UNIVERSITÁ DI PISA
DOTTORATO DI RICERCA IN INFORMATICA

THE GDPR COMPLIANCE THROUGH ACCESS CONTROL SYSTEMS

DOCTORAL THESIS

Author
Said Daoudagh

Supervisor (s)
Dr. Eda Marchetti
Prof. Anna Monreale

Internal Committee Members
Prof. Laura Ricci
Prof. Gabriele Lenzini

The Coordinator of the PhD Program
Prof. Antonio Brogi

Pisa, July 2021

XXXIII Cycle

This thesis is dedicated to the victims of COVID-19 and their families.

Acknowledgements

I would like to thank my supervisor Dr. Eda Marchetti, my co-supervisor Prof. Anna Monreale and the members of the Internal Committee Prof. Laura Ricci and Prof. Gabriele Lenzini for all their help and advice during my Ph.D. journey.
I thank the External Reviewers, Prof. Ana Rosa Cavalli and Prof. Ana Cristina Ramada Paiva for their time, effort, and valuable suggestions for improving the content of the thesis.
I would also like to thank my family, my life partner and my friends, whom without this would have not been possible. I also appreciate all the support I received from all my amazing colleagues.
The research presented in this thesis was partially supported by:

- The Luxembourg National Research Fund (FNR) CORE project C16/IS/11333956 “DAPRECO: DAta Protection REgulation COmpliance [1].
- CyberSec4Europe, H2020 Programme Grant Agreement No. 830929 [2] .
- Building Trust in Ecosystems and Ecosystem Components (BIECO), Programme H2020 Grant Agreement No 952702 [3].
- Being safe around collaborative and versatile robots in shared spaces (COVR), Horizon 2020 Agreement No. 779966 [4].

[1]https://www.fnr.lu/projects/data-protection-regulation-compliance/
[2]https://cybersec4europe.eu/
[3]https://www.bieco.org/
[4]https://www.safearoundrobots.com/home

Ringraziamenti

Le seguenti righe sono le più difficili da scrivere. Un sincero grazie a Eda, la tua presenza è fondamentale nella mia vita. Con te sono cresciuto non solo professionalmente, ma anche come persona. Ricordo il tuo lungo abbraccio in un momento molto difficile per me. Quindi, grazie per la tua amicizia. Mi viene in mente Giugno 2017, davanti a un caffè, chiacchieravamo, le tue parole mi risuonano ancora nella testa: Said ora puoi fare quello che vuoi, sei libero; se vuoi intraprendere la strada dottorato, lo sai io ti considero un amico e un collega, io ci sono qualsiasi tematica tu scelga, a patto che… La tua disponibilità e la tua schiettezza sono imparagonabili: è questo che io apprezzo molto di te; mi hai sempre detto belle cose e hai saputo sempre dirmi con tanta delicatezza e sincerità le cose difficili. Sei, Eda, una persona sulla quale posso sempre contare.

Senza il tuo supporto, non avrei mai avuto l'opportunità di conoscere Gabriele. I dieci mesi trascorsi in Lussemburgo, Gabriele, sono stati per me un periodo di grande riflessione e crescita personale. La tua amicizia, le nostre chiacchierate, le nostre passeggiate sono e saranno sempre una fonte di ispirazione. Ho imparato molto da te, non solo in ambito scientifico, ma soprattutto in ambito umano. G.K.P.K. mi avete fatto sentire a casa, in famiglia. Grazie per questo.

Una persona fondamentale in questo intreccio di relazioni è stato Cesare che mi ha supportato nei primi lavori sul GDPR in Lussemburgo. Le tue raccomandazioni, correzioni e spiegazioni delle norme legali le porto sempre con me. Porto con me la mia prima missione: Lussemburgo 2018. C'eri anche tu, Antonello. Grazie per essermi stato vicino all'inizio, durante e dopo questo periodo di dottorato. Mi ricordo le tue parole in aeroporto (mi avevano emozionato) e soprattutto quelle del giorno prima.

Grazie Prof.ssa Anna Monreale per aver accettato il primo anno di farmi da supervisore interno. Grazie Prof.ssa Laura Ricci per aver accettato di essere membro della commissione interna. Le vostre indicazioni e suggerimenti sono stati fondamentali durante tutto questo viaggio. Grazie per tutto il tempo dedicatomi per valutare rigorosamente i miei progressi durante i tre anni e mezzo di dottorato.

Senza la collaborazione di molte altre tante belle persone questo lavoro non sarebbe giunto a termine. Per la parte testing, grazie Francesca per la tua collaborazione.

Durante il periodo del primo lockdown ho conosciuto i membri del gruppo WN: Antonino, Francesco, Michele e Paolo, con il quale ci siamo visti di sfuggita quando si poteva ancora mangiare in bella compagnia senza distanziamento sociale. Spero di vedervi tutti di persona finalmente. Vorrei ringraziare Vincenzo Savarino per avermi dato la possibilità di conoscere le problematiche relative alla gestione del consenso e il mondo industriale.

Durante il periodo all'estero ho condiviso bei momenti con Arianna, Borce e Iraklis. Grazie per la vostra gentilezza nei miei confronti.

Deh, senza di te non avrei saputo come fare: Nice. Sei la colonna portante della mia vita. Gli ultimi 3/4 anni dei nostri 18 immagino siano stati molto duri per te. Non solo ti ringrazio per il tuo incondizionato quotidiano Amore, ma ti chiedo scusa per le mie eventuali (certe) mancanze. Sei la mia compagna e amica. Grazie di essere la persona meravigliosa che sei. Cercherò di ricordarmi le tue parole dopo aver consegnato la prima versione della tesi: mai più dottorato!!

Ogni volta che penso a te Papà, è un colpo al cuore. Sei sempre nei miei pensieri. È un colpo più profondo quando penso anche a te Mamma. Senza di te non sarei la persona che oggi sono. Grazie per l'amore materno, per i tuoi silenzi, per le tue premure e le parole di conforto che ogni volta sei capace di darmi. Fatiha, Naima, Samira, Mina e Laila: Grazie per essere delle sorelle fantastiche. Majid, mio fratello, Grazie. I miei nipotini che ultimamente ho visto poco: Sofia, unica femmina, e i maschietti Anwar, Gabriele e Fabian. Vi voglio un sacco di bene.

Last but not the least, grazie Daniele, sei il compagno ideale di questo fantastico viaggio chiamato Dottorato di Ricerca. Ti ringrazio per la tua gentilezza e amicizia. Ricorderò sempre le sedute del Senato.

Summary

THE GDPR is changing how Personal Data should be processed. It states, in Art. 5.1(f), that "[data] should be processed in a manner that ensures appropriate security of the personal data [...], using appropriate technical or organizational measures (integrity and confidentiality)". We identify in the Access Control (AC) systems such a measure. Indeed, AC is the mechanism used to restrict access to data or systems according to Access Control Policies (ACPs), i.e., a set of rules that specify who has access to which resources and under which circumstances. In our view the ACPs, when suitably enriched with attributes, elements and rules extracted from the GDPR provisions, can suitably specify the regulations and the AC systems can assure a by-design lawfully compliance with the privacy preserving rules. Vulnerabilities, threats, inaccuracies and misinterpretations that occur during the process of ACPs specification and AC systems implementation may have serious consequences for the security of personal data (security perspective) and for the lawfulness of the data processing (legal perspective). For mitigating these risks, this thesis provides a systematic process for automatically deriving, testing and enforcing ACPs and AC systems in line with the GDPR. Its data protection by-design solution promotes the adoption of AC systems ruled by policies systematically designed for expressing the GDPR's provisions. Specifically, the main contributions of this thesis are: (1) the definition of an Access Control Development Life Cycle for analyzing, designing, implementing and testing AC mechanisms (systems and policies) able to guarantee the compliance with the GDPR; (2) the realization of a reference architecture allowing the automatic application of the proposed Life Cycle; and (3) the use of the thesis proposal within five application examples highlighting the flexibility and feasibility of the proposal.

List of publications

International Journals

1. Paolo Barsocchi, Antonello Calabrò, Antonino Crivello, Said Daoudagh, Francesco Furfari, Michele Girolami and Eda Marchetti: COVID-19 & Privacy: Enhancing of Indoor Localization Architectures towards Effective Social Distancing. Array. https://doi.org/10.1016/j.array.2020.100051

2. Said Daoudagh, Francesca Lonetti and Eda Marchetti: An automated framework for continuous development and testing of access control systems. J Softw Evol Proc. 2020;e2306. https://doi.org/10.1002/smr.2306

3. Said Daoudagh, Francesca Lonetti, Eda Marchetti: XACMET: XACML Testing & Modeling. Softw. Qual. J. 28(1): 249-282 (2020)

International Conferences/Workshops with Peer Review

1. Said Daoudagh and Eda Marchetti: GRADUATION: A GDPR-based Mutation Methodology. QUATIC 2021.

2. Said Daoudagh, Eda Marchetti, Vincenzo Savarino, Roberto Di Bernardo, Marco Alessi: How to Improve the GDPR Compliance Through Consent Management and Access Control. ICISSP 2021.

3. Said Daoudagh, Francesca Lonetti and Eda Marchetti: Continuous Development and Testing of Access and Usage Control: A Systematic Literature Review. ESSE 2020

4. Paolo Barsocchi, Antonello Calabrò, Antonino Crivello, Said Daoudagh, Francesco Furfari, Michele Girolami, Eda Marchetti: A Privacy-By-Design Architecture for Indoor Localization Systems. QUATIC 2020: 358-366

5. Said Daoudagh, Francesca Lonetti, Eda Marchetti: Assessing Testing Strategies for Access Control Systems: A Controlled Experiment. ICISSP 2020: 107-118

6. Said Daoudagh, Eda Marchetti: A Life Cycle for Authorization Systems Development in the GDPR Perspective. ITASEC 2020: 128-140
7. Said Daoudagh, Eda Marchetti: Defining Controlled Experiments Inside the Access Control Environment. MODELSWARD 2020: 167-176
8. Said Daoudagh, Francesca Lonetti, Eda Marchetti: A Framework for the Validation of Access Control Systems. ETAA@ESORICS 2019: 35-51
9. Said Daoudagh, Francesca Lonetti, Eda Marchetti: A Decentralized Solution for Combinatorial Testing of Access Control Engine. ICISSP 2019: 126-135
10. Said Daoudagh, Francesca Lonetti, Eda Marchetti: A General Framework for Decentralized Combinatorial Testing of Access Control Engine: Examples of Application. ICISSP (Revised Selected Papers) 2019: 207-229
11. Cesare Bartolini, Said Daoudagh, Gabriele Lenzini, Eda Marchetti: Towards a Lawful Authorized Access: A Preliminary GDPR-based Authorized Access. ICSOFT 2019: 331-338
12. Antonello Calabrò, Said Daoudagh, Eda Marchetti: Integrating Access Control and Business Process for GDPR Compliance: A Preliminary Study. ITASEC 2019
13. Cesare Bartolini, Said Daoudagh, Gabriele Lenzini, Eda Marchetti: GDPR-Based User Stories in the Access Control Perspective. QUATIC 2019: 3-17
14. Antonia Bertolino, Said Daoudagh, Francesca Lonetti, Eda Marchetti: An automated model-based test oracle for access control systems. AST@ICSE 2018: 2-8

Peer Reviewed Posters

1. Said Daoudagh: GDPR Compliance Through Authorization Systems. PhD Forum@ITASEC 2020
2. Cesare Bartolini, Said Daoudagh, Gabriele Lenzini and Eda Marchetti: Testing of GDPR-based Access Control Policies. SessionPoster@ESORICS, 2019
3. Cesare Bartolini, Said Daoudagh, Gabriele Lenzini and Eda Marchetti: GDPR-based Access Control. CySep Summer School and IEEE EuroS&P 2019 Conference (2019)

Others

1. Antonia Bertolino, Said Daoudagh, Francesca Lonetti and Eda Marchetti: An automated model-based test oracle for access control systems. CoRR abs/1809.02724 (2018)

List of Acronyms

A

ABAC	Attribute-Based Access Control
AC	Access Control
ACM	Access Control Mechanism
ACP	Access Control Policy
ACS	Access Control System
ADLC	Authorization Development Life Cycle
ASD	Agile Software Development

C

CE	Controlled Experiment

D

DAC	Discretionary Access Control
DMS	Data Management System
DPIA	Data Protection Impact Assessment
DPO	Data Protection Officer
DSM	Digital Single Market

G

GDPR	General Data Protection Regulation
GQM	Goal-Question-Metric
GRADUATION	GdpR-bAseD mUtATION
GROOT	GdpR-based cOmbinatOrial Testing

H

HIPAA	Health Insurance Portability and Accountability Act
I	
ICT	Information and Communication Technology
ILS	Indoor Localization System
IT	Information Technology
M	
MAC	Mandatory Access Control
N	
NIST	National Institute of Standards and Technology
NLACP	Natural Language Access Control Policy
P	
PDP	Policy Decision Point
PEP	Policy Enforcement Point
PET	Privacy Enhancing Technology
PrOnto	Privacy Ontology
R	
RAccOnto	GDPR-based Access Control Ontology
RBAC	Role-Based Access Control
S	
SDLC	Software Development Life Cycle
SME	Small and Medium-sized Enterprise
SUT	System Under Test
W	
WP29	Article 29 Working Party
X	
XACMET	XACML Modeling & Testing
XACML	eXtensible Access Control Markup Language
XMF	XACML Mutation Framework
XP	eXtreme Programming

Contents

List of Figures

List of Tables

List of Algorithms

Part I

Objectives and Background

CHAPTER 1

Introduction: Objectives and Goals

THE General Data Protection Regulation (GDPR) is the EU legal framework for the protection of Personal Data of European citizens [90], which aims to harmonize the different data protection laws in Europe and to strengthen the rights of individuals. Thus, the GDPR precisely defines the involved concepts and roles: *Personal Data* is defined as any information about a *Data Subject*, i.e., an identified or identifiable natural person; data *Controller* and data *Processor* are defined as the persons involved into the data management and data processing Personal Data respectively. The GDPR imposes also several duties, and defines a system of fines to induce the Controller and the Processor to be compliant with the regulation. In particular, they need to:

i) ensure appropriate technical security level of personal data, as dictated by the "Integrity and Confidentiality" principle (Art. 5.1(f));

ii) demonstrate the compliance with the GDPR, as required by the "Accountability" principle (Art. 5.2); and

iii) adapt and rethink their data practices so as to be aligned with the "Data protection by design and by default" approach (Art. 25).

Despite the simplicity of these statements, their realization is not straightforward, especially when the role of data *Controller* and data *Processor* are taken inside Small and Medium-sized Enterprises (SMEs). Indeed, one of the most experienced difficulties is the GDPR's technical interpretation [17]. The simplicity of the natural language structure of the GDPR leaves the floor to a concrete difficulty for software architects, developers and security experts in translating the GDPR's provisions into technical requirements especially in case of lack or no sufficient legal expertise. If big organizations have the economic power to overcome this problem, by investing large amount of money both in technologies and legal consulting, usually this is not the same for SMEs. These look for low-cost, easy-to-use solutions for assuring their compliance with the GDPR and being prepared to comply with its provisions. Indeed, for all organizations being (by-design) compliant with the GDPR means having technical (and organizational) solutions that: (1) are general-purpose; (2) must take in consideration the regulation by-design; (3) must be easily integrated with the existing business processes; and finally, (4) must be rooted in the GDPR principles dictated in Art. 5.

At state of the practice, there is not a comprehensive ready-to-apply solution for all the above mentioned challenges.

Indeed, the problem is still far for being solved even for extremely large companies. Based on the data of CMS Legal Services EEIG[1], which monitors the GDPR enforcement, at the moment of writing this thesis data protection authorities have imposed 569 fines within the EU state members. Among them, Spain is the one having the highest number of imposed fines for a total of 205; whereas Italy is the state having the highest Sum of Fines for a total amount of 70M Euro for 50 fines imposed. Considering instead the severity of the fines ever imposed, France is the first in the list that imposed a fine to Google for 50M Euro in January 2019. Concerning the top 10 classification of the statistics regarding the "fines by type of violation", at the first place, there is "Insufficient legal basis for data processing" with a total of 218 fines; at second-place there is the "Insufficient technical and organizational measures to ensure information security" with a total of 129 fines. More than interesting, these data highlights that despite the economic power to invest to promote the GDPR initiatives for lawfully processing personal data, we are still far from being sufficiently compliant with the regulation.

Therefore, inspired primarily by the “Integrity and Confidentiality” principle, that calls for the adoption of Access Control (AC) to regulate the access to Personal Data, the underlining idea of this research:

> *leverage AC systems, the de facto mechanisms used to restrict data access, as a technical solution for protecting “personal data by-design”, and gaining legal compliance with the GDPR.*

The choice of AC systems has two important strengths: (1) their structure and (2) their applicability. Structurally, AC systems are based on Access Control Policies (ACPs), i.e., a set of rules that specify who has access to which resources and under which circumstances [205]. Because AC systems satisfy by construction the principle of *Integrity and Confidentiality* (Art. 5.1(f)), the idea of this research is to enrich them with policies elicited from the GDPR’s provisions. This lets the AC systems to realize the compliance by-design with the GDPR’s demands.

Considering the applicability, AC are general-purpose models supported by a standard and a reference architecture and easily integrable within the existing business processes, so as to decoupling the business logic from authorization.

From a technical point of view, deriving ACPs aligned with the GDPR’s provisions involves two conceptual mappings. First, the association of Resources to Personal Data and Controller (or Processor, or Data Subject) to who is requesting access to the data.

[1] https://www.enforcementtracker.com/ (Last Access 2021.03.28)

Then, it is necessary to associate the GDPR's provisions (e.g., *identify*, *extract*, *translate* and *encode*) to enforceable ACPs [238].

While the first association comes in a quite natural manner, the second one requires a careful translation process. Indeed, GDPR's provisions could have an ambiguous interpretation, could include implicit information, could be unstructured or could not be easily mappable into formal policies. Failing the correct association between GDPR's provisions and the ACPs [238] may have serious consequences not only on the protection of personal data but also for the lawfulness of the process adopted.

Thus, fundamental part of this thesis is to provide a systematic process for the realization of AC systems and ACPs compliant-by-design with the GDPR. As a result, the leveraged AC systems can protect personal data (*security perspective*) and process them lawfully (*legal perspective*).

1.1 Challenges

As already mentioned, in using AC as technical solution for protecting "personal data by-design" and gaining legal compliance with the GDPR, several challenges have to be faced up. We summarize in this section those strictly related with the thesis topics by referring to [210] for a detailed characterization. Considering in particular the data privacy aspects, the challenges are:

Performing Data Protection Impact Assessment. Performing Data Protection Impact Assessment (DPIA) in accordance with the GDPR is pivotal to promote and achieve privacy-by-design. For the different organizations, fulfilling the GDPR requirements is an integrated part of their business. The challenge here is that the GDPR's requirements are often too vague and open. This makes them subject to interpretation. Therefore, it might be difficult to correctly and completely comply with them [210].

GDPR-based development life cycle. The available development life cycles do not completely incorporate the privacy-by-design principles, and proposals targeting the GDPR's demands are still 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.

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

Considering in particular the access control aspects, the considered challenges are:

Modeling the law. For using Access Control elements and extensions to address concepts related to a given law requires formal translations in order to avoid misinterpretation or errors. Thus, the necessity of automatically enforceable matching of actual attributes gathered from legal use cases and the resulting policies in order to comply with the GDPR's obligation of "data protection by design and by default".

Enforcing privacy (security) policies. A reference access control architecture to support context-aware security policies should be defined so as to assure the enforcement of the privacy policies throughout different kind of systems and environment. Additionally, methods for leveraging the integration of the access control and business processes as well as mechanisms to guarantee the GDPR compliance during business activities of data management and analysis should be conceived.

Verification & Validation. The GDPR is changing how Personal Data should be processed. Part of the scientific and industrial worlds are replying to these exigencies by modifying the Access Control Mechanisms (ACMs) and the way of managing and writing their policies. Consequently, specific testing strategies or validation approaches should be defined so as to assure that the generated data protection based policies are aligned with the GDPR. Failing this task can lead in developing ACPs that allows an unauthorized user to access protected personal data (*security perspective*) and consequently resulting in an unlawful processing (*legal perspective*). Therefore, the need of developing facilities for verifying the compliance of the derived policy with respect to the requirements expressed in the GDPR.

1.2 Goal and Research Questions

The thesis focuses on data security and data (privacy) protection and follows software engineering procedures and best practices for join together Access Control (AC), Data Protection by-Design, and AC Testing into a unique *Privacy By Design* methodology. In particular, thesis main research goal can be summarized as:

> **Research Goal.** *To leverage AC systems, the de facto mechanisms used to restrict data access, as a technical means for protecting "personal data by-design", and gaining legal compliance with the GDPR.*

By clarifying the role of security measures, and in particular authorization systems, in the context of Data Protection, and by providing a systematic approach for implementing them, we can help achieving compliance with the GDPR. The results of the research presented in this thesis can be interpreted as guidelines of *data protection by design*, by using Access Control Systems (ACSs).
However, due to the variety of aspects included in the primary research goal, we have structured our broad investigation into the following open questions.

1.2.1 (RQ 1) How can authorization systems, and in particular AC, be used for guaranteeing compliance with the GDPR?

Authorization systems are a cornerstone of security, and they are being used for a long time to protect classified resources. They have also been used for dealing with different privacy concepts such as purpose and consent. Consequently, can they be leveraged for protecting personal data and satisfying the GDPR compliance? Is there already a comprehensive methodology or set of guidelines to make easier adoption in the state of the practice? Are there concepts and knowledge coming from other disciplines that can be exploited for systematically customize existing authorization systems? And how to encode the GDPR's obligations in the authorization systems? In order to provide a systematic approach for designing and using authorization systems in the context of the

GDPR, an accurate analysis of their currently adoption in other legal frameworks and in the industry need to be performed.

1.2.2 (RQ 2) To what extent can the GDPR's obligations be represented and enforced using Access Control Technologies?

Legal requirements are expressed in natural language and they are agnostic to the available technologies presented in our time. Therefore, they can be too vague to be automatically implemented within a reference system or technology. However, by defining the Integrity and Confidentiality principle, the GDPR implicitly calls for adopting ACSs. Indeed, ACSs are usually regulated by ACPs, which specify who, what, when, where, how, and why (i.e., the 5W1H) a user is granted or denied to access to a given asset. This information includes also the Personal Data. Thus, the question of how to identify, extract and define the ACPs that are by-design compliant with the GDPR is not straightforward. In particular, how to model AC policies in reference to the GDPR? How to identify AC requirements from the GDPR? How many AC requirements can the GDPR encode? This RQ can be reloaded also as: is it possible to model AC policies in reference to the GDPR? How can we identify AC requirements from the GDPR? How many AC requirements the GDPR encodes?

1.2.3 (RQ 3) Is it possible to gather technical requirements from the legal specifications defined in the GDPR?

The GDPR, as any other law, is intrinsically expressed in legal jargon, even if targets the organizations that process personal data. Its natural language provisions are far to be immediately interpreted as technical requirements, even if with the "personal data by design" obligation the GDPR forces organizations to implementing system's by-design aligned with the GDPR. This causes a general re-think of the organizations' data practices and a continuous and expensive research of ad-hoc technical solutions so as to guarantee the compliance with the GDPR's obligations. Therefore, a key aspect is the availability of facilities able to automatically extract from the legal specifications all and only the required information and to interpret them into technical requirements that can be easily implemented.

1.2.4 (RQ 4) For accomplishing compliance with the GDPR, which are the supporting technologies that could be integrated with AC?

The continuous growing of interest for the compliance with the GDPR is promoting the realization of different solutions in both industry and academia context. The solutions proposed into this thesis want to be in line with the current state of the practice. Therefore, an accurate analysis of the available proposals and an evaluation of their effectiveness in achieving compliance with the GDPR are necessary so as to select the most suitable ones for being profitably integrated into the ACS. Thus, another specific question is: are the available solutions sufficiently and suitably mature to be integrated into the AC (or into their reference architecture)? Additionally, in order to promote the adoption of the solutions of this thesis into real context: are the available proposals based on open standards? Can the available solutions help in achieving the compliance with the GDPR's demands? And in case, how is this possible?

1.2.5 (RQ 5) Which are the most suitable application domains for applying Access Control Technologies able to achieve the GDPR compliance?

The GDPR is potentially applicable into every domain: any context processing personal data is obliged to obey the GDPR's principles. At the same time, in Information and Communication Technologies (ICTs) systems also the ACSs are having a widespread adoption for ruling the resources and data access. Thus, the synergic union between GDPR and ACSs could be the crucial point for developing everywhere adoptable solutions. The feasibility of this idea need to be investigated considering the following questions: are AC really suitable for different application domains? Can ACSs be easily integrated in preexisting processes/environments? Is it possible to enable the authorization as a service paradigm? And more specific, is it possible to decouple business logic from the authorization one?

1.2.6 (RQ 6) Is it possible to realize an integrated test environment for the validation of (GDPR-aware) access control systems?

High security level is an important attribute for many environments. Thus, discovering the criticalities of a system is always a valid means for putting in practice efficacious and corrective actions to improve its overall security. This is extremely true and important for ACSs (both ACPs and Access Control Mechanisms (ACMs)), because their security and privacy vulnerabilities could insert either the risk of releasing inadequate security solutions that allow unauthorized access (*security perspective*) or to allow unlawful processing of personal data (*legal perspective*). At the state of the practice, most of the time the criticalities detection is achieved through the application of effective and efficient testing approaches. However, testing is a time consuming, error prone activity and it represents a critical step of the development process. In case of ACSs, due to their complexity, testing process becomes even more expensive, because accurate and specific validation approaches should be put in place. Thus the question: is it possible to realize a test environment specifically conceived for ACSs? And in particular: is it possible to develop specific test strategies? Is it possible to provide facilities for test cases generation and selection? Is it possible to develop an integrated environment for the automatic test cases execution and results collections? Is it possible to define an oracle for speeding up the test results evaluation? Is it possible to statistically evaluate the effectiveness of the applied testing strategies?

1.3 The Impact of COVID-19 Pandemic

As a response to the global outbreak of the COVID-19 pandemic, authorities have enforced a number of measures including *social distancing* and *travel restrictions* that lead to the *temporary* closure of activities ranging from public services, schools, industry to local businesses. Indeed, most of the activities, collaborations, projects [159] have been affected and slowed down by the new way of living, working and behaving [18]. Last but not least, the Ph.D. courses and thesis.

As for many others ongoing productions, COVID-19 pandemic forces this thesis to an unexpected adaption and revision of its schedule, targets and plans to face the delay in project collaborations, in collecting results, in the validation activities especially in relation with RQ 6 (see Sec. 1.2.6).

However, every cloud has a silver lining. New opportunities, never thought or scheduled before came into our path letting the exploration of the thesis proposals in different (and unexpected) application domains and contexts (RQ 5 in Section 1.2.5): the GDPR-based ACS as a mean for enhancing the Indoor Localization Systems (ILSs) towards effective privacy preserving social distancing.

Indeed, the rapid dynamics of COVID-19 "calls for quick and effective tracking of virus transmission chains and early detection of outbreaks." [171]. Authors in [171] recognize that Location (Big) Data [2] "should be seen as a potentially powerful weapon in combatting the pandemic". They also argue that the GDPR compliance, abiding Art. 25 (Data protection by design and by default), enables the benefits of them. Therefore, being compliant is no more a challenge but a daily reality; means and facilities to make all the users able to assert their rights are stringent requirements; the more the world becomes connected and operate in smart manner the more the GDPR becomes an important need in everyday.

During COVID-19 pandemic, the possibility to apply the thesis solutions in unusual, different situations confirmed our initial intuition: leverage AC systems, as a technical means for protecting “personal data by-design”, and gaining legal compliance with the GDPR is an hot topic of research, and an urgent need in any application domains and environments.

As a practical point of view, our reaction to the COVID-19 related delay has been to slightly change our initial research plan: we reserved more time to RQ 5, and we reduced the effort to invest in RQ 6 considering the already collected results sufficient enough to positively answer it. Therefore, we included two open research questions:

(RQ 5.1) How ILSs can benefit from the adoption of GDPR-based ACMs, to lawfully manage location (personal) data?

(RQ 5.2) How can we leverage them to *lawfully* guarantee the disruptive countermeasure imposed by COVID-19 pandemic, namely *preserving social distance* among people in indoor environments?

We report in Chapter 10 and Chapter 11 the results of RQ 5.1 and RQ 5.2 evaluation respectively.

1.4 Thesis Overview

The thesis is logically composed of the five parts as depicted in Figure 1.1. As in the figure, each chapter is related to one or more RQs previously presented. In the following, we summarize the content of each part, and the description of each chapter is provided.

(PART I) Objectives and Background. It illustrates the objectives of the thesis and discusses the main research questions the thesis aims to answer (Chapter 1, i.e., the current chapter), and it contains background about the main concepts used in the thesis, i.e., Data Privacy and Data Security (Chapter 2).

[2] Location Data are specific Personal Data. They are potentially able to describe the movement of people in greater detail. This is true in both the indoor and outdoor environments.

