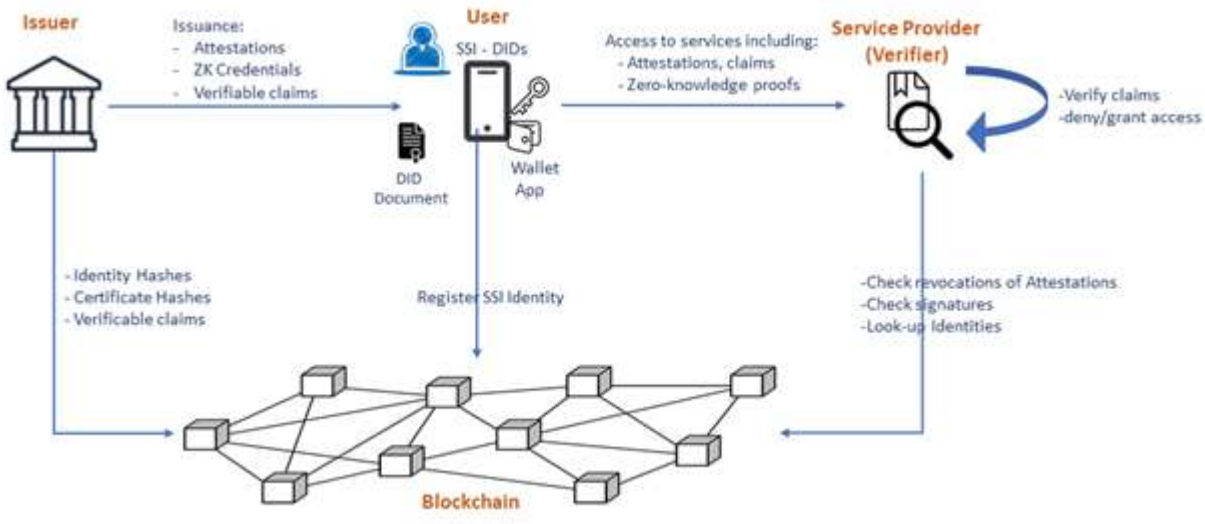

Figure 6: Trust infrastructure DLT enhanced

The provided enabler implies that a user is able to obtain in a decentralized way a valid credential with which she/he can operate in a privacy preserving way and managing her/his identity completely. The IdPs involved in the issuance of the credentials lose the ability to easily impersonate the user and thanks to DLT technology it is possible to perform audit processes, if necessary, on the actions that have occurred.

5.1.2 Relationship to WP3 Research and Assets

The SPeIDI component allows secure and strong user authentication to digital services, belongs to the groups of Identity & Trust Management services embedded in the Control and Management layer on the CyberSec4Europe architecture. The SPeIDI service eases European citizens the use of his/her eID against their origin country IdPs for cross-border authentication purposes., leveraging the eIDAS network infrastructure. In the context of the T3.2 Medical Data Exchange demonstrator, SPeIDI will provide a strong user authentication mechanism for facilitating and securing access to the Dawex data exchange platform.

5.1.3 Relationship to WP4 Roadmap

The roadmap defined in D4.3 envisages improving password systems, introduce unlinkability and minimal disclosure to Attribute-based systems, introduce privacy preservation technologies in Blockchain and add alternative methods of authentication. This Enabler goes in that way because it deals with the core issues to introduce the necessary improvements and technologies. It deals with p-ABC systems, zero knowledge cryptography, oblivious authentication and DLT systems. All these systems combined result in better management of user security and privacy.

5.1.4 Relationship to WP5 Demonstrators

Task T5.3 – the privacy-preserving identity management demonstrator – will leverage this asset. In more mature stages, the deployment of this enabler is envisioned in order to validate its performance and characteristics based on privacy management pilots using blockchain in an academic environment. In that sense, the demonstrators defined in T5.3, might make use of the proposed enabler. Thanks to the

incorporation of the DLT, it could be possible to obtain non-repudiation and traceability characteristics while, thanks to p-ABC technologies, privacy operations are improved.

5.2 Mobile Privacy-Attribute Based Credentials (Mobile p-ABC)

In the context of p-ABC systems, mobile compatibility is important [90]. Based on Idemix Anonymous Credential System [78, 79] and the implementation in the ABC4Trust project [151], Mobile p-ABC asset offers a minimal disclosure of personal information, through the use of zero knowledge proofs, for Android devices. This allows users to use their Android smartphones to present those ZK-Proofs against identity providers mitigating existing problems mentioned in IDP-02, IDP-04 and IDP-06.

The p-ABC implementation on Android enables users to perform online transactions with advanced privacy features. The system provides privacy preserving authentication through the cooperation of service providers, identity providers and users. The system prevents identity impersonation and helps to preserve user privacy.

5.2.1 Privacy-Preserving Properties

It strengthens link between physical identity (i.e. breeder documents) and derived digital identity (MobileID and mobile anonymous credentials), in order to mitigate identity-related threats. In addition, provides strict access control fully in hands of the user, while the data are stored in a secure component protected by anonymization and encryption.

Anonymous Credential systems (ACS) like Idemix allow minimal disclosure of personal attributes enabling privacy by design features to the identity management system. ACS rely on Attribute-based credentials and cryptography operations, such as Zero-knowledge crypto-proofs (ZKP), to provide pseudonymity, anonymity and minimal disclosure of attributes with diverse predicates. Nonetheless, ACS are still not widely used for their complexity and lack of user-friendly tools.

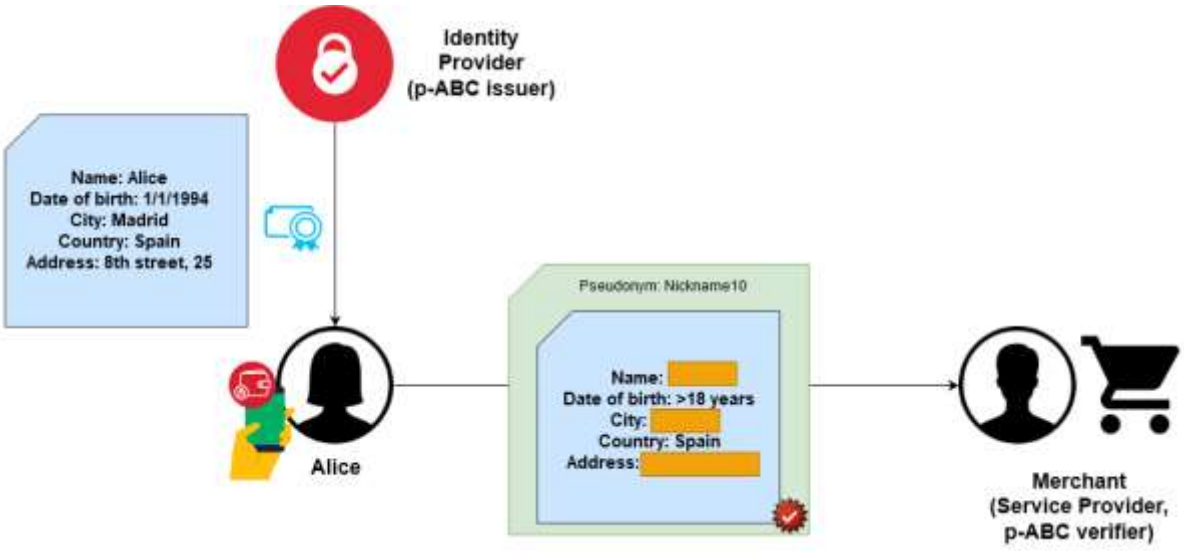

Figure 7: Mobile p-ABC

Figure 7 shows the operating diagram of mobile p-ABC. Alice is able to obtain from an identity provider a complete credential that is securely stored for later use in a secure wallet. Based on that credential Alice, will be able to test certain predicates without showing more information than necessary.

5.2.2 Relationship to WP3 Research and Assets

This component allows the use of p-ABC systems in android systems by enabling users to use advanced cryptographic technologies providing a strong user authentication mechanism to facilitate and secure access to various services. The mobile p-ABC maps the user domain that runs in the user's smartphone. The mobile wallet performs internally the crypto operations for obtain and keep securely protected the ABC credentials retrieved from the Issuer that is located online, in the control and management plane of the architecture. The service provider (e.g. a Merchant), acts as p-ABC verifier and is located in the Web-domain.

5.2.3 Relationship to WP4 Roadmap

The roadmap defined in D4.3 envisages improving password systems, introducing minimum disclosure to attribute-based systems and porting these features to a mobile environment. This enabler goes in that direction because it addresses the fundamental issues of introducing the improvements and technologies needed to be able to adapt p-ABC systems to mobile use by adding support for zero-knowledge cryptography along with authentication and authorization processes based on predicates. All these systems combined result in better management of user security and privacy in mobile environments.

5.2.4 Relationship to WP5 Demonstrators

This enabler will be applied on pilots in more advanced stages allowing to validate the operation of p-ABC systems in mobile systems. In this way it is possible to evaluate the performance of the solution in systems with limited capacities. Thus, the demonstrators defined in T5.3 will be able to make use of the mobile features in such a way that they will be able to extend the scenarios on which they operate.

5.3 Privacy Leakage Analysis Tool (PLEAK)

To verify compliance with most privacy principles the analyst needs to determine who has access to private data. In addition, one needs to study what are the conditions and extent of such access.

Business process models are a standard way to document and communicate the processes between parties and to agree on joint workflows. However, each business process involving multiple organizations or even multiple units inside one organization have their own privacy consequences. Each participant should verify that the data they share with other participants only contains what is intended and needed. Business processes document both the data processing inside the unit as well as the communication with other stakeholders. Especially, business processes captured in Business Process Model and Notation (BPMN) capture all this. Hence, they can be a good starting point for a thorough privacy analysis.

We have extended the BPMN language to capture details of privacy enhancing technologies in the processes [101]. Then this information can be used to advance the privacy analysis and to notice that some leakages are avoided by appropriate usage of PETs. This allows a Boolean privacy analysis to indicate which party really sees the contents of which data. In order to better understand the ramifications of seeing some data we also study data dependencies in the process. To support this BPMN models are annotated with our pseudocode or SQL to specify the data processing that is taking place. In the qualitative analysis level [122, 123] we describe how the output data of the process is derived from the input data of the process. Especially, if some input always is used to derive the output or if there are any conditions (predicates or filters) that need to be fulfilled. Finally, at the quantitative level [124, 125, 126] we compute the sensitivity of the workflow and estimate the noise needed to achieve differential privacy in this process.

All these analyses are brought together in the open source PLEAK tool (pleak.io) [127]. The source code of PLEAK is available in https://github.com/pleak-tools. It is a web-based tool allowing the user to model the process in BPMN, annotate the models with the details needed for the analysis and to run the automated privacy analysis.

5.3.1 Privacy-Preserving Properties

- PLEAK enables privacy-by-design approach already on the level of business process discovery and design.
- PLEAK allows to document overall decisions regarding privacy enhancing technologies in a BPMN model that is easily readable by a wide audience. PE-BPMN makes it clear which participants of the process have to carry out extra tasks to incorporate privacy technologies to the processes. This helps to deal with challenge LDP-02.
- PLEAK studies guessing advantage metric that is derived from sensitivity but designed to be easier to understand than it. The guessing advantage analysis gives as an output the differential privacy noise needed to make sure that the adversary seeing the output of the process does not benefit too much. Concretely, the user defines the maximum gain (the guessing advantage) that the adversary is allowed to get and also which private attributes of the input data the adversary is trying to learn. The analysis then finds the parameters of the noise such that the differentially private process would guarantee that the adversary does not learn more than allowed. Hence, PLEAK is working on building more intuitive metrics around differential privacy, dealing with challenge LDP-01.

5.3.2 Relationship to WP3 Research and Assets

PLEAK comprises two different assets: the BPMN tool and its analysers. The analysers are a separate asset – called “differential-privacy” – in T3.2 and the tool itself is an asset in T3.3. Currently, the enabler does not have interactions with other enablers.

With respect to CyberSec4Europe’s architecture (Figure 4 and Figure 5), PLEAK and the differential privacy analysers belong in the Security/Privacy preservation tools section of the Managed domain and in the user side security/privacy tools of the User Domain.

Finally, PLEAK is available at https://pleak.io/home. More information can be accessed at https://pleak.io/wiki/pleak. Note that to use the PLEAK installation managed by Cybernetica, a user name is required. In addition, one can set up an installation using the instructions and links given at https://pleak.io/wiki/pleak-installation.

5.3.3 Relationship to WP4 Roadmap

As PLEAK provides means for data leakage analysis, all challenges dealing with data privacy can benefit from using the tool to analyse their business processes during the design phase, to analyse data leakage risks in a more systematic manner. Currently our work in Tasks 5.5 (Maritime Transport) and T5.6 (Medical Data Exchange) is contributing to tackling the challenges described in Sections 8.4.1 (Challenge 1: Early identification and assessment of risks, threats and attack paths for critical maritime systems), 8.4.4 (Challenge 4: Maritime system communication security), 8.4.5 (Challenge 5: Securing autonomous ships) and 9.4.1 (Security and Privacy) of D4.3.

5.3.4 Relationship to WP5 Demonstrators

Pleak will be employed in the Maritime Transport (T5.5) and Medical Data Exchange (T5.6) demonstrators, where it will perform extensive privacy leakage analysis. With automatic leakage analysis, a more thorough

set of risks can be evaluated and mitigated already during the design phase, providing privacy-by-design not as an afterthought.

5.4 GENERAL_D (Gdpr ENforcEment of peRsonAL Data)

The GDPR is in charge of harmonize the regulation of Data Protection across the EU member states. At the same time, it enhances and arises business opportunities within the Digital Single Market space. However, the natural language nature of the GDPR makes most of the provisions to be expressed in generic terms and does not provide specific indication on how they should be actuated. Thus, applying and demonstrating the GDPR compliance, in order to avoid also the related penalties, becomes an important research challenge. As highlighted in the stat of the art (see Section 3.1.4), many businesses today are struggling in the definition of appropriate procedures and technical solutions for their development process so as to enforce and demonstrate the GDPR compliance. Currently, several proposals are trying to assist organizations in the deployment of adequate fine-grained mechanisms that take into account legal requirements, such as the data usage purpose, user consent and the data retention period. In particular, research attention has been devoted to authorization systems because they are recognised, by scientific communities and private companies, as the successful elements for the development of GDPR compliant solutions. Therefore, following this research line, we propose the GENERAL_D framework that integrates specific tools for the specification, deployment and testing of adequate fine-grained authorization mechanisms able to take into account legal requirements.

More precisely, GENERAL_D has been conceptualized around three axes of security and data (privacy) protection: Access Control, Data Protection by-Design and Access Control Testing & Monitoring. These axes are all related to the GDPR's demands that Controllers shall obey for being GDPR compliant-by-design. Additionally, in order to provide a ready-to-use, low-cost and effective solution for Small and Medium-sized Enterprises (SMEs), a set of supporting tools, methodologies and guidelines for developing Access Control Systems (ACSs) and Access Control Policies (ACPs) compliant-by-design with the GDPR are provided. Hence, the main outcomes of this enabler are: an Agile GDPR-based Authorization Development Life Cycle (ADLC) [100] and its supporting automation.

The novelty is that these outcomes are (by-design) conceived for *[t]aking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing [...]*, which are designed to implement data-protection principles, as in GDPR Art. 25.

Considering the Authorization Development Life Cycle (ADLC), GENERAL_D proposes an improvement of an existing approach for implementing authorization systems within enterprises [99].

The result is an Agile ADLC, which is profoundly rooted in the GDPR's "Data Protection by Design" approach (Art. 25) and the "Confidentiality and Integrity" principle defined in Art. 5.1 (f), composed of eight phases:

1. Define GDPR-based use case;
2. Gather authorization requirements;
3. Identify required attributes;
4. Author the authorization policies;
5. Test ACPs & AC mechanisms;
6. Deploy the architecture;
7. Deploy the policies; and
8. Run access reviews.

With the respect to D3.2 - "Cross Sectoral Cybersecurity Building Blocks", advancements in GENERAL_D consist of: a revision and improvement of steps 1 to 4 in order to be integrated with the other enables available in the CyberSec4Europe project (for instance CaPe); Improvement of the step 5 in order to be exploited for the Policy Decision Point (PDP) and policy testing; whereas steps from 6 to 8 are currently part of the integration with the Smart City demonstrators.

As in Figure 8, the ADLC implementation is composed of three main modules: 1) GDPR-Based Access Control Policies Management (module A); 2) Access Control System (module B); and 3) GDPR Analytics (module C).

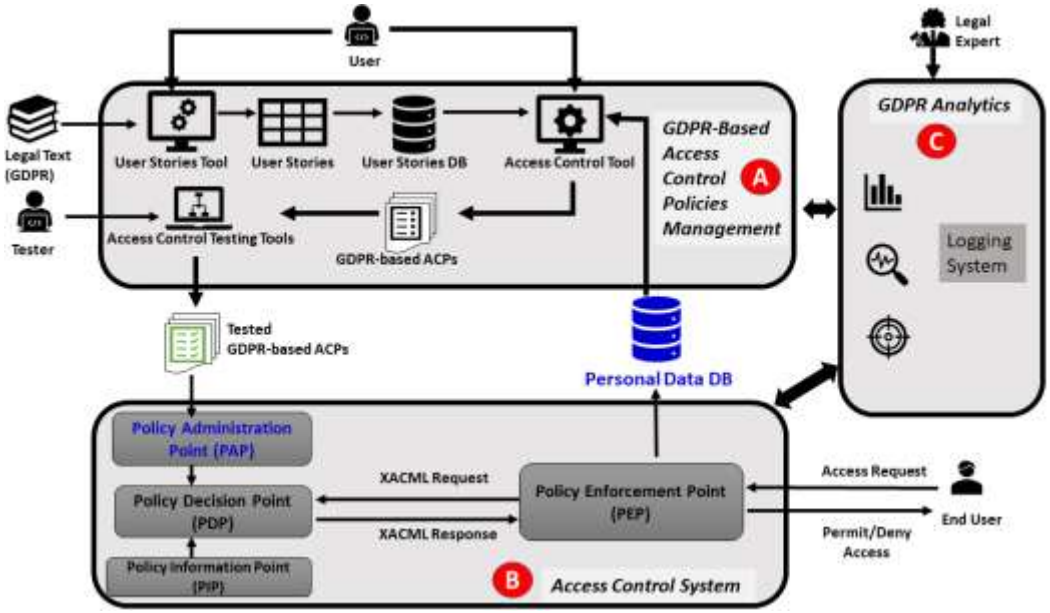

Figure 8: GENERAL_D Architecture

Very briefly, **module A** provides facilities to perform steps from 1 to 5 of the Agile ADLC. In particular, it takes as input a Legal Text (in our case the GDPR), analyses the articles related to AC and creates a Data Protection Backlog [54] containing a set of User Stories organized in Epics and Theme.