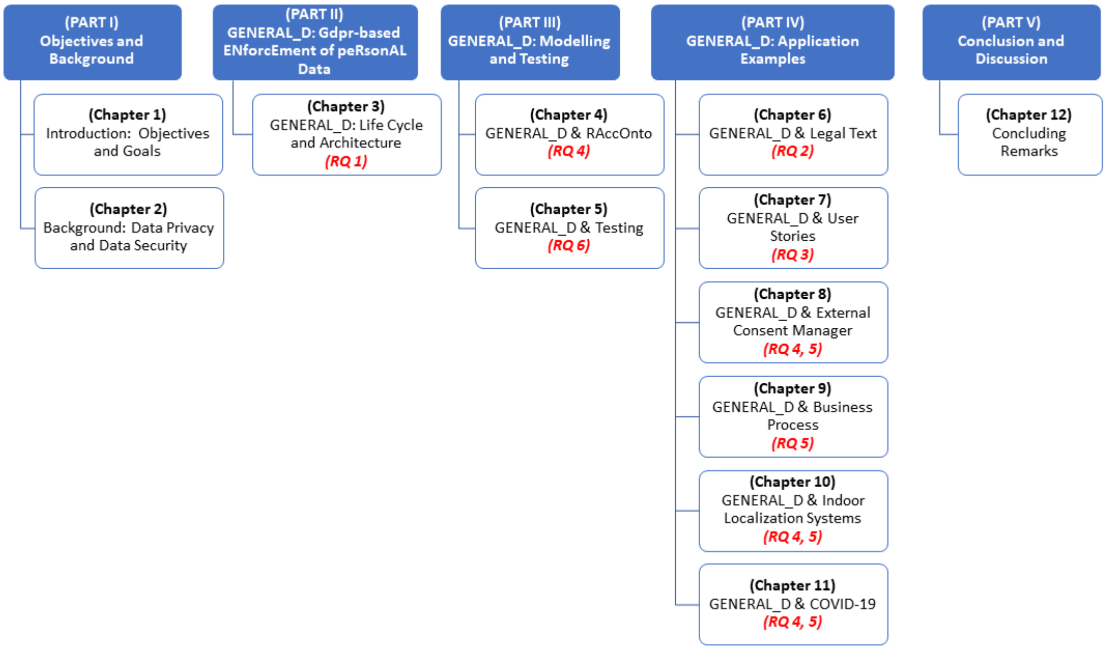


Figure 1.1: *Overview of the thesis content, chapters and the RQs they answer.*

(PART II) GENERAL_D: Gdpr-based ENforcEment of peRsonAL Data. This is the core part of the thesis where the GENERAL_D proposal is introduced. This is a flexible Life Cycle and its reference architecture for developing access control systems that are by-design compliant with the GDPR (Chapter 3). It also introduces the validation of the proposal through five examples that will be detailed in PART IV.

(PART III) GENERAL_D: Modelling and Testing. This part describes the ontology and the GDPR profile used for modelling AC in reference to the GDPR (Chapter 4). It also describes the features and methodologies useful to validate both (GDPR-based) access control policies and mechanisms (Chapter 5).

(PART IV) GENERAL_D: Application Examples. This part is composed of six chapters (Chapter 6-11) and focuses on the validation of the thesis proposal. In particular, it aims at illustrating the peculiarities of the proposal (i.e., GENERAL_D) that are: Generalization, Flexibility, Adaptability and Cost Reduction. In particular, each chapter refers to one of the application examples reported in Chapter 3.

(PART V) Conclusion and Discussion. This part concludes the thesis by reporting the final remarks (Chapter 12).

In detail the content of each chapter is summarized here below.

Chapter 2. Background: Data Privacy and Data Security. In this chapter, we give the main concepts related to the topic of the thesis: Data Privacy and Security. Firstly, we examine the concept of Privacy and how this concept became important as technology advance. We illustrate the first framework related to the privacy by design, then we discuss its integration into the EU legislation as Data Protection by Design. Afterwards, we illustrate the main concepts of the currently data protection legal framework, i.e., the GDPR. Finally, this chapter introduces the main properties of data security and their relation with Access Control (AC).

Chapter 3. GENERAL_D: Life Cycle and Architecture. This chapter illustrates a comprehensive Life Cycle for developing access control in compliance with the GDPR, that answers RQ 1 (1.2.1). After illustrating the main phases composing the Life Cycle, we illustrate five application examples highlighting the flexibility and feasibility of the proposal. These examples are then discussed in the following chapters. The conceived Life Cycle is tightly rooted into the sixth principle of the GDPR, i.e., Confidentiality and Integrity, and into the Data Protection by Design and by Default dictated in Art. 25. This chapter also briefly illustrates a reference architecture (named GENERAL_D) aiming at as much as automating the proposed Life Cycle.

The content of this chapter is adapted from the results published in [55, 78], and helps answer RQ 1 (1.2.1).

Chapter 4. GENERAL_D & RAccOnto. This chapter aims at defining an ontological representation of AC concepts in reference to the GDPR, and standard attributes to be used during modeling and development phase of policies. It firstly introduces GDPR-based Access Control Ontology (RAccOnto), a data protection ontology that models AC and leverages the Privacy Ontology (PrOnto) which is a legal ontology and it models the GDPR main concepts. RAccOnto contributes to write ACPs that are by-design compliant with the GDPR. To write however ACPs in reference to the GDPR, there are needs to explicitly refer to the selected GDPR's concepts within the policy. Therefore, the second contribution is an *XACML GDPR Policy Profile* proposal that provides standard attributes according to the GDPR's concepts. The results in this chapter, even not published yet, help answer RQ 4 (1.2.4), and the proposals are currently being under thorough validation within the CyberSec4Europe [3] project community.

Chapter 5. GENERAL_D & Testing. In this chapter, we introduce a comprehensive testing framework capable to formally validate both ACPs and ACMs, by enabling to conduct Controlled Experiments (CEs) in the context of AC. First, we introduce a typical Testing Process, and a reference architecture for its automation that can be customized with real artifacts. Then, we advance the notion of families of CEs in the context of AC, enabling conceiving well-defined testing goals to formally conduct Controlled Experiments in the context of AC. After that, we illustrate GdpR-based cOmbinatOrial Testing (GROOT), a general combinatorial strategy for testing systems managing GDPR's concepts; and XACML Modeling & Testing (XACMET), a testing framework which includes facilities for test cases generation, for automatically derivation of AC oracle and for measuring the coverage of test requests. For assessing GDPR-based test cases generation strategy, we define GdpR-bAseD mUtATION (GRADUATION), a generic methodology based on mutation analysis. The chapter concludes by illustrating a CE in the context of AC, detailing three main steps: definition of the experiment for the

[3] CyberSec4Europe H2020 Programme Grant Agreement No. 830929: https://cybersec4europe.eu/

comparison of two testing strategies; instrumentation and execution of the experiment; analysis of the results.

The chapter is based on several scientific contributions [31, 45, 69, 71–77, 79] and summarizes a long quest in AC testing, which helps answer RQ 6 (1.2.6). Moreover, part of the results of this chapter are included into the BIECO[4] Project and will be used during its validation activities.

Chapter 6. GENERAL_D & Legal Text. This chapter discusses how to protect personal data from unauthorised or unlawful processing, as dictated by the GDPR's sixth principle, *Integrity and Confidentiality*. By using AC as technical means to protect personal data, we observed that an initial trivial mapping (i.e., Personal Data can be considered the resources, whereas the Controller, the Processor, or the Data Subject are the subjects requesting access to the resources) could be not sufficient to guarantee compliance with the GDPR, by defining ACPs capturing only that mapping. Indeed, it may be challenging for ACPs designers to *identify*, to *extract*, to *translate* and to *encode* the GDPR's provisions into enforceable ACPs. Provisions can be ambiguous and can include implicit information. They are also unstructured and therefore not straightforwardly expressible in a formal policy. All these issues call for a systematic process for designing ACPs properly linked to the GDPR. Failing this task may have serious consequences: not only the AC system enforcing the ACPs can leave personal data unprotected, but the AC system may also become unlawful for the specific context of the GDPR. The risk can be mitigated by promoting the adoption of AC systems enforcing policies systematically designed for expressing GDPR's provisions. Consequently, the results of the research reported in this chapter are therefore a systematic approach for authoring access control policies that are by-design aligned with the provisions of the GDPR.

This chapter answers RQ 2 (1.2.2), and the results are reported in the related scientific contribution [32].

This research was performed in collaboration with Dr. Cesare Bartolini (SnT, University of Luxembourg) and Prof. Gabriele Lenzini head of IRiSC Lab[5] (SnT, University of Luxembourg), in the context of DAPRECO Project[6].

Chapter 7. GENERAL_D & User Stories. This chapter discusses the problem of how to translate the GDPR's provisions in technical requirements, in the AC perspective. Provisions are pieces of law and are not written to be immediately interpreted as technical requirements; the task is thus not straightforward. The *Agile software development methodology* can help untangle the problem. It promotes detailed procedure and form for describing requirements such as the specification of *User Stories*. These are concise yet informal requirement descriptions telling who, what and why something is needed by users. Additionally, User Stories are organized into prioritized lists, called *backlogs*. Therefore, inspired by the Agile development process, in this chapter we advance the notion of *Data Protection backlogs*, which are lists of User Stories about GDPR

[4]Building Trust in Ecosystems and Ecosystem Components (BIECO) Programme H2020 Grant Agreement No 952702. `https://www.bieco.org/`

[5]Interdisciplinary Research Group in Socio-technical Cybersecurity (IRiSC): `https://irisc-lab.uni.lu/`

[6]The Luxembourg National Research Fund (FNR) CORE project C16/IS/11333956 "DAPRECO:DAta Protection REgulation COmpliance:`https://www.fnr.lu/projects/data-protection-regulation-compliance/`

provisions described in terms of technical requirements. Consequently, for each User Story, we provided its corresponding ACP, so as to make easier the design and implementation of GDPR compliant AC systems incrementally. The results in this chapter help to answering RQ 3 (1.2.3), and they are published in [30]. This research was performed in collaboration with Dr. Cesare Bartolini (SnT, University of Luxembourg) and Prof. Gabriele Lenzini (SnT, University of Luxembourg), in the context of DAPRECO Project.

Chapter 8. GENERAL_D & External Consent Manager. This chapter reports the results help answering RQ 4 (1.2.4) and RQ 5 (1.2.5). It provides evidences of the flexibility of our proposal, i.e., GENERAL_D, in adapting and integrating pre-existing solutions. In particular, we consider the integration of an available Consent Manager (CM) and an Access Control (AC) to aid organizations to comply with the GDPR. The idea is to use GENERAL_D for converting the GDPR machine-readable format provided by an External CM into a set of enforceable ACPs. In this chapter, we defined a layered architecture able to make (potentially any) systems compliant by-design with the GDPR. To validate the feasibility of this proposal, we provide also a proof-of-concept by integrating an AC Manager, i.e., GENERAL_D, and an External Consent Manager coming from an industrial context.

This chapter is based on an industrial collaboration with Engineering [7] under the umbrella of CyberSec4Europe [8] EU Pilot project. The content presented here was published in [80].

Chapter 9. GENERAL_D & Business Process. Currently, the scientific communities and private companies are actively working to provide theoretical and practical solutions for enforcing the adoption of the GDPR and its compliance problem. In line with the *data protection by design* obligation, this chapter proposes an approach for the automation and enforcement of the GDPR requirements. The idea is to extend the currently adopted access control mechanisms, so as to leverage them to the enforcement of the GDPR compliance during business activities of data management and analysis. From a practical point of view, this means to integrate into the existing business processes specific facilities for assisting in the design, development, maintenance, and verification of the GDPR requirements, as well as to modify the language and architecture of the access control systems so as to let the management of the GDPR principles and obligations. For this, the basic steps of the proposed approach are provided as well as an example used to clarify the integrated use of access control systems and business process models.

The results of this chapter help answering RQ 5 (1.2.5) and are published in [55].

Chapter 10. GENERAL_D & Indoor Localization Systems. In this chapter, we discuss how the adoption of Consent Manager, based on open specification provide by Kantara initiative, and ACPs templates (e.g., those defined in Chapters 6 and 7) can help untangle the GDPR compliance in indoor environments. Therefore, in this chapter we show, for the first time, how to integrate GENERAL_D within an indoor positioning infrastruc-

[7] https://www.eng.it/

[8] CyberSec4Europe H2020 Programme Grant Agreement No. 830929: https://cybersec4europe.eu/

ture so as to internally guarantee by-design the enforcement of the GDPR's provisions. A prototype example is also provided for feasibility purposes.

The results of this chapter help answering RQs 4 (1.2.4) and 5 (1.2.5) and it is mainly based on the work in collaboration with the Wireless Networks (WN)[9] research group lead by Dr. Paolo Barsocchi (ISTI-CNR)[10]. The content presented here was published in [23].

Chapter 11. GENERAL_D & COVID-19. Because ILSs know your position and consequently could potentially know who is near to you, in this chapter we leverage the privacy-by-design ILS architecture proposed in Chapter 10 to lawfully measure the distance between people. This allows to address, in a privacy preserving way, the new simple yet disruptive requirement imposed by countries as countermeasures to fight COVID-19 pandemic: the so-called social distancing. Indeed, in this chapter we take the opportunity to show the flexibility and applicability of the thesis proposal in a new emerging and not fully explored context. More precisely, we explore the possibility of adopting the indoor localization technologies to measure the distance among users in indoor environments. We discuss how information about people's contacts collected can be exploited during three stages: before, during, and after people access a service. By enhancing the reference architecture for an Indoor Localization System (ILS), presented in Chapter 10, we illustrate three representative use-cases: Visiting a Museum, Airport Access, and Shopping Assistant. We derive some architectural requirements, and we discuss some issues that concretely cope with the real installation of an ILS in real-world settings. Therefore, we explore the privacy and trust reputation of an ILS, the discovery phase, and the deployment of the ILS in real-world settings. We finally present an evaluation framework for assessing the performance of the architecture proposed. This chapter helps answer RQs 4 (1.2.4) and 5 (1.2.5), and the results are reported in the related scientific contribution [24].

This chapter reports the work in collaboration with the Wireless Networks (WN) research group lead by Dr. Paolo Barsocchi (ISTI-CNR).

Chapter 12. Concluding Remarks. This chapter concludes the thesis. We present considerations taken in conclusion to this research quest, we revisit our primary objective and RQs, and we discuss how this thesis addresses them. We also present future works and some open problems which remain to be explored.

[9] `https://www.isti.cnr.it/en/research/laboratories/27/Wireless_Networks_WN`
[10] `https://www.isti.cnr.it/en/`

CHAPTER 2

Background: Data Privacy and Data Security

THIS chapter briefly introduces the main concepts related with Data Privacy and Data Security. In [158] “Data Security and Privacy” are recognized as one of the main research challenges in different domains.

2.1 Data Privacy

The right to privacy arose as the right "to be let alone", that is the right to confidentiality regarding individual’s personal information and private life. Over the years, the impact of the advance of ICT has defined the necessity of creating privacy-friendly technologies to protect the privacy of individuals, in particular personal data, from external interference. In this scenario, Ann Cavoukian, the Information and Privacy Commissioner of Ontario, Canada, introduced the term *privacy by design* formulating seven principles to apply [60]:

- 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 Life Cycle 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.

To notice that the principle *Privacy as the Default Setting*, i.e., *privacy by default*, 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).

Privacy by design, according to this first framework, refers to the application of data protection best practices at design time of building a system, while privacy by default refers to a default settings of a system providing the user protection against privacy risks.

In the legislative environment the Data Protection Directive 95/46/EC[1], at Recital 46 and Article 17, codified these two different concepts of privacy within the "Security of processing", without however mentioning the term *privacy by design*.

Later, in 2009, the Article 29 Working Party (WP29) and the Working Party on Police and Justice observed that the existing measures do not assure adequate privacy; for this reason, they supported the idea of including privacy by design and privacy by default into future frameworks[2].

Indeed, the GDPR embraces those terms in its Art. 25 (data protection by design and by default) which words: the controller shall "taking into account the state of the

[1] Data Protection Directive 95/46/EC: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A31995L0046

[2] 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

art [. . .] both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures [. . .] to integrate the necessary safeguards into the processing in order to meet the requirements of this regulation and protect the rights of data subjects", and " [. . .] by default, only personal data which are necessary for each specific purpose of the processing are processed. [. . .] In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons." (Art. 25).

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"[3]. 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.

2.2 GDPR Concepts

The General Data Protection Regulation (GDPR) [90][4] is the currently European Union Law (Regulation) for the protection of *personal data*. The regulation became into effect on May 2018 and has replaced the previous Data Protection Directive conceived in 1995. The aim of the Regulation is to harmonize the previous fragmented data protection laws across the EU, to strengthen the rights of the individual over their own data (so as to ensure equal protection of Human Rights of the European Citizens) and, at the same time, eliminate the barriers for the services to be delivered in the European Union and enhancing business opportunities within the Digital Single Market (DSM).

The GDPR is composed of 99 *articles* that represent the mandatory part of the regulation, and 173 *Recitals* that explain the motivation of the regulation and the intended achievements. The GDPR is applied to the processing of personal data, whether it is automated (even partially) or not.

In its Art. 4, the GDPR defines *Personal Data* as "any information relating to an identified or identifiable natural person ('data subject')". This means that a *Data Subject* is a natural person (a living human being) whose data are managed by a *Controller*.

The *purpose* of the *processing* of personal data is determined by the controller, and this "processing shall be lawful only if and to the extent that at least one of the" six legal bases "applies" (Art. 6). In particular, one of those legal bases is the consent given by the data subject "to the processing of his or her personal data for one or more specific purposes" (Art. 6.1(a)). *Consent* is defined as "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her" (Art. 4.11).

However, the GDPR sets the *Conditions for Consent* in Art. 7. On one hand, "the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data"; on the other, "the data subject shall have the right to

[3] 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

[4] 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).

withdraw his or her consent at any time" and "it shall be as easy to withdraw as to give consent"(Art. 7.3). The GDPR also sets other fundamental rights of the data subject, such as the right of access (Art. 15) and the right to data portability (Art. 20).

The core part of the GDPR is Art. 5 where the following principles are defined:

Lawfulness, fairness and transparency. Personal Data shall processed lawfully, fairly and transparently.

Purpose limitation. The processed data shall be collected for specific, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.

Data minimisation. Data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.

Accuracy. Collected Data must be accurate and, where necessary, kept up to date.

Storage limitation. They must be kept for no longer than is necessary for the purposes for which the personal data are processed.

Integrity and confidentiality. The controller shall use "appropriate technical or organisational measures" to "ensure appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage".

Accountability. The controller shall be responsible for, and be able to demonstrate compliance with the other principles.

2.3 Data Security

Data security has been a major issue in information technology. The term "information (data) security" is generally based on information (data) being considered as an asset. Indeed, organizations of all types and sizes, including SMEs, recognize that such data and related processes and systems, as well as the involved human being in this processes, are important assets for achieving business objectives. By collecting, processing and sharing information they face a range of risks that can affect the functioning of their assets. To reduce these risks, they implement specific information security controls.

The most used model to guide the development and implementation of a framework for managing information security within an organisation is represented by the so called CIA triad: confidentiality, integrity and availability of information[5].

1) *Confidentiality* is defined as the "property that data is not made available or disclosed to unauthorized entities or processes", while ensuring that those authorized have access to it.

2) *Integrity* is defined as the property of "accuracy and completeness". From a practical point of view, it means that data cannot be modified in an unauthorized manner.

3) *Availability* is defined as the property of "information being accessible and usable when an authorized party demands it".

[5]ENISA: Guidelines for SMEs on the security of personal data processing. DECEMBER 2016

It is recognized that data Security functions for guaranteeing the CIA triad are:

1) **Identification**: it occurs when a user or subject claims or declares an identity, e.g., username, or anything else that can uniquely identify a subject. Security systems use this identity to determine if a subject can access an object.

2) **Authentication**: it is defined as the process of proving the claimed identity, and it occurs when a subject provides appropriate credentials to prove his/her identity, e.g., when he/she provides the correct password with a username. Different methods of authentication exist and are known as: *Something you know*, the example of password; *Something you have*, e.g., a smart; or *Something you are*, e.g., using biometrics.

3) **Authorization**: Once a user is identified and authenticated, he/she can be authorized based on the proven identity.

2.3.1 Access Control

Access Control (AC) is a fundamental building block for secure information sharing [40], because it ensures that only the intended people can access security-classified data and that these intended users are only given the level of access required to accomplish their tasks. Several access control models have been proposed, including models taking into account time, location, and situation [39, 65, 132, 246] and models specific for privacy-sensitive data [175].

There have been several access control models proposed and formalized in the literature such as Discretionary Access Control (DAC) [205], Mandatory Access Control (MAC) [206], Role-Based Access Control (RBAC) [94, 207] and Attribute-Based Access Control (ABAC) [125]. In this thesis, we refer to ABAC [125], which is currently one of the mostly adopted in industrial environment [115] and "supplements and subsumes" the other models [125].

The National Institute of Standards and Technology (NIST) defines ABAC as "[a]n access control method where subject requests to perform operations on objects are granted or denied based on assigned attributes of the subject, assigned attributes of the object, environment conditions, and a set of policies that are specified in terms of those attributes and conditions" [115].

This definition lists various key concepts. In particular, *attributes* are characteristics of the subject or object, or environment conditions, containing information given by a name-value pair. A *subject* is a human user, legal entity or an abstract entity, such as a device that issues access requests to perform operations on objects/resource. Subjects are assigned one or more attributes.

An *object* is a resource for which access is managed by the ABAC system, such as devices, files, records, tables, processes, programs, networks, or domains containing or receiving information, records of processing activities.

An *operation* is the execution of a function upon an object. Operations include read, write, edit, modify and erase. *Environment conditions* represent the operational or situational context in which access requests occur (e.g., current time or location of a user).

AC is usually implemented through *Access Control Mechanism* (ACM), which is the system that provides a decision to an authorization request, typically based on pre-defined *Access Control Policy* (ACP). This is a specific statement of what is and is not allowed on the basis of a set of rules, defined in terms of conditions on attributes of subjects, resources (i.e., objects), actions (i.e., operations), and environment, and combining algorithms for establish the precedence among the rules. For instance, a policy contains a set of rules that specify who (e.g., Controller, Processor or Data Subject) has access to which resources (e.g., Personal Data) and under which circumstances (e.g., based on the Consent and Purpose) [205]. An ACP is often specified using Natural Language Access Control Policy (NLACP), which presents the following structure: [Subject] can [Action] [Resource] if [Condition].

2.3.2 XACML

The eXtensible Access Control Markup Language (XACML) [180] is a widespread standard implementation of ABAC model. It is a platform-independent XML-based language for the specification of access control policies. The main purpose of an XACML policy is to define the constraints that a subject needs to comply with for accessing a resource and doing an action in a given environment.

The structure of an XACML access control policy is sketched in Figure 2.1. An XACML policy has a tree structure whose main elements are: *PolicySet* (not presented in the figure), *Policy*, *Rule*, *Target* and *Condition*. The *PolicySet* includes one or more policies. A *Policy* contains a *Target* and one or more rules. The *Target* specifies a set of constraints on *attributes* of a given request. Typical categories of *attributes* are *Subject*, *Resource*, *Action* and *Environment*. The *Rule* specifies a *Target* and a *Condition* containing one or more boolean functions. If the *Condition* evaluates to true, then the Rule's *Effect* (a value of *Permit* or *Deny*) is returned, otherwise a *NotApplicable* decision is formulated. If an error occurs during the evaluation of a policy against a request, *Indeterminate* value is returned. The *PolicyCombiningAlgorithm* (not represented in the figure) and the *RuleCombiningAlgorithm* define how to combine the results from multiple policies and rules respectively in order to derive a single authorization access decision.

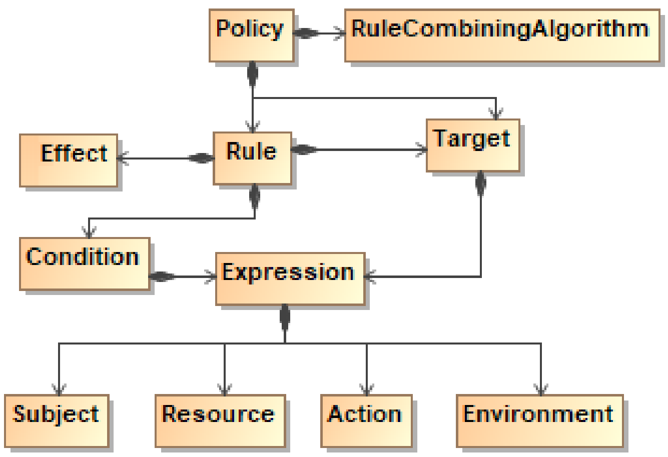


Figure 2.1: *XACML Policy Data Model.*

An *XACML Request* is composed of four main elements: 1. *Subject*, the entity requesting the access; 2. *Resource*, the requested object that is described in terms of attributes; 3. *Action*, the operation that the subject wants to perform; 4. *Environment*, the contextual information such as the request time and the location.

XACML defines also a reference architecture to allow the enforcement of XACML policies. As schematize in Figure 2.2, this is composed of different components, operating as follows:

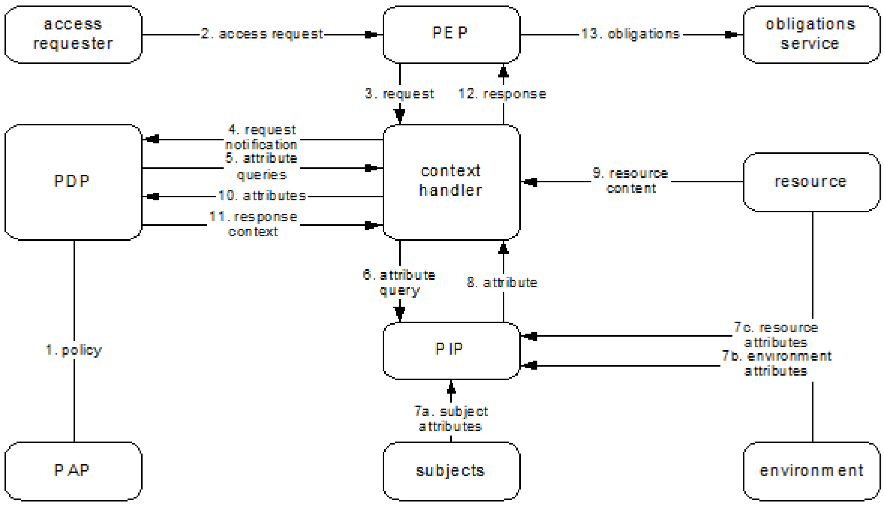


Figure 2.2: *XACML Reference Architecture (adopted from [180]).*

1. Policy Administration Point (PAP) writes policies and policy sets and make them available to the Policy Decision Point (PDP). These policies represent the complete policy for a specified target.

2. The Policy Enforcement Point (PEP) receives request for access from requester, and sends it to the context handler in its native request format, optionally including attributes of the subjects, resource, action, environment and other categories.

3. The context handler constructs an XACML request context, optionally adds attributes, and sends it to the PDP.

4. The PDP requests any additional subject, resource, action, environment and other categories (not shown) attributes from the context handler.

5. The context handler collaborates with the Policy Information Point (PIP) for obtaining the requested attributes and returns them to the PDP by optionally including the resource in the context.

6. The PDP therefore evaluates the policy and returns the response context (including the authorization decision) to the context handler. This translates the response context to the native response format of the PEP. The context handler returns the response to the PEP.

7. The PEP fulfills the obligations, and if access is permitted, then the PEP permits access to the resource; otherwise, it denies access.

Part II

GENERAL_D
Gdpr-based ENforcEment of peRsonAL Data

CHAPTER 3

GENERAL_D: Life Cycle and Architecture

THE GDPR defines the principle of Integrity and Confidentiality (Art. 5.1(f)), and implicitly calls for the adoption of authorization systems for regulating access to personal data. It also states in Art. 25 (Data protection by design and by default) that "[t]he controller shall [...] implement appropriate technical and organisational measures [...] which are designed to implement data-protection principles [...] in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation[...]." Inspired by these legal obligations, in this chapter, we present a process development Life Cycle, composed of eight phases, for the specification, development, and testing of authorization systems, as well as their deployment in a target environment. The Life Cycle targets legal aspects, such as the data usage purpose (defined by the Controller), the consent given by the data subject, and the data retention period, to cite a few. We also present a reference architecture for its (semi)-automation where available solutions for extracting, implementing and testing the data protection regulation are integrated. The objective is to propose for the first time a unique improved solution for addressing different aspects of the GDPR development and enforcement along all the Life Cycle phases.

To illustrate how the proposed Life Cycle could be effectively used in practice, we illustrate six application examples that highlight its flexibility and adaptability in different (real and) realistic contexts.

The current chapter is mainly based on and extends the following scientific contribution:

- [78] **Said Daoudagh** and Eda Marchetti: A Life Cycle for Authorization Systems Development in the GDPR Perspective. ITASEC 2020: 128-140

3.1 Introduction

The GDPR is the currently applicable EU legal framework concerning the protection of Personal Data of European citizens [90]. It is in charge of harmonizing the regulation of Data Protection across the EU member states, and at the same time, it enhances and arises business opportunities within the Digital Single Market (DSM) space. However, the natural language nature of the GDPR makes most of the provisions to be expressed in generic terms and does not provide a specific indication on how they should be actuated. Hence, applying and demonstrating the GDPR compliance, to avoid also the related penalties, becomes an important research challenge. Indeed, many businesses today are struggling in the definition of appropriate procedures and technical solutions for their development process to enforce and demonstrate the GDPR compliance [9,32,49,55,95]. In particular, following the *correct-by-design* principle, they are looking for effective and efficacious means for increasing the software's high-confidence and quality and, at the same time, reducing the cost and effort of development. Consequently, integrated solutions for designing and promptly testing their applications and systems are urgently necessary.

As for any other requirement, a fundamental step for any organization (e.g., a Small and Medium-sized Enterprise (SME)) is to guarantee the by-design compliant realization of the GDPR requirements. This means the integration of the data protection concepts into the overall software life cycle: from gathering of the requirements to deployment and subsequent maintenance of the system.

Currently, several proposals are trying to assist the 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 recognized, by scientific communities and private companies, as the successful elements for the development of privacy-by-design solutions in compliance with the GDPR [37,195,196]. However, to the best of our knowledge, most of the available proposals tend to target just single aspects of authorization system development, and no integrated solutions for the GDPR-by-design compliant development through the entire life cycle are provided. Therefore, a first objective of this chapter:

> **OBJ 1:** defining a GDPR-based Life Cycle for authorization systems.