User Stories are then translated into enforceable ACPs written in the XACML standard encoding the GDPR's principles, by taking into account real attributes contained in Personal Data DB [54].

Module A integrated also ACPs and the AC mechanisms testing tools that have the following main features:

1. Test Case Generation [128, 129];
2. Mutation Generation [130];
3. Test Cases Execution & Result Analyzer;
4. Testing Strategy Enhancement; and
5. Oracle Derivation [129, 131].

The Personal Data DB component of Figure 8 contains Personal Data, whose access is regulated by **module B**, i.e., an adapted and extended version of the XACML reference architecture with new features such as logging functionalities.

Finally, facilities for collecting and managing information for the GDPR compliance and audit purposes [57, 132] are included in **module C**, which is currently under development. The module contains logging systems, monitoring capabilities, and reporting functionalities of the proposed environment so that data mining and machine learning techniques can be adopted to construct behavioral models to discover and notify unwanted behaviors.

Considering the challenges reported in Section 2, GENERAL_D targets the following challenges:

- **DP-01** by providing the data protection backlog containing GDPR-based user stories each one connected with a specific article of the regulation [54].
- **DP-05** by providing automatic facilities for assessing and testing access control systems that regulate/limit access to personal data [129].
- **LDP-03** by providing an agile based authorization life cycle for the development of access cotrol systems which is rooted in the data protection by design principles [100, 57, 132, 133].
- **LDP-04** by providing a set of tools supporting the overall development life cycle. They enable the controller to assess and demostrate the compliance with the GDPR [129, 130, 133].

5.4.1 Privacy-Preserving Properties

Differently from the solutions available in literature (see Section 4), GENERAL_D does not focus on a single aspect of the development process, but it provides a unified environment able to: 1) model access control policies that are by-design compliant with the GDPR; 2) test those policies by means of state-of-the-art testing tools; and 3) to monitors their application during the production time, and eventually to suggest possible improvements in case of deviation of the expected behavior. Therefore, the solution proposed by GENERAL_D aims at providing, for the first time, a practical specification of the Authorization Development Life Cycle in the light of the GDPR covering all its stages.

The novelty of the GENERAL_D enabler is that its outcomes are (by-design) conceived for *[t]aking into account the state of the art, the cost of implementation and the nature,scope, context and purposes of processing [...]*, which are designed to implement data-protection principles, as in GDPR Art. 25.

5.4.2 Relationship to WP3 Research and Assets

The GENERAL_D has been designed as a standalone enabler for the design and enforcement of a privacy-by-design GDPR-based development life cycle. However the different GENERAL_D components can be customized for the integration into the CyberSec4Europe Architecture, for different goals such as specification, design and testing and assessment of GDPR compliance .

In accordance to the CyberSec4Europe Privacy-Preserving Functional Architecture introduced in D3.2 (Cross Sectoral Cybersecurity Building Blocks) the enabler can be part of the Regulatory Management component for extracting, implementing and testing the data protection regulation included in the Regulatory Management. It can be used for supporting the translation of the GDPR's provisions into Access Control Policies through specific User Stories collected into a Data Protection Backlog.

Additionally, the modularity of GENERAL_D lets its integration with different WP3 enablers. For instance:

- Data Protection Backlog can be used to conduct Data Protection Impact Assessment (DPIA) which is mandatory in the GDPR (Art. 35), and therefore, the enabler is related also with the Risk & Incident Management component of CyberSec4Europe Architecture. Consequently, GENERAL_D can collaborate with the DPIA template enabler;
- Testing facilities included in GENERAL_D enabler, can be used for validating either the Access Control Policies (or Access Control mechanisms) based on the XACML standard so as to verify

their compliance with the GDPR. Therefore, GENERAL_D can also collaborate with the Privacy Manager enabler.

5.4.3 Relationship to WP4 Roadmap

Considering the Challenges reported in deliverable D4.3, Section 10.4 Smart Cities, GENERAL_D focuses mainly on Challenge 4: Privacy by design by enhancing the integration of the privacy principle during the design of the architectures and systems. GENERAL_D provides means for the identification of the possible privacy violations and threats that could be encountered during the operation stage and provides specific solutions and facilities for demonstrating the privacy principle compliance.

However, GENERAL_D partially targets also the following challenges: D4.3, Section 10.4 Smart Cities Challenge 1, 3, 5, 6.

5.4.4 Relationship to WP5 Demonstrators

GENERAL_D can be integrated with all the demonstrators which rely on access control mechanism for regulating the (personal) data access. Thanks to its modularity, GENERAL_D can be customized for offering specific and suitable functionalities such as specification, design and testing and assessment of the GDPR compliance.

Inside CyberSec4Europe project, GENERAL_D will be integrated with the Smart-Cities demonstrators for privacy enhancing purpose. The ingegration with Murcia’s demostrator is going to be finalized while the integration with the Porto and Genoa demontrators is part of the next months work.

In the following, we briefly describe the integration with Murcia’s demostrator. In this context, GENERAL-D provides facilities for specifing and testing Access Control Policies (ACPs) written in the XACML standard.

In particular, GENERAL_D will collaborate with Murcia's Smart-City demonstrator and CaPe enabler so as to leverage the Smart-City system to be Privacy-by-Design compliant with the GDPR’s provisions. An overview of the proposed integration is presentd in Figure 9.

As in the figure, the enhanced Privacy-by-Desgn Smart City System will integrate CaPe enabler for managing the consent and purposes, GENERAL_D for the specification, management and validation of the ACPs, and the Murcia’s Smart-City demonstrator for the enforcement of the ACPs used by access control system.

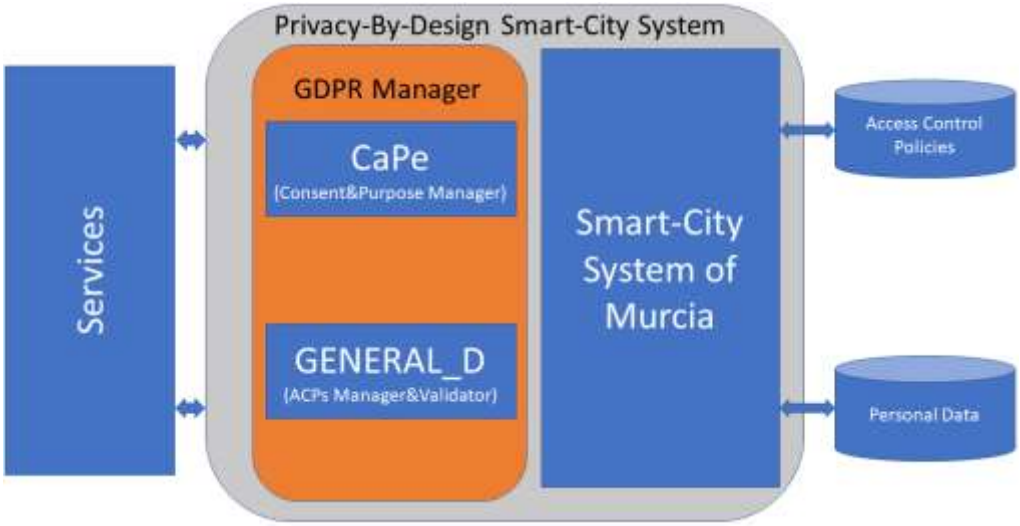

Figure 9: GENERAL_D - Privacy-By-Design Smart-City Solution

For this purpose, we have customized our standalone GENERAL_D proposal (depicted in Figure 8) to easily collaborate with the WP5 enablers. In this case, we integrated into the Privacy-By-Design Smart-City system only parts of the component of the the Module A of Figure 8.

In Figure 10 more details about the integration of the three components (GENERAL_D, CaPe and Murcia's Smart-City Platform) are provided. The conceived architecture allows performing the following steps:

1. Registration of end-users or services to the the Privacy-By-Design Smart City System. In this step, Murcia's Smart-City platform (Step 1 in Figure 10) will manage the authentication procedure of both the end-users and smart services. In this step, end-users will also interact with CaPe enabler so as to provide their consent and personal data for each specified purpose.
2. Generation of processable consent (Step 2 in Figure 10): in this step, CaPe enabler generates a processable consent in Json format.
3. Translation of the generated consent into a set of ACPs (Steps 3 and 4): In this step GENERAL-D enabler takes as input the Json consents and it generates a set of ACPs that are GDPR compliant. These policies are derived by the structures of the Data Protection Backlog and they are customized by using the data contained in the Json file (e.g., purpose, consent and personal data).
4. Testing of the conceived ACPs (Step 5): GENERAL_D tests the generated GDPR-based ACPs for verifying their integration into the smart city environment and highlights possible access control vulnerabilities.
5. Deployment of the validated ACPs into the AC system of Murcia's Smart-City platform (Step 6).
6. Enforcement of the GDPR's demands encoded in the deployed ACPs (Step 7 in Figure 10) during the operation phase.

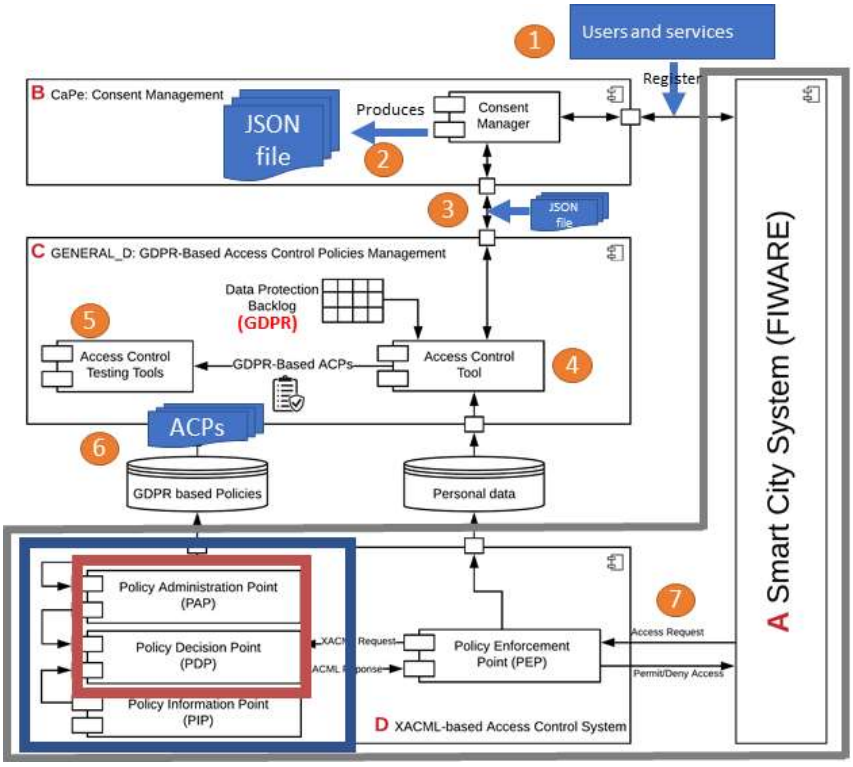

Figure 10: GENERAL_D, CaPe and Smart City Platform: The Proposed Architecture

5.5 DPIA Template

The digitalisation of almost every aspect of daily life and the use of the Internet in both private and business environments have dramatically increased data collection and accelerated the flow of information about individuals. As this vast amount of data can be used for various legitimate as well as illegal purposes, the EU authorities have, after years of negotiation, agreed on a single regulation – General Data Protection Regulation (GDPR) [77] that will strengthen the rights of individuals across the European Union and ensure uniform and coordinated action across Member States. The requirements set by the GDPR are sometimes vague or too open-ended and therefore subject to interpretation. Considering this and the number of requirements set forth by the GDPR, it can be challenging for organisations, especially smaller ones, to comply to them correctly and completely.

Guidelines for GDPR Compliant User Experience is a deliverable that was produced as D3.6 in the CyberSec4Europe project. As its name implies it is a collection of guidelines, best practices and recommendations for achieving GDPR compliance, based on a number of previous works form different organisations [134, 135, 136, 137, 138]. However, the focus and the privacy-preserving aspect of the deliverable and the enabler discussed in this section is the template for the process of performing a Data Protection Impact Assessment (DPIA). We will refer to this enabler as DPIA template. The DPIA template is just that – a template of to-do list, checklists and guidelines and is not meant to be the full and final solution. Users are encouraged to change, expand, and upgrade the given template to better suit their own organisation requirements and circumstances. By following the guidelines, data controllers and processors

can either execute data protection impact assessment and/or use the combined guidelines for GDPR compliance.

DPIA template is a combination of a guide and pre-prepared content in a form of table templates that personal data controllers can use to perform the Data Protection Impact Assessment. There are a few similar solutions you can find on the internet; however, what sets DPIA template apart from the rest is the simple construction, which users are encouraged to adapt to better suit their needs and circumstances, and a comprehensive instructions and a list of common threats to perform a required risk analysis. The DPIA template combines different global privacy frameworks into a single, unified requirements specification and at the same time combining process and material requirements. This enabler is particularly beneficial to small and medium organisations having problems performing or having questions about specific steps of the assessment by giving them a starting point on which they can build upon. This section includes a quick description and purpose of the DPIA. The remainder of the section is about the structure and functionality of the DPIA template.

DPIA is a requirement in the GDPR (Article 35) meant to analyse data processing by an organisation systematically and comprehensively with the goal of identifying and minimising personal data protection risks. Unlike most other risk analysis, DPIA is concentrated on the prevention of harm to data subjects, individuals, and overall society rather than the risk to the organisation itself. As mentioned, the DPIAs are a legal requirement under the GDPR when the processing is likely to result in a high risk to the rights and freedoms of natural persons. There is no absolute measure to determine when this is true, and it is this kind of situations that the DPIA template is trying to help with. The general recommendation for when one is not sure if the DPIA is necessary or even for when it is not required is to perform it anyway [138], because even when not necessary a DPIA can improve the trustworthiness and reputation of an organisation as well as have financial benefits (either form more business as a result of higher reputation/trust or from fewer risks that would otherwise not have been identified and appropriately addressed). Performing DPIA when not necessary, also assists in ensuring that the best practices for data security and privacy are being utilised and helps minimise the organisation’s liability. DPIA template includes an objective process to facilitate and document the decision on whether DPIA is legally required.

A DPIA must be performed before any type of processing is carried out and is an ongoing process that has to be regularly reviewed and brought up to date. A DPIA is not designed to prevent or remove all risk, but to identify and minimise them, as well as determine whether or not the level of risk is acceptable given the specific circumstances. A finished and properly performed DPIA will also help an organisation evaluate, document, and later show how you comply with all the personal data protection requirements.

The general structure of the DPIA template, and in turn of the DPIA process as well, is presented in Figure 11. The DPIA template describes and provides guidelines for DPIA related steps (blue shapes in Figure 11) as well as helps with the decision whether the assessment is necessary (green diamonds). The deliverable 3.6 we have mentioned before also includes other requirements/guidelines on GDPR compliance (orange shapes). These are spread out throughout the document and DPIA template and are pictured here separate and after the DPIA process just for the sake of simplicity.

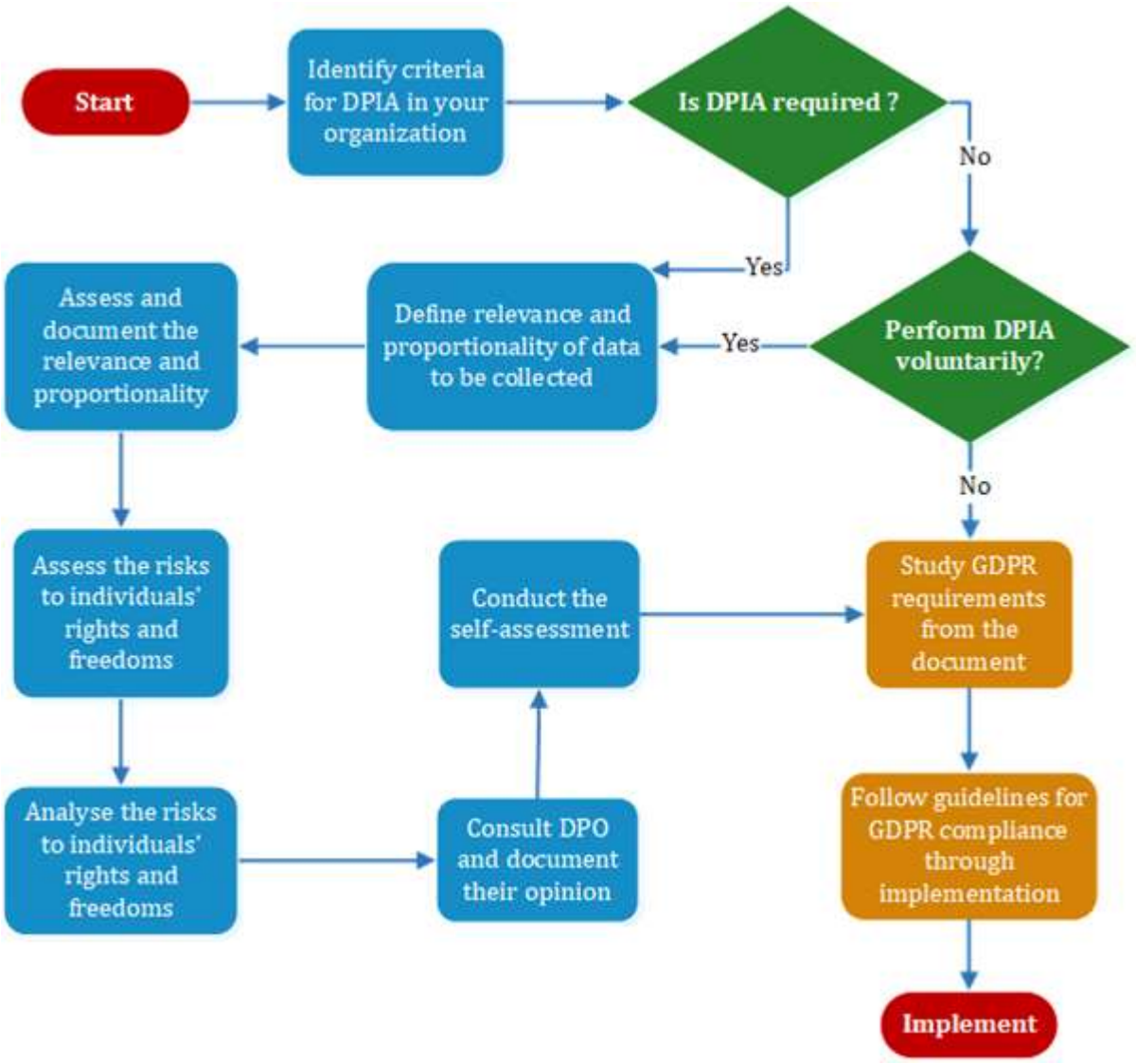

Figure 11: DPIA Template structure

DPIA template contains all the basic information about the assessment as well as many recommendations and good practices on how to perform it. Here we will just briefly outline the sections where DPIA template provides content that can directly be utilised by the users to perform their own assessment (the template part of the deliverable).

Before performing a DPIA, it is first important to establish whether the assessment is required by the GDPR. For this, the DPIA template provides a list of criteria based on the GDPR and recommendations by the Article 29 Working Party and endorsed by the European Data Protection Board. The user must only answer some questions about the type of processing they intend to perform. Based on the answers the template enables the user to make an easy judgment about the necessity of DPIA given the users' circumstances. This decision process can also be beneficial to show an organisation has performed the DPIA voluntarily.

Next major component of the DPIA is focused on risks arising from the processing of personal data. The first step is to establish what type of personal data will be processed, whether data processing is proportional/necessary, for how long it will be stored, and on what legal basis it will be collected. From this information compliance with the GDPR can be determined. Directions on how to establish compliance levels based on the collected information about the personal data are also given. The DPIA template provides instructions on how to measure risk based on the severity and probability of threats. The risk assessment methodology is aligned with ISO 31000:2018 and ISO 31010:2011 and can be directly used to assign risk levels to all identified threats. The DPIA template provides the users with a template to fill this data into, but more importantly, it already includes a long list of general personal data processing threats that can be present in any organisation. Users can freely add threats specific to their organisations, circumstances, used processes, etc. to the template. Finally, the analysis of the risks to an individual's rights and freedoms can be done.

During the DPIA it is often beneficial to receive an opinion on different aspects of the assessment from the Data Protection Officer (DPO), who should also adopt the completed DPIA into the organisation after it has been completed and oversee its implementation. The DPIA template suggests when a consultation with a DPO might be beneficial and provides guidelines for the DPO. It is worthwhile noting that a DPO is not required in all organizations and in such cases, the steps involving him/her can be omitted. However, the requirements for appointing a DPO are often related to the types of personal data processing that also require a DPIA. Therefore, when a DPIA is required, the organization will most likely also have to have a DPO. Notable exceptions are of course organizations that are performing a DPIA of their own accord without having to.

The final part of the enabler provides a template for self-assessment. This step of the DPIA is not required by the GDPR, but it is suggested, and it can help organisations track the work they have done and learn form it. Based on their performance they can improve future work on DPIAs for other processes/services

5.5.1 Privacy-Preserving Properties

In order to ensure data protection by design and by default the GDPR requires controllers to:

- Implement technical and organisational measures which will achieve the data protection principles of GDPR (most notably the minimisation and confidentiality) taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks to the rights and freedoms of natural persons posed by the processing.
- Integrate appropriate safeguards to protect the rights of data subjects.
- Ensure that only personal data necessary for a given purpose is processed by default (this includes how much data is collected, to what extent it is processed, for how long it is stored and its availability).
- The taken measures must ensure that by default personal data is not made accessible to indefinite number of natural persons, without the subject's agreement.

Performing a DPIA is strongly related to the data protection by design and by default as they both deal with; making sure only the necessary personal data is used, risks associated with the processing activities, utilising efficient and effective processes for handling personal data etc. DPIA can also help determine the technical and organisational measures that are required in order to ensure the processing complies with the data protection principles of the GDPR. The management of data protection risks is at the core of privacy by design and by default approach.

Controllers must be able to demonstrate that they have implemented measures and safeguards to achieve the desired effect in terms of data protection. To do so, the controller may determine appropriate key performance indicators to demonstrate compliance. Key performance indicators may include metrics to

demonstrate the effectiveness of the measures in question. Metrics may be quantitative, such as level of risk [139].

As such, the DPIA, as designed in the presented template, falls squarely under privacy by design and by default approach.

Having to be performed as early as possible, the DPIA serves as a safeguard towards achieving data protection by design [139, 140], even though DPIA is required only in some situations and data protection by design is a broader concept, as it applies organisationally and requires the data controllers to take certain considerations even before deciding whether the personal data processing is likely to result in a high risk or not [141].

Early consideration of privacy by design and by default is crucial for a successful implementation of the principles. From a cost benefit perspective, it would be in controllers' interest to take this into account sooner rather than later, as it could be challenging and costly to make changes to plans that have already been made and processing operations that have already been designed [139].

As we have shown, the DPIA is closely related to and can be a very important part of privacy by design. It facilitates early decision-making process in the design phase and supports documentation of the decision process, fulfilling accountability requirement. Further, documented risk analysis represents a basis for metrics to demonstrate the effectiveness of the measures in question, and that leads to the current state of the art personal data processing lifecycle. To support the achievability of privacy by design it is important to make the steps that lead to it as intuitive and as easy to take as possible. The DPIA template helps with the DP-01 challenge of performing a DPIA, which can be difficult and complex to carry out as the GDPR that requires it is not at all specific on its execution. The DPIA template building block reduces the complexity and explains the procedure of performing a DPIA and consequently, it also supports achieving or building a process/product with the privacy by design principles. As such, it also supports the LDP-04 challenge, where the process of performing a DPIA following the template enforces, while the resulting documentation demonstrates the privacy principles compliance.

5.5.2 Relationship to WP3 Research and Assets

In accordance with the CyberSec4Europe Global Architecture from deliverable D3.1 and the CyberSec4Europe Privacy-Preserving Functional Architecture introduced in D3.2 (we have touched upon them in Section 4–Figure 4 and Figure 5, respectively) the D3.6 (GDPR guidelines for compliant user experience) as a whole can be used in the Administration plane to construct privacy-compliant governance and management practices throughout adaptive security lifecycle. Primarily it is a part of the Regulatory Management, but it does also impact Risk and Incident Management, Policy-Based Security Management, and Security Modelling from the same Administration Plane.

Regulatory management indirectly impacts the Intelligence Plane and the Control and Management Plane. DPIA template, in particular, is very relevant in Intelligence Plane. It strongly ties in with the Risk Analysis/Assessment and Legal-Privacy Compliance Assessment

5.5.3 Relationship to WP4 Roadmap

The contributions of Guidelines for GDPR Compliant User Experience and specifically the DPIA template are included in the roadmap (described in D4.3) as part of the solutions to be produced until the end of the project under Medical Data Exchange (D4.3, section 9). While challenges relating to regulation, adoption of GDPR and implementation of a DPIA are relevant anywhere personal data is processed, they are especially important when processing medical data (or any other special types of data).

5.5.4 Relationship to WP5 Demonstrators

Like discussed, under the GDPR, the DPIA is not always required. The DPIA template will, therefore, be used in the demonstrator where such an assessment is definitely required. DPIA template will be used in Medical Data Exchange demonstrator in Tasks 5.6, where special categories of personal data will be processed. Processing of such data causes high risk to the rights and freedoms of individuals and consequently requires a DPIA. By using the DPIA template, the demonstrator will be able to more easily perform the DPIA including documenting and determining necessity, proportionality, nature, scope, context and purposes of the processing, identify and assess risks related to the processing of personal data in their demonstrator, and identify required measures to reduce the found risks. Building their demonstrator in this way will help ensure privacy by design. The final output of this template can also be used to show the demonstrator's compliance with the GDPR.

5.6 Edge-Privacy: A Privacy Manager for IoT using Edge Computing

Privacy Managers are useful entities for defining and enforcing data access control policies. They have been extensively used in different domains including, but not limited to, web environments, pervasive computing and cloud computing. However, very few works (see Section 3.1.2) have considered the idea of using Edge devices for the deployment of this type of service, which would enhance privacy protection of the data generated by personal devices thereby tackling challenge DP-05 defined in section 2.1.

By defining a Privacy Manager in the Edge users can retain control of the data collected by their own IoT devices. Users can decide which data are released to third parties and the level of detail at which these data are shared. Taking advantage of edge devices allows data to remain within the control sphere of the user, performing policy enforcement at the edge, rather than being outsourced to a possibly untrusted cloud service provider and relying on the policy enforcement mechanisms deployed there.

As such, some of the main features or requirements we imposed to the Privacy Manager we are implementing are as follows:

- To be easily deployable in Edge computing environments
- To implement interfaces for collecting and storing data from IoT devices
- To allow the authentication and authorization of users
- To enable data owners to define their privacy policies for controlling data access
- To enable authenticated third parties to query for the data
- To incorporate data filtering mechanisms with different granularity levels depending on who is requesting the data.

The aforementioned requirements lead to the general design of the Privacy Manager for IoT data and its core components, depicted in Figure 12. At the core of the system there is a database for storing the data collected from IoT devices and for the definition of user-defined privacy policies. There are components for handling access to the database, for communicating with IoT devices and for interacting with the users. We also envision the implementation of a northbound interface with cloud service providers in order to allow the secure storage of historical and backup data.

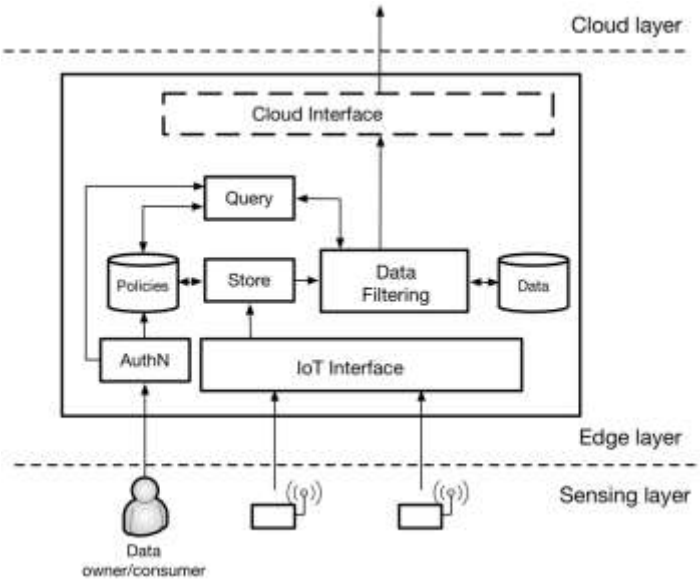

Figure 12: Edge Privacy Manager Architectural Design

5.6.1 Privacy-Preserving Properties

There is a need for tools and mechanisms for controlling and limiting access to personal data. Privacy Managers enable users to define and enforce their privacy preferences (or privacy policies) thereby empowering users to decide for themselves with whom, under what circumstances and with what level of granularity their personal information is shared with others.

This type of tools is becoming increasingly relevant as the number of devices that collect and share personally sensitive information about individuals continues to grow, from heart rate monitors to location trackers. These devices may be even scattered across different geographical locations or attached to the person thus making it difficult for users to control who has access to which data.

The recent development of the Edge Computing paradigm can be exploited for the deployment of IoT Privacy Managers. The benefits of taking advantage of Edge Computing for this purpose are three-fold: (i) data are stored next to where they are generated, (ii) user privacy policies are defined and enforced under the control of the user and (iii) privacy managers can be easily deployed anywhere due to the containerization capabilities of edge devices thus enabling the handling of data being collected by devices in remote locations and even while the user is on the move.

5.6.2 Relationship to WP3 Research and Assets

The Privacy Manager presented here has been designed as a standalone asset that belongs into one of the managed domains of the CyberSec4Europe Architecture. In particular, this asset is devised as part of the IoT domain to partially cover the goal of providing security and privacy services to deploy a basic Edge computing platform. Although the Privacy Manager is not dependent on other components of the CyberSec4Europe Architecture, its modular design allows to plug in new components so as to extend its functionalities.

This asset has been conceived as a software component based on containerization technologies, which notably facilitates its deployment on presumably any Edge computing environment. In particular, we have opted for using Docker containers inside a Kubernetes[10] cluster, which is the current trend in the development of most of the novel Edge platforms that are recently arising in the market. For example, Edge X Foundry[11], Mainflux[12] and KubeEdge[13] are rely on Kubernetes for the deployment management of containerized applications and services.

So far, the Privacy Manager allows the definition of privacy policies by data owners and the querying of data by data consumers using a REST interface. The authentication of users is currently done using a traditional username and password scheme but it can be easily extended to incorporate more advanced authentication mechanisms (e.g., attribute-based credentials or self-sovereign identities) like the ones provided in T3.2 within CyberSec4Europe. Upon successful authentication the user receives a token that will be later used to identify the clearance level the user has to perform actions. For example, it serves to determine the privacy filtering mechanism to be applied to any data requested by this user.

The data filtering module within the Privacy Manager is in charge of extracting and processing data from the database according to the privacy policies defined by the data owner. A privacy policy file is defined per role or user using a simple language consisting of a number of rules defining allowed actions over the data for a user/role. For the case of data access, the language allows to indicate filtering/transformation mechanisms to be applied before a particular type of data is returned to the data consumer. Extending this policy language to incorporate contextual information is considered as part of our future work.

Currently, the data filtering module considers three methods for the transformation of data although this module is designed to allow the incorporation of new filtering methods (e.g., some of the privacy models provided by DANS, the *Data Anonymization Service*). The first method allows the release of data exactly as it is stored in the database. This method is convenient for data owners and fully trusted entities in order to get access to raw data. The second method consists of the addition of noise to data, which will be consistent with the type of data over which noise is applied in order to balance data privacy and data utility. Finally, the third method is meant to return k-anonymous results to user data requests. This is done using by means of generalization and suppression techniques using an already existing library[14] based on the Mondrian algorithm to partition data into groups.

In addition to that, the Privacy Manager has a southbound interface for the collection of user's data coming from IoT devices. This interface is implemented using MQTT, which is a communication protocol extensively used in the IoT domain.

5.6.3 Relationship to WP4 Roadmap

The Roadmap defined in WP4 identifies a number of relevant challenges for each of the verticals considered in CyberSec4Europe and maps them onto research areas proposed by the European Commission's Joint Research Centre. Among the various areas considered in JRC's taxonomy of cybersecurity, the one that seems to be most relevant for the verticals in the project is the area of "Data Security and Privacy". This area includes several topics including "Design, implementation, and operation of data management systems

[10] Kubernetes:https://kubernetes.io/
[11] Edge X Foundry: https://wiki.edgexfoundry.org/
[12] Mainflux: https://www.mainflux.com
[13] KubeEdge: https://kubeedge.io/en/
[14] Nuclearstar https://github.com/Nuclearstar/K-Anonymity

that include security and privacy functions", which is precisely the goal of the Privacy Manager devised in this task.

5.6.4 Relationship to WP5 Demonstrators

The Privacy Manager presented in this section is being implemented as a containerized service inside a Kubernetes cluster deployed on edge computing devices. It is intended for enabling users to specify privacy policies that define fine-grained access control to the data collected from their personal IoT devices. As such, we believe the Privacy Manager could be of interest to the "Medical Data Exchange" demonstrator defined in T5.6 as it considers the case of people wearing self-monitoring health devices and the protection of these data during data sharing through a marketplace platform. In this scenario, the Privacy Manager could be used to control the granularity of the health-related data being collected by wearable devices and then sent to the data marketplace thereby empowering users to decide for themselves to what extent their information is being shared with others.

Moreover, the Privacy Manager could potentially be adapted and used to control access to information collected by other types of IoT devices even if this information is not personally sensitive. Therefore, the functionality offered by the Privacy Manager could be of interest to support some of the functionalities in the "Maritime Transport" and "Smart Cities" demonstrators.

The Privacy Manager is currently under development, and as such it has not been yet integrated with WP5 demonstrators.

5.7 Sharemind

The Sharemind MPC platform [203] is one of the better known MPC platforms due to its ease of programming [28] and real-world use cases [29]. Sharemind uses secret sharing to encrypt the data on site at the data provider. This way, the data never leaves the data provider in an unencrypted form. The data will be shared to three computation parties hosted by three independent non-colluding organisations. These computing parties will carry out predetermined queries without ever decrypting the values. The results are also kept as shares. Only the result parties have a right to decrypt the results.

Sharemind supports statistics through an R-like privacy-preserving statistical analysis environment Rmind [31]. This environment allows a data scientist to perform statistical queries on data without having access to individual values. The underlying cryptographic protocols are hidden from the user by the high-level language, hence making it possible for an analyst with no background in cryptography to make queries on secret shared data. Rmind was successfully used in a real-world study (PRIST) for analysing data combined from the education information system and the Tax and Customs Board in Estonia [37]. Figure 13 describes the setup of the Sharemind system for this study. None of the computing parties nor the result party were able to see the input values. This landmark study resulted in a legal research discussion on whether any personal data were actually processed in the meaning of the law [204]. This remains an open topic as it is still unclear how MPC legally relates to GDPR (challenge DP-04).

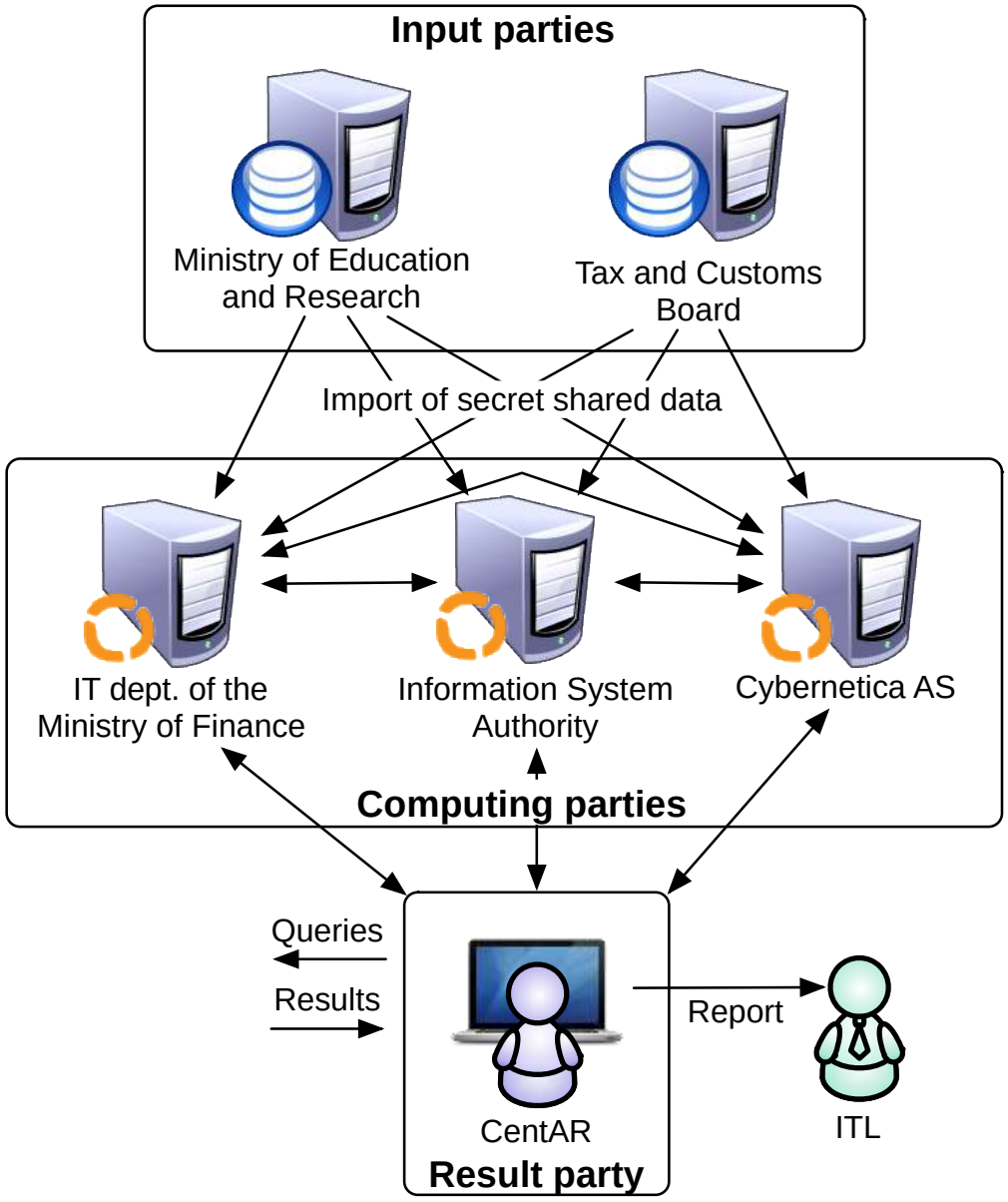

Figure 13: Setup of the Sharemind system for the PRIST study.