This means to define a specific and integrated process development life cycle for the specification, deployment, and testing of adequate fine-grained authorization mechanisms able to take also into account legal requirements. Additionally, to promote the applicability of the proposed life cycle into the business and industrial context, we also present its preliminary automation. Therefore, a second objective of this chapter:

> **OBJ 2:** providing an integrated environment for automatically enforcing the data protection or privacy regulations.

Indeed, we define an integrated environment where some of the available solutions are combined for: specifying the privacy requirements, controlling personal data, processing them, and demonstrating the compliance with the GDPR in collecting, using, storing, disclosing, and/or disposing of the personal data.

3.2 Related Work

As anticipated in the introduction of this chapter, most of the proposals available in the literature do not focus on providing a unique integrated solutions for the GDPR-by-design compliant development, but usually target just single aspects of authorization system development. In this section, considering the different development phases, we overview the most relevant currently available proposals.

Considering the use of access control as main means of protecting personal data, available solutions can target either the integration of specific privacy preserving rules into the access control mechanisms, or the translation law into specifications easily enforceable in the access control domain. Specifically, the former group uses Access Control 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 [61], whereas, a formal definition of the consent is introduced in [226]. The latter group explicitly refers a given law (e.g., the EU GDPR or the US Health Insurance Portability and Accountability Act (HIPAA)) in using access control. In particular, in [63] the authors have evaluated whether the XACML standard is adequate to express the constraints imposed in HIPAA, whereas in [92], the authors investigated the feasibility of translating the articles related to access control of the previous EU data protection directive.

Considering instead the implementation of ABAC solutions in the industrial environment, a systematic methodology is proposed in [53]. Here the authors discuss an approach for implementing ABAC policies tailored to the protection of resources in an industrial setting. Although the proposal is an example of systematic implementation of policies, it does not consider any legal framework.

Differently from the above contributions, our proposal does not focus on a single aspect of the development process and it is targeting the GDPR compliance. Indeed, we provide a unified environment able to: model ACPs that are by-design compliant with the GDPR; test those ACPs by means of state-of-the-art testing tools; and to monitor their application during the production time, and eventually to suggest possible improvements in case of deviation of the expected behaviour. Therefore, the proposed solution 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 result is an Agile Authorization Development Life Cycle (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).

3.3 GDPR-based Life Cycle for Authorization Systems

As for any other software application, the development of a GDPR-based Life Cycle for authorization systems involves different stages of software development that include activities for: (1) collecting and specifying legal requirements into formal representations; (2) defining and testing data protection policies; and (3) implementing AC-based mechanisms.

In presenting our proposal, among the different development processes, we refer to and modify the authorization policy life cycle introduced in [53], which is a systematic approach to implementing ABAC-based systems within enterprises. As for [53], our life cycle does not strictly depend on any specific ABAC implementation. However, in the remainder of this thesis, we customized our proposal to the widely adopted industrial XACML-based authorization system, because it is currently the available standardized specification for ABAC.

Figure 3.1 schematizes our proposal, which consists of the following steps:

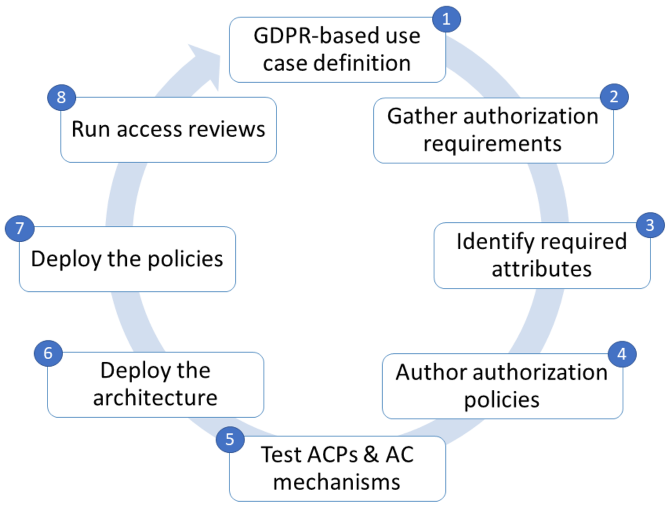


Figure 3.1: *The Authorization Policy Life Cycle (adapted from [53]).*

GDPR-based use case definition (step 1). i.e., define the context and an achievable scope to establish a common base to discuss with different stakeholders. In this case, the established use-cases need to be conceived according to GDPR implementation challenges. The stakeholders we have identified are those defined in the GDPR, i.e., Data Subject, Controller, Processor, Third-party, Data Protection Officer (DPO), and Supervisor Authority.

Gather authorization requirements (step 2). i.e., to gather all the requirements and the sources they come from. In our case, the primary source is the GDPR; therefore, authorization requirements should be defined in terms of statements or natural language authorization policies and must be rooted in the GDPR's articles. These statements must consider the stakeholders defined in the GDPR and the different Personal Data, which are the object of the GDPR protection, and their categories, the Purpose determined by the Controller, the Consent given by the Data Subject, and the record of processing activities the Controller and Processor shall maintain under their responsibility. Additionally, business requirements (e.g., working hours) and security best practices (e.g., encrypting data) need also to be defined.

Identify required attributes (step 3). i.e., to identify the attributes used in the selected requirements and their origin to make easier requirement reviews. The attributes should depend on the language or functionalities of the XACML reference architecture.

Author authorization policies (step 4). i.e., to transform the natural language statements into machine-interpretable statements, to eliminate any ambiguity introduced by natural language. Thus, a list of XACML policies encoding the GDPR's provisions needs to be defined, as well as the order in which those policies should be evaluated.

Test ACPs & AC mechanisms (step ⑤). i.e., to ensure that the implemented XACML policies meet the GDPR requirements. State-of-the-art and specifically conceived testing techniques should be used according to the different purposes. This step also involves the evaluation of the adequacy of the current AC mechanisms in the context of the GDPR.

Deploy the architecture (step ⑥). i.e., to define the contact point within existing systems in order to make the different applications able to interact with the authorization system. This step is usually business-dependent.

Deploy the policies (step ⑦). i.e., to deploy the authored XACML policies ac-cording to the selected (production) environment. This step is usually business-dependent.

Run access reviews (step ⑧). i.e., to analyze the deployed policies against a set of attributes to determine what these attributes grant. In the context of the GDPR, this should involve the simulation of realistic scenarios according to specific application use cases. Additionally, the data coming from the testing activities could be used to assess the implemented solutions and identify possible improvements.

In the specification of the different activities, we voluntarily conceived the life cycle as abstract: we defined only what each activity expects at the beginning in terms of input, and the obtained result (i.e., the expected output). This in order to provide the end-users (e.g., an SME) with the possibility to implement the different phases in the most suitable and profitable way.

For example, if an SME is only interested in the derivation of GDPR-based access control policies without testing them, the Life Cycle can be stopped at step ④. Oppositely, if an SME is only interested in the validation of its own policies, the step ⑤ is the one that should be executed skipping all the others. In Section 3.5, different examples of the application of proposed life cycle in realistic situations are presented in order remark its peculiarities, which are:

1. **Generalization** - to provide an abstract specification of the different life cycle's activities have been proposed. To this aim, the activities have been voluntary conceived as abstract, i.e., for each of them only what the activity expects at the beginning in terms of input, and the obtained result (i.e., the expected output) have been defined.

2. **Flexibility** - to provide a solution that gives the end-users (e.g., an SME) the possibility to implement the different phases in the most suitable and profitable way.

3. **Adaptability** - to provide an adaptable solution to the different end-users needs, in terms of both methodologies and technologies available.

4. **Cost Reduction** – to provide the end-users (e.g., an SME) a (Semi)-automated solution for their compliance needs and the possibility to include/reuse (the already) existing technologies.

3.4 Reference Architecture

This section presents GENERAL_D (Gdpr-based ENforcEment of peRsonAL Data), an abstract architecture that can be customized with several real tools, methodologies for assisting the development of GDPR-based Access Control (AC) systems, by following the principle of Data Protection by design and Data Protection by Default.

The general nature of the proposed GDPR-based life cycle does not constrain the environment to the specific tools, and different components implementations could be considered.

In order to propose an applicable and effective solution, the second objective of this chapter (**OBJ 2**) is to provide an integrated environment for the automatic enforcing of the GDPR-based life cycle presented in the previous section. To the best of our knowledge, this proposal is the first attempt to integrate, in a unique automated environment, different available solutions for extracting, implementing and testing the data protection regulation obligations.

Our proposal is depicted in Figure 3.2, and it 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)).

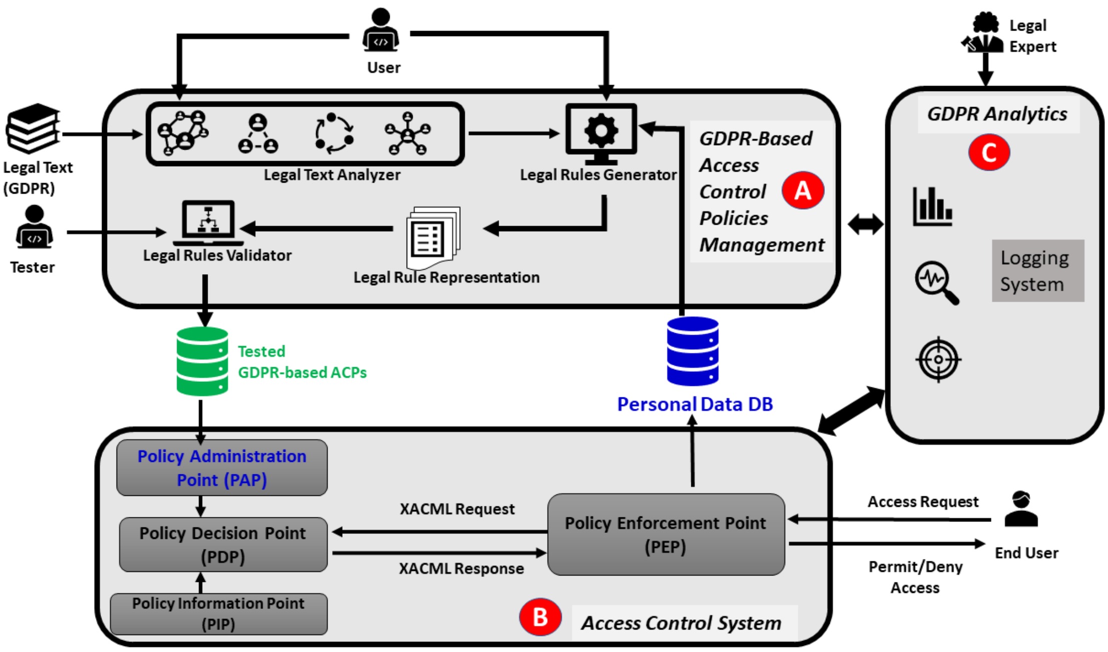


Figure 3.2: *The Proposed GDPR-based Environment.*

3.4.1 GDPR-Based Access Control Policies Management

Module (A) represents the core part of the GENERAL_D framework. It contains three main abstract components able to assist the (semi)-automation of the first five phases of the previously defined Life Cycle (i.e., from step (1) to step (5) in Figure 3.1).

In the remainder of this section, we detail how the modules have been implemented into the proposed environment and how they are related to the authorization Life Cycle.

In the proposed environment (see Figure 3.2), module (A) is in charge of supporting performing steps from (1) to (5). It is composed of three components: 1) Legal Text Analyzer; 2) Legal Rules Generator; and 3) Legal Rules Validator.

Legal Text Analyzer. Legal text analyzer helps performing mainly the first two step by providing a visual representation of the legal text, in our case the GDPR. It allows User to navigate the GDPR's article and enables gathering authorization requirements (Step (2)). It also provides functionalities for identifying the main GDPR's concepts and classify them according to the AC.

More precisely, the *Legal Text Analyzer* component takes as input a Legal Text (i.e., the GDPR text), analyses the GDPR's articles related to AC and creates an intermediate representation of the AC requirements.

Through the GUI of module (A) (see Figure 3.2), the User (in this case an authorization system developer) can select a set of predefined intermediate representation and proceed with their translation into ACPs.

Legal Rules Generator. Step (3) and Step (4) of the Life Cycle aim at transforming the intermediate representation into machine-readable statements. As a result, a list of XACML policies encoding the GDPR's obligations is defined. *Legal Rules Generator* component of module (A) is in charge of automating the two steps. Hence, the component takes as input a set of the intermediate representation of the GDPR's articles selected by the User, and provides the associated GDPR-based ACPs. Consequently, the component lets the User to identify GDPR-based attributes directly from the selected subset of the intermediate representations, so as to write meaningful ACPs in terms of GDPR's concepts.

In details, by considering Step (3) of the proposed Life Cycle (see Figure 3.1), first the component allows to classify the identified attributes into commonly-used access control entities (or categories), highlights relations between them and lets the mapping into the ABAC terms. Then, it automates the translation of the selected machine-readable representation into derived AC rules that corresponds to Step (4) of Figure 3.1. In particular, this step consists into the instantiation of the AC rules with actual attributes, and the translation of the resulting policies into a given formalism or language [1].

As in Figure 3.2, the final translation requires the interaction with the User and the *Personal Data DB*. Specifically, the User needs to identify in the *Personal Data DB* the real attributes to be considered.

Legal Rules Validator. Step (5) of Figure 3.1 aims at testing both the developed ACPs and the current AC mechanisms. Indeed, to ensure that the implemented XACML policies (or mechanisms) meet the GDPR requirements specific testing process should be adopted. Considering the environment of Figure 3.2, the *Legal Rules Validator* component of module (A) is in charge of implementing Step (5). In particular, it integrates available solutions for: assessment of test strategies, testing GDPR-based ACPs expressed in XACML and evaluating the adequacy of AC mechanisms with respect to

[1] In the current implementation, the XACML standard [180] is considered but other implementation of ABAC model can be equally adopted.

the GDPR's provisions. For the aim of completeness, we report here below the main features of the *Legal Rules Validator* component, and we refer to Chapter 5 for more details. Specifically:

1. *Test Case Generation*: starting from the developed ACPs, it is possible to generate AC requests able to test both the ACPs and AC mechanisms;

2. *Mutation Generation*: mutation analysis [187] can be applied on ACPs for measuring the adequacy of the generated test suites;

3. *Test Cases Execution & Result Analyzer*: it is an automated executor of test cases (i.e., access requests) able to collect the execution results and calculates either the effectiveness of the considered test suites, or the vulnerabilities detected;

4. *Testing Strategy Enhancement*: it suggests possible hints for enhancing the applied test suites;

5. *Oracle Derivation*: it is an automatic oracle able to associate the expected result for a given AC request based on a given ACP.

The *Tester* can interact with the *Legal Rules Validator* component for realizing specific testing purposes. For instance, for testing GDPR-based ACPs expressed in XACML, the Tester can run the following facilities: first, the *Test Case Generation* for deriving a set of AC requests (in this case a test strategy can be selected from available ones); then, through *Test Cases Execution & Result Analyzer*, the Tester can execute the test cases on the GDPR-based ACPs and collect the results; whereas, through the *Oracle Derivation* component she/he can associate the expected result to each of the executed test case; finally, the *Testing Strategy Enhancement* component can be used to visualize the results and suggestions for possible improvement of the test case generation strategies.

In the following next section, we briefly provide some hints for targeting the last three phases of the proposed authorization Life Cycle that involve the deployment of the AC architecture (Step 6 of Figure 3.1), the deployment of the developed and tested policies (Step 7), and the final analysis of the process development data (Step 8).

3.4.2 Access Control System

The idea behind Step 6 is to decouple the authorization functionalities from the business logic. This enables to adapt and extend the XACML reference architecture with new features without modifying the business logic of the applications that use and process Personal Data. This separation of concerns helps to propose scalable, manageable and extensible authorization solutions.

Once the architecture is deployed (module B of Figure 3.2), Step 7 involves the deployment of the tested GDPR-based ACPs within the *Policy Administration Point (PAP)* component of the *AC system* in order to assure the GDPR compliance. This allows the *Policy Decision Point (PDP)* to retrieve and to evaluate the right ACP when the system receives an access request, from the *End User* (e.g., Data Subject or Controller), to the Personal Data hosted into the *Personal Data DB*.

3.4.3 GDPR Analytics

Additionally, by referring to Step 8, facilities for collecting and managing information for the GDPR compliance and audit purposes [29, 55] should be included. To this purpose, module C of Figure 3.2 is the proposal that we are currently finalizing. The module extends with logging systems, monitoring capabilities, and reporting functionalities of the proposed environment [67], so that data mining and machine learning techniques can be adopted to construct behavioral models based on data coming from the logging and testing activities, and to discover and notify unwanted behaviors.

3.5 Examples of Application of the Proposed Life Cycle

In this section, without the pretend to be exhaustive, we specify five possible realistic scenarios in which our life cycle can be customized for different end-users needs.

In particular, the first two examples focus on a situation in which the end-user (e.g., an SME) wants to leverage its available access control system to manage the GDPR's articles. The third scenarios refers to the GDPR obligations enforcement. In this case, the end-user is looking for a machine-readable representation of the GDPR to enforce in its environment concepts of Personal Data, Data Subject, Controller, or third-party, and their relationships. The fourth scenario represents a situation in which the end-user wants to leverage its Business Process (BP) with activities specifically conceived for managing Personal Data according to the GDPR requirements. Finally, the last scenario refers to the adoption of the proposed life cycle for the development of new systems.

For aim of clarity, in the following sections we voluntarily do not refer to 5 (i.e., Test ACPs & AC mechanisms). Indeed, all the developed ACPs and ACM components need to be carefully validated, and we refer to Chapter 5 for more details about this step. We also demand to Chapters from 6 to 11 for more details about the implementation of each of the proposed scenario.

3.5.1 Example 1: The GDPR's Articles as Reference Use Cases

This example focuses on a situation in which the end-user (e.g., an SME) wants to leverage its access control system to manage the GDPR articles. According to this scenario the preconditions are:

- The SME needs to manage Personal Data in compliance with the regulation;
- The SME already uses an access control system to regulate access to its resources and data (assets);
- The SME lacks of internal expertise about the GDPR enforcement.

We claim that the application of the first five phases of the proposed Life Cycle can help the SME to obtain a GDPR-based access control system able to automatically manage personal data. Indeed, the process can help the SME to focus on the most suitable set of GDPR obligations and to define their relation with Access Control (AC) rules. This allows to (1) monitor the evolution of the compliance over time; (2) trace back which obligations are already covered, i.e., enabling the definition of a traceability function or feature; and (3) easily perform the review process, when necessary.

Therefore, the first five steps can be customized as follows:

GDPR-based use case definition (step 1). The first step is the definition of its typical SME use-case scenarios. The aim is to help the SME in the definition and collection of the following information: the targeted customers (e.g., Data Subjects in terms of the GDPR), the Personal Data the service requires (e.g., Name and Permanent Address attributes, as well as the associated values), and the purpose of processing the collected data (e.g., Marketing or Profiling). During this activity, the interactions between the end-users and the system and the possible actions should be also envisaged (e.g., Read, Write and Delete). As a final result, each of the identified use-cases will be focused on a specific obligation so as to making the GDPR compliance process iterative and controllable.

Gather authorization requirements (step 2). For each selected article, one or more natural language authorization requirement can be identified and represented according to the following specific form: *[Subject] can [Action] [Resource] if [Condition]*. Successively, the GDPR terms of Controller, Processor, Personal Data (and their categories), can be associated to the collected elements. Thanks to this simple structure, sufficiently detailed access control rules can be elicited from each article.

Identify required attributes (step 3). Considering each identified use case and the collected data , this step is composed of three main sub-steps:

(a) identify the GDPR's concepts involved in the authorization requirements;

(b) identify the concrete attributes defined in the reference scenario; and

(c) classify the identified attributes into the commonly used entities (or categories) of the AC specification: namely, Subject, Resource, Action, and Environment.

As a result, a precise mapping between the GDPR's concepts, the concrete entities involved in the reference use case, and the access control attributes can be identified. This enables writing concrete and precise access control policies.

Author authorization policies (step 4). An access control policy can be defined for each use case, i.e., for each of the selected GDPR articles. This can be performed through the following activities:

(a) define a set of abstract access control rules, each related to a specific access control requirement identified in step (2), by using the mapping results obtained from sub-steps 3.a and 3.c.

(b) combine the obtained rules into a single abstract access control policy by (i) defining the order in which the rules will be evaluated, (ii) adding a default rule, and (iii) selecting the rule conflict resolution algorithm. As a result, an abstract policy (i.e., ACP not directly enforceable by the AC system), expressed in terms of GDPR's concepts can be derived.

(c) replace each GDPR's concept in the abstract policy with the corresponding one (i.e., the concrete attribute gathered from the reference scenario) according to the precise mapping result in the previous step.

As a result, an enforceable access control policy expressing the GDPR's demand stated in each selected article can be defined.

3.5.2 Example 2: User Stories Extracted from the GDPR as Use Cases

The objective of this example is to illustrate the integration of the activities of the proposed life cycle into a generic development process. In this case, we consider the Agile one. For this, similarly to the previous example, we consider a situation in which the end-user (SME) wants leverage its access control system to manage the GDPR articles. According to this scenario, the preconditions are:

- The SME uses an Agile approach for developing its systems, services, and products.
- The SME uses a backlog repository containing a prioritized set of structured requirements to be developed, tested, and released.
- The SME already uses an access control system to regulate access to its resource and data (assets).
- The SME lacks of internal expertise about the GDPR enforcement.

In the Agile perspective, usually a backlog containing a structural description of a set of target functionalities is adopted. The backlog contains a list of User Stories (USs), i.e., concise yet informal descriptions telling who, what and why something is required by end-users of a system. This simple but structured sentences can express stakeholders' needs and expected functionalities from the end-users perspective. In this example, we illustrate the integration of the activities of the proposed Life Cycle into the Agile perspective by defining specific GDPR-focused User Stories, i.e., a list of Stories about GDPR's provisions told as technical requirements. Indeed, in the context of the GDPR, having a ready-to-use set of User Stories, focused on GDPR's provisions and associated with specific ACPs, represents an important means to minimize development effort and assure high quality of the final product. Consequently, when an authorization system needs to be implemented, developers could pick up the necessary predefined User Stories, and their associated ACPs, and exploit them to easily implement the required policies into the Access Control Mechanism.

From the Agile perspective, the first four steps of the proposed life cycle can be customized as follows:

GDPR-based use case definition (step 1). As in the previous example, the use-cases are defined in terms of articles of the GDPR and define one or more User Stories, organized in Epics. This modus operandi helps in defining a common vocabulary or language that the Agile DevOps team could better understand. The basic idea is to substitute the legal requirements, often misunderstood by technicians (for instance the AC policy architects), with technical requirements enabling a concrete implementation of the GDPR's obligations. This can contribute to overcome the SME's lack of legal expertise.

A typical User Story template presents the following structure: *As a [end user], I want to achieve [goal] so that [I realize the following benefit of].* This template is then used and customized for the selected articles by defining the technical requirements encoded. Therefore, each gathered User Story (or Epic) could represent our reference use case. Proceeding in this way, we gain the same benefits as in Example 1, i.e., controllability of the incremental compliance with the GDPR and enabling traceability functionalities.

Gather authorization requirements (step 2). This step is customized from an Agile perspective by gathering the authorization requirement directly from the reference User Story. The idea here is to leverage the typical structure of the User Story and the natural language structure of the authorization requirement (see Example 1, Step 2), by mapping them so as to translate the US into a specific AC requirement. This helps to gather simple yet sufficiently precise access control requirements from the User Story told using the main concepts defined in the GDPR.

Identify required attributes (step 3). This step is similar to Step 3 defined in Example 1. As a result, a precise mapping between the GDPR's concepts, involved in the current User Story, can be derived so as to enable writing a simple yet concrete and precise access control policy.

Author authorization policies (step 4). An access control policy can be derived from each User Story. To this end, the following activities should be performed:

(a) Define a set of abstract access control rules related to the access control requirement identified in Step 2, by using the mapping results obtained from sub-steps 3(a) and 3(c) (see Step 3 of Example 1).

(b) Define an abstract access control policy by (i) adding a default rule, and (ii) selecting the rule conflict resolution algorithm. As a result, an abstract policy expressed in terms of the GDPR's concepts, associated with the User Story can be defined.

(c) As in Example 1, each GDPR's concept in the abstract policy can be replaced with the corresponding one according to the precise mapping result in the previous step.

3.5.3 Example 3: External Consent Manager as Use Case

The primary objective of the third application example is to show the flexibility of the proposed Life Cycle to adapt and integrate pre-existing solutions. We consider in this case the situation in which an SME wants to integrate into its access control system an existing Consent Manager in order to make easier the management of the consent given by the data subject, and the processing of her or his personal data for one or more specific purposes. According to this scenario, the preconditions are:

- The SME wants to integrate an External Consent Manager within its system.
- The consent should be given "freely, specific, informed and unambiguously".
- The Consent Manager should allow data subjects to view, edit, download and delete their personal and consent data.
- The SME already uses an access control system to regulate access to its resources and data (assets).

- The SME lacks internal expertise about the GDPR enforcement.

Providing the data subjects means for managing personal and consent data is not only a strict adherence to the GDPR requirement: it is one of the cornerstones for building trusted business relationships. Thus, in this scenario, the SME can exploit Consent Manager systems to obey the Regulation because it represents a justification for the processing of personal data.

Among the best practices adopted in the Consent Management systems development, there is the definition of the consent according to specific formats, such as the one proposed by Kantara initiative [198]. Indeed, the initial "Consent Receipt Specification"[2] is currently being extended for satisfying the GDPR's obligations. In this draft version, named "GDPR Explicit Consent Record & Receipt Extension for Kantara CISWG: Consent Receipt", the consent specification allows controllers to clearly specify, in a human-readable format, the requirements for: linking the consent to existing privacy notices and policies; describing which "information has been or will be collected, the purposes for that collection as well as relevant information about how that information will be used or disclosed." The peculiarity of this format is the possibility to be represented in standard JSON format. An example of a concrete consent representation, in the context of the GDPR, is reported in Figure 3.3.

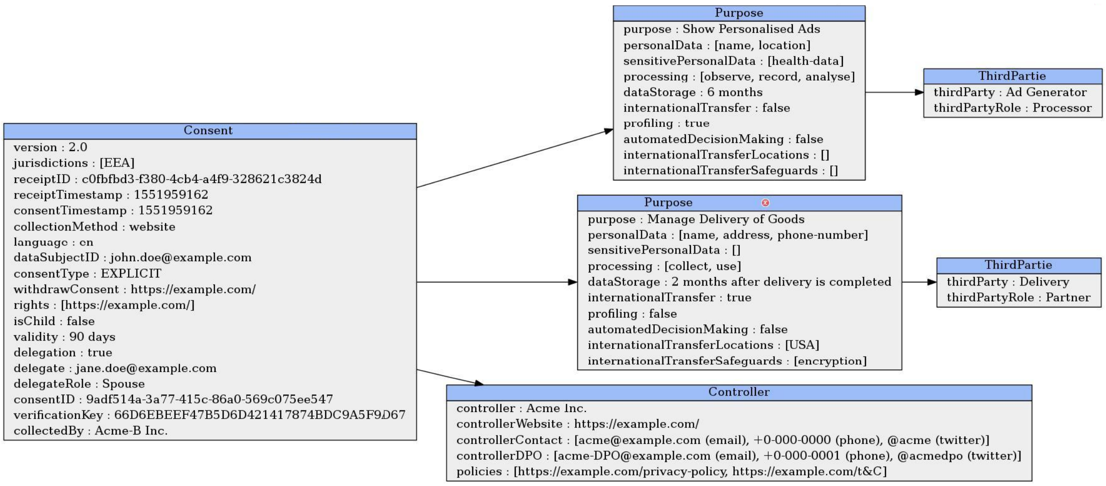


Figure 3.3: *A Kantara GDPR Explicit Consent Record Example.*

However, despite the defined consent structure, the lack of a standardized data collection process does not ensure the effectiveness and correctness of the personal data gathering and storage. Usually, the SME's service developer is in charge to embed into the business logic the consent requirements. If in the SME there is not a sufficient and consolidated expertise about the GDPR enforcement different problems can be encountered: i) the encoded legal requirements could hide privacy risks; ii) technical differences between the different services could not be taken in correct consideration; iii) an accurate validation process able to highlight privacy and security vulnerabilities could not be put in place.

We claim that, with the implementation of the first five steps of the proposed Life Cycle, SMEs can easily integrate a pre-existing consent manager, make the derived

[2] https://kantarainitiative.org/file-downloads/consent-receipt-specification-v1-1-0/

consent enforceable through an authorization system, and verify the correctness of the final integrated system. More precisely, for each consent provided by the data subject, we define its translation as a standardized access control policy. In this way, we decouple the business logic and authorization requirements gathered from that consent.

Consequently, the proposed Life Cycle is customized as follows:

GDPR-based use case definition (step 1). As motivated thereof, the consent is usually represented in a structured manner. Even there is no specific standard format defined yet, most of the solutions proposed from both academia and industry (e.g., the emerging consent specification provided by Kantara initiative), specify clearly the stakeholders, i.e., Data Subject, Controller, Processor, DPO, and Third Parties; Personal Data and their categories, the purposes for which personal data were collected, when they are collected and the period of the validity of the consent. Moreover, for each of them, it is also provided a concrete value. By having these mappings, the next steps could be easily performed.

Gather authorization requirements (step 2). This step aims at gathering all the authorization requirements from the current consent, by specifying who is the requester (i.e., Controller, Processor, or third-party), the protected resources (i.e., Personal Data), and the envisaged actions (i.e., processing) over those resources, as well as the environmental information represented for instance by the date of obtaining the consent and its validity period.

Identify required attributes (step 3). The consent file is analyzed to identify all the involved attributes in the gathered authorization requirements. This step involves only the classification of the identified attributes into the commonly used entities (or categories) in AC, namely, Subject, Resource, Action, and Environment; whereas the mapping between the concrete attributes and the GDPR's concepts are already given by the current consent.

Author authorization policies (step 4). This step aims at providing a concrete and enforceable access control policy, encoded in a given language, e.g., the XACML standard, associated with the current consent. More precisely, based on the mapping performed in the previous step:

(a) for each of the authorization requirement (specified in Step 2) an access control rule is generated;

(b) the order in which those rules are evaluated is defined; and finally,

(c) a default rule is added to the policy.

3.5.4 Example 4: Business Process as Use Case

In enterprise reality, it is quite common to use Business Process (BP) models to manage the different business activities. To model BPs, one of the most popular and widely adopted languages is Business Process Model and Notation (BPMN), which provides a visual representation supported by a formal XML specification. The extensible standard nature of the BPMN makes it possible to empower it to express activities related to data protection. Indeed, BP models visually represent the flow of business activities. They allow to smoothly manage the assignment of tasks, the interactions between the different roles, and the changes in the organization or the business activities. Moreover, BP models are also capable to precisely define for each activity the necessary data, action needed to accomplish the related task as well as the different roles involved.

Therefore, motivated by thereof, BPs can represent a meaningful scenario for investigating the adoption of the proposed Life Cycle. According to this scenario, the preconditions are:

- The SME wants to integrate privacy concepts and obligations, expressed through access control policies, into the BP execution.
- The SME already uses an access control system to regulate access to its resource and data (assets).
- The SME lacks internal expertise about the GDPR enforcement.

In this application example, however, for the sake of simplicity we assume the availability of access control policies templates each related to the GDPR's articles (e.g., the result of step 4.b of Example 1) or policies templates related to User Stories gathered from the GDPR's technical requirements (e.g., those defined in Step 4.b of Example 2).

In the context of BPs, the first four steps of the proposed Life Cycle can be customized as follows, by considering templates related to GDPR's articles[3]:

GDPR-based use case definition (step ①) and Gather authorization requirements (step ②). This steps aim at analyzing the business process activities, possibly expressed through BPMN, so as to establish a common basis to discuss with different stakeholders. The purpose is to leverage the business process to be compliant with the GDPR implementation challenges. Consequently, the BP model should be defined as accurately as possible to precisely identify the different activities, the role of each participant in the process, the data processed during the activities as well as their flow. This allows to identify only those activities that can be affected by the GDPR requirements, and for each of them the GDPR articles affecting it are detected[4]. Consequently, the activity is extended/substituted with specific sub-processes that are compliant with the spotted GDPR specifications so as to enforce the defined provisions and make easier requirement reviews. To aid in performing this phase, a pre-defined set of sub-processes will be provided according to the GDPR's demands and their related abstract access control policies templates. Additionally, all the required attributes and their values will be identified so as to fill the predefined ACPs templates, as well as their classification in the commonly used categories in AC. Moreover, depending on the different (industrial) environments, that set will also include specific activities necessary to allow the integration with access control systems.

Identify required attributes (step ③). This step consists of (a) identifying the attributes involved in the (extended) activity by the GDPR, and (b) classifying them into the commonly used categories in AC (i.e., Subject, Resource, Action, and Environment).

[3]The same steps are easily customized by considering ACPs Templates related to User Stories.

[4]Note that we are interested in those articles that spot some relation with access control, and it is out of the scope of this thesis analyzing activities affected by GDPR's provisions not related to access control.

As a result, we will have a precise mapping of the concrete entities involved in the activity and the access control attributes by allowing writing enforceable access control policies.

Author authorization policies (step 4). This step is aiming at leveraging the abstract access control templates obtained in step 4.b of Example 1[5] which are related to BP activities affected by the GDPR. As a result, we will be able to define GDPR-based ACPs related to the activities affected by the GDPR, by filling the related access control policy templates with the already identified attributes in the previous step.

3.5.5 Example 5: Internal Consent Manager and Indoor Localization Systems as Use Case

This example focuses on a situation in which the SME wants to leverage an existing Cyber-physical system to be compliant with the GDPR. For this use case we consider the specific case of an Indoor Localization System (ILS) [192]. By construction, location-based services of the ISL will support sharing and using a huge amount of information and data. Among them personal data, such as MAC or IP address, user localization with related timestamp, storage of location visited, devices used, or personal preferences, are of particular interests due to their nature of exposing privacy risk in case specific security protection is lacking. According to this scenario the preconditions are:

- The SME wants to leverage its Indoor Localization System (ILS) to be compliant with the GDPR provisions.
- The SME wants to adopt a data protection by design approach;
- The SME wants to embed its own consent manager so as to have more control for accountability purpose.
- The SME wants to include an AC systems to regulate access to personal data collected in the indoor environment to comply with the Integrity and Confidentiality principle.
- The SME lacks of internal expertise about the GDPR enforcement.

We claim that the adoption of the life cycle lets the SME to define its consent manager, (based for instance on the consent specification provided by Kantara initiative), and GDPR-based ACPs templates so as to conceive a privacy-by-design architecture. Specifically, this example includes Example 1 and Example 3 described in the previous section so as to come out with a more complex and more realistic use case scenario.

The customization of the proposed life cycle in the context of indoor localization system is therefore as follows:

GDPR-based use case definition (step 1). As motivated thereof, the consent is usually represented in a structured manner. Even there is no a specific standard format defined yet, most of the solutions proposed from both academia and industry (e.g.,

[5] We can also leverage the result of step 4.b of Example 2; i.e., considering the policies templates related to the User Stories in Agile perspective.

the emerging consent specification provided by Kantara initiative), specify clearly the stakeholders, i.e., Data Subject, Controller, Processor, DPO and third-parties; Personal Data and their categories, the purposes for which personal data were collected, when they are collected and the period of the validity of the consent. Moreover, for each of them is also provided a concrete value. By having these mappings, the next steps could be easily performed.

Gather authorization requirements (step 2). This step is the same of Step 2 of Example 3, i.e., it aims at gathering all the authorization requirements from the current consent, by specifying who is the requester, what are the protected resources, the envisaged actions and the environmental.

Identify required attributes (step 3). This step is the same of Step 3 of Example 3, i.e., the consent file is analyzed to identify all the involved attributes in the gathered authorization requirements, as well as their classification into the commonly used entities (or categories) in AC.

Author authorization policies (step 4). This step aims at providing a concrete and enforceable GDPR-based access control policy by leveraging the abstract access control templates obtained in step 4.b of Example 1 [6]. More precisely, based on the attributes contained current consent and their mapping performed in the previous step:

1. for each template derived in step 4.b of Example 1:
   (a) replace each GDPR's concept in the abstract policy with the corresponding one, and in accordance of their classification performed in the previous step.

As a result, we will obtain enforceable access control policies, each related to a selected template by using the concrete attributes defined in the current consent.

Test ACPs & AC mechanisms (step 5). Before deploying the developed ACPs and the authorization architecture, testing techniques should be used so as to validate their correctness. We refer to Chapter 5 for more details about testing techniques and available tools for performing this step.

Deploy the architecture (step 6). i.e., to define the contact point within the ILS in order to make the developed services able to interact with the reference authorization system.

Deploy the policies (step 7). i.e., to deploying the authored XACML policies into specific database within the ILS.

Remark. The GDPR represents a significant breakthrough in the digital economy and brings a lot of changes to the way in which online services are offered. This scenario calls for new approaches for developing systems where legal requirements are taken into account, just like the other requirements that a system must respond to. This proposal focused on data protection requirements and, in particular, on the development of

[6] We can also leverage the result of step 4.b of Example 2; i.e., considering the policies templates related to the User Stories in Agile perspective.

authorization systems able to enforcing the GDPR provisions. The idea is to provide, for the first time, a specific GDPR-based Life Cycle, capable to assure the by-design compliance of the developed access control systems. We also provide a reference architecture for a semi-automation of the proposed Life Cycle. To illustrate the flexibility of our proposal, we provided its customization in five different application examples.

Part III

GENERAL_D: Modelling and Testing

CHAPTER 4

GENERAL_D & RAccOnto

THIS chapter aims at providing two supporting facilities for performing the first four steps of the proposed Life Cycle introduced in Chapter 3. It introduces RAccOnto, a data protection ontology that models Access Control (AC) and leverages the PrOnto which models the GDPR main concepts. The goal of RAccOnto is to support Access Control Policies (ACPs) modeler to write ACPs that are by-design compliant with the GDPR. Furthermore, RAccOnto can be used also for ACPs testing and validation purposes. To write however XACML-based ACPs in reference to the GDPR, we need to explicitly refer to the selected GDPR's concepts within the policy. Therefore, the second facility is an *XACML GDPR Policy Profile* proposal that provides standard attributes according to the GDPR's concepts.

4.1 Introduction

The GDPR is changing how *Personal Data* should be processed. It states, in Art. 5.1(f), that "[data] should be processed in a manner that ensures appropriate security of the personal data [...] using appropriate technical or organisational measures (integrity and confidentiality)".

In this thesis, we claim that AC systems can be such a measure. AC is a mechanism used to restrict access to data or systems according to ACPs, i.e., a set of rules that specify who has access to which resources and under which circumstances [205]. By implementing them, one can gain compliance with the principle of Integrity and Confidentiality, but when enriched with policies elicited from the GDPR's provisions, we believe, AC systems can realize a compliance by-design with the GDPR's provisions expressed in the policies.

According to the GDPR, resources are Personal Data while the Controller, the Processor, or the Data Subject are those requesting access to them. But, besides this simple mapping, it may be challenging for ACPs designers to *identify*, to *extract*, to *translate* and to *encode* the GDPR's provisions into enforceable ACPs [238]. Provisions can be ambiguous and can include implicit information. They are also unstructured and therefore not straightforwardly expressible in a formal policy. This calls for a systematic process, following which one can design ACPs properly linked to the GDPR. Failing this task may have serious consequences: not only the AC system enforcing them will leave unprotected personal data but, in the specific context of the GDPR, it will also become unlawful.

The risk can be mitigated by promoting the adoption of AC systems with policies which are systematically designed for expressing GDPR's provisions.

However, designing ACPs in reference to the GDPR requires referring, within a policy, to GDPR concepts and to relationships among them. It also demands for a consistent vocabulary along the whole life cycle of the development of the ACPs. An help in this direction comes from Semantic Web technologies and in particular from the legal ontologies.

In this light, there is the urgent need to model legal access control ontology of the data protection regulation, which must not be limited to the GDPR and which can be extended to other legal frameworks, in order to define legal concepts and the relationships among them. Therefore, this chapter presents a preliminary first draft of an Access Control ontology on the GDPR, called RAccOnto, which is built on top of PrOnto. PrOnto is a state-of-the-art legal ontology aiming to provide a legal knowledge modelling of the privacy agents, data types, processing operations, rights and obligations. The goal of RAccOnto is to support Access Control Policies (ACPs) modeler to write ACPs that are by-design compliant with the GDPR.

Moreover, depending on access control language used to concretely develop policies, there is also a need to equip that language with the necessary elements so as to clearly identify the GDPR concept withing the policy. One of the most used languages is the XACML standard, which provides among others an XACML Privacy Policy Profile for referring within an ACP privacy related concepts in a standardized way.

The reminder of this chapter is organized as follows. Section 4.2 reports briefly the main concepts related to the semantic web and ontologies. Section 4.3 describes the MelOn ontology we used for developing the RAccOnto ontology. An overview of RAccOnto is reported in Section 4.4, whereas Section 4.5 describes it in details.

4.2 Semantic Web and Ontologies

Nowadays, Web and Semantic Web (SW) [38] are commonly used terms with two different purposes. Web is designed to connect from the one hand humans who make knowledge available in Web pages from the one other humans who consume that knowledge. Semantic Web is defined as a global information space populated by pages containing formal knowledge, i.e., knowledge expressed using a formal language. These information can be consumed by artificial web agents to carry out tasks that would hardly be possible to automate otherwise. Therefore, SW is not expected to replace the Web, but rather to extend it with formal knowledge: the ultimate goal of SW is to make

Internet data machine-readable.

To enable the encoding of semantics with the data, different technologies can be used: Resource Description Framework (RDF) [157], Resource Description Framework Schema (RDFS) [52] and Web Ontology Language (OWL) [169]. These technologies are used to formally represent a domain of interest. For example, ontology can describe concepts, relationships between entities, and categories of things. These embedded semantics offer significant advantages such as reasoning over data and operating with heterogeneous data sources.

In this chapter, we use SW as a means to formally model access control domain in reference to legal domain.

4.3 Methodology and Implementation

In this section, we introduce a preliminary first draft of an Access Control ontology on the GDPR, called RAccOnto, which is developed through MeLOn (Methodology for building Legal Ontology). MeLOn is an interdisciplinary approach explicitly designed for legal ontologies and the related difficulties encountered by the legal community during the definition of a model of reality through ontological techniques [183]. The specialization and implementation of the ten recursive steps of the MeLOn methodology to support the writing of ACPs that are by-design compliant with the GDPR is reported in the following.

1. **Describe the goal of the ontology.** This step includes the definition of the research questions, i.e., the ontology intends to cope with, and the selection of use-cases where the ontology is can be applied. In the RAccOnto specialization the following goals are defined:

   - model data protection legal access control from legal texts;
   - build a legal ontology that is usable for personal data enforcement through access control systems;
   - build a legal ontology that is usable for access control development, reasoning, testing, monitoring and auditing.

   From the application point of view, the above listed goals leverage RAccOnto to be:

   a) *a tool for modelling ACPs*: During the modeling phase the AC architect/developer can be guided by RAccOnto to develop ACPs that are by-design compliant with the GDPR.

   b) *a tool for validation and verification of ACPs*: Once the ACPs are developed, the AC validator can perform different type of analysis so as to validate the developed ACPs, and to assess their correctness.

   c) *a tool for auditing purpose*: RAccOnto can aid the Data Protection Officer (DPO) or a Supervisor Authority to audit the actual accesses to the Personal Data performed during the processing activities. This can be done by integrating also the access control requests delivered to the access control systems and their authorization responses.

2. **Evaluation indicators.** According to the goals (in step 1) specific indicators must be defined to evaluate the developed ontology. In RAccOnto, we have selected the following subset of the criteria defined in [183]:
    - *Coherence*: this means that the axioms defined in RAccOnto should not create inconsistencies and/or contradictions;
    - *Completeness*: RAccOnto should include the main concepts and adequately cover the targeted domain;
    - *Efficiency*: RAccOnto should be technically sound and concise; moreover, it must allow performing reasoning in reasonable time;
    - *Effectiveness*: RAccOnto must be helpful in practice, and covers the most important queries about the targeted domain;
    - *Usability*: the end-users should find RAccOnto clear, understandable, easy to use, and self-explained. The terminology used in RAccOnto should be as close as possible to the main one used inside of the legal and ICT security communities;
    - *Agreement*: this refers to grade of acceptance of RAccOnto in the legal expert and ICT security communities.
3. **State of the art survey.** This is related to finding the most suitable state of the art ontologies for the aim to reuse existing legal and access control ontologies and vocabularies. Among the available existing legal ontologies, we are relying on PrOnto [183]. The domain vocabulary we are referring to is the XACML standard.
4. **List all the relevant terminology**. We produce a glossary with the most relevant legal terms extracted directly from PrOnto ontology. In particular, we included all the legal definitions that are of interest in the domain of access control.
5. **Use usable tools.** We used tools that are close to the legal and ICT security experts, such as tables or UML diagrams, in order to model the knowledge-base of the access control domains in reference to the legal one. ICT security experts, with the aid of legal expert, can use Graffoo tool[1] that allows to use graphical instruments and to transform the UML into OWL/XML serialization.
6. **Refine and optimize.** An ontology expert collaborated to the writing of RAccOnto axioms so as to check and assess the coherence criteria. Successively, axioms have been manually added into RAccOnto.
7. **Test the output.** Legal and access control experts evaluated the criteria of completeness, effectiveness and usability so as to test the RAccOnto properties.
8. **Evaluate the ontology.** All the criteria defined in the second step of the methodology were applied to simple instances so as to provide useful metrics for evaluating the ontology. For this purpose, a set of SPARQL queries was defined and the related outputs were measured.
9. **Publish the document.** RAccOnto will be published using LODE tool[2].

[1] https://essepuntato.it/graffoo/
[2] https://essepuntato.it/lode/

10. **Collect feedback.** At this stage, feedback has been partially collected from both legal and access control communities to reach an adequate grade of the agreement criteria.

4.4 RAccOnto Overview

The aim of RAccOnto is to help access control architects gather AC requirements directly from the GDPR, and consequently enabling the definition of ACPs that are by-design compliant with the GDPR. As a result, a GDPR profile for access control that contains a consistent vocabulary can be defined. Additionally, a mean for reasoning over the developed policies to demonstrate their compliance with the regulation, in accordance with the *Accountability* principle can be obtained.

As in Figure 4.1, the general idea behind RAccOnto ontology is to connect access control and GDPR concepts. Thus, RAccOnto components are: the ACPs model, defined by XACML Standard [179], and a subset of PrOnto [183] ontology elements.

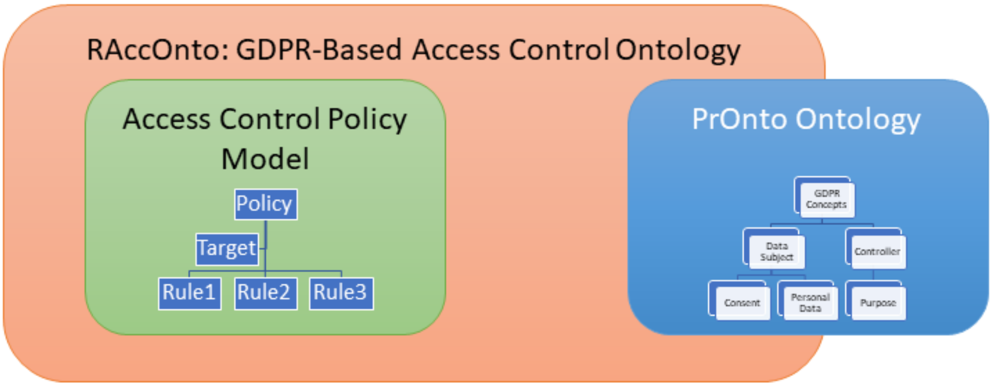


Figure 4.1: *Overview of RAccOnto.*

As in the figure, the starting point of the RAccOnto is the PrOnto [183, 184] ontology. It represents a first step towards IT solutions for indexing information in the EU data protection domain, thus facilitating their navigation and search. However, PrOnto ontology does not provide the possibility to express fine-grained access control rules, representing obligations and permissions from the GDPR perspective, useful for to systematically developing GDPR-based ACPs. Therefore, the aim of this chapter is to build a knowledge base of such policies on top of the PrOnto ontology.

4.5 RAccOnto Modules

RAccOnto reuses extensively the PrOnto ontology, which is composed of five modules, each containing specific legal concepts and relations between them. Figure 4.2 reports an overview of these modules that are: (1) documents and data (Data in Figure 4.2); (2) actors and roles (Agents in Figure 4.2); (3) processing and workflow (Processing in Figure 4.2); (4) purposes and legal bases (Purposes in Figure 4.2); and (5) legal rules and deontic formula (Rights in Figure 4.2).

As in Figure 4.2 an additional module has been added to PrOnto ones: the Access Control which models the ACPs concepts based on the XACML standard.

The idea is to start from the main concepts modeled in PrOnto, extract those that are of interest and enhance them by adding access control concepts so as to provide a comprehensive knowledge base to model access control policies in the context of the GDPR.

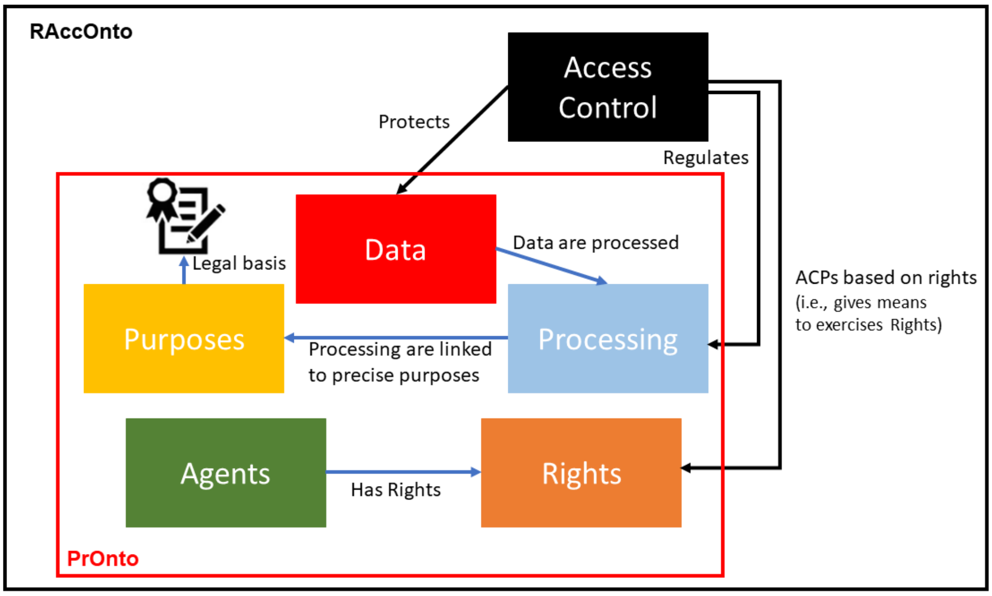


Figure 4.2: *Modules of PrOnto ontology enhanced with Access Control. Adopted and enhanced from [184].*

In the following subsections more details about each of the modules of Figure 4.2 are provided.

4.5.1 Access Control Module

RAccOnto ontology will be used mainly within the access control community, enabling the access control architects and developers to model and implement ACPs in compliance with the GDPR. Therefore, to build a knowledge base about access control policy specification and legal requirements, we have looked at the currently available access control models as DAC, MAC, RBAC, and ABAC. In this work, we are referring to the latter one, i.e., ABAC model. More precisely, we have selected its standard implementation called XACML [179] as main source for modelling Access Control module. Figure 4.3 depicts the access control ontology (i.e., the Access Control module reported in Figure 4.2) we have defined. More precisely, the access control concepts we are interested in, and we have identified are:

1. Policy: is the top element of an XACML policy, and it is composed of a set of Rules, containing actual constraints ruling access to a protected Resource, and a Target, which defines its applicability to a given request. The Policy has also a combining algorithm.

2. Target: in XACML, this element defines a set of simple constraints in terms of

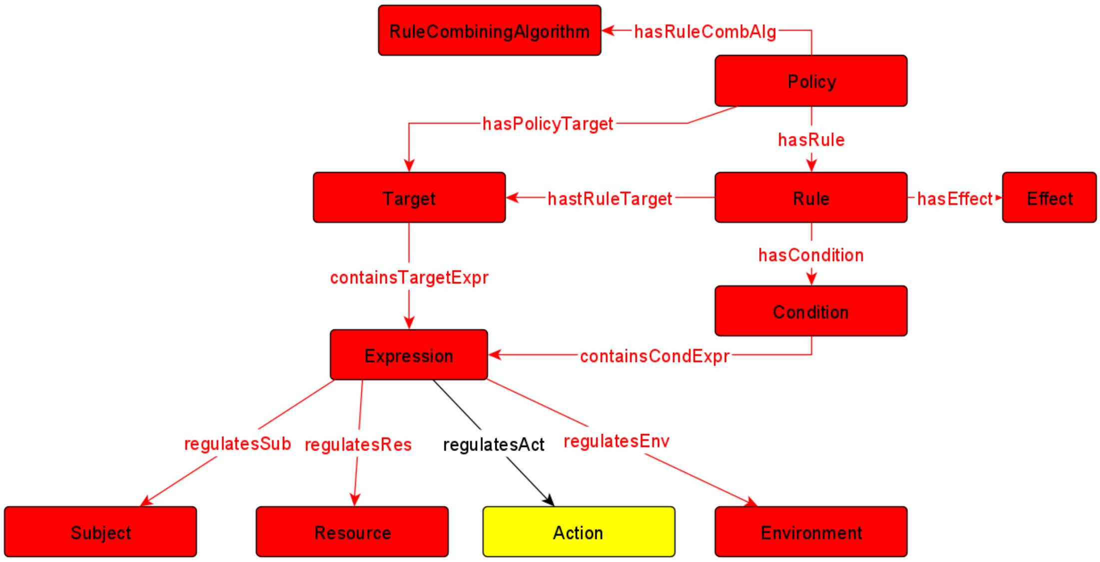


Figure 4.3: *RAccOnto: Access Control Module.*

boolean expression over Subject, Resource, Action and Environment elements, and establishes the applicability of a given request to the policy and the rule the target is referring to.

3. RuleCombiningAlgorithm: defines the strategy adopted for solving inconsistencies during the evaluation of the rules contained in the policy. XACML defines the following algorithms: first-applicable, deny-overrides, permit-overrides and their ordered versions.

4. Rule: in XACML, a rule is the basic element evaluable within a policy. A Rule is composed of a Target (as the Policy) and a Condition. It has an Effect, which is returned as result when the Rule is evaluated.

5. Effect: represents the authorization response when the associated Rule is evaluated true. Effect can assume the following values: Deny, Permit, Not Applicable and Indeterminate.

6. Condition: this element is used to apply complex constraints in terms of boolean expression over the Subject, Resource, Action and Environment elements.

7. Resource: defines the protected resource by the Policy.

8. Subject: it is the entity who requests access to a resource to perform a given action over it.

9. Environment: often this element contains contextual information that must be held at the time of the evaluation of the policy for a given request.

10. Expression: this class encodes the authorization specification as set of constraints over the values of subjects, resources, actions and environment elements.

Another core element of access control is the Action, that the requester wants to perform over the protected resource. We do not include Action concept in the Access

Control module, rather we reuse the definition provided by PrOnto ontology within the Processing module.

As in Figure 4.3, the above defined concepts are involved in different properties, all used in the access control domain for specifying meaningful access control policies.

4.5.2 Documents and data

The ontology in this module models (personal) data that is the object of the GDPR and it is target of its protection. According to the GDPR, in PrOnto, data are defined in categories: personal data, non-personal data, anonymized data, pseudonymised data. In this module we are interested in the data sub-module which is connected to the Access Control through the Resource. Being Personal Data the object of the GDPR, in access control domain it represents the valuable resource the access control mechanism is regulating the access to.

4.5.3 Agents

PrOnto distinguishes agents and roles. Agents are defined as Physical persons, organizations, IT organizations or artificial intelligence and software or robots. Whereas roles are related to contexts and processing activities that an agent is taking part in. For instance, a third-party could act as a controller or processor with respect to different processing activities. PrOnto models role in subclasses, e.g., data subject controller, processor, supervisory authority and the new introduced figure the DPO. This module is connected to the Access Control module through the *isRepresentedBy* property that connects access control Subject concept and Agent concept defined in this module.

4.5.4 Processing and workflow

PrOnto models workflow as a sequence of steps, and it is composed of two parts: the plan to do something (e.g., workflow) and the concrete sequence of actions actually performed (e.g., execution of the workflow). Actions are modeled in PrOnto in subclasses such as delete, transmit and store. As reported previously, we connect this module with Access Control one throughout the *regulatesAct* property that connects Expression (in access control module) and Action (in PrOnto) concepts.

4.5.5 Purposes and legal bases

The GDPR allows processing of personal data according to several lawful purposes, each involving a set of personal data. In this module, we are interested in the purpose and in only one legal basis, i.e., the Consent. We connected indirectly the Access Control module to the Purpose throughout the *isRepresentedBy* property. More precisely, a Subject is represented by an Agent that *determines* a specific *Purpose*.

4.5.6 Legal rules and deontic formula

This includes deontic operators such as right, obligation, permission and prohibition, allowing to model the necessary predicates to implement legal rules. In this module we are interested only in the concept *Right* and its subclasses. *Right* concept therefore is connected to the Access Control module through the *isRelatedToRight* property that involves the *Policy* concept depicted in Figure 4.3.

4.6 GDPR-based XACML Profile

The X in XACML stands for eXtensible and one way to extend XACML-based authorization is to use XACML profiles. A profile can extend the functionality of a policy server in a number of ways. This can be as simple as the addition of a classification or terminology from an existing standardized domain, it can also include more advanced features, such as new data types or user defined functions.

By referring to the privacy, XACML already defined a specific profile named *XACML Privacy Profile*. This profile "provides standard attributes and a standard <Rule> element for enforcing the" Purpose specification and Use limitation principles[3], "related to the purpose for which personally identifiable information (PII) is collected and used".

The XACML Privacy Profile defines the following two standard attributes:

"urn:oasis:names:tc:xacml:2.0:resource:purpose" This attribute, of type “string”[4], indicates the purpose for which the data resource was collected. The owner of the resource SHOULD be informed and consent to the use of the resource for this purpose. The attribute value MAY be a regular expression. The custodian’s privacy policy SHOULD define the semantics of all available values.

"urn:oasis:names:tc:xacml:2.0:action:purpose" This attribute, of type “string”, indicates the purpose for which access to the data resource is requested. Action purposes MAY be organized hierarchically, in which case the value MUST represent a node in the hierarchy

For the matching purpose, XACML Privacy Profile provides an example of rule exemplifying the usage of the above

```
<Rule xmlns="urn:oasis:names:tc:xacml:3.0:core:schema:wd-17"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance";
xsi:schemaLocation="urn:oasis:names:tc:xacml:3.0:core:schema:wd-17
xacml-core-v3-schema-wd-17.xsd"
RuleId="urn:oasis:names:tc:xacml:2.0:matching-purpose"
Effect="Deny">
    <Condition>
        <Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:not">
            <Apply FunctionId="urn:oasis:names:tc:xacml:1.0:function:any-of-any">
                <Function
                    FunctionId="urn:oasis:names:tc:xacml:1.0:function:string-regexp-match"/>
                <AttributeDesignator MustBePresent="false"
                    Category="urn:oasis:names:tc:xacml:3.0:attribute-category:resource"
                    AttributeId="urn:oasis:names:tc:xacml:2.0:resource:purpose"
                    DataType="http://www.w3.org/2001/XMLSchema#string"/>
                <AttributeDesignator MustBePresent="false"
                    Category="urn:oasis:names:tc:xacml:3.0:attribute-category:action"
                    AttributeId="urn:oasis:names:tc:xacml:2.0:action:purpose"
                    DataType="http://www.w3.org/2001/XMLSchema#string"/>
            </Apply>
        </Apply>
    </Condition>
</Rule>
```

Listing 4.1: *Example rules.*

[3]These principles are described by OECD in: Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, OECD, 1980

[4]For the aim of clarity all the data types are reported in short version. The actual long version should contain the following prefix:"http://www.w3.org/2001/XMLSchema#". For instance type "string" should be "http://www.w3.org/2001/XMLSchema#string".