As discussed in challenge DP-03, there is a significant computational overhead to working with encrypted or secret shared data. Sharemind has shown quite good benchmark results even with floating-point numbers. Depending on the analysis task and its urgency, however, this can be a barrier. We are currently researching specific use-cases and algorithms, and their feasibility. For instance, in the satellite collision task [205], the computations took approximately 10 hours, however, they need to be run only once in 24 hours, making the algorithm still usable despite the long computation time. Also, hybrid solutions can be used in some cases. For example, a machine learning model can be trained on each partner's data and then shared with others in secret-shared form. These can then be combined in secret shared form and depending on the sensitivity of the data, applied in either secret shared or open format.

5.7.1 Privacy-Preserving Properties

Sharemind allows the analyst to compute on data without seeing individual values. This preserves the privacy of data subjects, as only the aggregated results will be revealed. As data are stored and analysed in a secret-shared format, even system administrators do not have access to the values.

The Rmind statistical analysis environment allows the analyst to query summary statistics of the data, including data distribution, quantiles, and heatmaps. This allows them to see what the data is like without actually looking at the individual values in a dataset (dealing with challenge DP-02). The system also allows for computations with or without outliers (the cut-off is based on quantiles, which can be determined by the user).

5.7.2 Relationship to WP3 Research and Assets

Within CyberSec4Europe's privacy-preserving architecture, Sharemind belongs in the Security/Privacy preservation tools section of the Managed domain and in the PET clients of the User domain.

Work on Sharemind is still ongoing. We are currently working on several feasibility studies concerning secure GWAS taking into account population stratification, detecting financial fraud in a privacy-preserving manner, and privacy-preserving machine learning for image analysis.

The Sharemind MPC SDK and Emulator (not an actual MPC engine) can be downloaded from https://sharemind.cyber.ee/sharemind-mpc/ This SDK allows to get acquainted with the C-like Sharemind MPC language secrec. A developer can compile the program and run it in an emulator which will estimate the running time. The other licencing options and their differences can also be seen on the same page.

5.7.3 Relationship to WP4 Roadmap

Our work in financial fraud detection contributes to Task 5.1 (Open Banking), Challenge 3 (Cross-border cooperation under different legislation and security controls) described in Section 4.4.3 of D4.3.

In Task 5.6 (Medical Data Exchange), Sharemind can be one enabler that can help deal with challenges described in 9.4.1 (Security and Privacy) of D4.3.

Our work with smart grids contributes to Task 5.7 (Smart Cities), namely to the challenges described in Challenge 4 (privacy by design) of Section 10.4.1 (High Priority Challenges) of D4.3.

5.7.4 Relationship to WP5 Demonstrators

Currently we are not planning on integrating this enabler to any demonstrator. However, we will show, as a proof of concept, that the challenges can alternatively be addressed using secure multi-party computation, and also measure the feasibility of the MPC solution.

5.8 pTASC: Privacy-Preserving Middleware

The ever-increasing number of interconnected devices in smart environments, i.e., homes and cities, is bolstering the amount of data generated and exchanged. These devices can range from small embedded platforms, such as those included in home appliances, to critical operational systems, such as traffic lights. However, this increasing adoption is raising significant security and privacy concerns. Although some researchers have already solved some of these issues, data privacy still lacks a viable solution, especially when considering a flexible, decentralized approach to avoid a central overseer. One of the biggest challenges regarding privacy is the lack of transparency about how data flows are mediated and regulated as, often, these resources share data with external entities without the users' knowledge. We argue that a

novel data-sharing control mechanism is required to properly control users' privacy and their respective Internet of Things (IoT) devices. This work focuses on a middleware layer solution for the IoT devices, which allows the control of the data generated by the device by its owner. The platform places the user as an active participant in the data market, behaving as its own data intermediary for potential consumers by monitoring, controlling, and negotiating the usage of their data.

The middleware consists of a secure data sharing model. In Figure 14 we have the architecture of the system. It consists on multiple IoT devices, the owner's smartphone (that acts as Permissions Controller), and a router that acts as a manager and controls data sharing for external entities..

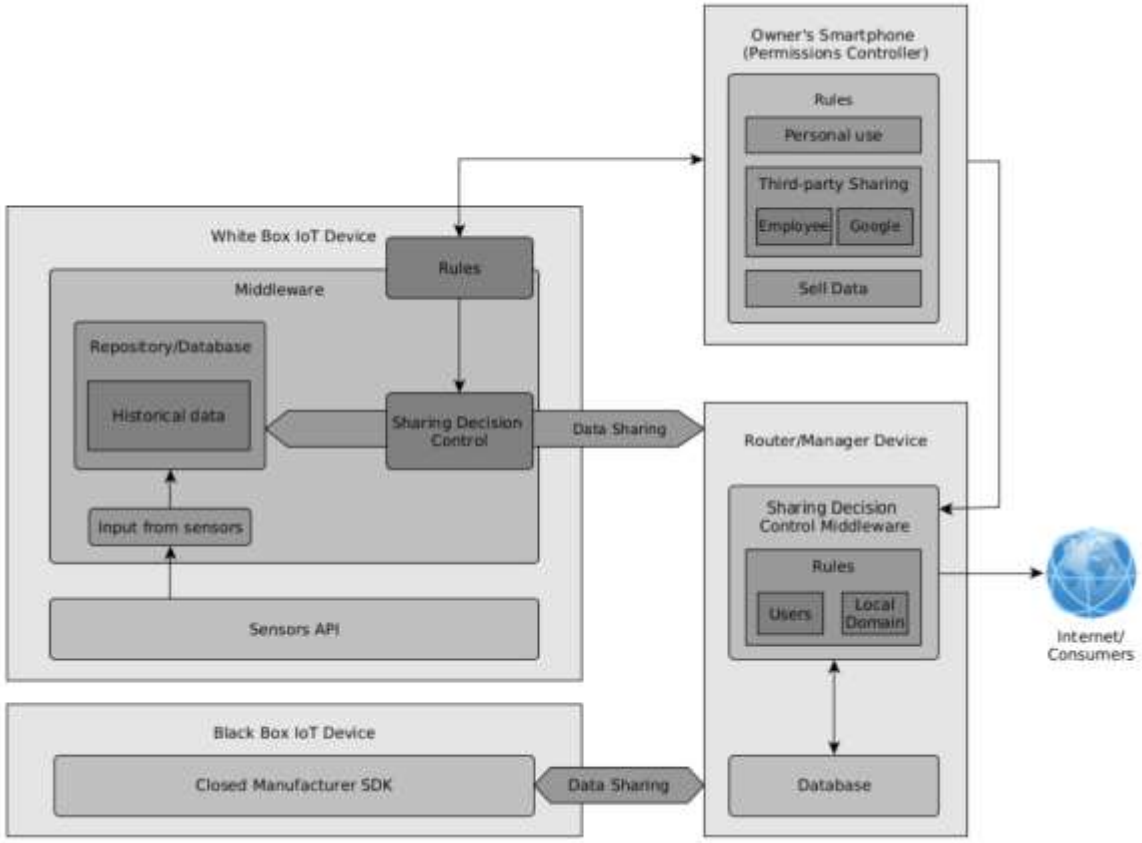

Figure 14: pTASC architecture

In our approach, users can decide the data and traffic exchanged according to their preferences. Users have the option to block all traffic by default and to make exceptions for some specific domains. Therefore, for example, users can block marketing/advertising sites and only communicate with the manufacturer's domains. Note that if users choose to block all communications from their devices to the Internet, some of the features may stop working, as some of these devices will not work in offline mode. Finally, users will have to choose between usability and privacy.

On existing routers, users can change the set of rules and block specific traffic, without the need for a middleware. However, there are no configurable privacy platforms that show connections made by default by the device (configured by the manufacturer), and that allow blocking those connections "on the fly". The

middleware allows users to monitor incoming and outgoing connections so that users can verify that the device is running an untrusted program and block or disable updates to specific resources.

Along with this network traffic monitoring and, depending on the manufacturer's device firmware, users can store data offline on their home router for future reference. As users can connect with multiple routers (at home or work, for example), they can choose different permissions for their data depending on the device context, managing the data's life cycle. Users can also monetize data by selling them to external entities.

Authentication

The solution focuses on using a PKI where the CA is represented by a manager device that can be switched on/off to reduce Single Point of Failure (SPOF) problems. It combines public-key cryptography and symmetric keys with the One Time Password (OTP) concept using a secure token. After the discovery process, devices need to authenticate with each other. For mutual trust, both devices must exchange manager's signed certificates. After verifying the authenticity of certificates, a symmetric key is generated between both devices to establish secure communication. This authentication method will be used to guarantee the establishment of secure TLS channels and thus prevent potential impersonation attacks. This method can handle multiple managers that can potentially agree on exchanging information about the devices associated with them.

Considering the challenges reported in Section 2, pTASC targets the following challenges:

- **DP-05** by allowing users to keep their personally sensitive information under control.
- **DP-06** by providing mechanisms for control how the information is disseminated, and control the private data on the communication.
- **IDP-05** by providing a mechanism for guarantee a proper identity management of things for authentication end-to-end

5.8.1 Privacy-Preserving Properties

- In IoT, every device or appliance will be connected to the internet and collecting data about consumers. This challenges some of the most fundamental principles of privacy: informational self-determination, data minimization, consent and the right to individual access. Therefore it is mandatory to find mechanisms to ensure an adequate level of privacy protection, user empowerment, through new applications and services based on access to personal information.
- pTASC manages the identity management of the connected things without a PKI focused on context-aware approaches to control the things, exploring end-to-end secure communications.
- pTASC empowers users with mechanisms to control and manage consent to share their data, giving users the possibility to configure their own policies, preferences and terms of adherence, and do it in ways that can be automated both for them and for the organizations they engage. Data owners are able to understand the complete picture of what is going to happen behind the scenes to their data and what will they be able to receive in return by allowing access to that data.
- The platform places the user as an active player in the data market, behaving as its own data broker for the potential data end users.

- Users can store data to be controlled offline and to analyze current connections, discarding them according to the user's preferences, without delay, extensible to network communications.
- Unlike previous work, the developed middleware is independent of the device's SDK, as control is placed at the device and router layers, allowing users to fully control the shared data.

5.8.2 Relationship to WP3 Research and Assets

The pTASC has been designed as a standalone enabler for the design and enforcement of a privacy-by-design GDPR-based data life cycle, where the user decided which are the company accessing and managing it's data. However the different pTASC components can be customized for the integration into the CyberSec4Europe Architecture, for different goals such as specification, design and testing and assessment of control about the user data and consent.

This asset has been conceived as a software component based on containerization technologies, which notably facilitates its deployment on presumably any Edge computing environment. In particular, we are focused in the implementing measures to be capable of integrating PTASC in the Fiware infrastructure.

Additionally, the modularity of PTASC lets its integration with different WP3 enablers. For instance:

I. The system can be integrated with DANS (Section 5.1.3) in order to allow the user to have a control and impact assessment about the data that is selling or sharing with other parties.
II. Testing facilities included in PTASC enabler, can be used for validating either the Access Control Policies (or Access Control mechanisms) it will allow collaborate with the Privacy Manager Enabler (Section 5.6).

5.8.3 Relationship to WP4 Roadmap

Considering the Challenges reported in the D4.3, Section 10.4 Smart Cities, pTASC focuses mainly on Challenge 4: Privacy by design by enhancing the Identity Management component with protocols and frameworks for authentication, authorization, and rights management, along with privacy and identity management. The secure token generates OTPs and also stores the public key of the manager to be transmitted to trusted devices. This prevents the manager from being impersonated, even if the attacker can change its name, the attacker needs to have their data (authenticated OTP and its public key) in the physical secure token, which becomes impractical. The use of secure token helps protect against hackers, as physical access to the secure token is necessary to generate the OTP.

pTASC also focus on Data Security and Privacy in the communications. After both exchange their certificates to prove that are trusted, a symmetric key is generated between the two devices to establish communication.

For symmetric key generation, we can use some type of algorithm such as Diffie-Hellman if both nodes can send messages to each other. To authenticate both peers, we use ECSDA for signing and verification and ECIES for encryption. Diffie-Hellman (DHE, where final 'E' stands for "ephemeral") or Elliptic-curve Diffie-Hellman (ECDHE, where final 'E' stands for "ephemeral") is used for key exchange. The ephemeral is important to use in this case, because it helps in the security. If it is used always the same key in the generation of the previous master secret, and if this generation key is compromised, the master secrets of all communication sessions may be compromised. Compromised. Using the key pair for an extended period of time creates the need for storing the keys in some secure place since devices can be shutdown. There is always some risk a stored key pair can be compromised, although a wide variety of methods can be and are used to try to prevent compromise. The ephemeral mode avoids this risk of attack not storing the key pairs on the device. If, on the other hand, it is used a new key pair every millisecond, Perfect Forward Secrecy (PFS) is said to exist. In brief, as the peer already know each other, it is established a shared secret using

ECDHE. When the devices have obtained the shared secret, they can exchange data with symmetric encryption, in this case, it is used AES256.

However, for all the characteristics that pTASC has, it partially targets also the following challenges: 6 and 10, mainly on Data Security and Privacy and Identity Management.

5.8.4 Relationship to WP5 Demonstrators

pTASC can be integrated in both **Porto** and **Murcia** systems (WP5, Task 5.7). The integrations must be created for third party be capable of buying data and the user have the options of sell or not that information (if not anonymized). For reference we can check Figure 14, where should be added some controls to perform the Management by the user and control of outgoing traffic to the network.

In Porto the connections will be added on top of the architecture as a middleware that limits the access to the data for all the users behaviour as a central point of data aggregation.

For both of this cases we are starting the integrations between our approach and Fiware.

5.9 ARGUS: Cloud of Clouds Storage System

Cloud storage allows users to store their data remotely, giving access anywhere and to anyone with an Internet connection. The accessibility, lack of local data maintenance, and absence of local storage hardware are the main advantages of this type of storage. Its convenience is driving the adoption of this type of storage. However, one of the main barriers to its widespread adoption is the sovereignty issues originated by a lack of trust in storing private and sensitive information in such a medium. Recent attacks on cloud-based storage show that current solutions do not provide adequate security levels and, subsequently, fail to protect users' privacy. Usually, users rely solely on the security supplied by the storage providers, which will ultimately lead to data leakage in the presence of a security breach. In this system, we propose and implement a broker (ARGUS) that acts as a proxy to the existing public cloud infrastructures by performing all the necessary authentication, cryptography, and erasure coding. ARGUS uses erasure code as a way to provide efficient redundancy (opposite to standard replication) while adding an extra layer to data protection in which data is broken into fragments, expanded and encoded with redundant data pieces that are stored across a set of different storage providers (public or private). ARGUS's key characteristics are confidentiality, integrity, and availability of data stored in public cloud systems.

ARGUS is a broker that acts as a proxy to the public cloud infrastructures by performing all the necessary authentication, cryptography, and erasure coding. By doing so, it offloads the computational workloads from clients. Our approach ensures confidentiality, integrity, and availability of the data in the public cloud systems.

To ensure user data security and privacy, the user can opt to encrypt everything locally and be responsible for the key management. In this way, the user does not need to rely on a third party. Also, we can delegate the encryption to a third party, which has the benefit of reducing the cost of execution on a limited device. Confidentiality, Integrity and

ARGUS maintains the integrity of the data as it stores an HMAC of all files. The confidentiality of the data is ensured as the data are encrypted, and the user can save their private key locally. ARGUS provides high-availability through the redundancy that is assigned in the different cloud providers; that is, information is redundant on the three public clouds.

The system uses Intel's CPU SGX extensions to cipher user credentials (access tokens that give access to the user's public cloud storage). This is an improvement over current implementations in systems that use the Google Drive API because the credentials are stored locally in the file system.

Considering the challenges reported in Section 2, ARGUS targets the following challenges:

- **DP-03** by providing sharing mechanism with MPC, with pre-processing capabilities to reduce the complexity of processing with a hybrid solution.
- **DP-07** by providing anonymization mechanism for control on how the information is stored, and control the private data on the communication.
- **DP-08** by introducing the possibility of the user store the files locally or in a private cloud depending in the user requirements.
- **IDP-06** by ensuring that the encryption keys of the HTTPS protocol are manipulated in clear text only inside a trust zone (negotiating all the cryptographic material only inside the enclave).

5.9.1 Privacy-Preserving Properties

- ARGUS maintains the integrity of the data as it stores an HMAC of all files. The confidentiality of the data is ensured as the data are encrypted, and the user can save their private key locally. ARGUS provides high-availability through the redundancy assigned in the different cloud providers; that is, information is redundant on the three public clouds. \
- Given the amount of data present in the cloud, particularly for sharing purposes, we integrate an adjustable FEC mechanism that uses privacy tools, such as the Health Insurance Portability and Accountability Act of 1996 (HIPAA). The use of multiple cloud providers, including public and private, allows for a level of redundancy that can be adjusted on-the-fly to face privacy conditions better;
- When the system detects a HIPPA identifier or other privacy alert through the runtime privacy schemes, it uses a private cloud to store that particular data, which is essential before sharing information with other parties;
- The cloud-of-clouds deployment ensures that the data at rest is not accessible to any public cloud provider. The different clouds only have the vision of a part of the encrypted data, even if they uncover the encryption keys, they only access a portion of the file;
- We use public key infrastructure (PKI) as the basis for our secure infrastructure, which is especially essential for the usage in the https communication. For this, a module that secures the private key behind a secure enclave is being integrated, never exposing it to attackers since it never leaves the enclave, which works as an extra layer of security to ensure that even the system is compromised, there is no leakage of data through a server attack.