Inspired by this profile, in the following we provide possible standard XACML attributes that should be used when writing access control policies in reference with the GDPR, i.e., when the GDPR's concepts are involved.

"urn:oasis:names:tc:xacml:3.0:profile:subject:gdpr:datasubject" This attribute, of type "string", indicates the data subject, i.e., identified or identifiable natural person. An identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity. The attribute value MAY be a regular expression.

"urn:oasis:names:tc:xacml:3.0:profile:subject:gdpr:controller" This attribute, of type "string", indicates the controller, i.e., the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data. The attribute value MAY be a regular expression.

"urn:oasis:names:tc:xacml:3.0:profile:subject:gdpr:processor" This attribute, of type "string", indicates the processor, i.e., a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller. The attribute value MAY be a regular expression.

"urn:oasis:names:tc:xacml:3.0:profile:subject:gdpr:dpo" This attribute, of type "string", indicates the data protection officer. The attribute value MAY be a regular expression.

"urn:oasis:names:tc:xacml:3.0:profile:subject:gdpr:supervisorauthority" This attribute, of type "string", indicates the supervisor authority. The attribute value MAY be a regular expression.

"urn:oasis:names:tc:xacml:3.0:profile:resource:gdpr:personaldata" This attribute, of type "string", indicates personal data i.e., "any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person".

"urn:oasis:names:tc:xacml:3.0:profile:resource:purpose:consent" This attribute, of type "boolean", indicates the consent of the data subject meaning any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her. The consent is related to a specific purpose.

Remark. In this chapter, we introduced two supporting felicities to aid performing the first four steps of the proposed life cycle. In particular, we have introduced RAccOnto ontology which leveraged PrOnto in the context of access control. We started from the main concepts modeled in PrOnto, by extracting those that are of interest and enhancing them by adding access control concepts. Then, we linked them so as to provide a comprehensive knowledge base to model access control policies in the context of the GDPR. As a result, we obtained a GDPR profile for access control. The second facility refers to the definition of an XACML GDPR Policy Profile, that provides a set of standard attributes to be used within the policy to identify the GDPR's concepts.

As future work, we are planning to thoroughly test both RAccOnto and the XACML GDPR Policy Profile by legal experts and access control experts so as to have sufficiently feedback from the two communities. As part of future work, we will also investigate the possibility of advancing the conceived XACML GDPR Policy Profile as a reference standard for the GDPR.

CHAPTER 5

GENERAL_D & Testing

In the recent years, one most promising approach for replying to the strict exigencies and rules imposed by the GDPR about the management of *Personal Data* is to adopt customized Access Control (AC) mechanisms and policies. Due to the critical and crucial role of this kind of systems and the deployed Access Control Policies (ACPs), effective and efficient validation methods, taking in consideration the peculiarity of the reference legal framework (i.e., the GDPR), should be applied before their deployment into a given production environment. Indeed, in the context of access control systems, testing activity is among the most adopted means to validate that sensitive information or resources are correctly accessed. Therefore, in this chapter, we propose comprehensive testing process for validating both ACPs and ACMs, and a reference architecture for its automation. The peculiarity of our solution is that it compensates the lack of legal knowledge of testers, by providing them automatic methodologies focusing by-design on the legal aspects. Indeed, the proposed testing framework lets the end user (i.e., the AC tester or validator) to focus only on performing AC testing activities without the need to know the GDPR's provisions. Moreover, we illustrate how to conduct Controlled Experiments (CEs) in the context of AC for thoroughly assessing different kinds of artifacts by considering well-defined metrics.

By referring to GENERAL_D (i.e., the Life Cycle and its supporting architecture) presented in Chapter 3, this chapter refers to Step 5 (i.e., *Test ACPs & AC mechanisms* step in Figure 3.1), and it is related to the component *Legal Rules Validators* of module A depicted in Figure 3.2.

The results presented in this chapter are partially based on the following related publications:

[79] Said Daoudagh and Eda Marchetti: GRADUATION: A GDPR-based Mutation

Methodology. QUATIC 2021

[69] Said Daoudagh, Francesca Lonetti and Eda Marchetti: An automated framework for continuous development and testing of access control systems. J Softw Evol Proc. 2020;e2306. https://doi.org/10.1002/smr.2306

[76] Said Daoudagh, Francesca Lonetti, Eda Marchetti: XACMET: XACML Testing & Modeling. Softw. Qual. J. 28(1): 249-282 (2020)

[74] Said Daoudagh, Francesca Lonetti and Eda Marchetti: Continuous Development and Testing of Access and Usage Control: A Systematic Literature Review. ESSE 2020

[73] Said Daoudagh, Francesca Lonetti, Eda Marchetti: Assessing Testing Strategies for Access Control Systems: A Controlled Experiment. ICISSP 2020: 107-118

[77] Said Daoudagh, Eda Marchetti: Defining Controlled Experiments Inside the Access Control Environment. MODELSWARD 2020: 167-176

[72] Said Daoudagh, Francesca Lonetti, Eda Marchetti: A Framework for the Validation of Access Control Systems. ETAA@ESORICS 2019: 35-51

[75] Said Daoudagh, Francesca Lonetti, Eda Marchetti: A Decentralized Solution for Combinatorial Testing of Access Control Engine. ICISSP 2019: 126-135

[71] Said Daoudagh, Francesca Lonetti, Eda Marchetti: A General Framework for Decentralized Combinatorial Testing of Access Control Engine: Examples of Application. ICISSP (Revised Selected Papers) 2019: 207-229

[45] Antonia Bertolino, Said Daoudagh, Francesca Lonetti, Eda Marchetti: An automated model-based test oracle for access control systems. AST@ICSE 2018: 2-8

[31] Cesare Bartolini, Said Daoudagh, Gabriele Lenzini and Eda Marchetti: Testing of GDPR-based Access Control Policies. SessionPoster@ESORICS, 2019

5.1 Introduction

Nowadays, quality of ICT systems and modern applications is strictly tied with the security and privacy. Among security mechanisms, a critical role is played by Access Control Systems (ACSs), which aim to ensure that only the intended subjects (e.g., Data Subject, Controller and Processor) can access the protected data (e.g., Personal Data or special Categories of Personal Data) and get the permission levels required to accomplish their tasks and no much more. Due to the complexity of ACSs, for ensuring the required security level, a key factor becomes the application of effective and efficient testing approaches: knowing in advance the criticality of the systems lets to put in practice efficacious and corrective actions so as to improve the overall security of the system. However, testing phase is a time consuming, error prone and it represents a critical step of the development process, which involves different activities: from test strategy selection to the test case derivation, from execution to the final test results evaluation. Bad choices in each stage of the testing phase may compromise the entire process, with the risk of releasing inadequate security solutions that allow unauthorized access from the *security perspective* or unlawful processing from the *legal perspective*.

Therefore, thorough testing of ACSs represents a key activity to guarantee the trustworthiness of sensitive data and protect information technology systems against inappropriate or undesired user access. Several strategies for the generation of test cases (i.e., access requests) for access control systems have been defined in scientific literature. They leverage the application of combinatorial approaches to XACML policies values for generating test inputs [41,47,160]; or exploit change-impact analysis for test cases generation starting from policies specification [161]; or are based on the representation of policy-implied behavior by means of models [66,170].

The need for effective and efficient evaluation methods of test cases generation strategies is growing in order to gain confidence that a system meets its security requirements. Indeed, an effective test generation strategy for the access control systems allows: i) on one side to exercise all the security-critical aspects and discover all the possible faults of the access control systems; ii) on the other side to develop a successful and cost-effective testing phase. Thus, the challenge of how to select the most promising approach. Indeed, the assessment of the most effective test cases generation strategies for XACML based access control systems usually relies on several techniques, e.g., coverage [163] or mutation analysis [42]. It also exploits specific metrics and evidences gathered or from formal assurance techniques or from experimental evaluations. In particular, if not properly formalized, the testing activity can have the following consequences: impossibility of replicating and controlling the process especially in case of regression testing [247], difficulties in the generalization of the testing results and consequent derivation of statistical significance values; and problems in defining and sharing a common testing knowledge so as to avoid recurring failures and speeding up the corrective process.

A reply to these issues comes from the software engineering context, where Controlled Experiments (CEs) [35,128,234] are commonly used to investigate the cause-effect relationships of introducing new methods, techniques or tools and to build a body of knowledge supported by observation and empirical evidence. Therefore, CEs let to validate the different activities of the testing process by means of the identification of important variables, the definition of specific testing models and objectives, and the derivation of empirical evidence. In the CE different treatments can be applied to, or by, different subjects, while other variables are kept constant and the effects on response variables are measured.

Authors in [128,234] categorize experiments as either technology-oriented or human-oriented, depending on whether artifacts or human subjects have given various treatments. In this work, we revise and customize the technology-oriented experiments in order to provide general guidelines for correctly end effectively performing the testing of the AC systems. Therefore, we provide the characterization of the first three (over the five) steps of the *Experiment Process*, namely *Scoping*, *Planning*, and *Operation*.

The main contributions of this chapter can be summarized in the following six main objectives:

Contr. ① Defining the steps of a typical Testing Process, and a reference architecture for its automation. The architecture is composed of different components that can be customized with real artifacts for: the test cases generation (i.e., access control requests); the mutation generation for both access control policies and access control mechanism (in our case the Policy Decision Point (PDP)); the execution of the test suites and their analysis.

Contr. ② Advancing and defining the notion of families of CEs in the context of AC. This allows conceiving well-defined testing goals, a formally defining Controlled Experiments in the context of AC, e.g., XACML domain.

Contr. ③ Introduce the GROOT strategy. This is a general combinatorial strategy for testing systems managing GDPR's concepts (e.g., Data Subject, Personal Data or Controller). In this chapter GROOT is customized for generating XACML requests useful for testing GDPR-based ACPs.

Contr. ④ Defining the XACMET testing framework which includes:

- *A new test case generation strategy based on path coverage*: the strategy is able to reduce, in most of the cases, the size of the test suite (in terms of number of requests), guaranteeing the same effectiveness of the mostly adopted approaches. Briefly, XACMET testing strategy builds from the XACML specification a typed graph, called the XAC-Graph, representing the XACML policy evaluation and use it for test case generation. Even though the test strategy cannot *a priori* guarantee a low number of test requests, it is able to exercise different policy elements guaranteeing test cases diversity.
- *An automatic derivation of an XACML oracle*: the XAC-Graph used for test cases generation can be used also for oracle definition. Indeed the graph includes the expected output corresponding to each test request and therefore can be exploited for the automatic definition of test verdicts.
- *Capabilities for measuring the coverage of test requests*: the test execution coverage can be evaluated in terms of the different paths on the derived XAC-Graph. The coverage measure can also be used for reducing the set of test requests so that all the paths are executed at least once.

Contr. 5 Introduciung GRADUATION, a generic methodology for assessing the fault detection effectiveness of GDPR-based testing strategies by means of mutation testing. In particular, GRADUATION provides a set of mutation operators specifically based on a GDPR-based fault model. A preliminary implementation of GRADUATION methodology in the XACML context is also presented.

Contr. 6 Conducting a CE in the context of AC, detailing in particular its main three steps: definition of the experiment for the comparison of two testing strategies; instrumentation and execution of the experiment; and, analysis of the results.

The peculiarity of our solution is that it compensates the lack of legal knowledge of the testers, by providing them automatic methodologies focusing by-design on the legal aspects. Indeed, the proposed testing framework lets the end user (i.e., the AC tester or validator) to focus only on performing AC testing activities without the need to know the GDPR's provisions.

The reminder of this chapter is organized as follows. Section 5.2 illustrates the necessary background about testing and discusses related work in the context of AC. We define a standard testing process in Section 5.3.1 and we illustrate a reference architecture for its automation in Section 5.3.2; whereas we advance and define the notion family of CEs in the context of AC in Section 5.4. We illustrate GROOT, a test cases generation strategy based on combinatorial approach in Section 5.5, and XACMET a testing framework offering test cases generation, oracle derivation and coverage measurements facilities in Section 5.6. Afterward, Section 5.7 introduces GRADUATION, a GDPR-based mutation generator strategy. Finally, we conclude this chapter by discussing a CE in the context of AC in Section 5.8.

5.2 Background and Related Work

In this section, we overview some of the mostly adopted testing strategies, we provide details about the mutation testing techniques and oracle problem and we briefly introduce the Controlled Experiment main concepts. Related works conclude this section by providing details about the testing strategies and activities targeting the specific XACML environment.

5.2.1 Testing Approaches

Testing is an important and critical part of software development, consuming even more than half of the effort required for producing deliverable software. Indeed, test cases execution represents the biggest part of software cost that can be evaluated in terms of: the cost of designing a suitable set of test cases which can reveal the presence of bugs; the cost of running those tests, which also requires a considerable amount of time; the cost of detecting them, i.e., the development of a proper "oracle" which can identify the manifestation of bugs as soon as possible; the cost of correcting them. Testing includes different methods and approaches. For aims of completeness in this section we briefly described only those close to the topic of this thesis. We refer to [149] for a survey of the recent testing proposals.

Among the proposal for the black-box testing, which relies only on the input/output behaviour of the system, the testing techniques considered in this chapter include [149]:

- Testing from formal specifications: in this case it is required that specifications be stated in a formal language, with a precise syntax and semantics, for instance the XACML. The tests are hence derived automatically from the specification, which are also used for deriving inductive proofs for checking the correct outcome.
- Equivalence partitioning: in this case the input domain is partitioned into equivalence classes so that elements in the same class behave similarly. In this context, the Category Partition is a well-known and quite intuitive method, which provides a systematic, formalized approach to partition testing.
- Boundary-values analysis: this is a complementary approach to equivalence partitioning, and concentrates on the errors occurring at boundaries of the input domain. The test cases are thus chosen near the extremes of the class.
- Random methods: they consist of generating random test cases based on a uniform distribution over the input domain. It is a low-cost technique because large sets of test patterns can be generated cheaply without requiring any preliminary analysis of software.
- Combinatorial testing: in combinatorial testing, test cases are designed to execute combinations of input parameters. Because providing all combinations is usually not feasible in practice, due to their extremely large numbers, combinatorial approaches able to generate smaller test suites for which all combinations of the features are guaranteed, are preferred. Among them, common approach is all-pair testing technique, focusing on all possible discrete combinations of each pair of input parameters.

One of the points against the black-box testing is its dependence on the specification's correctness and the necessity of using a large amount of inputs in order to get good confidence of acceptable behaviour. Thus, alternative to black-box is the white-box testing, which requires complete access to the object's structure and internal data, i.e., the visibility of the source code. In white-box testing, the tests are derived from the program's structure, which is also used to track which parts of the code have been executed during testing. Some testing techniques considered in this chapter include [149]:

- Control flow-based criteria: these techniques use the control flow graph representation of a program, in which nodes correspond to sequentially executed statements while edges represent the flow of control between statements. The aim of white box testing criteria is to cover as much as possible the control flow graph, limiting the number of selected test cases. In particular, they differentiate in: *statement coverage* which is based on executable statements, *branch coverage* which focuses on the blocks and case statements that affect the control flow, *condition coverage* which relies on subexpressions independently of each other, *path coverage* which is based on the possible paths exercised through the code.
- Data-Flow coverage: in data-flow testing, a data definition of a variable is a location where a value is stored in memory (definition) and a data use is a location where the value of the variable is accessed for computations use (c-use) or for predicate use (p-use). The data-flow testing goal is to generate tests that execute program subpaths from definition to use. Traditional data-flow analysis techniques work on control flow graphs annotated with specific information on data usage.

5.2.2 Mutation Testing

Mutation testing is a technique in which syntactic faults, simulating typical programmer's mistakes, are seeded in the original program in order to produce a set of faulty programs, called mutants, each one containing one fault. The main purpose of mutation testing is to assess the adequacy of a given test suite. Each test case is executed on the original program (also called Gold program) and its mutants; then, outputs are collected: if the mutant's output is different from the original program's one, the fault is detected and the mutant is said to be killed.

The mutation score is the ratio of the number of detected faults over the total number of seeded faults and indicates the effectiveness of the test suite. Since mutation testing was proposed in the '70s, it has been applied to many programming languages, such as Java, Fortran, Ada, C, SQL, and many mutation tools have been developed to support automated mutation analysis. We refer to [187] for an extensive survey of software mutation testing. The general process of mutation analysis consists of two steps: first, change the original program with predefined mutation operators and generate a set of mutated program, called mutants; then, the mutants are executed against a test suite, and information is collected during the execution for various purpose of analysis.

In the context of access control systems, some proposals address mutation techniques to assess the fault detection effectiveness of test sets for security policies and provide specific mutation operators [42]. The defined mutation operators manipulate the target and condition elements of the XACML policy, in order to generate a set of faulty policies. The policy under test and the faulty policies are evaluated by the PDP against the same access requests, then the test outputs, represented by the access responses, of the original and the mutated policies are compared to get the mutation score.

5.2.3 Oracle Problem

An important component of testing is the oracle. Indeed, a test is meaningful only if it is possible to decide about its outcome. Ideally, an oracle is any (human or mechanical) agent that decides whether the program behaved correctly on a given test. The oracle is specified to output a reject verdict if it observes a failure (or even an error, for smarter oracles), and approve otherwise. Not always the oracle can reach a decision: in these cases the test output is classified as inconclusive [149].

In a scenario in which a limited number of test cases is executed, the oracle can be the tester himself/herself, who can either inspect a posterior the test log, or even decide a priori, during test planning, the conditions that make a test successful and code these conditions into the employed test driver.

When the tests cases are automatically derived, or also when their number is quite high, in the order of thousands, or millions, a manual log inspection or codification of the test results is not thinkable. Automated oracles must then be implemented.

In literature different proposals are available for oracle specification. In this chapter we briefly recall some of them referring to [149] for a recent overview. Specifically we consider solutions based on:

- Model-based specification languages: the purpose is to exploit the models and a syntax that define desired behavior in terms of its effects on the model so to derive the expected oracle. However, the models or the documents describing the specification could be very abstract, quite far from concrete execution output and consequently, oracle definition could result quite problematic.

- State transition systems: this kind of approaches focuses on the formal modeling of the system through state transition systems. In particular, they focus on the reaction of a system to stimuli, i.e., *transitions*, and abstract a property of the states.

- Assertions and contracts: if assertions could be embedded into the program so to provide run-time checking capability, conditions are instead expressly specified to be used as test oracles. As consequence, the produced execution traces could be logged and analyzed so to derive the oracle verdicts.

In view of these considerations, it should be evident that the *oracle* might not always judge correctly. So the notion of coverage of an oracle is introduced to measure its accuracy. It could be measured for instance by the probability that the oracle rejects a test (on an input chosen at random from a given probability distribution of inputs), given that it should reject it [21], whereby a perfect oracle exhibits a 100% coverage, while a less than perfect oracle may yield different measures of accuracy.

5.2.4 Controlled Experiments

Experiments (or Controlled Experiments (CEs)) are used in software engineering to investigate the cause-effect relationships. They consist of a well-defined *Experiment Process* including five specific phases: *(i)* Scoping, *(ii)* Planning, *(iii)* Operation, *(iv)* Analysis and Interpretation, and *(v)* Presentation and Package. However, in the Experiment Process it is not mandatory to finish an activity before starting the next one. As a consequence, it is possible to go back and refine a previous activity before continuing with next one. In this sense it is partially iterative.

The purpose of a CE is therefore to systematically define the elements necessary for ensure the integrity and replicability of the obtained results. Very briefly, the main element are:

(1) *objects* on which the experiment is run are the *experimental units* and can involve all the system or part of it;

(2) *subjects* that represent artifacts on which the methods or techniques are applied;

(3) the outcome of an experiment is referred to a quantitative *response variable* (also called *Dependent Variable*);

(4) each considered characteristic target of the experiment to be studied, that can affect the response variable, is called a *factor* (also called *Independent Variables*);

(5) the possible values of the factors are called *levels*; and

(6) *parameter*, i.e., any other invariable (qualitative or quantitative) characteristic of the software project that does not influence the result of the experiment.

Consequently, in each experiment a combination of alternatives of factors are applied by a subject on a unit. A defined and precise specification of the experiment guarantees both: the *External replication* [126], i.e., reproducing the experiment in different contexts and environments so as to increase the confidence in experiment results; the *Internal replication*, i.e., the repetition of the experiment more time in the same environment or condition to increase the reliability of the experiment results.

In Software Engineering field, CEs are gaining a lot of attention [135, 217] and different proposals are trying to give guidance on how to conduct CEs [128, 234]. Following this tendency, our proposal want to come up with a Goal Definition Framework that enables one to conduct technology-oriented experiments in the Access Control (AC) context. More precisely, the novelty of our proposal is to provide general guidelines for correctly end effectively performing the testing of AC systems.

5.2.5 Related Work

Part of the scientific and industrial worlds are relying on Access Control mechanisms for the GDPR enforcement. However, the proposed solutions may be exposed to the risk of encoding data protection vulnerabilities or threats, and therefore accurate and specific testing process should be adopted.

In this section an overview of the most important proposals concerning the use of AC and the GDPR's principles, as well as empirical validation for assessing the testing results are provided. Thus in the following we identified four main research fields: 1) Analysis and modeling of policy specification; 2) Test cases and oracle derivation; 3) Coverage assessment; and 4) Supporting controlled experiment.

Analysis and modeling of policy specification. Available proposals include different verification techniques [242], such as model-checking [251] or SAT solvers [225]. Well-known analysis and verification tools for access control policies are: i) Margrave [96], which represents policies as Multi-Terminal Binary Decision Diagrams (MTBDDs) and can answer queries about policy properties; and ii) ACPT (Access Control Policy Testing) tool [118] that transforms policies into finite state machines and represents static, dynamic and historic constraints into Computational Tree Logic. The capabilities and performances of such tools are analytically evaluated in [143]. The authors of [190] provide an optimized approach for XACML policies modeling based on tree structures aimed at fast searching and evaluation of applicable rules derivation.

Test cases and oracle derivation. Considering the automated test cases generation, solutions have been proposed for testing either the XACML policy or the PDP implementation [46, 47]. Among them, the most referred ones such as X-CREATE [43, 47] and the Targen tools [160] use combinatorial approaches for test cases generation. Specifically, the X-CREATE tool and the Targen tool generate test inputs using combinatorial approaches of the XACML policies values and the truth values of independent clauses of policy values, respectively. However, combinatorial approaches are shallow with respect to policy semantics. Model-based testing has already been widely investigated for policy testing, e.g., [194, 239]. Such approaches provide methodologies or tools for automatically generating access control test models from functional models and access control rules. A different approach is provided by Cirg [116, 144] that is able to exploit change-impact analysis for test cases generation starting from policy specification. An overview of verification and testing methods for access control policies and models is provided in [116].

Other approaches leverage existing symbolic execution techniques for generating test cases. Specifically, in [146], first the access control policy under test is converted into semantically equivalent C Code Representation (CCR). Then, the CCR is symbolically executed to generate test inputs. About the automated oracle, notwithstanding the huge interest devoted to this topic, reducing the human activity in the evaluation of the testing results is still an issue [21]. With respect to existing approaches, the solutions adopted in this chapter focus on the one hand to enrich the expressiveness and diversity of the generated test cases, and on the one other to automatically derive the verdict associated with each of them. Indeed, the automated oracle derivation is a key aspect in the context of XACML systems and testers need usually to manually verify the XACML responses. To this purpose in [68], we provided an integrated toolchain including test cases generation as well as policy and oracle specification for the PDP testing, while in [56] we addressed the use of monitoring facilities for the assessment of the run-time execution of XACML policies. In this thesis, we focus mainly on our more recent works (i.e., [45, 76]) where an automated model-based oracle is presented. In this work, the expected behaviour of the evaluation of a given XACML policy is modeled as a labeled graph and the proposed testing strategies is able to guarantee the full path coverage of such graph and the automatic derivation of the expected verdicts. Thus, the proposal of this chapter exposes the following peculiarities: i) the derivation of XACML requests explicitly takes into account the semantics of XACML functions as well as the policy and rule combining algorithms; ii) the derivation for each test request is automatically associated to its expected verdict.

Coverage assessment. Coverage assessment is an important feature to focus the testing activity on the generation or selection of the test cases that cover the most important elements and/or policy constructs [211]. Considering coverage assessment of a test suite with respect to an XACML policy, in literature there are few works facing this problem. Seminal works as [163] and [48] present some coverage criteria for XACML policy. Specifically, in the former, the authors define three structural coverage metrics targeting XACML policies, rules and conditions respectively and use them for reducing test sets and measure the effects of test reduction in terms of fault detection. In the latter, the authors also address the policy set and do not require the policy execution and PDP instrumentation. More recently, in [254] the authors extend the introduced concept by presenting a combinatorial testing approach based on data-flow coverage, in [241] the test execution information is used to determine which policy element is faulty, while in [240] a family of coverage criteria for XACML policies is formalised and used for test cases generation. The authors of [150] propose an access control policy infrastructure, based on an external monitoring facility, for enabling the coverage measurement of XACML policies and evaluation of different testing strategies.

As emerged from the state-of-the-art, if from the security point of view different solutions are currently available for testing access control mechanisms and their policies, there are not evidences for considering them effective also from the privacy one. Indeed, there are very few proposals specifically focused on testing that sensitive information or resources are correctly accessed and managed. Thus, an assessment of existing testing approaches according to the GDPR requirements are still needed.

To this purpose, as detailed more in Section 8.3, we propose a possible solution for assessing software testing techniques in line with the GDPR specifications.

Supporting controlled experiment. Empirical validations play a key role in the evaluation of a software system. Validation in software engineering discipline, as in other research fields, relies on building different models of this discipline, i.e., modeling the objects of the domain, the processes manipulating the objects, the relations between processes and objects [235]. The authors of [235] give first an overview of empirical strategies such as surveys, case studies, experiments and then define the main steps of experiment process such as scoping, planning, operation, analysis and interpretation, presentation and package.

The work in [51] discusses specific challenges and issues of performing empirical studies with software testing techniques whereas the authors of [86] identify two main complementary classes of empirical studies addressed in software testing: case studies [202] and controlled experiments [122].

Controlled experimentation in software testing leverages numerous software artifacts, including for instance different versions of software systems, test suites, test object, fault data, mutated software. Obtaining such artifacts and organizing them in an environment able to support controlled experimentation is a difficult task.

The work in [85, 86] presents first a survey of papers on testing that provide empirical studies identifying the main challenges of experimentation in software testing. These identified challenges are then used for designing and constructing an extensible infrastructure able to support controlled experimentation with software testing and regression testing techniques. This infrastructure provides guidelines for object selection, organization, and setup processes with the aim of reducing the costs of executing and replicating controlled experiments as well as aggregating results across experiments.

In this chapter, we aim to leverage the advantages of the controlled experiments in the context of access control by defining and executing a controlled experiment for the evaluation of two test cases generation strategies.

5.3 Defining and Implementing Testing Process (Contr. 1)

In this section, we present the testing process integrated in the GENERAL_D proposal as well as its reference architecture.

5.3.1 Workflow of the Testing Process

A typical AC systems testing process, shown in Figure 5.1, consists of at least four main steps: (A) *test cases generation*; (B) *mutants generation*; (C) *execution of the test cases*; and finally, (D) *results analysis*.

Step A The first step is related to the generation of test cases (step A) for the aim of generating a test suite starting from a given ACP.

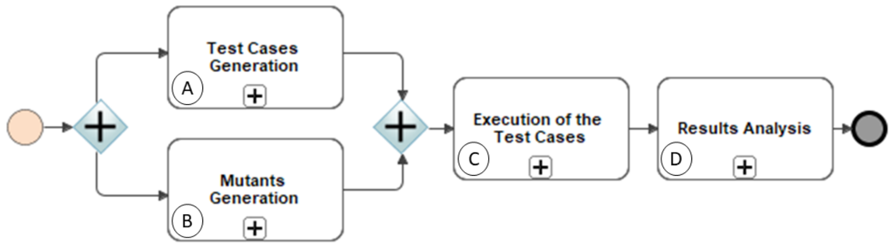


Figure 5.1: *Workflow of the Testing Process.*

Step B The next step (B in Figure 5.1) is related to the generation of ACPs mutants; the mutated versions of the ACPs can be generated by applying a set of mutation operators. The basic idea of mutation testing is to simulate typical programmer's mistakes, by seeding syntactic faults in the original program in order to produce a set of faulty programs, called mutants, each one containing one fault. The main purpose of mutation testing is to assess the adequacy of a test suite. Each test case is executed on the original program and its mutants (in our case the ACP and its mutated versions), then outputs are collected (i.e., the authorization responses): if the mutant's output is different from the original program's one, the fault is detected and the mutant is said to be killed. The mutation score is the ratio of the number of detected faults over the total number of seeded faults and indicates the effectiveness of the test suite.

Step C The results of steps A and B are then used in the next phase (step C in Figure 5.1), that allows the execution of test cases on the original ACP and on its mutated versions.

Step D Finally, the result of step C is used in step D of the testing process (see Figure 5.1).
In order to speed-up the testing phase, the above steps have been integrated into the XACML Mutation Framework (XMF) as described in the following section.

5.3.2 Reference Architecture

In this section, we present the *XACML Mutation Framework (XMF)* useful both for testing the PDP component and ACPs, and for assessing the test cases generation strategies [72]. The framework provides three main functionalities: *1)* test case generation, execution and assessment; *2)* mutants generation; and *3)* a data mart for OLAP analysis [105].

Very briefly, considering the testing of PDP (or the ACP), XMF lets the execution of a set of access requests on the PDP (or on using the ACP) and the consequent comparison of the collected responses against the expected ones. Either in case of PDP or policy testing, a PDP needs to be configured to use a selected policy, the requests have to be sent to the PDP under test, and the responses (permit, deny, not applicable or indeterminate) collected.

Considering instead the assessment of the test generation strategies, XMF lets first the execution of the requests (test cases) on the original PDP (or the ACP) and the collection of the associated set of responses. Then, XMF replaces the original PDP (or the ACP) with one of its mutated versions, each of the test cases re-executed on this mutant, and responses are collected again. Finally, the responses are analyzed and compared so as to discover the killed mutants. In the case of PDP, a mutant is considered killed when an exception is raised or when the returned response is different from the expected one. As final step, the mutation score for the whole test suite is calculated by dividing the number of killed mutants by the number of mutants.

According to the literature, a test suite is considered of high quality if it is able to reach a high mutation score, i.e., the test suite has a high fault detection capability.

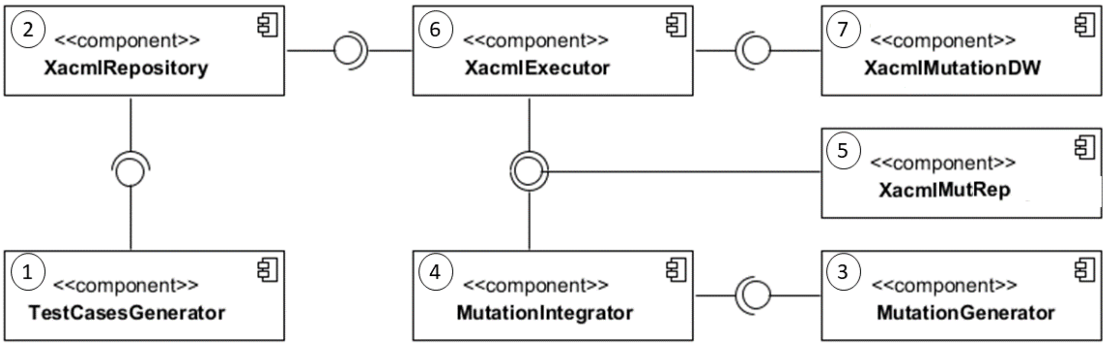


Figure 5.2: *The proposed XACML Mutation Testing Framework.*

Figure 5.2 schematizes the architecture of XMF framework which mainly consists on the following seven components:

1. ***TestCasesGenerator*** is an automated XACML requests generator, which implements and/or integrates different testing strategies or tools so as to reduce as much as possible the time and the effort required for the test cases specification;

2. ***XacmlRepository*** is a database that contains XACML policies, XACML requests, i.e., test cases, and XACML decisions defined by the XACML language, i.e., Permit, Deny, NotApplicable and Indeterminate. The data are organized so as to be able to associate the requests to the policies from which they are generated and to keep track of the generator used for their generation;

3. ***MutationGenerator*** is a generator that automatically derives mutated versions of the original PDP or the ACP. In case of the PDP, these are generated by applying a set of Java based mutation operators producing set of mutated java classes, each one containing only one fault. In case of ACP, the mutants are generated by applying specific XACML-based mutation operators;

4. ***MutationIntegrator*** works in direct collaboration with the MutationGenerator for seeding the faults in the code of PDP (or the access control policy) and producing executable mutated versions of the original PDP (or the access control policy);

5. ***XacmlMutRep*** maintains all the original PDPs (or the access control policies) and the associated mutated versions. It also contains the mutation operator applied to the original PDP (or the access control policy) to obtain the mutated version;

6. ***XacmlExecutor*** is an automated executor of test cases on the original PDP (or the access control policy) and the associated set of mutated PDPs (or access control policies);

7 ***XacmlMutationDW*** contains a data mart for storing the collected data derived from test cases and mutants generation activities as well as the evaluation activity.

5.4 Controlled Experiment Family (Contr. 2)

In literature, different solutions are currently available for testing AC systems and their behavior [46, 47, 116]. They can be mainly divided into the following research fields: i) test strategies definition [45, 47]; ii) test strategy assessment [48, 72, 150]; iii) test cases generation and execution [43, 116]; and iv) test execution and oracle derivation which are focused on approaches for evaluating the AC replies to specific inputs [45,56,68,76].
Unfortunately, there are not standardized guidelines for correctly and systematically performing the testing process in order to avoid errors and improve the effectiveness of the validation. In particular, the lack of a formalized specification of the testing activity can have the following consequences: impossibility of replicating and controlling the process especially in case of regression testing [247], difficulties in the generalization of the testing results and consequent derivation of statistical significance values; and problems in defining and sharing a common testing knowledge so as to avoid recurring failures and speeding up the corrective process. Differently, our work wants to contribute to formally and thoroughly conduct CEs in the context of AC.

Indeed, a reply to these issues comes from the software engineering context, where CEs [35, 128, 234] are commonly used to investigate the cause-effect relationships of introducing new methods, techniques or tools and to build a body of knowledge supported by observation and empirical evidence. Therefore, controlled experiments let to validate the different activities of the testing process by means of the identification of important variables, the definition of specific testing models and objectives, and the derivation of empirical evidence. In the controlled experiment different treatments can be applied to, or by, different subjects, while other variables are kept constant and the effects on response variables are measured.

Authors in [128,234] categorize experiments as either technology-oriented or human-oriented, depending on whether artifacts or human subjects have given various treatments. In this work, we revise and customize the technology-oriented experiments in order to provide general guidelines for correctly end effectively performing the testing of the AC systems. Therefore, we provide the characterization of the first three (over the five) steps of the *Experiment Process*, namely *Scoping*, *Planning*, and *Operation*. We refer to Section 5.8 for a concrete application example and a detailed checklist of the required implementation steps.

The reminder of this section is organized as follows. Section 5.4.1 introduces the Goal-Question-Metric paradigm, and Section 5.4.2 illustrates our proposal of a family of Controlled Experiments in the context of Access Control. Finally, in Sections 5.4.3, we detail the first phase of the Controlled Experiment.

5.4.1 Goal-Question-Metric

Originally presented in [35], the Goal-Question-Metric (GQM) paradigm proposes a top-down approach to define measurement: goals lead to questions, which are then answered by metrics. A GQM model is a hierarchical structure as presented in Figure 5.3

starting with a goal by specifying purpose of measurement, object to be measured, issue to be measured, and viewpoint from which the measure is taken (*Conceptual level*). The goal is refined into several questions that usually break down the issue into its major components (*Operational level*). Each question is then refined into metrics, some of them objective and others subjective (*Quantitative level*). The same metric can be used to answer different questions under the same goal as well as different goals [36].

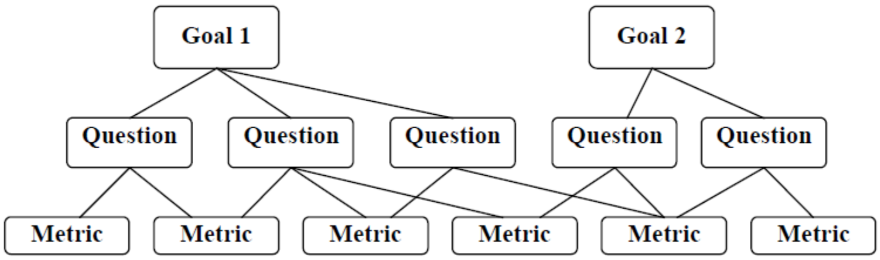


Figure 5.3: *The Goal Question Metric (GQM) model (adopted from [36]).*

In security domain there are a few proposals using the GQM and they are used to mainly identify security requirements and metrics. For example, authors in [120] used GQM approach to define clear and comprehensible measures for a set of established security requirements. The GQM approach based on Standard security metrics and on Service Oriented Architecture (SOA) maturity is presented in [131], where scholars aimed at supporting organizations to assess SOA Security as well as to ensure the safety of their SOA based collaborations. To assessing the security of data stored in cloud storage, authors in [244] attempt to provide practical guidance and example of measurements using GQM. A more recent work is presented in [232] where the authors presented a quantitative evaluation approach for defining security assurance metrics using two perspectives, vulnerabilities and security requirements.

Differently from the above works, our proposal aims at enabling the derivation of metrics for answering questions related to investigation goals in the context of AC. In particular, the intention is to enable CEs in the context of AC by covering all the phases of the process. In this work however we focus on the first three phases of the process and we refer to Section 5.8 for more details about the remaining phases.

5.4.2 A GQM Proposal for Access Control Testing

The general idea behind our proposal is to provide a set of CE families useful for formally and thoroughly describing scientific investigations in the context of AC systems. Indeed, our intuition is to use the standard and consolidated GQM template [35], as guidance to select, and consequently classify, concepts of interest in the domain of AC. Then, by exploiting the knowledge and techniques typical of the software testing scientific environment, a concrete AC-based goal definition framework can be derived. This set will be well-defined, specific and achievable AC testing goals to be exploited for different experimentations.

The proposal of the present work, although grounded in a domain-related AC testing, represents an example of realization of CE families, that can be easily applied in all the domains where a scientific investigation in which a formal and rigorous fashion should be performed.

As in Figure 5.4, the proposal is composed of five conceptual components: the *Goal-Question-Metric* ①, the *Access Control Context* ② and *Software Testing* ③, which represent the conceptual models of the target experiment.

These models are integrated in the *Goal Definition Framework* component ④ so as to define a specialized GQM, which is the common basic knowledge for the AC families. Then, the GQM exploited in the *Main Research Goal* component ⑤ for defining scientific testing goals in AC testing process, and therefore for selecting specific and achievable AC testing goals ⑥ to be evaluated in real contexts.

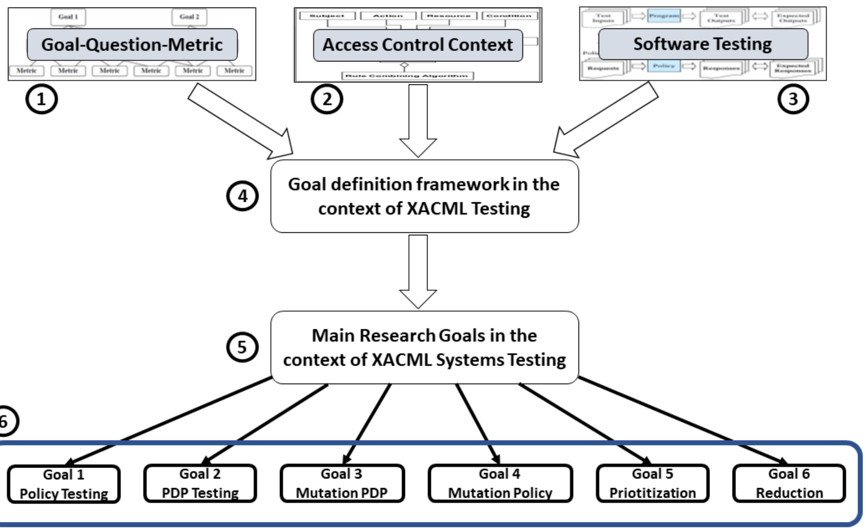


Figure 5.4: *GQM Access Control Model.*

In the following, we illustrate how the use of a specialized GQM can be an important innovation for the development of CEs in AC context. In particular, by referring to the structure of a CE presented in Section 5.2.4, we detail the execution of the first three steps of the process (i.e., *Scoping*, *Planning* and *Operation*), which are those that need to be specialized for the AC domain. As previously mentioned, we refer to Section 5.8 for a complete example including also the last two phases.

5.4.3 Experiment Scoping

The purpose of the scoping phase is to determine the foundations of the experiment by defining goals according to a specific framework. As described in the previous section, the idea here is to use the Goal-Question-Metric (GQM) method, integrated with concepts of AC and Software Testing for deriving a specialized template for the definition of CEs goals in the context of AC testing. By referring to Figure 5.4, the scoping phase exploits the domain specific concepts of components ①, ② and ③ so as to define a reference framework, i.e., Goal Definition Framework (component ④).

According to [35], the GQM template consists of five elements: (1) *object of study* is target entity of the experiment. It can be a product, process, resource, model, metric or theory; (2) *purpose* defines the intention of the experiment. It may be to evaluate the impact of two different techniques or to characterize the learning curve of an organization; (3) *quality focus* is the primary effect under study in the experiment. It can be effectiveness, cost or reliability; (4) *perspective* describes the viewpoint from which the experiment results are interpreted. Examples are developer, project manager, customer and researcher; (5) *context* is the environment in which the experiment is run. It defines which personnel is involved in the experiment (subjects) and which software artifacts, called objects [1], are used in the experiment.

Consequently, the intention of the GQM template is to:

> *Analyze* <Object(s) of study> *for the purpose of* <Purpose> *with respect to their* <Quality focus> *from the point of view of the* <Perspective> *in the context of* <Context>.

Table 5.1: *AC concepts.*

GQM elements	**AC concepts**
Object of study	XACML-based PDPs XACML-based ACPs
Purpose	-
Quality focus	-
Perspective	ACP Architect AC System Developer AC System Administrator
Context	Subjects (XACML Policies) Objects (XACML-based PDPs)

AC Model. The objective here is to characterize the CE in the context of AC by gathering the main concepts, terms and components that can be used to formulate interesting goals from the scientific point of view. The selected elements are then used in the GQM template for the *object of the study*, the *purpose*, the *perspective* and the *context*. The classification we propose in this work is summarized in table Table 5.1. In particular, the first column (`GQM elements`) lists the GQM element, while the second one (`AC concepts`) reports concepts useful for defining meaningful research investigation in the context of AC.

Software Testing. In literature different proposals exist that leverage well-known software techniques to test ACPs and AC mechanisms. By analyzing current literature, we summarize in Table 5.2 in the column `Software Testing concepts` some of the main software testing concepts useful in generic controlled experiment. We also classify them according to the GQM template elements (column `GQM elements`).

However, without the pretend to be exhaustive, the table reports a simplification of a possible classification. Indeed, we limit ourselves to the definition and assessment of the test case generation strategies because they are recognized as the most crucial activities of the testing process. In the assessment of the effectiveness of a test strategy, concepts as coverage criteria and mutation analyses or test oracle are often used, and therefore included in Table 5.2. We also add the prioritization and reduction concepts because they are commonly adopted techniques for reducing the number of test cases to be executed and consequently the effort and time due to the overall testing phase.

[1] Note that the *objects* here are generally different from the *objects of study*

Table 5.2: *Software Testing concepts.*

GQM elements	Software Testing concepts
Object of study	Test case generation strategy Test case prioritization technique Mutation Generators Test case reduction technique Oracle Derivation
Purpose	Characterize Evaluate
Quality focus	Effectiveness Cost Size APFD Performance
Perspective	Researcher Tester Project manager User
Context	-

Goal Definition Framework. On the bases of the concepts reported in Table 5.2, the specialized Goal Definition Framework is derived. This is a comprehensive framework based on GQM for defining research investigation goals for testing tools, methodologies and strategies in the AC (both ACPs and AC mechanisms) context. To the best of the authors' knowledge, this proposal is the first attempt to provide a formally and thoroughly solution for the definition of Controlled Experiments in AC domain. Table 5.3 reports the conceived framework, which represents the output of component 4 of our proposal depicted in Figure 5.4. Specifically, Table 5.3 has a column for each of the five

Table 5.3: *Goal definition framework in the context of XACML Testing.*

Object of study	Purpose	Quality focus	Perspective	Context
Test case generation strategy	Characterize	Effectiveness	Researcher	Subjects (XACML Policies)
Test case prioritization technique	Evaluate	Cost	Tester	Objects (XACML-based PDPs)
Mutation Generators	Assess	Size	Project manager	
Test case reduction technique		APFD	User	
XACML-based PDPs		Performance	ACP Architect	
XACML Policies			AC System Developer	
XACML-based Oracle Derivation			AC System Administrator	

GQM elements, where the identified AC and Software Testing concepts are reported: namely `Object of study`, `Purpose`, `Quality focus`, `Perspective`, and `Context`.

Research Goals in AC context. By combining the elements of the five columns of Table 5.3, a well-defined and focused scientific investigation goal, that enable the specification of CE in the context of AC, can be identified. Thus, the Goal Definition Framework lets the definition of families of goals for access control systems testing. In Table 5.4 a non-exhaustive list of the mostly adopted research goals is reported: the first column (`Research Goal`) reports a label associated to each defined goal; the second column (`Goal Definition`) contains the definition of the goal using the GQM template customized with a specific combination of elements of Table 5.3; whereas in the last column (`Co-Authored Publications`), for the aim of completeness, we recall the scientific contributions where the goal has been applied using real context data.

As evidenced in the table, not all the combinations of the elements of Table 5.3 enable the definition of an interesting and well-defined goal. According to the different situations, the user should select the correct combination, depending on the concrete objective.

Table 5.4: *Main Research Goals in the context of XACML Systems Testing and Related Publications.*

Research Goal	Goal Definition	Co-Authored Publications
Goal 1: Policy Testing	***Analyze*** test case generation strategies ***for the purpose of*** evaluation ***with respect to their*** effectiveness and size of test suite produced ***from the point of view of the*** researcher ***in the context of*** XACML policy testing.	[31,79]
Goal 2: PDP Testing	***Analyze*** test case generation strategies ***for the purpose of*** evaluation ***with respect to their*** effectiveness and size of test suite produced ***from the point of view of the*** researcher ***in the context of*** XACML policy decision point testing.	[45,69,71,73,75,76,79]
Goal 3: Mutation PDP	***Analyze*** mutation generators ***for the purpose of*** evaluation ***with respect to their*** applicability ***from the point of view of the*** researcher ***in the context of*** XACML policy decision point testing.	[70,72,73,76,79]
Goal 4: Mutation Policy	***Analyze*** mutation generators ***for the purpose of*** evaluation ***with respect to their*** effectiveness and size of test suite produced ***from the point of view of the*** researcher ***in the context of*** XACML policy testing.	[42,79]
Goal 5: Prioritization	***Analyze*** test case prioritization techniques ***for the purpose of*** evaluation ***with respect to their*** effectiveness (rate of fault detection, using APFD (Average Percentage Faults Detected) metric) ***from the point of view of the*** researcher ***in the context of*** XACML policy testing.	[44]
Goal 6: Oracle	***Analyze*** oracle derivation techniques ***for the purpose of*** evaluation ***with respect to their*** correctness and cost ***from the point of view of the*** researcher ***in the context of*** XACML policy and PDP testing.	[45,71,73,75,76]
Goal 6: Efficiency	***Analyze*** testing frameworks ***for the purpose of*** evaluation ***with respect to their*** efficiency (in terms of cost and time metrics) ***from the point of view of the*** researcher ***in the context of*** XACML policy and PDP testing.	[71,75]

For the aim of completeness, among the goal listed in Table 5.4, in Section 5.8 we detail the execution of all the phases of the proposed Controlled Experiment considering the *Goal 2*.

Remark. In this section, we presented a family of controlled experiments in the context of AC testing. The idea is to define a set of standardized guidelines for correctly and systematically performing the testing process to avoid errors and to improve the effectiveness of the validation. The proposal relies on a characterization of the first step of the experiment process (i.e., Scoping) by leveraging the Goal-Question-Metric template.

5.5 Test Cases Generation: GROOT (Contr. 3)

In this section, we target the third testing objective depicted in the introduction of this chapter that concerns the specification of testing strategies specifically conceived for validating GDPR provisions. In particular, in this section, we introduce GROOT, a GdpR-based cOmbinatOrial Testing strategy. This is a general combinatorial strategy for testing systems managing GDPR's concepts (e.g., Data Subject, Personal Data or Controller).

Inspired by [176], in illustrating the GROOT approach, we use the following definitions:

Definition 1 (GDPR-based SUT Model)**.** *A GDPR-based SUT model is a tuple* $Model_{GDPR}(P, V)$*, where:*

- $P \subseteq \{DataSubject, Controller, PersonalData, Processor, Consent, Purpose, ProcessingActivity, ThirdParty\}$ *is the set parameters that affect the GDPR-based SUT, and*
- $V = \{V_i \mid \ i \in P$ *and* V_i *is the set of values for the parameter i* $\}$ *is the set of sets of the values that can be selected for each parameter.*

Definition 2 (GDPR-based Test Case)**.** *Given a GDPR-based SUT mode* $Model_{GDPR}(P, V)$*, a GDPR-based Test Case is tuple* $TC_{GDPR}(ATT)$ *where:*
$ATT = \{ATT_i \mid ATT_i \subseteq V_i$ *, i* $\in$ *P and* $V_i \in V\}$

An example, in tabular form, of a GDPR-based SUT Model is reported in Table 5.5: columns represent the set pf parameters P; and the content of the table represent the values (V) associated to each parameter. For instance, the set associated with parameter $Purpose$ is $V_{Purpose} = \{MyCholesterol, UntargetedMarketing\}$.

DataSubject	Controller	PersonalData	Purpose	ProcessingActivity	ThirdParty
DSName	orgName	Age	MyCholesterol	AGREGATE	orgName
	piiController	Gender	Untargeted Marketing	DERIVE	
	address	Blood Cholesterol		QUERY	
	e-mailC	e-mailDS		COLLECT	
	phone number			SEND	

Table 5.5: *Parameters (P) and Values (V) Associated to the Use Case Scenario in Section 5.7.5.*

In the following, we briefly illustrate GROOT which is composed of four main steps (see Figure 5.5): Reqs Implementation (Step ①); Parameters Derivation (Step ②); Parameters Combination (Step ③); and Test Cases Generation (Step ④).

(1) *Reqs Implementation (Step ①)*: Starting from Data Protection requirements, in our case the GDPR's reqs, an implementation of them is obtained. In the context of this thesis, an implementation refers to GDPR-based ACP implementing GDPR's demands. We do not discuss here either how to express Data Protection Requirements or how to implement them. We refer to Chapter 6, Chapter 7 and Chapter 8 for possible strategies to implement ACPs in compliance with the GDPR. The hypothesis of our work is that a GDPR-based implementation is available in terms of a specification language, for instance an ontology, a UML instance or an access control policy.

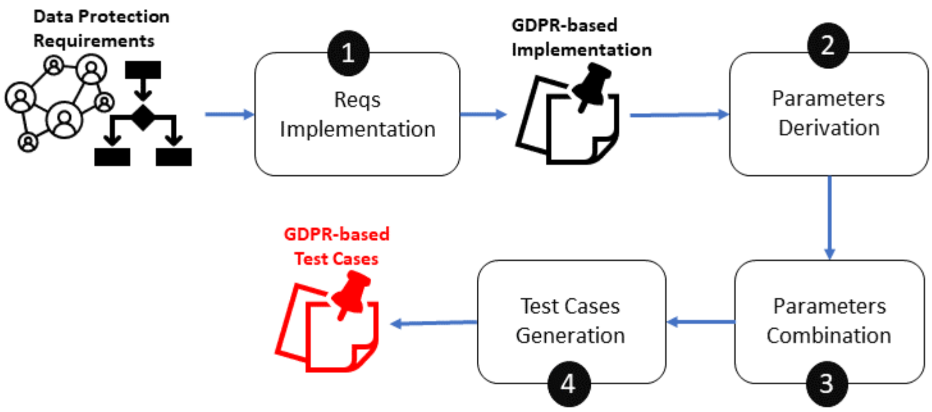


Figure 5.5: *GROOT Methodology: A Combinatorial Approach for Test Cases Generation in the Context of the GDPR.*

(2) *Parameters Derivation (Step 2)*: The GDPR-based implementation is parsed in order to identify the set of parameters P, and the associated set of sets V as per Definition 1. More precisely, for each parameter P_i, this step is able to associated the correct subset V_i containing values taken from the considered implementation. An example of result of this step is reported in Table 5.5, and it is based on the use case scenario described in Section 5.7.5.

(3) *Parameters Combination (Step 3)*: In this step, the combination of the parameters' values is computed. In particular, according to the Definition 2, for each i we firstly derive all possible subsets of V_i called $P(V_i)$ i.e., power set of V_i. Then we combine all the obtained subset so as to derive the tuple (ATT), which in turn each represents a possible test case. The obtained set n-tuples represent all possible combinations of the considered parameters, and this allows performing the exhaustive testing.

(4) *Test Cases Generation (Step 4)*: For each of the obtained combination in Step 3, in this step a specific executable test case in generated. In the context of AC, a test case is represented as an AC request that can be evaluated by the ACM.

It is well-known that combinatorial testing suffers for the explosion problem of exhaustive testing and lack of oracle. To tackle both of them, in [71, 75] we have explored the possibility of using decentralized solutions, where we have provided:

- a general framework able to leverage the available distributed computational resources for exploiting all the power of combinatorial approaches for the generation of XACML requests; and
- a decentralized framework for the PDP oracle problem. Well-known approaches for automated XACML oracle derivation rely on voting mechanisms. Our distributed framework leverages these approaches to provide a more efficient oracle solution able to reduce the high computational costs related to the derivation of an authorization response associated to an XACML request.

It is out of the scope of this chapter discussing in details these results, and we refer [71, 75] for more details.

5.5.1 Combinatorial Testing for Traditional XACML-based AC

GROOT includes our two state-of-the-art combinatorial test cases generation strategies previously conceived [47]: Simple Combinatorial and Multiple Combinatorial.

Simple Combinatorial. The Simple Combinatorial testing strategy applies a combinatorial approach to the policy values, and more precisely, it combines values of the four main categories defined in XACML: *Subject*, *Resource*, *Action* and *Environment*. Consequently, this methodology requires the definition of the SubjectSet, ResourceSet, ActionSet, EnvironmentSet, each containing set of entities of the four main categories contained in the policy under test. For instance, SubjectSet contains entities that describe the different subjects that the policy rules the access of; ResourceSet contains entities describing all the protected resources by the policy; similarly, ActionSet contains entities identifying the permitted actions. Specifically, a subject entity is a combination of the values of `<AttributeId>` and `<DataType>` attributes and the value of the `<AttributeValue>` element of the SubjectSet set. Resource, action, and environment entities are similarly derived considering the ResourceSet, ActionSet, and EnvironmentSet values. Random entities are also included in each set so that the resulting test plan could be used also for robustness and negative testing purposes.

Starting form the above-defined sets, Simple Combinatorial derives all combinations of subject entities, resource entities, action entities and environment entities. Each combination is then translated in a Simple XACML request containing the entities of that combination. The generated test suite guarantees a coverage of all pairs (by applying pairwise approach), then of all triples (by applying the three-wise approach) and finally of all quadruples of values entities derived by the policy, by applying the four-wise approach that contains all possible combinations.

The main advantage of the proposed strategy is its simplicity and its ability to achieve the full coverage of the policy input domain represented by the policy values combinations.

Multiple Combinatorial. Similar to the Simple Combinatorial, this strategy relies on a combinatorial approach, and for each policy, four sets are generated, the SubjectSet, ResourceSet, ActionSet, and EnvironmentSet, containing the values of elements and attributes of the subjects, resources, actions and environments respectively.

We define for each set S $\in$ {SubjectSet, ResourceSet, ActionSet, EnvironmentSet}:

- the power set of S, called P(S), as the set of all possible subsets of S;
- the cardinality of P(S) as $\sharp P(S) = 2^n$, where n is the cardinality of S;
- the *subset entity* as each element in P(S). For instance, the element is called *subject subset* if S=SubjectSet.

The possibly exponential cardinality of P(S) is reduced by fixing the number of its subset entities. Indeed, the necessary condition for an XACML request to be applicable on a field of the XACML policy (rule, target, condition) is that this request simultaneously includes all the entities that are specified in that policy field. Thus, the XACML policy provides the minimum and maximum number of entities of the same type that have to be included in a request. For instance, if in an XACML policy there is never a condition or a target in which not less than two and not more than three subject entities are required for its evaluation, the minimum and maximum number of subject entities is 2 and 3 respectively. We use these numbers to (optionally) decrease the subject subsets.

The test requests are then generated by combining the subject, resource, action and environment subsets as in the following:

- apply the pair-wise combination to cover all pairs (a, b) where: a $\in$ A, b $\in$ B such that A, B $\in$ {P(SubjectSet), P(ResourceSet), P(ActionSet), P(EnvironmentSet)} and A$\neq$B, we obtain the *PW* set;

- similarly apply the three-wise, to cover all triples (a, b, c) where a $\in$ A, b $\in$ B and c $\in$ C, such that A, B, C $\in$ {P(SubjectSet), P(ResourceSet), P(ActionSet), P(EnvironmentSet)} and A$\neq$B$\neq$C, we obtain the *TW* set;

- apply the four-wise, i.e. all possible combinations of the subject subsets, resource subsets, action subsets and environment subsets, we obtain the *FW* set.

Because the inclusion propriety is $PW \subseteq TW \subseteq FW$, duplicated combinations have been eliminated considering the following sets: *PW* called *Pairwise*, $TW \setminus PW$ called *Threewise* and $FW \setminus (TW \cup PW)$ called *Fourwise*.

Considering first Pairwise set, then Threewise set and finally Fourwise set, for each combination an XACML request containing the subset entities is generated. The maximum number of requests derived by this strategy is equal to the cardinality of FW set.

5.6 Test Cases Generation: XACMET (Contr. 4)

In XACML-based access control systems, incoming access requests are transmitted to the Policy Decision Point (PDP) that grants or denies the access based on the defined XACML policies. The criticality of the PDP component requires an intensive testing activity consisting in probing such component with a set of requests and checking whether its responses grant or deny the requested access as specified in the policy. Existing approaches for improving manual derivation of test requests such as combinatorial ones do not consider policy functions semantics and do not provide a verdict oracle.