5.9.2 Relationship to WP3 Research and Enablers

The ARGUS has been designed as a framework that enables the design and enforcement of a privacy policy introduced by the GDPR-based data life cycle. In this case, ARGUS allows the usage of MPC, ARX, and other tools. All of this integration will enable us to build a reliable system that addresses the current challenges of data sharing and management using the low price of public clouds. The different ARGUS components are customized for integrating into the CyberSec4Europe architecture or integrating other tools in the architecture for various goals such as specification, design, and testing.

This asset has been conceived as a software component based on JAVA and keycloak, which notably facilitates its deployment on presumably any IoT device. In particular, we are focused on the implementation of mechanisms to be capable of integrating ARGUS in the FIWARE infrastructure.

Additionally, the modularity of ARGUS lets its integration with different WP3 enablers. For instance:

I. The system can integrate to DANS (Section 5.1.3) to allow users to have a control and impact assessment about the data that they intend to share, replacing the ARX framework.
II. The system can also integrate with PLEAK (Section 5.3) to allow a privacy-by-design approach already on the level of business process discovery and design to generate policies adopted by the users.

5.9.3 Relationship to WP4 Roadmap

Considering the Challenges reported in the D4.3, Section 10.4 Smart Cities, pTASC focuses on Challenge 1: Digital Trust Platform mainly on Data Security and Privacy with design, implementation, and operation of data management systems that include security and privacy functions that provide anonymity, pseudonymity, unlinkability, undetectability, or unobservability. Also, ARGUS includes Privacy Enhancing Technologies (PET) to the data. Also, Challenge 4 and 10 are partial included on the ARGUS, mainly on the topics of Data Security and Privacy and Trust Management and Accountability.

5.9.4 Relationship to WP5 Demonstrators

ARGUS can be integrated in both Porto and Murcia systems (WP5, Task 5.7). In both demonstrators, this system can be used for persistent data storage in the cloud using the ARGUS to speed the data transfer and security, allowing for a simple implementation at the edge. Also, with the privacy sharing policies (MPC and anonymization features), it will allow for sharing the data between both cities. For both of these cases, we are starting the integrations between our approach and FIWARE.

5.10 Cloud Based Anonymous Credential Systems

Anonymous credential systems as introduced by Chaum [142] and later refinded by Camenisch et al [78, 79, 143] allow one to strongly authenticate oneself without identification. The user can prove to the relying party that she owns certain attributes (e.g., she is sufficiently old to enter a resource), without revealing her identity or her precise birth date.

Yet, traditional anonymous credential systems are very expensive on the end-user's device, and therefore not suited for many applications, e.g., using low-cost devices such as smart cards. This is overcome by so-called cloud-based (encrypted) anonymous credential systems [144, 145], where virtually all computation is outsourced to a semi-trusted cloud provider, and only very minimal computations need to be performed on the end user side.

For a more detailed description, we refer to the original literature and to D3.2 (Section 3.1).

5.10.1 Privacy-Preserving Functionalities

Cloud-based anonymous credential systems contribute to increasing the user's privacy in multiple ways, in particular by addressing the risk of over-identification (IDP02).

- Following the main ambition of anonymous credential systems, this enable puts the user back into full control over her own data. That is, the user has fine-granular control over which data goes where, and which third party is able to access which data. Furthermore, the user can only disclose

parts of her attributes to the third party, e.g., her birth date but not her name, or her country of residence but not her full address.

- Except for some low-cost computations on the user side (basically corresponding to one Diffie-Hellman key exchange), no computations need to be performed on the user side. All the remaining computations are outsourced to a so-called cloud wallet. By this, the usability aspect of (IDP02) is directly addressed, as the technology can directly be deployed on most hardware components. Furthermore, the message flow allows high compatibility with existing industry standards such as OpenID Connect.
- In order to guarantee privacy also against the cloud wallet provider, several measures have been developed and refined by Krenn et al [144, 145]. On the one hand, all information stored in the cloud wallet is protected using advanced encryption schemes (proxy re-encryption [146]), such that the wallet provider does not require plaintext access to the user's sensitive attributes at any time of the computation. Furhermore, with the improvements provided in [145], even colluding service providers and wallets can not access any information that the user did not intentionally decice to share with the service provider. Finally, if the network level is anonymized using appropriate measures (e.g., TOR [147]) a variant of the protocol even allows the user to hide from the cloud wallet provider for which service provider it is currently computing an authentication token.
- Multiple actions by the same user, e.g., authentications to the same or different service providers, cannot be linked if the network level is appropriately anonymized.

5.10.2 Relationship to WP3 Research and Enablers

As described in D3.2, this enabler is in particular related to the following WP3 privacy enablers:

- Self-Sovereign & Privacy-preserving SS-PP-IdM (see Section 5.1)
- Mobile Privacy-Attribute Based Credentials (see Section 5.2)

While aiming for other specific aspects of the underlying technology, these enablers are also improving and extending the core concept of anonymous credentials.

5.10.3 Relationship to WP4 Roadmap

Cloud-based anonymous credentials directly relate to the challenge of "Unlinkability and minimal disclosure" presented in D4.3 (Section 6.4.2). Future extensions might include the computation of predicates on attributes also in this cloud-based setting, or distributing the wallet across multiple service providers in order to increase availability and further reduce trust (cf. also the challenge of "Distributed oblivious identity management" in Section 6.4.3 of D4.3).

5.10.4 Relationship to WP5 Demonstrators

While the Privacy-Preserving Identity Management demonstrator (T5.3) is directly addressing the same challenge as this enabler (IDP02), the demonstrator will mainly contributing to showcasing the availability of mature solutions and thus raising awareness. Due to the specific selected use case, cloud-based anonymous credentials will not be deployed in the first piloting round of the demonstrator, and a final decision for the second piloting round has not yet been made.

5.11 ArchiStar Distributed Storage

Secret sharing [148] allows one to decompose a piece of information into a variety, say *n*, shares, such that only predefined subsets of these shares can later be used to recover the secret information, while no other set of shares contains any information about the data. This protection can be computational or even

information-theoretical, where in the latter case not even a computationally unbounded adversary would be able to reconstruct the secret. The most promiment is the scenario of *threshold* secret sharing, where any set of at least *t* shares is qualified to reconstruct, while no set of *t-1* shares can do so.

ArchiStar/SECOSTOR [149] is a distributed storage solution based on secret sharing, where the different shares can be stored to different cloud providers. As explained in detail in [149] and D3.2 (Section 10.1), ArchiStar/SECOSTOR can signifciantly increase the availability of cloud storages at low overhead while at the same time positively impacting the confidentiality and privacy of the data. Furthermore, due to it's open source code and compatibility with any storage provider (up to possiblly additional interfaces required on the ArchiStar/SECOSTOR side), the tool also prevents against vendor-lockins. Finally, due to it's keyless nature, high usability and reduced key management efforts can be guaranteed.

5.11.1 Privacy-Preserving Properties

The privacy-preserving aspects of ArchiStar/SECOSTOR can be summarized as follow:

- No cloud storage provider gains any information about the data stored by the user. At the cost of higher storage overheads, this benefit can even be made information theoretic, such that even unbounded malicious storage nodes can gain any information. This also holds in the case of leaked data, e.g., because of attacks, inadequate erasure before disposal, etc. as the leaked data does not contain any information about its content.
- Users are given long-term privacy and confidentiality guarantees for their data, as regular proactive re-sharing steps can be deployed in order to invalidate any shares that might have been leaked without notice. By doing so, the time windows within which an adversary needs to compromise at least *t* shares can be adjusted arbitarily, though there might be practical limitations on the frequency of this step due to the resulting communication overhead.

ArchiStar/SECOSTOR supports the user in enforcing their right for erasure, as it is sufficient if at least *n-t+1* storage nodes securely erasure the user's data upon request, as the remaining *t-1* shares could not be used for reconstruction.

5.11.2 Relationship to WP3 Research and Assets

ArchiStar/SECOSTOR is related to the following privacy enablers within WP3:

- Sharemind MPC - Privacy-preserving data analysis (see Section 5.7)
- FlexProd - Integrity-Preserving Data Analytics (see Section 5.12)

Both these enablers deploy secret sharing as an initial step before performing privacy- and integrity-preserving data analytics based on secure multi-party computation. While ShareMind MPC was developed independently, FlexProd directly leverages ArchiStar/SECOSTOR for the decomposition and storage of the user's input data.

5.11.3 Relationship to WP4 Roadmap

Secure and privacy-preserving storage of potentially sensitive information with high availability guarantees yet low overhead - even compared to classical replication - can provide added value to numerous application domains and use cases. Specifically considering the CyberSec4Europe's research roadmap developed in D4.3, ArchiStar/SECOSTOR can contribute to at least the following challenges. For martime transport, the availability and redundancy guarantees, as well as the lack of any single point of failure in the storage system, directly contributes to increasing the resilience of critical systems against data loss or compromise

of a single storage entity. In the medical data exchange domain (Section 9.4.1 of D4.3), ArchiStar/SECOSTOR directly relates to the challenges of privacy, in particular against the storage provider, and in case of compromise of a storage node. Furthermore, the risk of data loss can be minimized due to the built-in redundancy of the storage concept. In addition, related to the challenge of trust (Section 9.4.2), ArchiStar/SECOSTOR avoids any single point of failure, thereby reducing the necessary trust in the data sharing platform itself.

5.11.4 Relationship to WP5 Demonstrators

As outlined in D3.2, ArchiStar/SECOSTOR addresses challenges sketched by various demonstrator cases of CyberSec4Europe (specifically, T5.1, T5.2, T5.5, T5.6, and T5.7), cf. also D5.1, in particular regarding the protection of storage of sensitive information. However, at this point, the deployment in the actual demonstrator cases is still unclear.

5.12 FlexProd: Integrity-Preserving Data Analytics

Secure multi-party computation (MPC) allows a set of entities to jointly compute a function on private inputs. That is, the parties can agree on a function to be evaluated, and each party only learns the output intended for this party, but nothing else. In particular, the inputs of all parties are kept confidential (up to the degree that they are protected by the function to be evaluated).

Besides protecting the privacy and confidentiality of the inputs, FlexProd further aims at also guaranteeing the authenticity and correctness of the result. While in plain MPC these properties follow from the assumption that sufficiently many computation nodes behave honestly - or at least do not collude - FlexProd aims at giving formal cryptographic guarantees of the correctness of the computation.

This is achieved by combining secret sharing (cf. ArchiStar/SECOSTOR, Section 5.11) with zero-knowledge proofs of knowledge [150], which allow one to proof the validity of a claim without revealing anything else than what is already revealed by the claim itself.

For a detailed description of the enabler, we refer to D3.2 (Section 10.3).

5.12.1 Privacy-Preserving Properties

The privacy-preserving aspects of FlexProd can be summarized as follow:

- Secure multi-party computation allows for privacy-preserving analytics of user data, potentially from different data sources, without revealing the user's input data. For instance, FlexProd was initially designed for realizing a notary services for online auctions; in such cases, only the highest bid will be disclosed, without revealing the bids of the defeated bidders.
- At no point during the computation, the user needs to rely on policies or privacy agreements, but has technical guarantees that her data is not revealed to any of the compute nodes.

5.12.2 Relationship to WP3 Research and Assets

FlexProd is closely related to the following WP3 enables:

- Sharemind MPC - Privacy-preserving data analysis (see Section 5.7)
- ArchiStar/SECOSTOR Secure Distributed Storage (see Section 5.11)

Sharemind MPC provides a mature and high-TRL framework for privacy-preserving data analysis. While FlexProd does not yet have this high TRL, Sharemind MPC does not come with cryptographic guarantees regarding the integrity and authenticty of the result. ArchiStar/SECOSTOR is used as a building block for FlexProd, providing the secret sharing mechanism needed in the initial phase of MPC.

5.12.3 Relationship to WP4 Roadmap

Secure multiparty computation in general, and specifically FlexProd, can contribute to overcoming privacy and integrity challenges in numerous domains where data from different stakeholders needs to be combined or analyzed without revealing single data points for confidentiality reasons. Looking at CyberSec4Europe's research roadmap, e.g., the privacy and trust challenges in the medical data exchange domain (cf. Sections 9.4.1 and 9.4.2 in D4.3) could be addressed using FlexProd. For instance, by offering means for analyzing sensitive patient data across numerous patients for research purposes without the requirement to ever access the information of a specific patient in the plain, the privacy of patient data is protected by cryptographic means, and the patients' willingness to privately "donate" their data for research purposes might increase.

5.12.4 Relationship to WP5 Demonstrators

As outlined in D3.2, FlexProd could be deployed in a variety of demonstrators. Specifically, requirements from the following demonstrators have been identified:

- T5.1 - Open Banking
- T5.2 - Supply Chain Security Assurance
- T5.6 - Medical Data Exchange

We refer to D3.2 for the specific lists of requirements.

5.13 Data Anonymization Service (DANS)

The large amount of medical information generated by citizens, wearable companies, hospitals, laboratories, pharmaceutical industry or health public organizations need to be protected (accomplish GDPR) and the data owner privacy must be preserved. Data anonymization is a privacy protection technique allowing sensitive data analysis but preserving user privacy.

Due to the high value of this information de-anonymization methods can be exploited by attackers revealing personal and sensitive information from the data owner for malicious purposes. DANS enabler will facilitate the anonymization of any set of data, so it will help with the anonymization of the sensitive data shared through the medical data exchange platforms. By using generalization and masking techniques for transforming data, and privacy and risk models such as k-anonymity, this sensitive data can be protected properly without compromise their final use for analytic purposes. As indicated in deliverable D3.2 [233] DANS enabler is based on a data anonymization Java open-source tool (ARX[15]), which is comprehensive tool for anonymizing biomedical data [234].

5.13.1 Privacy-Preserving Properties

Among the different techniques for protecting sensitive data during the sharing process the data anonymization is a common practice. DANS enabler provides different models that analyse the re-identification risks, and *k-anonymity* and *l-diversity* privacy models for managing medical data, as well as generalization for data transformation. Additionally, it is able to be anonymized datasets with a large amount of records, which makes it really scalable [234].

[15] https://arx.deidentifier.org/

These features address the following challenges identified in Section 2:

- **DP-07** by providing anonymization methods that detect an protect personal and sensitive identifiabale information when data are managed out of the user's control;
- **IDP-01** by providing models that inform about the re-identification risks;
- **IDP-04** by transforming medical data the ability to obtain information beyond the necessary is limited;
- **LDP-01** is partially covered, as DANS is providing a statistical report regarding the anonymization process.

5.13.2 Relationship to WP3 Research and Assets

The Edge-Privacy enabler is a privacy manager for IoT. The privacy manager allows the users to retain control of the data collected by the IoT device, and also the definition of privacy policies and privacy filtering mechanisms for data transformation. The data filtering module could include some privacy models provided by the DANS enabler.

The DANS enabler is included in the web domain of the CyberSec4Europe global architecture and is part of the security and privacy preservations tools for preserving the user privacy data (Section 4, Figure 5) [233].

The DANS enabler will provide an anonymization procedure for managing different categories of medical data as follows:

- **Identifying** attributes will be removed from the input, as they can be re-identified easily;
- **Quasi-identifying** attributes will be transformed accordingly with the specified transformation procedures (generalization, micro aggregation, …). If combined can be used for re-identification attacks;
- **Sensitive** attributes will be kept as-is, but they can be protected using privacy models, such as t-closeness or l-diversity;
- **Insensitive** attributes will be kept unmodified.

Basically, the transformation model used by DANS is the generalization. This procedure needs **hierarchies** associated to the attributes of the input data. A hierarchy to be applied to an attribute can be defined by means of a *csv* file, in order to ease the specification. The k-anonymity privacy model will be used for managing the quasi-identifying attributes. This privacy model is the most used for managing health data. l-diversity is the selected privacy model for sensitive attributes treatment.

DANS enabler is provided as a service and can be deployed in different environments (data exchange platform, data provider, third parties).

The DANS version 0.2 provides a REST API with two services:

- POST /dans/uploadFile/
- POST /dans/file/

POST /dans/uploadFile/		
Interface	**uploadFile** (file)	
	Description	This operation will be used to upload not only the file to be anonymized but also upload the files specifying the required *hierarchies*
	Inputs	• **inputFile** parameter contains the file to be anonymized • **additionalMetadata** parameter indicates the file format ("csv" format or "xls"/"xlsx"
	Output	**identifier** for the file just uploaded
Constraints / Observations		• As this component requires to operate with a plain version of the data provider file key, this component must be "trusted" by the data provider

POST /dans/file/		
Interface	**file** (fileId)	
	Description	This operation anonimyzes the identified file, following the directives provided for the anonymization process
	Inputs	• **fileId** parameter contains the identifier returned by the uploadFile service • **daNSinput** parameter provides the specification for anonymizing that file
	Output	**file** anonymized csv/excel file or the anonymized process report pdf file
Constraints / Observations		It is necessary to select the **Response Content Type**: • Text/plain: if you want to obtain the anonymized data in a cvs format. • Application/pdf: if you want to get a pdf report of the anonymization process.provider

5.13.3 Relationship to WP4 Roadmap

The roadmap described in D4.3 shows the adoption of the anonymization enabler by the medical data exchange platform. This anonymization enabler faces the challenge 2 "Mechanisms for preserving user data privacy" described in D4.3 (Section 9.5.2). In the first stage this enabler is provided as a service, allowing the users to anonymize their data before the sharing process. In a second stage the anonymization enabler will be provided for using in smart devices.

5.13.4 Relationship to WP5 Demonstrators

DANS will be used for the use case MD-UC2 "Sharing Sensitive Health Data through Files" belonging to the medical data exchange demonstrator of Task 5.6. The use of DANS enabler will provide to the users of the medical data exchange platform a privacy-preserving tool of sensitive medical information.

Figure 15 depicts how the DANS enabler is interacting with the medical data exchange platform and how the anonymization process is performed.

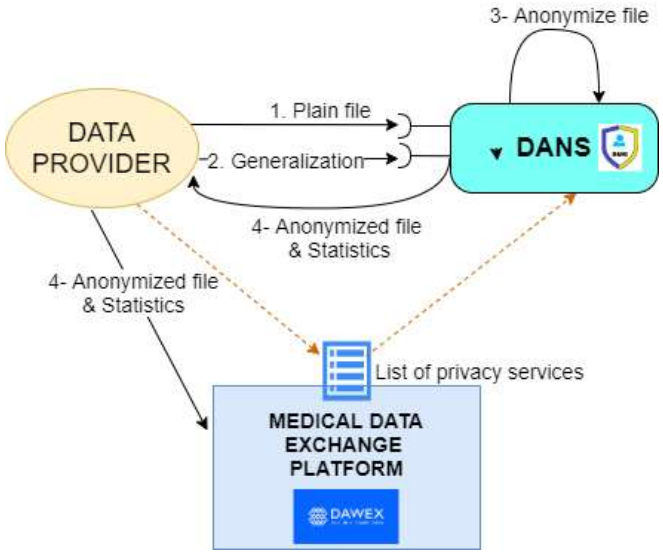

Figure 15: Anonymization Process and DANS - Data exchange platform interaction.

5.14 Service Provider eIDAS Integrator (SPeiDI)

The European Commission main strategy for Europe digital transformation is "*empowering and including every citizen, strengthening the potential of every business and meeting global challenges with our core values*"[16]. Building a secure and trust digital environment is crucial for increasing the use of online services by European citizens, In this context the use of national eID means issued by the trusted bodies in each EU Member States allows citizens and business to access online public services and doing business in a cross-border way[17]. eIDAS regulation[18] addresses this issue forcing the cross-border recognition of eID means for public trusted services on September 2018.

Currently the EC is trying to facilitate the use of national trusted eID means when citizens accessing online private services requiring authentication mechanisms based on eID[19]. Moreover, the use of eID means based

[16] https://ec.europa.eu/digital-single-market/en
[17] https://cyberwatching.eu/policy-landscape/privacy/eidas
[18] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2014.257.01.0073.01.ENG
[19] https://ec.europa.eu/futurium/en/system/files/ged/draft_principles_eid_interoperability_and_guidance_for_online_platforms_1.pdf

on trusted EU Member States schemas in the financial sector are helpful for fighting against money laundering[20].

The citizen's risk and trust perception of public and private online services are very important for accepting their use [236] [237], being the trust the powerful aspect. Additionally, the use of the eID means and the identity management systems is affected by the citizen's risk and trust perception [238].

SPeIDI enabler is based on based on the eID building block[21] provided by CEF following the eIDAS technical specifications [239] and allows to private digital services the integration with the eIDAS network. In this way the online platforms offer a secure access to their digital services by using strong authentication mechanisms based on the use of trusted eID means- On the other hand citizens are accessing to trusted online services offering a secure access.

5.14.1 Privacy-Preserving Properties

SPeIDI enabler facilitates the user identification when accessing online services, providing the minimum user data needed for a proper identification. Moreover, the strong authentication mechanism provided by the eIDAS network assures the person on the other side of the channel is who claims to be. These features address the following challenges identified in Section 2:

- **DP-06:** by providing mechanisms assuring the data are realeased to trusted parties by using keys and secure protocols;
- **IDP-02** by providing the minimum user data set provided by the eIDAS network. The user has control over the optional data to be disclosed.
- **IDP-04:** The service provider connected to the eIDAS network must be register in advance on the organization managing the country eIDAS node where the service provider belongs to. A secure channel and a protocol of recognition is established between the service provider and their country eIDAS node. Additionally, the country identity provider is a trusted organization authorized by the Member State authorities.
- **LDP-03:** SPeIDI enabler is based on eID blockchain and eIDAS technical specifications, which are following the eIDAS regulation that reinforces the rules of GDPR.

5.14.2 Relationship to WP3 Research and Assets

The Self-sovereign privacy-preserving IdM in blockchain (SS-PPIdM) enabler allows users a total control and management of their personal identity attributes in a decentralized manner, it means that no central authority is needed. The data owner acts as a data controller and take the decision on how to use their data during the online transactions [233]. The SS-PPIdM enabler is based on the SSI model and the blockchain technology. and associated entities such as issuer authorities. One of these issuer authorities could be the eIDAS network. In this case the user can receive the eIDAS credential through the SPeIDI enabler and store and manage this credential by their own.

[20] https://ec.europa.eu/info/business-economy-euro/banking-and-finance/financial-supervision-and-risk-management/anti-money-laundering-and-counter-terrorist-financing_en
[21] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eID

The SPeIDI enabler is included in the web domain of the CyberSec4Europe global architecture and is part of the Identity & Trust Management services which is included in the Control and Management plane (Section 4, Figure 5) [233] .

The SPeIDI enabler receives an authentication request from the service provided as a JWT token for facilitating the integration with the eIDAS network. This JWT token will be translated to a SAML request, which will be provided to the country eIDAS node. Once the user is authenticated against the identity provider the country eIDAS node provides a SAML response which will be translated to a secure JWT token containing the user information and provided to the service provider.

The SPeIDI enabler offers two interfaces:

- A unique endpoint is offered to the service provider for starting the authentication process: *POST /authenticate {JWT token}*
- A ReturnPage endpoint is offered to the eIDAS country node for receiving the SAML response the country eIDAS node sends: *POST /ReturnPage.*

POST /authenticate/		
Interface	**authenticate** (JWT token)	
	Description	This end point will provide a JWT token containing the authentication request and the requested user identity information and signed with a secret key
	Inputs	• **JWT request token**
	Output	• **JWT response token**
Constraints / Observations		• As this component requires to connect to a trusted country eIDAS node, this component must be “registered” in advance on the country eIDAS node system

POST /ReturnpPge/		
Interface	**ReturnPage**	
	Description	This servlet receives an asynchronous SAML response, which must be validated and deciphered.
	Inputs	• **SAMLResponse**
	Output	• **Response code**
Constraints / Observations		As this component requires to connect to a trusted country eIDAS node, this component must be “registered” in advance on the country eIDAS node system

5.14.3 Relationship to WP4 Roadmap

The roadmap described in D4.3 shows the integration of the eIDAS authentication enabler by the medical data exchange platform. According to the challenges described in D4.3 (section 9.5 Research Challenges for medical data exchange), the use of this enabler addresses the challenge 1 “Security and privacy” (Section 9.5.1), protecting the access to sensitive data such as medical records.

5.14.4 Relationship to WP5 Demonstrators

SPeIDI will be used for the use case MD-UC3 “Enhancing the Security of On-Boarding and Accessing the Dawex Platform” belonging to the medical data exchange demonstrator of Task 5.6 [235]. The use of SPeIDI enabler will provide to the medical data exchange platform a strong authentication mechanism with a high level of assurance. Therefore, only the allowed users get access to the data.

The next figure depicts how the SPeIDI enabler is integrated with the medical data exchange platform.

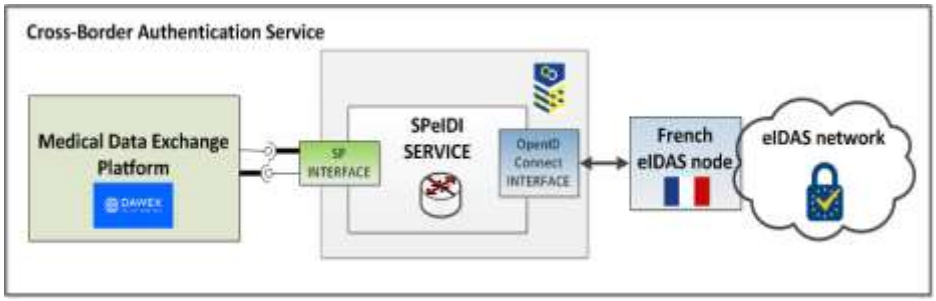

Figure 16: SPeIDI - Data exchange platform interaction.

6 Conclusions

This document presented D3.11, titled “Definition of Privacy by Design and Privacy Preserving Enablers”. It focused on CyberSec4Europe’s enablers, also known as “assets”, that deal with privacy and privacy-preserving functionalities.

The discussion started with the identification of a number of challenges in today’s privacy research landscape that the project wants to tackle, and an exhaustive analysis of the state of the art in selected categories of privacy research.

The description of CyberSec4Europe’s privacy-preserving architecture served as an introduction to the assets’ catalogue. Assets come with an explanation of their inner workings an privacy-preserving functionalities, as well as their relationship to CyberSec4Europe’s core research and development work packages, that is, WP3 - “Blueprint Design and Common Research”, WP4 - “Research and Development Roadmap”, and WP5 - “Demonstration Cases”. Namely, the document highlights how they relate to WP3 lines of research, as well as other assets; how they contribute to the advancement of WP4 research roadmap; and how they could be integrated in WP5 demonstrators.

In summary, this document provided a complete overview of CyberSec4Europe’s privacy-preserving technologies and their importance in the context of today’s privacy research challenges.

Bibliography

[1] Information Commissioner's Office, Data protection impact assessments, (n.d.). https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/?q=DPIA (accessed 8 July 2020).

[2] European Union Agency for Cybersecurity, Online tool for the security of personal data processing, (2019). https://www.enisa.europa.eu/risk-level-tool/ (accessed 8 July 2020).

[3] CNIL, The open source PIA software helps to carry out data protection impact assesment, (2019). https://www.cnil.fr/en/open-source-pia-software-helps-carry-out-data-protection-impact-assesment (accessed 8 July 2020).

[4] Edinburgh Business School, Heriot Watt University, Privacy by design and data protection impact assessment (DPIA) toolkit, (2018). https://www.hw.ac.uk/documents/privacy-by-design-dpia-toolkit.pdf (accessed 13 July 2020).

[5] Family Links Network, Template for Data Protection Impact Assessment (DPIA), (n.d.). https://iapp.org/resources/article/template-for-data-protection-impact-assessment-dpia/ (accessed 8 July 2020).

[6] Digital data protection impact assessment (DPIA) tool, (n.d). Local Digital. https://localdigital.gov.uk/funded-project/digital-data-protection-impact-assessment-dpia-tool/ (accessed 8 July 2020).

[7] Datenschutz Folgenabschätzung, A Data Protection Impact Assessment (DPIA) Tool for Practical Use in Companies and Public Administration (n.d.). https://www.dsfa.eu/index.php/en/home-en/ (accessed 8 July 2020).

[8] J. Anhalt et al., "Toward context-aware computing: experiences and lessons," in IEEE Intelligent Systems, vol. 16, no. 3, pp. 38-46, May-June 2001, doi: 10.1109/5254.940025.

[9] Langheinrich M. (2002) A Privacy Awareness System for Ubiquitous Computing Environments. In: Borriello G., Holmquist L.E. (eds) UbiComp 2002: Ubiquitous Computing. UbiComp 2002. Lecture Notes in Computer Science, vol 2498. Springer, Berlin, Heidelberg

[10] Könings, B., Wiedersheim, B. and Weber, M. "Privacy Management and Control in ATRACO"'Lecture Notes in Computer Science', Springer Berlin Heidelberg, 2010, pp. 51—60.

[11] B. Carminati and E. Ferrari, "Trusted Privacy Manager: A System for Privacy Enforcement," 21st International Conference on Data Engineering Workshops (ICDEW'05), Tokyo, Japan, 2005, pp. 1195-1195, doi: 10.1109/ICDE.2005.299.

[12] L. Cranor, et al. "A P3P preference exchange language 1.0 (APPEL 1.0). World Wide Web Consortium, Working Draft April 2002. https://www.w3.org/TR/P3P-preferences/.

[13] P3P: The Platform for Privacy Preferences. https://www.w3.org/P3P/.

[14] Elie Raad, Richard Chbeir. Privacy in Online Social Networks. Security and Privacy Preserving inSocial Networks, Springer-Verlag Wien, pp.3-45, 2013.

[15] Jakob, M., Moler, Z., Pechoucek, M. and Vaculin, R. "Content-Based Privacy Management on the Social Web"'2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology', IEEE, 2011.

[16] Squicciarini, A., Paci, F. and Sundareswaran, S. "PriMa: An Effective Privacy Protection Mechanism for Social Networks" Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security – ASIACCS '10. ACM Press, 2010.

[17] Squicciarini, A. C., Paci, F. and Sundareswaran, S. "PriMa: a comprehensive approach to privacy protection in social network sites," annals of telecommunications – annales des telecommunications (69:1-2), 2013, pp. 21–36.

[18] Mowbray, M. and Pearson, S. "A client-based privacy manager for cloud computing"'Proceedings of the Fourth International ICST Conference on COMmunication System softWAre and middlewaRE - COMSWARE 09', ACM Press, 2009.

[19] Pearson, S., Shen, Y. and Mowbray, M. "A Privacy Manager for Cloud Computing"'Lecture Notes in Computer Science', Springer Berlin Heidelberg, 2009, pp. 90–106.

[20] S. Pearson and M. Casassa-Mont, "Sticky Policies: An Approach for Managing Privacy across Multiple Parties," in Computer, vol. 44, no. 9, pp. 60-68, Sept. 2011, doi: 10.1109/MC.2011.225.

[21] Schwarzbach, B., Franczyk, B., Petrich, L., Schier, A. and Hompel, M. T. "Cloud Based Privacy Preserving Collaborative Business Process Management"'2016 IEEE International Conference on Computer and Information Technology (CIT)', IEEE, 2016

[22] Standards – OASIS: eXtensible Access Control Markup Language(XACML) v3.0.https://www.oasis-open.org/standards/#xacmlv3.0

[23] Fhom, H. S., Kuntze, N., Rudolph, C., Cupelli, M., Liu, J. and Monti, A. "A user-centric privacy manager for future energy systems"'2010 International Conference on Power System Technology', IEEE, 2010.

[24] Torre, I., Koceva, F., Sanchez, O. R. and Adorni, G. "A framework for personal data protection in the IoT"'2016 11th International Conference for Internet Technology and Secured Transactions (ICITST)', IEEE, 2016.

[25] Das, A., Degeling, M., Smullen, D. and Sadeh, N. "Personalized Privacy Assistants for the Internet of Things: Providing Users with Notice and Choice," IEEE Pervasive Computing (17:3), 2018, pp. 35–46.

[26] Fernandez, M., Jaimunk, J. and Thuraisingham, B. "Privacy-Preserving Architecture for Cloud-IoT Platforms"'2019 IEEE International Conference on Web Services (ICWS)', IEEE, 2019.

[27] Davies, N., Taft, N., Satyanarayanan, M., Clinch, S. and Amos, B. "Privacy Mediators: Helping IoT Cross the Chasm"'Proceedings of the 17th International Workshop on Mobile Computing Systems and Applications – HotMobile 16', ACM Press, 2016.

[28] Marcella Hastings, Brett Hemenway, Daniel Noble and Steve Zdancewic, "SoK: General Purpose Compilers for Secure Multi-Party Computation," 2019 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 2019, pages 1220-1237.

[29] David W. Archer, Dan Bogdanov, Liina Kamm, Yehuda Lindell, Kurt Nielsen, Jakob I. Pagter, Nigel P. Smart, and Rebecca N. Wright. From Keys to Databases – Real-World Applications of Secure Multi-Party Computation. The Computer Journal, Volume 61, Issue 12, 1 December 2018, pages 1749-1771.

[30] Dan Bogdanov, Liina Kamm, Sven Laur, Pille Pruulmann-Vengerfeldt, Riivo Talviste, and Jan Willemson. Privacy-Preserving Statistical Data Analysis on Federated Databases. In Privacy Technologies and Policy, volume 8450 of LNCS, pages 30-55. Springer International Publishing, 2014.

[31] Dan Bogdanov, Liina Kamm, Sven Laur and Ville Sokk, "Rmind: A Tool for Cryptographically Secure Statistical Analysis," in IEEE Transactions on Dependable and Secure Computing, vol. 15, no. 3, pp. 481-495, May-June 2018.

[32] Dan Bogdanov, Margus Niitsoo, Tomas Toft, and Jan Willemson. High-performance secure multi-party computation for data mining applications. International Journal of Information Security, 11(6):403-418, September 2012.

[33] Wenliang Du and Mikhail J. Atallah. Privacy-preserving cooperative statistical analysis. In Computer Security Applications Conference, 2001. ACSAC 2001. Proceedings 17th Annual, pages 102-110. IEEE, December 2001.

[34] Wenliang Du, Yunghsiang S. Han, and Shigang Chen. Privacy- Preserving Multivariate Statistical Analysis: Linear Regression and Classification. In Proceedings of the 4th SIAM International Conference on Data Mining, pages 222-233, 2004.

[35] Yehuda Lindell and Benny Pinkas. Secure Multiparty Computation for Privacy-Preserving Data Mining. The Journal of Privacy and Confidentiality, 1(1):59-98, 2009.

[36] Emmanuel A. Abbe, Amir E. Khandani, and Andrew W. Lo. Privacy-Preserving Methods for Sharing Financial Risk Exposures. American Economic Review, 102(3):65-70, May 2012.

[37] Dan Bogdanov, Liina Kamm, Baldur Kubo, Reimo Rebane, Ville Sokk, and Riivo Talviste. Students and taxes: a privacy-preserving study using secure computation. Proceedings on Privacy Enhancing Technologies, 3:117-135, 2016.

[38] Dan Bogdanov, Riivo Talviste, and Jan Willemson. Deploying Secure Multi-Party Computation for Financial Data Analysis. In Financial Cryptography and Data Security, volume 7397 of LNCS, pages 57-64. Springer Berlin Heidelberg, 2012.

[39] Mark Flood, Jonathan Katz, Stephen Ong, and Adam Smith. Cryptography and the Economics of Supervisory Information: Balancing Transparency and Condentiality. Technical Report 0011, Office of Financial Research, September 2013.

[40] Hyunghoon Cho, David J Wu, and Bonnie Berger. Secure genome-wide association analysis using multiparty computation. Nature Biotechnology, 36(6):547, 2018.

[41] Erika Check Hayden. Extreme cryptography paves way to personalized medicine. Nature, 519(7544):400-401, March 2015.

[42] Brian Hie, Hyunghoon Cho, and Bonnie Berger. Realizing private and practical pharmacological collaboration. Science, 362(6412):347-350, 2018.

[43] Karthik A. Jagadeesh, David J. Wu, Johannes A. Birgmeier, Dan Boneh, and Gill Bejerano. Deriving genomic diagnoses without revealing patient genomes. Science, 357(6352):692-695, 2017.

[44] Somesh Jha, Louis Kruger, and Vitaly Shmatikov. Towards practical privacy for genomic computation. In Security and Privacy, 2008, pages 216-230. IEEE, 2008.

[45] Liina Kamm, Dan Bogdanov, Sven Laur, and Jaak Vilo. A new way to protect privacy in large-scale genome-wide association studies. Bioinformatics, 29(7):886-893, April 2013.

[46] Nicholette Zeliadt. Cryptographic methods enable analyses without privacy breaches. Nature Medicine, 20(6):563, June 2014.