In this section, target the forth testing objectives depicted in the introduction of this chapter that concern the specification a model based testing strategy. In particular we introduce the XACML Modeling Testing (XACMET) , a novel approach for systematic generation of XACML requests as well as automated model-based oracle derivation. Referring to [66] for a more detailed description, we briefly presents in this chapter its main features, that are: a new test case generation strategy based on path coverage (Section 5.6.2); an automatic derivation of an XACML oracle (Section 5.6.3); and, capabilities for measuring the coverage of test requests (Section 5.6.4).

With aim of readiness, in illustrating XACMET, we refer to an example of a simplified XACML policy useful for ruling library access. We report in Figure 5.6 the target policy. As in the Figure, the policy target says that the policy applies to any subject, resource and action. The policy is composed of two rules: *ruleA* and *ruleB*.

```
Policy ................................ root element
 |                                      rule-combining-algorithm:deny-overrides
 |__Target ............................ Sample Policy Applicable to any Requests and rules
 |                                      the access to Four resources
 |__Rule .............................. RuleId = ruleA, Effect = Deny
 |   |__Target ........................ The write action cannot be done by owner of the re-
 |   |   |                              sources book, document and documententry
 |   |   |__Resource .................. resource-id = book
 |   |   |__Resource .................. resource-id = document
 |   |   |__Resource .................. resource-id = documententry
 |   |   |__Action .................... action-id = write
 |   |__Condition ..................... The owner of the resources can request the access to
 |       |                              only one resource per time.
 |       |__And ....................... And Operator
 |           |__string-one-and-only ... type-One-And-Only Function.
 |           |                          # resource-id = 1
 |           |__string-is-in .......... type-Is-In Function.
 |                                      {resource-id} ⊆ {subject-id1}.
 |__Rule .............................. RuleId = ruleB, Effect = Permit
     |__Target ........................ The read action can be done by Julius on resource
         |                              journals
         |__Subject ................... subject-id = Julius
         |__Resource .................. resource = journals
         |__Action .................... action = read
```

Figure 5.6: *An Access Control Policy.*

ruleA, with effect *Deny*, has a target specifying that this rule applies only to the access requests of a “write” action of “book”, “document” and “documententry” resources. The rule condition will be evaluated true when the request resource value is contained into the set of request subject values.

The effect of the second rule *ruleB* is *Permit*, and it is returned as a response when the subject is “Julius”, the action is “read”, and the resource is “journals”. The rule combining algorithm of the policy is *deny-overrides*, which means that in case the same requests satisfies both the rules a *Deny* effect is returned as an authorization decision.

5.6.1 XACML Policy Modeling

The basic idea behind XACMET is to derive all possible evaluation paths of given XACML policy that a possible correct Policy Decision Point (PDP), conforms to the XACML specification, could generate during an XACML request evaluation. In doing so, we started modelling the XACML policy as a XML tree (called XAC-tree); we transformed the obtained tree into a well-defined graph (called XAC-Grapg). The transformation was guided taking into account the both the XACML specification and the semantic of all the element of of the policy (e.g., rules, functions and combining algorithms). Finally, by unfolding the obtained graph we were able to obtain a set of all possible paths (called XAC-Path), which represents a possible evaluation of a given access control request. The set of XAC-Paths can therefore be leveraged for different purposes: test cases generation; oracle derivation; and policy coverage measurement.

In the following we informally describe the steps involved for deriving XAC-Tree, XAC-Graph and XAC-Path, and we refer to the related publications for a formal definition [45, 76].

XAC-Tree Derivation

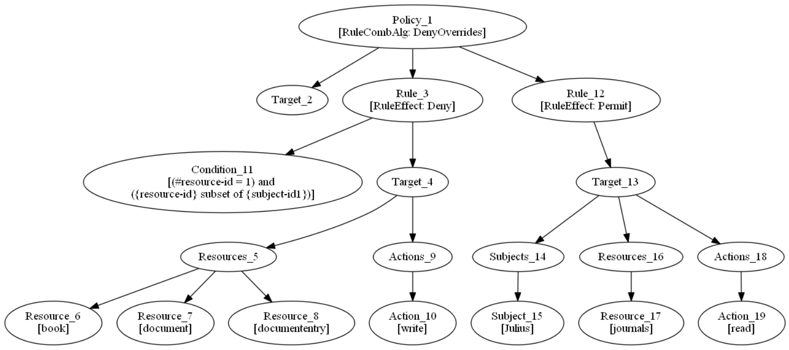


Figure 5.7: *XAC-Tree. Label T_P means node of type T and parameter P. The attributes are within square brackets.*

Considering the XACML policy as an XML document, we can represent it as a tree, called the XAC-Tree.

In particular, the following concepts can be used:

- *Contained:* Element i is contained within element j if i is between the start-tag and the end-tag of j.
- *Parent:* Element i is the parent of element j when j is contained within i and i is exactly one level above j.
- *Sibling*: The siblings in an XML document are the elements that are on a same level of the tree and have the same parent. In particular, given parent i of elements j and k, j is left (right) sibling of k if j is contained just before (after) k within element i.

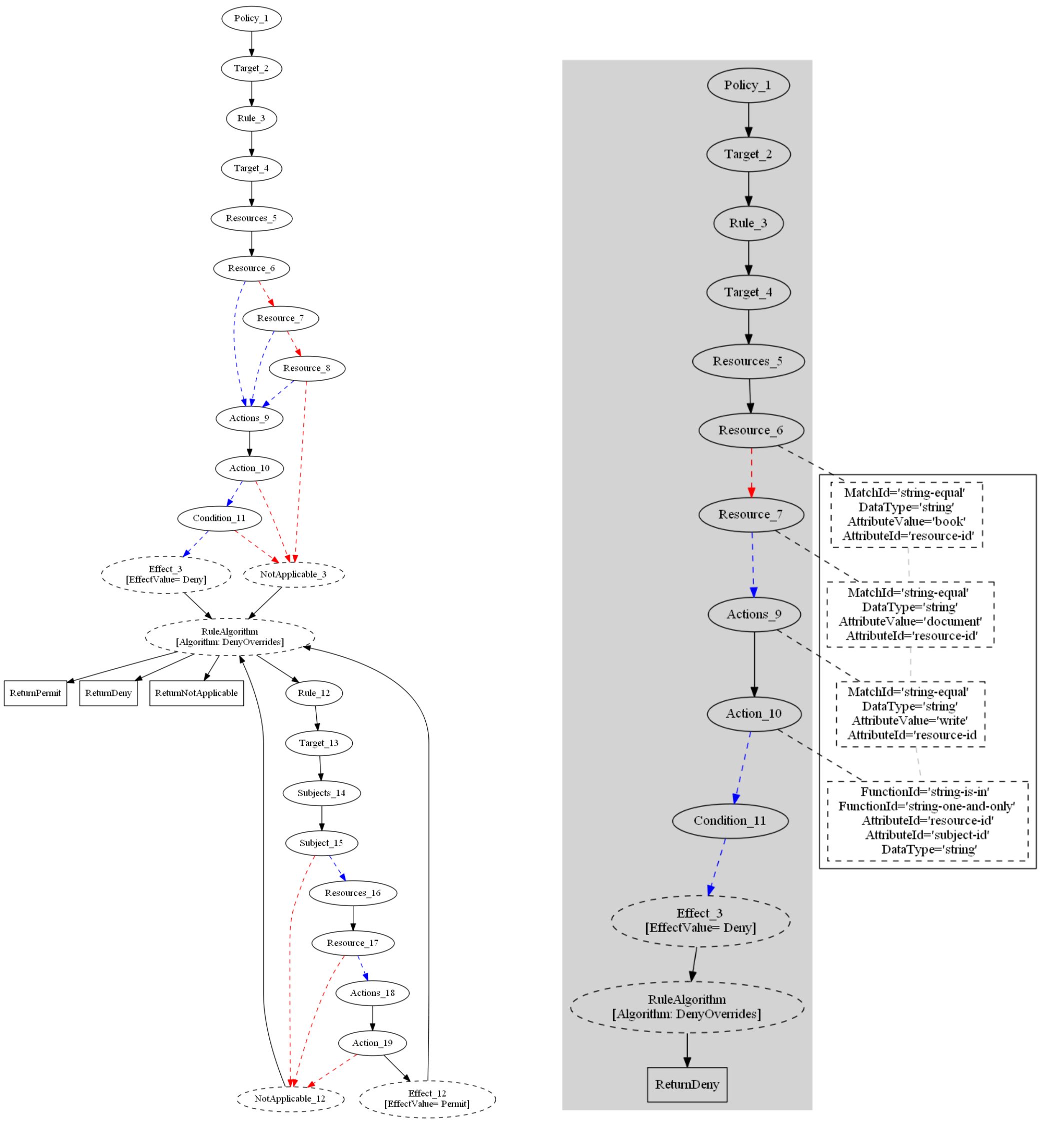
The XAC-Tree derivation exploits the parent relationship of the XACML policy and uses the following sets of types and values:

- T_V = {Policy, Target Rule, Subjects, Subject, Resources, Resource, Actions, Action, Environments, Environment};
- T_{V_a} = {RuleAlgorithm, Effect, NotApplicable};
- T_{V_v} ={ReturnPermit, ReturnDeny, ReturnNotApplicable};
- `RCA`= {FirstApplicable, DenyOverrides, PermitOverrides};
- `RE` = {Permit, Deny}.

By considering the policy of Figure 5.6, the associated XAC-Tree is shown in Figure 5.7. As depicted in the figure, for instance, *ruleA* becomes the node `Rule_3` into the XAC-Tree, and it child of node `Policy_1` and sibling of both nodes `Target_3` and `Rule_12`. In this case, 3 is the suffix of the node and the EffectRule attribute of the node `Rule_3` is set to Deny as specified in the Effect of *ruleA*.

XAC-Graph Derivation

The representation of the XACML policy as XAC-Tree is then used to derive a model of the XACML evaluation, i.e., the XAC-Graph, by considering the semantic of each node and taking into account the XACML specification. XAC-Graph is a labelled and typed graph, and can basically be derived applying a depth-first search approach to the XAC-Tree.



(a) *XAC-Graph. Label T_P means node of type T and parameter P. The attributes are within square brackets.*

(b) *A path of XAC-Graph. The boxes connected to the nodes contain the functions and values.*

Figure 5.8: *XAC-Graph and Example of Derived Path.*

By visiting XAC-Graph, we can derive both the set of test requests that allow for the full coverage of the paths of XAC-Graph and the derivation of the oracle verdicts. In both cases, the process adopted is divided into two main steps: *coloring* and *unfolding*.

Coloring XAC-Graph. During the coloring step, the concept of *Forward Node* is adopted. In particular, given the XAC-Graph G = (V_g, E_g, Entry) for each node i $\in V_g$ it is possible to identity the Forward Node $FN(i) \in V_g$ as the set of nodes j such that i is an XAC-GraphParent of j. Consequently, it is possible to define for each node i $\in V_g$ the Forward Star $FS(i) \in E_g$ as the set of edges (i, j), where j $\in FN(i)$.

Thus, given an XAC-Graph for each node b and c $\in V_g$, with $t_b \in$ {Subject, Resource, Action, Environment}, the cardinality of $FN(b)$ = 2, and $t_c \in$ {Subject, Resource, Action, Environment, NotApplicable}, the coloring process marks each edge (b,c) $\in FN(b)$: with red dashed line if $t_b = t_c$ or t_c = NotApplicable; with blue dotted line otherwise.

In practice, the blue dashed edges represent a successful evaluation of the node b. Figure 5.8(a) shows the XAC-Graph of the XAC-Tree of Figure 5.7.

Unfolding XAC-Graph. During the unfolding process, the paths are obtained by visiting the XAC-Graph from the Entry node to each node in V_v. The cycles are due to the presence of the node typed `Rule_Algorithm`. In XACMET, the order of the paths strictly depends on the order in which the rules are evaluated, which in turn is guided by the FirstApplicable, DenyOverrides, PermitOverrides algorithms.

In Figure 5.8(b), we show a path of XAC-Graph.

5.6.2 XACMET as Test Cases Generator

The main idea under the XACMET test cases generation approach relies on the semantics of XACML functions. Indeed, usually XACML rules to be satisfied may require that some conditions are met in the request. Another important aspect to be considered during the test case generation is the evaluation of the rule combining algorithms. Indeed, the rule combining algorithm prioritizes rules evaluation and therefore it pilots the result of the evaluation of an XACML policy.

The XACMET approach expressly takes into account the application of the rule combining algorithm during the request generation. Thus, it provides requests that systematically exercise all possible combinations of rule evaluations so as to guarantee the detection of this kind of faults. Intuitively, for each path identified during the unfolding step, the generation of requests is based on the data associated to each node. In particular, given *P* a path of *k* nodes for each node *i*, the functions and the values collected are translated into properties and constraints. If the outgoing edge is a blue dotted line (red dashed line), the set of values satisfying (not satisfying) the properties are identified using a constraint satisfaction approach. For all the nodes in a path *P*, the collected set of values are integrated and the values for test requests generation are successively extracted.

During the tests generation two different situation can occur:

- *Standard derivation*: the values in the various sets are selected and combined so as to generate the set of requests;
- *Alerting derivation*: possible inconsistences between the selected values can be detected, which hint at the potential presence of unfeasible paths in the XAC-Graph.

Thus, not all XAC-Paths guarantee the generation of an XACML request due to the presence of two or more contradictory constraints within the same path. This peculiarity can be useful to highlight issues in the policy specifications and to improve the expressiveness of the policy itself.

The algorithm we conceived for the XACML requests generation purpose is reported in Algorithm 1. It receives the XACML policy as input and returns a set of XACML requests. Informally, the algorithm generates an XACML request for each feasible XAC-Path by satisfying all the constraints encoded in that XAC-Path.

Algorithm 1 XACML Requests Generation

```
1: input: XACMLPolicyP
2: output: XR                                                        ▷ A set of XACML requests
3: XR ← {}
4: XacTree ← createXacTree(P)
5: XacGraph ← createXacGraph(XacTree)
6: XacPaths ← createXacPaths(XacGraph)
7: Foreach p_i ∈ XacPaths do
8:     JM ← JaCoPConstraintModel(p_i)
9:                  ▷ Create a Java Constraint Model from the node of p_i containing XACML constraints
10:    xacRequest ← JaCoPConstraintSolver.solve(JM)
11:    if xacRequest ≠ NULL then
12:        XR.add(xacRequest)
13:    else
14:        infeasiblePath.add(p_i)
15:    end if
16: end for
17: return XR
```

Firstly, the algorithm derives an XAC-Tree (Algorithm 1, line 4). Then, staring from the XAC-Tree, it derives the corresponding XAC-Graph (Algorithm 1, line 5). Afterward, it derives an ordered set of XAC-Paths by applying the Coloring XAC-Graph and Unfolding XAC-Graph procedures introduced previously. In this step, (Algorithm 1, line 6) it also takes into account the combining algorithm defined in the XACML policy. The idea is to transform the computation of a XAC-Path into a constraint satisfaction problem and to use constraint solving techniques to generate an XACML request (Algorithm 1, lines 8-10).

A Constraint Satisfaction Problem (CSP) consists of a set of variables and a set of constraints over the values of these variables [16]. Each variable has a domain of possible values. The constraints define a set of restrictions over the possible variables values with respect to the already assigned ones. An assignment is a state of a CSP, in which some or all the variables have an assigned value. An assignment is called *consistent* if the assigned values of the variables do not violate any constraint. It is called *complete* if a value is assigned to every variable. An assignment consistent and complete is called a *solution*.

Among the existing techniques for solving CSPs [16], in the current implementation, we adopt the JaCoP (Java Constraint Programming) solver [139] (Algorithm 1, line 10) to create a CSP instance (Algorithm 1, line 8), determine the feasibility of XAC-Path and consequently generate XACML requests (Algorithm 1, line 10).

Finally, if a solution exists for the created CSP instance, the XAC-Path is considered feasible and the XACML request is derived by that solution (Algorithm 1, lines 11-15).

5.6.3 XACMET as PDP Oracle

The second feature provided by the XACMET approach is the derivation of a PDP oracle. Usually, random and combinatorial solutions can automatically generate the test requests, but do not provide their expected responses. By contrast, XACMET derives together with each test request its expected verdict, decreasing as a consequence the costs and risks of manual result inspection [117]. Given a generic request, the evaluation of an XACML policy with that request strictly depends on: the request values, the policy constraints as well as the combining algorithm that prioritizes the evaluation of the policy rules. Specifically, we define an *evaluation path* as the sequence of policy elements that are exercised by a generic request during its evaluation against an XACML policy and the final verdict associated to it. Thus, the general idea of the XACMET approach is to derive all possible evaluation paths from the policy specification and order them according to the rule combining algorithm. For instance, let us consider the policy of Figure 5.6, having as elements, the rules *ruleA* and *ruleB* and *deny-overrides* as the combining algorithm, the possible evaluation paths are:

1. *ruleA* evaluated to true and *ruleB* evaluated to false: the associated verdict is *Deny*, i.e., the effect of the first rule;

2. *ruleA* evaluated to false and *ruleB* evaluated to true: the associated verdict is *Permit*, i.e., the effect of the second rule;

3. *ruleA* and *ruleB* both evaluated to false: the associated verdict is *NotApplicable*;

4. *ruleA* and *ruleB* both evaluated to true: the associated verdict is *Deny*, because it takes the precedence regardless of the result of the second rule.

This set of paths is ordered according to the semantics of the rule combining algorithm, and then according to the verdict associated to each path. For instance, in case of *deny-overrides* combining algorithm, first the paths having *Deny* are evaluated, then those having *Permit*, and finally those having *NotApplicable*. For paths having the same verdict, the evaluation order of the paths is based on their length, namely the shortest path takes the precedence. For the policy of Figure 5.6, the order of the evaluated paths is (1), (4), (2) and (3).

The ordered set of paths is then used for the requests evaluation and the verdicts association. For each request, the first path for which all the path constraints are satisfied by the request values is identified and the final verdict associated to the request is derived.

The algorithm conceived for the oracle derivation is reported in Algorithm 2. Informally, starting from an XACML policy, the it derives a set of ordered XAC-Paths, each containing an expected result. Therefore, given a request, the algorithm is able to identify the first path satisfied by the request and then associates an expected result to it.

More precisely, Algorithm 2 takes as input an XACML policy *P* and a set of XACML requests $XR = \{xr_1, \ldots, xr_n\}$ and returns a set of expected results for these requests.

Firstly, it derives a set of XAC-Paths $XacPaths = \{xp_1, \ldots xp_k\}$, using: i) XAC-Tree (Algorithm 2, line 4); ii) XAC-Graph (Algorithm 2, line 5); and, iii) Coloring and Unfolding XAC-Graph procedures (Algorithm 2, line 6).

Algorithm 2 XACML Oracle

```
1: input: P, XR                       ▷ A set XACML Policy P and a set of XACML requests XR
2: output: ER                         ▷ A set of expected decisions
3: ER ← {}
4: XacTree ← createXacTree(P)
5: XacGraph ← createXacGraph(XacTree)
6: XacPaths ← createXacPaths(XacGraph)
7: Foreach xr_i ∈ SR do
8:     Foreach p_j ∈ XacPaths do
9:         JM ← JaCoPConstraintModel(p_j, xr_i)
10:            ▷ Create a Java Constraint Model from the node of p_j and values in xr_i
11:        result ← JaCoPConstraintSolver.solve(JM)
12:        if result = TRUE then
13:            xr_i.setDecision(p_j.DECISION)
14:            ER.add(p_j.DECISION)
15:            continue
16:        end if
17:    end for
18: end for
19: return ER
```

Then, for each request xr_i and for each XAC-Path xp_j, the algorithm derives a constraint model (Algorithm 2, line 9). This model takes into account both the constraints encoded within the constrained nodes of the path xp_j and the values contained in the request xr_i (Algorithm 2, line 9). This model represents a CSP instance. If a solution exists for this instance, the request xr_i satisfies the XAC-Path xp_j, and then the associated verdict is returned (Algorithm 2, lines 9-15).

For instance, by referring to the XACML policy reported in Figure 5.6, the XAC-Path of Figure 5.8(b) and a request having *resource* and *subject* values equal to *document* and *action* value equal to *write*, the algorithm firstly derives a constraint model including: i) the constraints of the boxes in the right side of Figure 5.8; and ii) the values of the request. Then, the algorithm solves the constraint model (Algorithm 2, line 9-11) and derives the final verdict for the request that in this case is *Deny*.

5.6.4 XACMET and Measuring Path Coverage

XACMET can be used to measure the path coverage of a generic set of requests. Indeed, each of the evaluation paths represents the set of constraints that should be satisfied by some specific request values so as to reach the final verdict. This information can be useful to improve the policy itself and avoid possible security flaws.

For instance, considering the ordered set of paths of policy of Listing 9.6, listed in Section 5.2.3 (i.e., (1), (4), (2) and (3)), a request asking to `write` a `documententry` does not match paths (2), (3) and (4), but covers (satisfies) path (1) since it satisfies only *ruleA*, so it reaches a path coverage equal to 25%.

In literature, coverage metrics are the most used approaches for evaluating the quality of a test suite [211]. In the access control context, among the different coverage criteria, the path coverage is one of the hardest to be adopted; identifying all the possible paths of an XACML policy has the same complexity of the definition of a policy evaluation engine. In case of XACMET, the paths identification is the basis of the XACMET proposal itself and consequently the path coverage is the easiest coverage metrics to be adopted.

5.7 Mutation Generation: GRADUATION (Contr. 5)

In this section, we focus on the fifth objective mentioned in the introduction of this chapter, i.e., providing a methodology for assessing the fault detection effectiveness of GDPR-based testing strategies. For this, we introduce here the GdpR-bAseD mUtA-TION (GRADUATION) methodology based on the mutation testing approach.

In the context of GDPR, and data privacy management in general, only few proposals are targeting the definition of mutation operators able to deal with the specific privacy characteristics and requirements of the privacy standards [26]. In these cases, the proposed mutation operators do not exhaustively cover all the important criticalities of the GDPR. For instance, they do not consider mutation operators concerning the erroneous use of the purpose defined by the controller and the consent given by the data subject.

In this section, we move a step ahead in this research direction by presenting the GRADUATION methodology, supported by a prototype tool, for:

1. analysing and managing model-based specifications of legal text (like the GDPR), so as to extract main concepts and useful data;

2. selecting and applying a set of mutation operators to a specific GDPR-based model instance, so as to derive its mutated versions;

3. selecting and executing a given test suite in order to compute its fault detection effectiveness;

4. analysing the obtained mutation results.

For better clarify the methodology application, we present the specialization of the GRADUATION in the context of GDPR-based authorization systems. Indeed, privacy legislation require organizations to deploy adequate fine-grained AC mechanisms that take into account additional legal requirements, such as the data usage purpose, user consent and the data retention period. Consequently, this rises up the problem of developing effective and efficient test strategies able to guarantee the lack of unauthorized access to personal data (*security perspective*) and unlawful processing (*legal perspective*).

In is important to notice that even if the specialization of the GRADUATION tool refers to the AC mechanisms based on the XACML specification [179], the GRADUATION methodology, and in particular its mutation operators set, is agnostic with respect to the AC mechanisms specification language, and can be applicable to any system that dealing with GDPR.

With the aim of providing a comprehensive assessment environment, the specialization of the GRADUATION tool presented in this section includes also all the XACML-based mutation operators available in literature. Thus, the XACML-based GRADUATION instance includes two available sets: 1) traditional XACML-based mutation operators [42], and 2) the new conceived operators based on the GDPR's peculiarities.

Summarizing the main contributions of this section are:

- a generic methodology, called GRADUATION for assessing the fault detection effectiveness of GDPR-based testing strategies by means of mutation testing.

- A set of mutation operators specifically based on a GDPR-based fault model.
- A preliminary implementation of GRADUATION methodology in the XACML context.

5.7.1 Related Works

Even if mutation testing is considered a fundamental step for evaluating the fault detection effectiveness of a test strategy, its application in the access control, and mainly in GDPR context, is still an open research issue. Indeed, considering access control testing the most noteworthy proposals are: the fault models and relative set of mutation operators simulating syntactic faults of XACML access control policies proposed by [162]; the generic metamodel for the specification rule-based security policy and the relative set of mutation operators provided by [170]; the XACMUT tool [42], which includes and enhances the mutation operators of [162] and [170] addressing specific faults of the XACML 2.0 language; and the proposal of [70] which implements mutation analysis at the level of the policy evaluation engine instead of applying it at the level of access control policy.

Considering the mutation testing in the context of GDPR, in the best of our knowledge the only proposal currently available is represented by [26]. Indeed, this paper is the first attempts of extending mutation operators for validating ontologies expressing GDPR provisions. However, even if generic, the mutation operators proposed in the paper do not cover all the specific aspects of the privacy standard.

Considering the available literature, our proposal from the one hand, extends the set of mutation operators, so as to validating the test suites or strategies against the GDPR peculiarities, from the other, provides an implementation able to integrating into a unique environment all the existing approach for mutation testing in the area of access control system.

5.7.2 Methodology for GDPR-based Mutants Derivation

GRADUATION methodology is composed of four main steps (see Figure 5.9): (1) Model Derivation; (2) Model Parsing; (3) Implementation Parsing; and (4) Mutation Application.

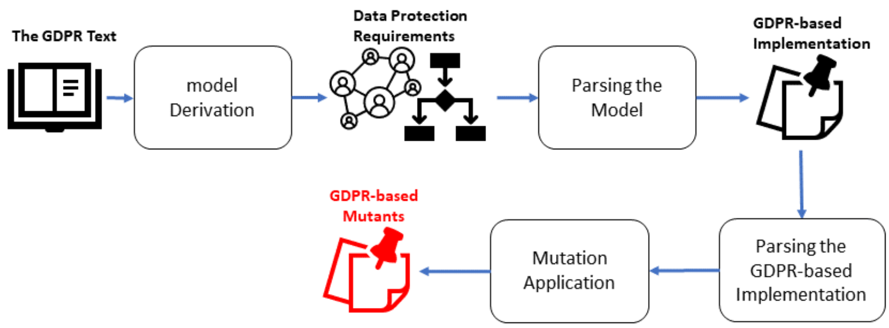


Figure 5.9: *GDPR-based Mutation Methodology.*

(1) *GDPR-based model derivation (Step 1)*: starting from a legal test, in this case the GDPR one, the model representing the main concepts and the relations between

them is obtained. To this purpose, in literature different proposals focused on the derivation of a formal representation of legal text are available [29, 147, 183, 184, 201]. It is out of the scope of this work investigating the most suitable approach for this purpose. The hypothesis of our work is that a GDPR-based model is available in terms a specification language, for instance an ontology, a UML model or an access control model. With the aim of clarifying the proposed methodology and without loosing in generality, we refer to the PrOnto [183, 184] ontology or RAccOnto (see Chapter 4) as examples of GDPR-based models. However, other proposals can be also considered.

(2) *Parsing of the GDPR-based model (Step 2)*: the GDPR-based model is parsed in order to identify the concrete entities and their input domain. These data are useful for the customized mutants generation. According to PrOnto ontology the GDPR entities can be classified as: (a) *Data* that is the object of the GDPR and it is target of its protection. Data can be: Personal Data, non-personal data, anonymized data and pseudonymised data; (b) *Agents and Roles* such as data subject, controller, processor, supervisory authority and the new introduced figure the DPO, as well as third-party; (c) *Processing activities* expressed as a set of actions such as delete, transmit and store; (d) *Purposes and legal bases* such as the consent; and finally, (e) *Legal rules* such as right, obligation, permission and prohibition.

(3) *Parsing the GDPR-based implementation (Step 3)*: According to the mutation testing approach, an original GDPR-based implementation, called *gold* implementation of the legal model is selected. The gold implementation is analyzed for: i) identifying the set of data entities, such for instance the current id of the Processor, the name of a Data Subject and so on; ii) instrumenting the gold implementation for the automatic derivation of its set of mutated versions.

(4) *Applying the mutation operators (Step 4)*: A set of GDPR-based mutation operators is applied to the gold implementation so as to derive the mutants set. In this step two kinds of mutations are considered: *intra-implementation* and *inter-implementation mutations*. The former set refers to the application of mutation operators managing only the information and data extracted from the the gold implementation (i.e. during the step named *Parsing the GDPR-based implementation*. The latter set refers to mutation operator managing the information relative to the GDPR-based model (i.e. derived during the step named *Parsing of the GDPR-based model*. The conceived GDPR-based mutation operators are reported in Section 5.7.3.

5.7.3 GDPR-based Mutation Operators

The GDPR-based mutation operators can be classified in three main categories: i) operators targeting the purpose of processing and the consent given by a data subject; ii) mutation operators targeting the roles defined in the GDPR such as Data Subject, Controller and Processor; iii) and finally, operators focusing on Personal Data, i.e., the object of the EU legal framework, and their categories.

These operators have the ability to be applied to different domains, because voluntarily conceived as generic. Therefore, depending on the specific language or formalism used for defining the GDPR's requirements, they can be implemented and applied accordingly.

The generic GDPR-based mutation operators are as the following:

Giving Consent (GC) this operator changes the value of the Consent given by the data subject.

Withdraw Consent (WC) this operator is dual to GC, and it changes the value of the consent element in the targeted implementation.

Change Purpose (CP) this operator replaces a purpose with other defined in the considered implementation. In case there is only one purpose, CP operator changes the purpose with a random one defined in the GDPR model or in other available supporting sources;

Change Controller (CC) this operator replaces a Controller with another. In case missing candidates, CC changes the current Controller with a randomly generated Controller. This operator is applied also when Joint Controllers exit and involved in the processing of Personal Data, i.e., in defining the Purpose of processing, obtaining the consent and using Personal Data accordingly;

Replace Data Subject (RDS) this operator is able to replace a Data Subject with another. Similar to CC and CP operators, i.e., in case of missing candidates, RDS choose random Data Subject that replaces the current one;

Replace Controller with Processor (RCP) this operator changes a Controller with a Processor presented in the current implementation;

Replace Processor (RP) this operator replaces a Processor with another Processor.