[47] Brett Hemenway, Steve Lu, Rafail Ostrovsky, and William Welser IV. High-precision secure computation of satellite collision probabilities. In SCN, pages 169-187. Springer, 2016.

[48] Liina Kamm and Jan Willemson. Secure floating point arithmetic and private satellite collision analysis. International Journal of Information Security, pages 1-18, 2014.

[49] M. Davari and E. Bertino, “Access Control Model Extensions to Support Data Privacy Protection based on GDPR,” 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 4017-4024.

[50] Kaniz Fatema, Christophe Debruyne, Dave Lewis, Declan O’Sullivan, John P Morrison, and Abdullah-Al Mazed. A semi-automated methodology for extracting access control rules from the european data protection directive. In Security and Privacy Workshops (SPW), 2016 IEEE, pages 25-32. IEEE, 2016.

[51] Francesco Di Cerbo, Fabio Martinelli, Ilaria Matteucci, and Paolo Mori. Towards a declarative approach to stateful and stateless usage control for data protection. In WEBIST, pages 308-315. SciTePress, 2018.

[52] Max-Robert Ulbricht and Frank Pallas. Yappl – A lightweight privacy preference language for legally sufficient and automated consent provision in iot scenarios. In DPM 2018 and CBT 2018 – ESORICS 2018 International Workshops, Barcelona, Spain, September 6-7, 2018, pages 329-344, 2018.

[53] Bartolini, C.; Daoudagh, S.; Lenzini, G.; Marchetti, E. Towards a lawful authorized access: A preliminary GDPR-based authorized access. In Proceedings of the ICSOFT 2019, Prague, Czech Republic, 26-28 July 2019; pp. 26-28.

[54] Bartolini, C.; Daoudagh, S.; Lenzini, G.; Marchetti, E. GDPR-Based User Stories in the Access Control Perspective. In Quality of Information and Communications Technology, Proceedings of the 12th International Conference, QUATIC 2019, Ciudad Real, Spain, 11-13 September 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 3-17.

[55] M. Davari and E. Bertino: "Reactive access control systems". Proceedings of the 23nd ACM on Symposium on Access Control Models and Technologies, pp. 205-207, 2018.

[56] Munoz-Arcentales, A., López-Pernas, S., Pozo, A., Alonso, Á., Salvachúa, J., & Huecas, G. (2020). Data Usage and Access Control in Industrial Data Spaces: Implementation Using FIWARE. *Sustainability*, *12*(9), 3885.

[57] Calabró, A.; Daoudagh, S.; Marchetti, E. Integrating Access Control and Business Process for GDPR Compliance: A Preliminary Study. In Proceedings of the ITASEC 2019, Pisa, Italy, 13-15 February 2019.

[58] Arfelt, E.; Basin, D.; Debois, S. Monitoring the GDPR. In Comp. Sec.-ESORICS 2019; Sako, K., Schneider, S., Ryan, P.Y.A., Eds.; Springer: Berlin/Heidelberg, Germany, 2019; pp. 681-699.

[59] Pietro Ferrara and Fausto Spoto. Static analysis for GDPR compliance. In Elena Ferrari, Marco Baldi, and Roberto Baldoni, editors, Proceedings of the Second Italian Conference on Cyber Security (ITASEC), February 2018.

[60] Silvio Ranise and Hari Siswantoro. Automated legal compliance checking by security policy analysis. In Computer Safety, Reliability, and Security – SAFECOMP 2017 Workshops, ASSURE,DECSoS, SASSUR, TELERISE, and TIPS, Trento, Italy, September 12, 2017, Proceedings, vol-ume 10489 ofLecture Notes in Computer Science, pages 361-372. Springer, 2017.

[61] Bartolini, C.; Daoudagh, S.; Lenzini, G.; Marchetti, E. Testing of GDPR-based Access Control Policies. Presented at ESORICS (Poster-Session) 2019. Available at: http://security.isti.cnr.it/.

[62] Cadwalladr, C., Graham-Harrison, E.: Revealed: 50 million facebook profiles harvested for cambridge analytica in major data breach. The Guardian https://www.theguardian.com/news/2018/mar/17/ cambridge-analytica-facebook-influence-us-election.

[63] Goel, V., Perlroth, N.: Yahoo says 1 billion user accounts were hacked. The New York Times https://www.nytimes.com/2016/12/14/technology/yahoo-hack. html.

[64] Hong, N., Hoffman, L., Andriotis, A.: Capital one reports data breach affecting 100 million customers, applicants. The Wall Street Journal https://www.wsj.com/ articles/capital-one-reports-data-breach-11564443355.

[65] Newman, L.H.: Equifax officially has no excuse. Wired https://www.wired.com/ story/equifax-breach-no-excuse/.

[66] Sweeney, L. (2002). k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05), 557-570.

[67] Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007). l-diversity: Privacy beyond k-anonymity. ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1), 3-es.

[68] Narayanan, A., & Shmatikov, V. (2006). How to break anonymity of the netflix prize dataset. arXiv preprint cs/0610105.

[69] Voigt, P., & Von dem Bussche, A. (2017). The eu general data protection regulation (gdpr). A Practical Guide, 1st Ed., Cham: Springer International Publishing.

[70] F. Bélanger and R. E. Crossler, "Privacy in the digital age: A review of information privacy research in information systems," MIS Q., vol. 35, pp. 1017-1042, 12 2011. http://dl.acm.org/citation.cfm?id=2208940. 2208951.

[71] D. J. Solove, "A taxonomy of privacy," U. Pa. L. Rev., vol. 154, p. 477, 2005.

[72] R. L. Finn, D. Wright, and M. Friedewald, "Seven types of privacy," in European data protection: coming of age, pp. 3-32, Springer, 2013.

[73] H. J. Smith, T. Dinev, and H. Xu, "Information privacy research: an interdisciplinary review," MIS quarterly, vol. 35, no. 4, pp. 989-1016, 2011.

[74] E. F. Stone, H. G. Gueutal, D. G. Gardner, and S. McClure, "A field experiment comparing information-privacy values, beliefs, and attitudes across several types of organizations.," Journal of applied psychology, vol. 68, no. 3, p. 459, 1983.

[75] G. Danezis and S. Gürses, "A critical review of 10 years of privacy technology," Proceedings of surveillance cultures: a global surveillance society, pp. 1-16, 2010.

[76] A. Pfitzmann and M. Hansen, "A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management," 2009.

[77] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," Official Journal of the European Union, vol. L119, pp. 1-88, 5 2016. http: //eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC.

[78] J. Camenisch and A. Lysyanskaya, "An efficient system for nontransferable anonymous credentials with optional anonymity revocation," in Advances in Cryptology—EUROCRYPT 2001, pp. 93-118, Springer, 2001.

[79] J. Camenisch and E. Van Herreweghen, “Design and implementation of the idemix anonymous credential system,” in Proceedings of the 9th ACM Conference on Computer and Communications Security, CCS ’02, (New York, NY, USA), pp. 21-30, ACM, 2002.

[80] C. Ribeiro, H. Leitold, S. Esposito, and D. Mitzam, “Stork: a real, heterogeneous, large-scale eid management system,” International Journal of Information Security, vol. 17, pp. 569-585, Oct 2018.

[81] D. Recordon and D. Reed, “Openid 2.0: a platform for user-centric identity management,” in Proceedings of the second ACM workshop on Digital identity management, pp. 11-16, ACM, 2006.

[82] J. Hughes and E. Maler, “Security assertion markup language (saml) v2. 0 technical overview,” OASIS SSTC Working Draft sstc-saml-techoverview-2.0-draft-08, pp. 29-38, 2005

[83] A. Czeskis and J. Lang, “Fido nfc protocol specification v1. 0,” FIDO Alliance Proposed Standard, pp. 1-5, 2015.

[84] A. Tobin and D. Reed, “The inevitable rise of self-sovereign identity,” tech. rep., The Sovrin Foundation, 2016. https://sovrin.org/wpcontent/uploads/2017/06/The-Inevitable-Rise-of-Self-SovereignIdentity.pdf.

[85] C. Allen, “The path to self-sovereign identity,” tech. rep., 2017.

[86] F. van den Broek, B. Hampiholi, and B. Jacobs, “Securely derived identity credentials on smart phones via self-enrolment,” in International Workshop on Security and Trust Management, pp. 106-121, Springer, 2016.

[87] J. B. Bernabe, M. David, R. T. Moreno, J. P. Cordero, S. Bahloul, and A. Skarmeta, “Aries: Evaluation of a reliable and privacy-preserving european identity management framework,” Future Generation Computer Systems, vol. 102, pp. 409 - 425, 2020.

[88] R. T. Moreno, J. B. Bernabe, A. Skarmeta, M. Stausholm, T. K. Frederiksen, N. Martínez, N. Ponte, E. Sakkopoulos, and A. Lehmann, “Olympus: towards oblivious identity management for private and user-friendly services,” in 2019 Global IoT Summit (GIoTS), pp. 1-6, June 2019.

[89] J. B. Bernabé, J. L. H. Ramos, and A. F. Gómez-Skarmeta, “Holistic privacy-preserving identity management system for the internet of things,” Mobile Information Systems, vol. 2017, pp. 6384186:1- 6384186:20, 2017.

[90] J. L. C. Sanchez, J. B. Bernabe, and A. F. Skarmeta, “Integration of anonymous credential systems in iot constrained environments,” IEEE Access, vol. 6, pp. 4767-4778, 2018.

[91] D. Reed, M. Sprony, D. Longley, C. Allen, R. Grant, and M. Sabadello, “Decentralized identifiers (dids) v0. 11 data model and syntaxes for decentralized identifiers (dids). w3c,” 2018.

[92] M. Sporny and D. Longley, “Verifiable claims data model and representations,” tech. rep., W3C, 2017.

[93] J. Bonneau, A. Narayanan, A. Miller, J. Clark, J. A. Kroll, and E. W. Felten, "Mixcoin: Anonymity for bitcoin with accountable mixes," in International Conference on Financial Cryptography and Data Security, pp. 486-504, Springer, 2014.

[94] K. Wagner, B. Némethi, E. Renieris, P. Lang, E. Brunet, and E. Holst, "Self-sovereign identity, a position paper on blockchain enabled identity and the road ahead," tech. rep., Blockchain Bundesverband, 2018.

[95] Dumortier, J. (2017). Regulation (EU) No 910/2014 on Electronic Identification and Trust Services for Electronic Transactions in the Internal Market (eIDAS Regulation). In EU Regulation of E-Commerce. Edward Elgar Publishing.

[96] Amir Shayan Ahmadian, Daniel Strüber, Volker Riediger, and Jan Jürjens. 2018. Supporting privacy impact assessment by model-based privacy analysis. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing (SAC '18). Association for Computing Machinery, New York, NY, USA, 1467-1474. DOI: https://doi.org/10.1145/3167132.3167288.

[97] David Basin, Søren Debois, and Thomas Hildebrandt. On purpose and by necessity. In Proceedings of the Twenty-Second International Conference on Financial Cryptography and Data Security (FC), February 2018.

[98] Q. Ramadan, M. Salnitriy, D. Strüber, J. Jürjens and P. Giorgini, "From Secure Business Process Modeling to Design-Level Security Verification," 2017 ACM/IEEE 20th International Conference on Model Driven Engineering Languages and Systems (MODELS), Austin, TX, 2017, pp. 123-133, doi: 10.1109/MODELS.2017.10.

[99] David Brossard, Gerry Gebel, and Mark Berg. A systematic approach to implementing ABAC. In Proceedings of the 2Nd ACM Workshop on Attribute-Based Access Control, ABAC '17, pages 53-59, New York, NY, USA, 2017. ACM.

[100] Said Daoudagh, and Eda Marchetti. A Life Cycle for Authorization Systems Development in the GDPR Perspective. Proceedings of the Fourth Italian Conference on Cyber Security ITSASEC2020, Ancona, Italy, February 4th to 7th, 2020, pages 128-140, CEUR Workshop Proceedings.

[101] Pille Pullonen, Jake Tom, Raimundas Matulevicius, Aivo Toots: Privacy-enhanced BPMN: enabling data privacy analysis in business processes models. Software and Systems Modeling 18(6): 3235-3264 (2019).

[102] Accorsi, R., Lehmann, A.: Automatic information flow analysis of business process models. In: 10th International Conference on Business Process Management (BPM), pp.172-187. Springer (2012).

[103] Accorsi, R., Lehmann, A., Lohmann, N.: Information leak detection in business processmodels: Theory, application, and tool support. Information Systems 47, 244-257 (2015).

[104] Ladha, W., Mehandjiev, N., Sampaio, P.: Modelling of Privacy-Aware Business Pro-cesses in BPMN to Protect Personal Data. In: Proceedings of the 29th Annual ACMSymposium on Applied Computing, pp. 1399-1405 (2014).

[105] Ayed, G., Ghernaouti-Helie, S.: Processes View Modeling of Identity-Related PrivacyBusiness Interoperability: Considering User-Supremacy Federated Identity Technical Model and Identity Contract Negotiation. In: Proceedings of the ASONAM (2012).

[106] Ramadan, G., Str¨uber, D., Salnitri, M., J¨urjens, J., Riediger, V., S., S.: A Semi-Automated BPMN-Based Framework for Detecting Conflicts Between Security, Data-Minimization, and Fairness Requirements. Software and Systems Modeling (2020).

[107] Noah M. Johnson, Joseph P. Near, Dawn Song: Towards Practical Differential Privacy for SQL Queries. Proc. VLDB Endow. 11(5): 526-539 (2018).

[108] Kotsogiannis, Ios, Yuchao Tao, Xi He, Maryam Fanaeepour, Ashwin Machanavajjhala, Michael Hay, and Gerome Miklau. "Privatesql: a differentially private sql query engine." Proceedings of the VLDB Endowment 12, no. 11 (2019): 1371-1384.

[109] Wilson, Royce J., Celia Yuxin Zhang, William Lam, Damien Desfontaines, Daniel Simmons-Marengo, and Bryant Gipson. "Differentially private sql with bounded user contribution." Proceedings on Privacy Enhancing Technologies 2020, no. 2 (2020): 230-250.

[110] Myrto Arapinis, Diego Figueira, and Marco Gaboardi. Sensitivity of counting queries. In Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi, editors, 43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy, volume 55 of LIPIcs, pages 120:1-120:13. Schloss Dagstuhl – Leibniz-Zentrum fuer Informatik, 2016.

[111] Catuscia Palamidessi and Marco Stronati. Differential privacy for relational algebra: Improving the sensitivity bounds via constraint systems. In Herbert Wiklicky and Mieke Massink, editors, Proceedings 10th Workshop on Quantitative Aspects of Programming Languages and Systems, QAPL 2012, Tallinn, Estonia, 31 March and 1 April 2012., volume 85 of EPTCS, pages 92-105, 2012.

[112] Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. Smooth sensitivity and sampling in private data analysis. In David S. Johnson and Uriel Feige, editors, Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007, pages 75-84. ACM, 2007.

[113] Marco Gaboardi, Andreas Haeberlen, Justin Hsu, Arjun Narayan, and Benjamin C. Pierce. Linear dependent types for differential privacy. In Roberto Giacobazzi and Radhia Cousot, editors, The 40th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL '13, Rome, Italy – January 23 – 25, 2013, pages 357-370. ACM, 2013.

[114] Hamid Ebadi and David Sands. Featherweight PINQ. Journal of Privacy and Security, 7(2):159-184, 2016.

[115] Daniel Kifer and Ashwin Machanavajjhala. 2014. Pufferfish: A framework for mathematical privacy definitions. ACM Trans. Database Syst. 39, 1, Article 3 (January 2014), 36 pages. DOI: https://doi.org/10.1145/2514689.

[116] M. Crosby, P. Pattanayak, S. Verma, and V. Kalyanaraman: *Blockchain technology: Beyond Bitcoin.* Appl. Innov., vol. 2, 6-10 (2016).

[117] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou: *Hawk:The blockchain model of cryptography and privacy-preserving smartcontracts*. IEEE Symp. Secur. Privacy (SP) 2016, 839-858.

[118] I. Miers, C. Garman, M. Green, and A. D. Rubin: *Zerocoin: Anonymous distributed E-cash from bitcoin*. IEEE Symp. Secur. Privacy (SP) 2013, 397-411.

[119] S. Alboaie and D. Cosovan: *Private data system enabling self-sovereign storage managed by executable choreographies*. Distributed Applications and Interoperable Systems 2017, 83-98.

[120] Moreno, R. T., Bernal Bernabe, J., García Rodríguez, J., Frederiksen, T. K., Stausholm, M., Martínez, N., ... & Skarmeta, A. (2020). The OLYMPUS Architecture—Oblivious Identity Management for Private User-Friendly Services. Sensors, 20(3), 945.

[121] Camenisch, J., Drijvers, M., Lehmann, A., Neven, G., & Towa, P. (2020). Short Threshold Dynamic Group Signatures. IACR Cryptol. ePrint Arch., 2020, 16.

[122] Marlon Dumas, Luciano García-Bañuelos, Peeter Laud: Differential Privacy Analysis of Data Processing Workflows. GraMSec@CSF 2016.

[123] Marlon Dumas, Luciano García-Bañuelos, Peeter Laud: Disclosure Analysis of SQL Workflows. GraMSec@CSF 2018.

[124] Martin Pettai, Peeter Laud: Combining Differential Privacy and Mutual Information for Analyzing Leakages in Workflows. POST 2017: 298-319.

[125] Peeter Laud, Alisa Pankova, Martin Pettai: Achieving Differential Privacy using Methods from Calculus, https://arxiv.org/abs/1811.06343.

[126] Peeter Laud, Alisa Pankova: Interpreting Epsilon of Differential Privacy in Terms of Advantage in Guessing or Approximating Sensitive Attributes, https://arxiv.org/abs/1911.12777.

[127] Aivo Toots, Reedik Tuuling, Maksym Yerokhin, Marlon Dumas, Luciano Garcia-Banuelos, Peeter Laud, Raimundas Matulevicius, Alisa Pankova, Martin Pettai, Pille Pullonen, Jake Tom: Business Process Privacy Analysis in Pleak. FASE, 2019. Full version available in https://arxiv.org/abs/1902.05052.

[128] Antonia Bertolino, Said Daoudagh, Francesca Lonetti, Eda Marchetti, and Louis Schilders. 2013. Automated testing of eXtensible Access Control Markup Language-based access control systems. IET Software 7, 4 (2013), 203-212.

[129] S. Daoudagh, F. Lonetti, and E. Marchetti. 2020 XACMET: XACML Modeling &Testing. In Software Quality Journal (2020).

[130] A. Bertolino, S. Daoudagh, F. Lonetti, and E. Marchetti. 2013. XACMUT: XACML2.0 Mutants Generator. In Proc. of Mutation 2013. 28-33.

[131] Antonia Bertolino, Said Daoudagh, Francesca Lonetti, and Eda Marchetti. 2018. An Automated Model-based Test Oracle for Access Control Systems (AST '18). ACM, New York, NY, USA, 2-8.

[132] Cesare Bartolini, Antonello Calabrò, and Eda Marchetti. 2019. GDPR and business processes: an effective solution. In Proc. of APPIS 2019, Las Palmas de Gran Canaria, Spain, January 07-09, 2019. 7:1-7:5.

[133] Paolo Barsocchi, Antonello Calabrò, Antonino Crivello, Said Daoudagh, Francesco Furfari, Michele Girolami and Eda Marchetti. 2020. A Privacy-By-Design Architecture for Indoor Localization Systems. In Proc. of QUATIC 2020, Online conference September 8-10.

[134] APEC Privacy Framework, (2015). https://www.apec.org/Publications/2017/08/APEC-Privacy-Framework-(2015).

[135] Canadian Institute of Chartered Accountants (CICA) and American Institute of Certified Public Accountants, Generally Accepted Privacy Principles, (2009).

[136] ISACA, GDPR Data Protection Impact Assessments, 2017. http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/GDPR-Data-Protection-Impact-Assessments.aspx.

[137] ISO/IEC 29100:2011 Information technology — Security techniques — Privacy framework, (n.d.). https://www.iso.org/standard/45123.html.

[138] Article 29 Data Protection Working Party, Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679, (2017). https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=611236.

[139] European Data Protection Board, Guidelines 4/2019 on Article 25 Data Protection by Design and by Default, (2019). https://edpb.europa.eu/our-work-tools/public-consultations-art-704/2019/guidelines-42019-article-25-data-protection-design.

[140] European Data Protection Supervisor, Preliminary Opinion on Privacy by Design, (2018). https://edps.europa.eu/data-protection/our-work/publications/opinions/privacy-design_en (accessed 18 May 2020).

[141] Information Commissioner's Office, Data protection by design and default, (n.d.). https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-by-design-and-default/ (accessed 15 July 2020).

[142] D. Chaum: Security without identification: Transaction systems to make big brother obsolete. Commun.ACM 28: 1030-1044 (1985).

[143] Jan Camenisch, Stephan Krenn, Anja Lehmann, Gert Læssøe Mikkelsen, Gregory Neven, Michael Østergaard Pedersen:Formal Treatment of Privacy-Enhancing Credential Systems. SAC 2015: 3-24.

[144] Stephan Krenn, Thomas Lorünser, Anja Salzer, Christoph Striecks:Towards Attribute-Based Credentials in the Cloud. CANS 2017: 179-202.

[145] Ulrich Haböck, Stephan Krenn: Breaking and Fixing Anonymous Credentials for the Cloud. CANS 2019: 249-269.

[146] Matt Blaze, Gerrit Bleumer, Martin Strauss:Divertible Protocols and Atomic Proxy Cryptography. EUROCRYPT 1998: 127-144.

[147] Tor Project: https://www.torproject.org/

[148] A. Shamir: How to Share a Secret. Commun. ACM 22(11): 612-613 (1979).

[149] T. Lorünser, An. Happe, B. Rainer, F. Wohner, C. Striecks, D. Demirel, G. Traverso: Advanced architecture for distributed storage in dynamic environments (SECOSTOR Tool). PRISMACLOUD project deliverable D5.3. 2017.

[150] S. Goldwasser, S. Micali, and C. Rackoff: The knowledge complexity of interactive proof systems", SIAM Journal on Computing, 18 (1): 186-208, 1989.

[151] Sabouri, A., Krontiris, I., & Rannenberg, K. (2012, September). Attribute-based credentials for trust (abc4trust). In *International Conference on Trust, Privacy and Security in Digital Business* (pp. 218-219). Springer, Berlin, Heidelberg.

[152] L. T. Vaszar, M. K. Cho, T. A. Raffin, Privacy issues in personalized medicine, Pharmacogenomics 4 (2) (2003) 107–112.

[153] R. B. Altman, T. E. Klein, Challenges for biomedical informatics and pharmacogenomics, Annual review of pharmacology and toxicology 42 (1) (2002) 113–133.

[154] M. Naveed, E. Ayday, E. W. Clayton, J. Fellay, C. A. Gunter, J.-P. Hubaux, B. A. Malin, X. Wang, Privacy in the genomic era, ACM Computing Surveys (CSUR) 48 (1) (2015) 6.

[155] X. Jiang, Y. Zhao, X. Wang, B. Malin, S. Wang, L. Ohno-Machado, H. Tang, A community assessment of privacy preserving techniques for human genomes, BMC medical informatics and decision making 14 (Suppl 1) (2014) S1.

[156] B. Malin, L. Sweeney, Determining the identifiability of dna database entries., in: Proceedings of the AMIA Symposium, American Medical Informatics Association, 2000, p. 537.

[157] B. Malin, L. Sweeney, How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems, Journal of biomedical informatics 37 (3) (2004) 179–192.

[158] Z. Lin, M. Hewett, R. B. Altman, Using binning to maintain confidentiality of medical data., in: Proceedings of the AMIA Symposium, American Medical Informatics Association, 2002, p. 454. 3.

[159] B. Malin, Protecting dna sequence anonymity with generalization lattices, Carnegie Mellon University, School of Computer Science [Institute for Software Research International], 2004.

[160] M. Kantarcioglu, W. Jiang, Y. Liu, B. Malin, A cryptographic approach to securely share and query genomic sequences, IEEE Transactions on information technology in biomedicine 12 (5) (2008) 606–617.

[161] N. Homer, S. Szelinger, M. Redman, D. e. a. Duggan, Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays, PLoS Genet 4 (8) (2008) e1000167.

[162] S. Shringarpure, C. Bustamante, Privacy risks from genomic datasharing beacons, The American Journal of Human Genetics 97 (5) (2015) 631–646.

[163] M. T. Goodrich, The mastermind attack on genomic data, in: Security and Privacy, 2009 30th IEEE Symposium on, IEEE, 2009, pp. 204–218.

[164] M. Humbert, E. Ayday, J.-P. Hubaux, A. Telenti, Addressing the concerns of the lacks family: quantification of kin genomic privacy, in: Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security, ACM, 2013, pp. 1141–1152.

[165] M. C. Schatz, “Cloudburst: highly sensitive read mapping with mapreduce,” Bioinformatics, vol. 25, no. 11, pp. 1363–1369, 2009.

[166] R. V. Pandey and C. Schlotterer, “Distmap: a toolkit for distributed ¨ short read mapping on a hadoop cluster,” PLoS One, vol. 8, no. 8, pp. 1363–1369, 2013.

[167] Y. Huang, D. Evans et al., “Faster secure two-party computation using garbled circuits.” in USENIX, vol. 201, no. 1, 2011.

[168] E. De Cristofaro, S. Faber et al., “Secure genomic testing with size-and position-hiding private substring matching,” in WPES, 2013.

[169] Y. Chen, B. Peng et al., “Large-scale privacy-preserving mapping of human genomic sequences on hybrid clouds.” in NDSS, 2012.

[170] V. Popic and S. Batzoglou, “A hybrid cloud read aligner based on minhash and kmer voting that preserves privacy,” Nat. Commun., vol. 8, 2017.

[171] C. Lambert, M. Fernandes et al., “Maskal: Privacy preserving masked reads alignment using intel sgx,” in SRDS, 2018.

[172] E. Ayday, J. L. Raisaro et al., Privacy-preserving processing of raw genomic data, 2014.

[173] V. V. Cogo, A. Bessani et al., “A high-throughput method to detect privacy-sensitive human genomic data,” in WPES, 2015.

[174] S. O. Dyke, E. S. Dove et al., “Sharing health-related data: a privacy test?” NPJ genomic medicine, vol. 1, pp. 1–6, 2016.

[175] Resende, João S., Rolando Martins, and Luís Antunes. "Enforcing Privacy and Security in Public Cloud Storage." 2018 16th Annual Conference on Privacy, Security and Trust (PST). IEEE, 2018.

[176] Mendes, Ricardo, et al. "Charon: A secure cloud-of-clouds system for storing and sharing big data." IEEE Transactions on Cloud Computing (2019).

[177] Bessani, Alysson, et al. "SCFS: A Shared Cloud-backed File System." 2014 USENIX Annual Technical Conference (USENIX ATC 14). 2014.

[178] Pontes, Rogério, et al. "SafeFS: a modular architecture for secure user-space file systems: one FUSE to rule them all." Proceedings of the 10th ACM International Systems and Storage Conference. 2017.

[179] Conti, Mauro, et al. "Internet of Things security and forensics: Challenges and opportunities." (2018): 544-546.

[180] Wang, Hua, Zonghua Zhang, and Tarek Taleb. "Special issue on security and privacy of IoT." World Wide Web 21.1 (2018): 1-6.

[181] Yang, Chaowei, et al. "Big Data and cloud computing: innovation opportunities and challenges." International Journal of Digital Earth 10.1 (2017): 13-53.

[182] A. B. Brush, E. Filippov, D. Huang, J. Jung, R. Mahajan, F. Martinez, K. Mazhar, A. Phanishayee, A. Samuel, J. Scott and R. P. Singh, “Lab of Things: A Platform for Conducting Studies with Connected Devices in Multiple Homes,” in Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, New York, NY, USA, 2013.

[183] LogMeIn, I. "Xively." (2015).

[184] Perera, Charith, et al. "Privacy of big data in the internet of things era." IEEE IT Special Issue Internet of Anything 6 (2015)

[185] Sinha, Nitin, Korrapati Eswari Pujitha, and John Sahaya Rani Alex. "Xively based sensing and monitoring system for IoT." Computer Communication and Informatics (ICCCI), 2015 International Conference on. IEEE, 2015

[186] Popescu, Mihaela, et al. "Consumer surveillance and distributive privacy harms in the age of big data." Digital media: Transformations in human communication (2017): 313-327.

[187] Brewster, T. "Meet Datacoup–the company that wants to help you sell your data." The Guardian (2014).

[188] Trust, identity & data management — solved (2017). https://mydex.org/

[189] Axeda - AXEDA MACHINE CLOUD SERVICE (2016). http://www.ptc.com/internet-of-things/solutions

[190] J. Mineraud, O. Mazhelis, X. Su, and S. Tarkoma, "Contemporary Internet-of-Things platforms" http://arxiv.org/abs/1501.07438, Jan 2015, technical report.

[191] Kim, Jaeho, and Jang-Won Lee. "OpenIoT: An open service framework for the Internet of Things." Internet of Things (WF-IoT), 201.

[192] Ramparany, Fano, et al. "Handling smart environment devices, data and services at the semantic level with the FI-WARE core platform." Big Data (Big Data), 2014 IEEE International Conference on. IEEE, 2014.

[193] Zahariadis, Theodore, et al. "FIWARE lab: managing resources and services in a cloud federation supporting future internet applications." Utility and Cloud Computing (UCC), 2014 IEEE/ACM 7th International Conference on. IEEE, 2014.

[194] Fazio, Maria, et al. "Exploiting the FIWARE cloud platform to develop a remote patient monitoring system." Computers and Communication (ISCC), 2015 IEEE Symposium on. IEEE, 2015.

[195] Glikson, Alex. "Fi-ware: Core platform for future internet applications." Proceedings of the 4th annual international conference on systems and storage. 2011.

[196] Schmitt, Corinna, Claudio Anliker, and Burkhard Stiller. "Pull support for IoT applications using mobile access framework WebMaDa." Internet of Things (WF-IoT), 2016 IEEE 3rd World Forum on. IEEE, 2016.

[197] Fiware pep-proxy (2017). http://fiware-pep-proxy.readthedocs.io/en/latest/user_guide/

[198] European Union Agency for Cybersecurity (ENISA), Privacy and Data Protection by Design, 2014. https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design

(accessed 18 May 2020).

[199] A. Cavoukian, Privacy by Design - The 7 Foundational Principles, n.d. https://iapp.org/resources/article/privacy-by-design-the-7-foundational-principles/ (accessed 20 May 2020).

[200] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, 1995. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31995L0046 (accessed 20 May 2020).

[201] ARTICLE 29 Data Protection Working Party and Working Party on Police and Justice, The Future of Privacy: Joint contribution to the Consultation of the European Commission on the legal framework for the fundamental right to protection of personal data, (2009).

https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2009/wp168_en.pdf (accessed 20 May 2020).

[202] L.F. de la Torre, What does 'data protection by design and by default' mean under EU Data Protection Law?, (2019). https://medium.com/golden-data/what-does-data-protection-by-design-and-by-default-mean-under-eu-data-protection-law-fc40f585c0c5 (accessed 1 June 2020).

[203] Dan Bogdanov. Sharemind: programmable secure computations with practical applications. PhD thesis, University of Tartu, 2013.

[204] Gerald Spindler and Philipp Schmechel. Personal Data and Encryption in the European General Data Protection Regulation. JIPITEC, 7(2):163–177, 2016.

[205] Liina Kamm and Jan Willemson. Secure floating point arithmetic and private satellite collision analysis. International Journal of Information Security, pages 1 18, 2014.

[206] N. Sartor, "Data Anonymization Software – Differences Between Static and Interactive Anonymization", January 14, 2019.

[207] B. Malle, P. Kieseberg and A. Holzinger, "Interactive Anonymization for Privacy aware Machine", in Krempl, Georg, Lemaire, Vincent, Polikar, Robi, Sick, Bernhard, Kottke, Daniel & Calma, Adrian (eds) IAL@ ECML PKDD 2017 Workshop and Tutorial on Interactive Adaptative learnin, Skopje, pp 15-26.

[208] D. Xiao, G. Wang, J. Gehrke, "Interactive Anonymization of Sensitive Data", in Proc. ACM SIGMOD 2009, pp. 1051-1053, SIGMOD'09, June 29–July 2, 2009.

[209] http://cs.utdallas.edu/dspl/cgi-bin/toolbox/index.php

[210] https://sourceforge.net/projects/anony-toolkit/

[211] http://dl.acm.org/citation.cfm?id=1687607

[212] http://users.uop.gr/~poulis/SECRETA/index.html

[213] https://sourceforge.net/projects/openanonymizer/

[214] https://www.tmf-ev.de/Themen/Projekte/V08601_AnonTool.aspx

[215] http://ppsf.ikelab.net/

[216] https://arx.deidentifier.org/

[217] https://amnesia.openaire.eu/

[218] http://neon.vb.cbs.nl/casc/mu.htm

[219] https://cran.r-project.org/package=sdcMicro

[220] https://realrolfje.github.io/anonimatron/

[221] https://github.com/google/rappor

[222] https://arxiv.org/abs/1407.6981

[223] https://www.cs.umd.edu/~elaine/docs/gupt.pdf

[224] https://www.microsoft.com/en-us/research/project/privacy-integrated-queries-pinq/

[225] https://www.vldb.org/pvldb/vol7/p637-proserpio.pdf

[226] https://privacytools.seas.harvard.edu/publications/psipaper

[227] https://arxiv.org/pdf/1706.09479.pdf

[228] https://cran.r-project.org/web/packages/diffpriv/index.html

[229] https://aircloak.com/solutions/features-en/

[230] J. Schwenk, "Welcome to the Future of Trust", July 2019.

[231]] D. Du Seuil and C. Pastor Matut, "Understanding the European Self Sovereign Identity Framework (ESSIF)", Webinar 32 July 7 2019.

[232] Working for development, integration and adoption of Self-Sovereign Identities (SSI) technologies https://essif-lab.eu/

[233] S. Krenn, CyberSec4Europe D3.2 “Cross Sectoral Cybersecurity Building Blocks”. European Commission 2020: https://cybersec4europe.eu/wp-content/uploads/2020/06/D3.2-Cross_sectoral_cybersecurity-building-blocks-v2.0.pdf

[234] F. Prasser, F. Kohlmayer, R. Lautenschläger, and K. A. Kuhn, “ARX- A Comprehensive Tool for Anonymizing Biomedical Data”, in AMIIA Annual Symposium Proceedings Archive, 2014, pp. 984-983,published online 2014 Nov 14.

[235] A. Sforzin, CyberSec4Europe D5.2 “Specification and Set-up Demonstration case Phase 1”. European Commission 2020. https://cybersec4europe.eu/wp-content/uploads/2020/05/D5.2-Specification-and-Set-up-of-Demonstration-Case-Phase-1-v1.0_Submitted.pdf

[236] Mou, J., Shin, D. & Cohen, J.F. Trust and risk in consumer acceptance of e-services. *Electron Commer Res* **17,** 255–288 (2017). https://doi.org/10.1007/s10660-015-9205-4

[237]] E. Colesca, "Increasing e-trust: a solution to minimize risk in e-government adoption", in Journal of Applied Quantitative Methods, pp. 31-44, JAQM Vol. 4, 1, Spring 2009. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.517.8969&rep=rep1&type=pdf

[238] Halperin, R., & Backhouse, J. (2012). Risk, trust and eID: Exploring public perceptions of digital identity systems. *First Monday*, *17*(4). https://doi.org/10.5210/fm.v17i4.3867

[239]] J.C. Pérez, LEPS D3.3 <2Operational and Technical Documentation of SP integration", European Commission 2020. http://www.leps-project.eu/node/345

[240] D.G. Berbecauru, A. Lioy & C. Cameroni. Providing Login and Wi-Fi Access Services Withthe eIDAS Network: A Practical Approach. IEEE Access, pp. 126186-126200, Volume 8, 2020 . DOI 10.1109/ACCESS.2020.3007998. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9136679

[241] https://en.wikipedia.org/wiki/PRISM_(surveillance_program)

[242] https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal

[243] https://www.schneier.com/blog/archives/2020/04/security_and_pr_1.html

[244] https://www.schneier.com/blog/archives/2020/04/secure_internet.html

[245] https://www.schneier.com/blog/archives/2020/06/zooms_commitmen.html

[246] https://www.schneier.com/blog/archives/2020/06/zoom_will_be_en.html

[247] A. C. Yao. How to generate and exchange secrets. In 27th Annual Symposium on Foundations of Computer Science (SFCS 1986), pages 162–167, Oct 1986.

[248] Fully homomorphic encryption using ideal lattices. In Proceedings of the 41st ACM Symposium on Theory of Computing, STOC'09, pages 169–178. ACM, 2009.

[249] Shay Gueron. A Memory Encryption Engine Suitable for General Purpose Processors. Cryptology ePrint Archive, Report 2016/204, 2016. https://eprint.iacr.org/2016/ 204.

[250] Brendan McMahan and Daniel Ramage. Federated Learning: Collaborative Machine Learning without Centralized Training Data. https://ai.googleblog.com/2017/04/ federated-learning-collaborative.html, 04 2017.