Change Personal Data (CPD) this operator is able to change a personal data with another.

Change Personal Data Category (CPDC) this operator changes the category of of given personal data with another.

5.7.4 GRADUATION Tool

In this section we describe the contextualization of the GRADUATION methodology in the context of XACML-based access control systems. By referring to the methodology previously presented, we provide here a reference architecture and a preliminary implementation of the proposed steps, focusing in particular on step 4 named *Applying the mutation operators*, which is the one that need to be specialized more in the XACML context.

In Figure 5.10 the customized GRADUATION architecture is depicted. In particular, Box A represents the first three steps of the GRADUATION methodology. In this case the GDPR-based model is represented through a RAccOnto ontology, while the GDPR-based implementation is a GDPR-based access control policy written in the XACML language.

Box B instantiates Step 4 of the proposed methodology in the XACML-based access control context, and it is composed of the following components:

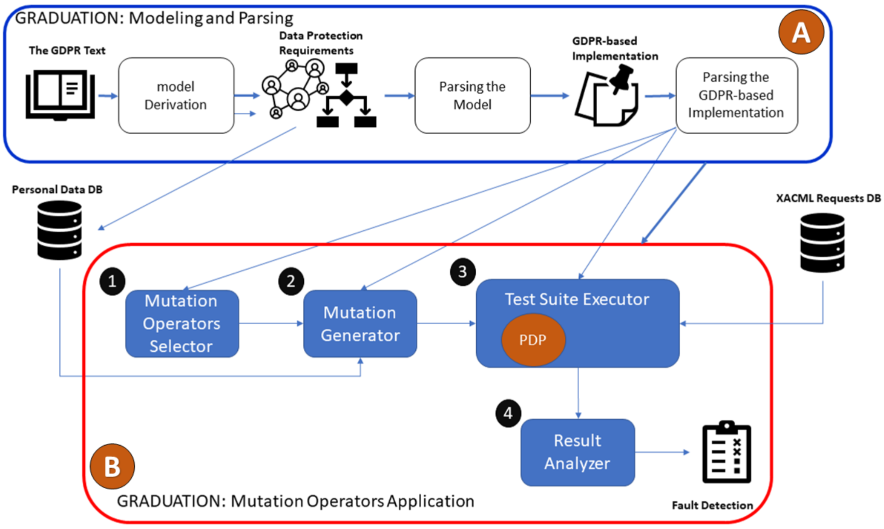


Figure 5.10: *Overview of GRADUATION.*

- *Mutants Operators Selector*: As mentioned in the introduction of this section the current implementation of GRADUATION methodology contains two set of mutation operators: Standard XACML Mutation Operators, and the GDPR-based Mutation Operators customized for XACML domain. More details about the two proposed set of operators are provided here below in this section.

- *Mutation Generator*: this components has the responsibility of generating mutated versions of the Gold (GDPR-based) XACML policy by applying the selected mutation operators (both standard and GDPR-based) by end-user.

- *Test Suite Executor*: this component executes the XACML requests provided by the user on the original XACML policy (Gold Policy) and on the generated set of mutated policies. For requests evaluation this component integrates an XACML-based PDP engine[2], which is able to provide the corresponding result (Permit, Deny, NotApplicable or Indeterminate) for a given policy *P* and a request *Req*.

- *Results Analyzer*: this component takes as input all the results obtained by the execution of the test suite on the original XACML policy and on its set of mutants and computes the fault detection effectiveness. It works as follows: for each request the result obtained by its execution on the original XACML policy is compared with those obtained on its mutants set. If the results are different, the mutant is classified as killed. The component provides as output the list of killed mutants, survived mutants, and the percentage of fault detection effectiveness obtained by the requests execution. It also provides functionalities allowing to filter by mutation operators, by test cases (i.e., the XACML requests), and by the expected authorization decision. This is useful for providing different perspective of the data and for analyzing deeply the different aspects of these mutation data views.

[2]There are different open-source implementations of the PDP available as such: Sun PDP (http://sunxacml.sourceforge.net/), HERAS-AF (https://bitbucket.org/herasaf/herasaf-xacml-core/) and Balana (https://github.com/wso2/balana).

Standard XACML Mutation Operators. The current standard mutation operators can be categorized based on the XACML elements they emulate potential faults. There are operators emulating fault at: 1) PoliySet element level such as Policy Set Target True (PSTT), Policy Set Target False (PSTF) and Change Policy Combining Algorithm (CPC); 2) Policy element level, e.g., Change Rule Combining Algorithm (CRC) and Policy Target False (PTF); 3) Rule element level, such as Rule Target True (RTT), Rule Condition False (RCF) and Change Rule Effect (CRE); and finally, 4) XACML Functions level, for instance RemoveUniquenessFunction (RUF), ChangeLogicalFunction (CLF) and AddNotFunction (ANF). For a more detailed description and comprehensive overview of the standard XACML-based mutation operators, we refer the reader to [42].

5.7.5 Using GRADUATION Tool

In the following, we briefly detail the application of the GRADUATION methodology by considering a simple use case scenario taken from fitness environment: Alice, a Data Subject, wants to use a smart fitness application to monitor her daily activities to achieve a predefined training objective. In this case we suppose that a customized (mobile) application is provided by a generic myFitness company (Controller). To meet Alice's needs, myFitness has so far defined two purposes (MyCholesterol and Untargeted Marketing), each related to a specific data set of Personal Data and achieved by allowing access to perform a specific set of Actions. Specifically, the MyCholesterol purpose is achieved by performing AGGREGATE, DERIVE and QUERY actions; whereas the Untargeted Marketing purpose is achieved by performing COLLECT, QUERY and SEND actions At the time of subscribing to the myFitness application, Alice provided her personal data (i.e., e-mail, Age, Gender, and Blood Cholesterol) and gave her consent to process her e-mail and Age for Untargeted Marketing purpose, and her Blood Cholesterol for MyCholesterol purpose. Additionally, Alice withhold her consent to share her personal data with a third-party company named xxx-HealthOrg company. In turn, myFitness gave to Alice controller's contacts that include: piiController, orgName, address, e-mail, and phone number.

Based on the above scenario, the following access control policy, that allows a lawfulness of processing of personal data related to Alice, can be defined:

LawfulnessOfProcessingPolicy:

R1: permission(Controller=myFitness, DataSubject=Alice PersonalData=Blood Cholesterol, purpose=MyCholesterol, Action=DERIVE Consent=TRUE)

R2: permission(Controller=myFitness, DataSubject=Alice, PersonalData=Email, purpose=UntargetedMarketing, Action=SEND Consent=TRUE)

According to the GRADUATION methodology, during the first three steps of the methodology, i.e., during the activities included in the box A of Figure 5.10, customized data o are retrieved from the access control policy and the RAccOnto representation of the GDPR.

In Table 5.6, the result of these activities is represented. In particular, the table reports the set of GDPR-based entities (column *GDPR Entity*), their classification (column *Category*), their names (column *Name*) and related values (column *Value*).

Without going into the details of the XACML language, we describe in the following

GDPR Entity	Category	Name	Value
Controller	Agent	orgName	myFitness
Controller	Biodata	piiController	myFitnessID
Controller	Biodata	address	-
Controller	Biodata	e-mailC	-
Controller	Biodata	phone number	-
Third-party	Agent	orgName	xxx-HealthOrg
Data Subject	Agent	DSName	Alice
Personal Data	Biodata	Age	-
Personal Data	Biodata	Gender	-
Personal Data	Biodata	Blood Cholesterol	-
Personal Data	Biodata	e-mailDS	-
Purpose	-	Purpose	MyCholesterol
Purpose	-	Purpose	Untargeted Marketing
Processing	-	Action	AGREGATE
Processing	-	Action	DERIVE
Processing	-	Action	QUERY
Processing	-	Action	COLLECT
Processing	-	Action	SEND

Table 5.6: *GDPR Entities Extracted from the Model.*

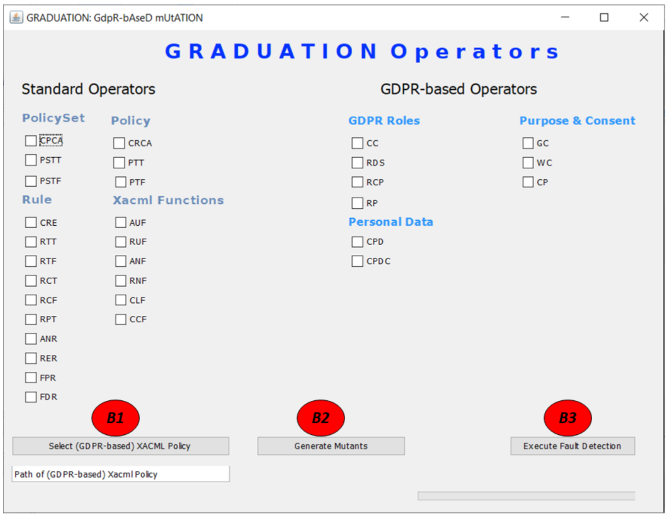


Figure 5.11: *GRADUATION Main GUI.*

the application of the mutation operators defined in GRADUATION by considering the above *LawfulnessOfProcessingPolicy*, i.e., the derived Alice's access control policy.

In the proposed instantiation of the GRADUATION methodology, the end-user interaction is managed through an User Interface (UI) as depicted in Figure 5.11. Through this interface, the end-user can: (1) select the GDPR-based XACML policy (button **B1** in Figure 5.11); (2) select the GDPR-based mutation operators and the standard ones, and apply them to the selected policy (button **B2**); (3) execute the policy mutants against a given test suite (button **B3**). In particular, this step involves the selection of the set of XACML requests that will be evaluated; the execution of the policy and the derived mutants against test suite; and the evaluation of which mutants (both standard and GDPR-based) have been killed by the application of the selected test suite.

Considering the execution of the selection of GDPR-based XACML policy (button **B1**) and the application of GDPR-based mutation operators (button **B2**) here below some example of mutants related to *LawfulnessOfProcessingPolicy* are reported. In particular in bold-italics text we report the name of the Mutation operator applied while in bold-blue we highlited the mutation applied within R1 and R2 rules.

WC MUTANT
LawfulnessOfProcessingPolicy-WC1:

R1: permission(Controller=myFitness, DataSubject=Alice PersonalData=Blood Cholesterol, purpose=MyCholesterol, Action=DERIVE **Consent=FALSE**)

R2: permission(Controller=myFitness, DataSubject=Alice, PersonalData=Email, purpose=UntargetedMarketing, Action=SEND Consent=YES)

CP MUTANT
LawfulnessOfProcessingPolicy-CP2:

R1: permission(Controller=myFitness, DataSubject=Alice PersonalData=Blood Cholesterol, **purpose=UntargetedMarketing**, Action=DERIVE Consent=TRUE)

R2: permission(Controller=myFitness, DataSubject=Alice, PersonalData=Email, purpose=UntargetedMarketing, Action=SEND Consent=TRUE)

CPD MUTANT
LawfulnessOfProcessingPolicy-CPD:

R1: permission(Controller=myFitness, DataSubject=Alice **PersonalData=AGE**, purpose=MyCholesterol, Action=DERIVE Consent=TRUE)

R2: permission(Controller=myFitness, DataSubject=Alice, PersonalData=Email, purpose=UntargetedMarketing, Action=SEND Consent=YES)

Remark. In this section we have introduced GRADUATION, a comprehensive methodology for defining and applying mutation operators specifically conceived in the context for the GDPR. The methodology and the proposed mutation operators have been voluntarily conceived independent for any modeling language, used for formally represent the GDPR. The applicability of the GRADUATION has been exemplified in the context of XACML-based access control domain. Thus, the XACML-based GRADUATION implementation has been used to generate mutated versions of GDPR-based access control policies expressed in XACML formalism. The current version of the proposal is in advanced implementation stage, and currently we are working to extend the GDPR-based mutation operators set so as to cover other GDPR's demands as well as to improve its validation with real case studies. Ongoing work includes also the specialization of the GRADUATION methodology considering other formalism and languages such as UML and Semantic Web Technologies.

5.8 Execution & Result Analysis: A Controlled Experiment (Contr. 6)

The aim of this section is to target the sixth testing objective presented in the introduction of this chapter, i.e., provide details of how to conducting a CE experiment in the context of AC. In particular, we focus on the definition of the experiment for the comparison of two testing strategies, instrumentation and execution of the experiment, and analysis of the results.

For this we consider two generation strategies: GROOT and XACMET [66]. We use the Goal Question Metric (GQM) template [35] to formalize the goal of the experiment and define three metrics:

1. effectiveness,
2. size and

3. average percentage faults detected (APFD).

Moreover, as infrastructure for performing the controlled experiment, we leverages the XACML Mutation Framework (XMF) [72], previously presented in Section 5.3, that allows for replicability of the experiment as well as generalization of results, finding aggregating, and finally reducing of experimentation costs.

Therefore, the main contribution of the section is the formal definition of a Controlled Experiment within the access control domain. Indeed, according to [128, 235], we present the thee main steps:

- the definition of a controlled experiment for the evaluation of two test cases generation strategies;
- the instrumentation and execution of the experiment;
- the analysis of the results.

5.8.1 Experiments Definition and Planning

The controlled experiment definition consists of three main steps:

1. defining the goal of the experiment, the research questions and the associated Hypotheses that have to be formally tested as in Section 5.8.1;
2. introducing the context of the experiment as well as the variables, the subjects and the object of the experiment as in Section 5.8.1;
3. designing and instrumenting the experiment, i.e., the realization of the means for performing the experiment and monitoring it, without affecting the control of the experiment as in Section 5.8.1.

Goal and Hypotheses Formulation

According to the Goal Question Metric (GQM) template [35], our research goal is as follows:

> ***Analyze*** two Test Generation Strategies (TGS_1 and TGS_2) ***for the purpose of*** evaluation ***with respect to their*** Effectiveness, Size and APFD of test suite produced ***from the point of view of the*** researcher ***in the context of*** XACML policy decision point testing.

In order to address the goal of our experiment, we defined three research questions and their associated hypotheses:

- **RQ1 Effectiveness**: How much does the quality of a test suite produced by $Strategy_1$ (TGS_1) differ from the quality of test suite produced by $Strategy_2$ (TGS_2) in terms of Effectiveness, i.e., the mutation score? For evaluating the effectiveness of the strategies we consider full test suites derived from each strategy;
- **RQ2 Size**: How much does the cost of a test suite produced by $Strategy_1$ differ from the cost of test suite produced by $Strategy_2$ in terms of Size, i.e., the number of test cases? For evaluating the cost of the strategies in terms of the number of XACML requests generated, we assume that all requests have the same cost in terms of generation and evaluation as well as verdict verification;

- **RQ3 APFD**: How much does the Average Percentage Faults Detected (APFD) of a test suite produced by $Strategy_1$ differ from the APFD of test suite produced by $Strategy_2$?

To answer the above research questions, the following Null Hypotheses have been defined:

- $H_{0Eff} : \mu_{EffTGS_1} = \mu_{EffTGS_2}$ the $Strategy_1$ finds on average the same number of faults, i.e., the effectiveness, as the $Strategy_2$, where μ denotes the average percentage of the killed mutants using the complete test suites generated by the two strategies;
- $H_{0Size} : \mu_{SizeTGS_1} = \mu_{SizeTGS_2}$ the size of test suite *is equal* for $Strategy_1$ and $Strategy_2$;
- $H_{0APFD} : \mu_{APFDTGS_1} = \mu_{APFDTGS_2}$ the average APFD *is equal* for $Strategy_1$ and $Strategy_2$.

A null hypothesis states that there are no real underlying trends or patterns in the experiment setting; the only reasons for differences in the observations are coincidental. This is the hypothesis that we want to reject with a higher significance. When the null hypothesis can be rejected with relatively high confidence, it is possible to formulate an alternative hypothesis, as follows:

- $H_{1Eff} = \neg H_{0Eff}$;
- $H_{1Size} = \neg H_{0Size}$;
- $H_{1APFD} = \neg H_{0APFD}$.

Context, Variables and Subjects Selection

In the context of Access Control Systems (ACSs), the aim of our controlled experiment is the evaluation of test cases generation strategies by means of the mutation analysis at the level of the Java based PDP engine. The comparison involves the effectiveness of the test suite generated by each strategy, the cost associated to each test suite in terms of its size and the velocity at which that effectiveness is reached. According to the classification of the experiment context in [235], the comparison of the selected test strategies has been conducted through a *Multi-test within object study*, i.e., a kind of controlled experiment that examines a single object across a set of subjects. In this experiment the object is the PDP engine, while the subjects are the XACML policies.

The variables involved in the experiment are: one ***Independent variable***, i.e., the *test case generation strategy* with two levels or alternatives (treatments) for the main factor (i.e., Level 1 and Level 2) and three ***Dependent Variables***, i.e., the *Effectiveness*, the *Size* (or the *cost*) of the test suites and the *APFD* metrics.
The ***Object of the experiment*** is the Policy Decision Point developed by Sun, Sun-PDP [223].

According to [128], the ***Parameter***, that could influence the result of the experiment or, alternatively, the response variable, is the *Mutation Generator* tool or strategy used to generate the mutated versions of the PDP. Among the currently available tools, we selected the μJava tool [154, 209], but other solutions could be selected, such as Javalanche [208], Major [127] or Judy [155].

Finally, we defined the ***Subjects*** involved in the experiment, i.e., the XACML Policies. According to the recommendation in [128], the selection is closely connected to the generalization of the results from the experiment, thus the selection must be representative for that population.

Table 5.7: *XACML Policies Subjects.*

Xacml Policy	Functionality					
	# Rule	# Cond	# Sub	# Res	# Act	# Funct
2_73020419964_2	6	5	3	3	0	4
create-document-policy	3	2	1	2	1	3
demo-5	3	2	2	3	2	4
demo-11	3	2	2	3	1	5
demo-26	2	1	1	3	1	4
read-document-policy	4	3	2	4	1	3
read-informationunit-policy	2	1	0	2	1	2
read-patient-policy	4	3	2	4	1	3
Xacml-Nottingham-Policy-1	3	0	24	3	3	2

For this, we considered a set of real-world XACML policies taken from real contexts and European projects as summarized in Table 5.7. In particular, the columns represent the number of rules, conditions, subjects, resources, actions and distinct functions within each policy. As in the table, policies named demo-5, demo-11 and demo-26 have been taken from the Open Source repository software Fedora (Flexible Extensible Digital Object Repository Architecture) [93] for controlling the access to the administered digital contents; the remaining six are those released by the TAS3 European project [224].

The GQM template, considering the context of Access Control, is as the following:

> ***Analyze*** GROOT and XACMET Strategies ***for the purpose of*** evaluation ***with respect to their*** effectiveness and size of test suite produced ***from the point of view of the*** researcher ***in the context of*** XACML policy decision point testing.

Experiment Design and Instrumentation

The comparison between the performances of the different test strategies has been defined using the Paired Comparison design, a particular kind of one factor with two treatments [235]. In this design, each subject uses both treatments on the same object. In the context of this work it can be translated into: both test strategies (GROOT and XACMET) have to be applied to each XACML policy and both the obtained test suites evaluated using the SunPDP and its mutants.

According to [235], an important step for the controlled experiment is the instrumentation, i.e., the realization of the means for performing the experiment and monitoring it. In particular, the instruments considered are of three types, namely objects, guidelines and measurement instruments.

Object. When planning for an experiment, it is important to choose objects that are appropriate. The object of our experiment is an XACML-based PDP named Sun PDP [223], which is an open source implementation of the OASIS XACML standard, written in Java. We decided on Sun's PDP engine because it is currently one of the most mature and widely used engine for XACML policy evaluation, which provides complete support for all the mandatory features of XACML 2.0 as well as a number of optional features. This engine supports also all the standard attribute types, functions and combining algorithms and includes APIs for adding new functionalities as needed. The Sun PDP source code is broken into ten packages: seven packages include the core implementation, two packages include classes used for the configuration code, rarely used by programmers, and one package contains test code samples.

The comparison of the selected test strategies required the definition of the SunPDP Mutants. Therefore, through Mutation Generator and Mutation Integrator components, mutation operators (both class-level and method-level operators) have been applied to the SunPDP code and executable mutant versions derived.

Guidelines and Measurement. Guidelines are needed to guide the participants in the experiment and they include process descriptions and checklists. Measurements are conducted via data collection that in human-intensive experiments are generally performed by manual forms or interviews. Since we conducted a technology-oriented experiment, we embedded both aspects in the automation process. Specifically, we leveraged the functionalities of XMF framework, i.e., test cases generation, mutants generation, test cases execution and results analysis, to perform our controlled experiment and automatically collect the data.

5.8.2 Experiment Operation

The experiment operation mainly consists of three steps: preparation, execution and data validation. Specifically, the preparation step focuses on the preparation of the subjects, object and the material needed.

In our experiment, the preparation step consists of: XACML policy selection, XACML requests generation, XACML based PDP selection, and finally the mutants generator selection.

During the execution step, the XACML requests are evaluated and obtained data are collected.

Finally, during the data validation step, the dependent variables are calculated, the collected data are managed and analyzed so to provide a valid picture of the experiment. In the following sections, we report the results of our data validation.

Data Validation

The data validation phase includes:

1. i) a former descriptive statistics, which allows the visualization of the information using informal representations. In our experiment, these statistics span from the number of generated requests, to the distribution of the mutants on the different Java PDP classes, to the evaluation of the different test case executions. Even

if simple and informal, these descriptive statistics let both to highlight an important criticality on the data set considered, that could have compromised the entire experiment, and to find out the corrective actions for solving it. For the sake of brevity, in the following we report only the number of executions and the number of distinct mutants.

2. ii) a latter formal description of the experiment results. In our experiment we focused on the Paired T-Test with Null Hypothesis and present the results in next subsections.

Number of Executions. Table 5.8 reports for each of the nine XACML policies, the number of executions for the test suites derived by the XACMET and the GROOT test strategies (second and third column respectively). The total amount of executions is reported in the last column of the table. From the data collected, as reported in the last row, only 16% of the executions is performed by XACMET strategy, meaning that testing cost of such strategy is very low compared to the one of the GROOT strategy.

Table 5.8: *Number of Executions by XACML Policy and Strategy.*

Xacml Policy	# of Executions		All
	XACMET	GROOT	
2_73020419964_2	12320	481800	494120
create-document-policy	13560	96360	109920
demo-11	104390	321200	425590
demo-26	64240	128480	192720
demo-5	128480	674520	803000
read-document-policy	4014	240900	244914
read-informationunit-policy	27752	48180	75932
read-patient-policy	48180	240900	289080
Xacml-Nottingham-Policy-1	38892	55072	93964
All	441828	2287412	2729240

For the aim of completeness, for each XACML policy, Figure 5.12 reports the percentage of executions for the two testing strategies considered. In particular, the blue bars (black in black and white printing) refer to the XACMET testing strategy, while the orange bars (light gray in black and white printing) report the percentage of executions of GROOT strategy.

Number of distinct Mutants. The information about the number of executions can be analyzed also from the point of view of how many distinct mutated PDPs are evaluated by each XACML policy. In particular, Table 5.9 reports, in the second column, the number of distinct mutated PDPs considering the XACMET strategy, while in the third column those related to the GROOT ones are presented. By analyzing the data in the table only four of the nine XACML policies are evaluated by all mutated PDPs set. Specifically, for the first, second, sixth and seventh policies listed in the Table, only the test suite derived by the GROOT strategy is able to execute all the mutated PDPs; for the last policies neither of the two test suites is able to execute all the mutated PDPs.

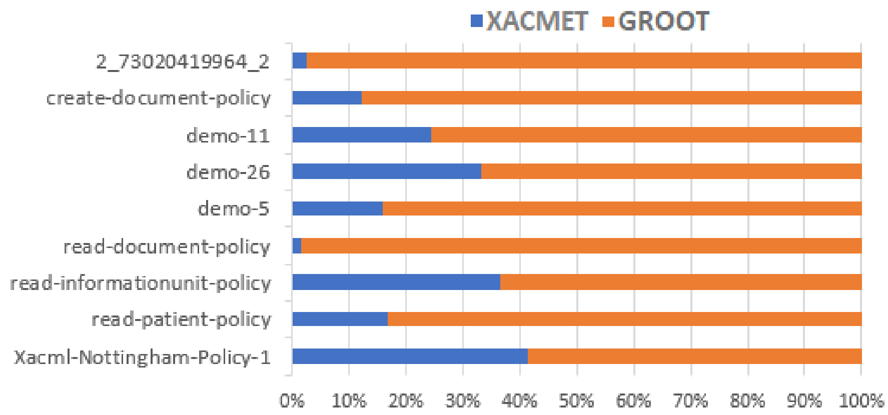


Figure 5.12: *% of Executions by XACML policy, by XACMET and GROOT strategy.*

Table 5.9: *Number of Distinct MutatedPDP Evaluated by XACML Policy and Strategy.*

Xacml Policy	# of Distinct MutatedPDP		All
	XACMET	GROOT	
2_73020419964_2	1540	8030	9570
create-document-policy	2712	8030	10742
demo-11	8030	8030	16060
demo-26	8030	8030	16060
demo-5	8030	8030	16060
read-document-policy	669	8030	8699
read-informationunit-policy	6938	8030	14968
read-patient-policy	8030	8030	16060
Xacml-Nottingham-Policy-1	1389	3442	4831
All	45368	67682	113050

The results reported in this table highlighted an important criticality for the evaluation of the controlled experiment metrics (Effectiveness and APFD). Indeed, because there is a difference in the number of distinct mutated PDPs for the two test suites, the Hypotheses testing could be invalidated and consequently also the answers to the target RQs. For a fair experiment and evaluation, it is important to guarantee that both strategies are evaluated using the same set of mutated PDPs. Therefore, the data sets have been reduced considering only the minimal common set of mutated PDPs for each test strategy.

As natural consequence such reduction has also an impact on the number of executions for the test suites derived by the XACMET and the GROOT strategies as reported in Table 5.10. The comparison of the data of this table with those reported in the previous one (Table 5.8) shows that the reduction has the biggest impact on the GROOT strategy. For the sake of completeness, we report, in Table 5.11, the percentage of reduction of executions associated to each XACML policy.

5.8.3 Results: Effectiveness (RQ 1)

According to [128, 235], we applied the Paired T-Test to formally verify the Null Hypothesis with the confidence level of 95%. This choice was a natural consequence of the

Table 5.10: *Number of Reduced Executions by XACML Policy and Strategy.*

Xacml Policy	# of Executions		All
	XACMET	GROOT	
2_73020419964_2	12320	92400	104720
create-document-policy	13560	32544	46104
demo-11	104390	321200	425590
demo-26	64240	128480	192720
demo-5	128480	674520	803000
read-document-policy	4014	20070	24084
read-informationunit-policy	27752	41628	69380
read-patient-policy	48180	240900	289080
Xacml-Nottingham-Policy-1	38892	22224	61116
All	441828	1573966	2015794

Table 5.11: *% of Reduced Executions by XACML Policy and Strategy.*

Xacml Policy	# of Distinct MutatedPDP		% of Reduction	
	XACMET	GROOT	GROOT	All
2_73020419964_2	1540	1540	81%	79%
create-document-policy	2712	2712	66%	58%
demo-11	8030	8030	0%	0%
demo-26	8030	8030	0%	0%
demo-5	8030	8030	0%	0%
read-document-policy	669	669	92%	90%
read-informationunit-policy	6938	6938	14%	9%
read-patient-policy	8030	8030	0%	0%
Xacml-Nottingham-Policy-1	1389	1389	60%	35%
ALL	45368	45368	31%	26%

type of design adopted, i.e, the paired comparison. Following the standard best practices, we decided to accept a probability of 5% of committing a Type-1-Error [235], i.e., the Null Hypothesis is rejected if the computed p-value is less or equal to 0,05 ($alpha = 0.05$). Stating from the collected data, therefore, we generated the necessary sample data so as to test each Null Hypothesis formulated in Section 5.8.1.

In this subsection, we illustrate results related with RQ 1, whereas those related with RQ 2 and RQ 3 are discussed in the next subsections.

As presented in Section 5.8.1 the aim of RQ 1 is to ***Analyze*** GROOT and XACMET Strategies ***for the purpose of*** evaluation ***with respect to their*** test suite effectiveness ***from the point of view of the*** researcher ***in the context of*** XACML PDP testing without constraints (i.e., considering the whole test suite generated).

A general attribute for the evaluation of the quality of a test cases generation strategy is its effectiveness, defined in terms of number (or percentage) of mutated PDPs killed. Therefore the effectiveness is calculated as:

$$Effectiveness = \frac{\#mutatedPDPsKilled}{\#mutatedPDPs}$$

It is important to remark that in order to correctly compute this measure the number of distinct killed mutants by each strategy and on each policy must be considered. The samples relative to the effectiveness of GROOT and XACMET test strategies for Null Hypothesis testing (H_{0Eff}) are reported in Table 5.12 (columns 2 and 3). In particular, the upper part of the table reports the samples associated to each test cases generation strategy, while the lower part reports some statistical information of each sample. Both the strategies have a similar behaviour: the test suites are able to kill less than 20% of the mutated PDPs.

Table 5.12: *RQ 1: Effectiveness and RQ 2: Size*

Xacml Policies Subjects	Effectiveness		Size	
	GROOT	XACMET	GROOT	XACMET
2_73020419964_2	2,14	13,57	60	9
create-document-policy	15,30	14,93	16	5
demo-11	8,78	8,89	40	13
demo-26	7,67	8,97	16	8
demo-5	9,18	9,14	84	16
read-document-policy	19,88	19,43	30	6
read-informationunit-policy	8,59	8,91	6	4
read-patient-policy	7,68	8,79	30	6
Xacml-Nottingham-Policy-1	19,65	19,65	16	18
	Samples Statistics			
N	9	9	9	9
Missing Count	0	0	0	0
Mean	10,987	12,4766	33,1111	9,4444
Standard Deviation	5,988	4,6064	24,9822	5,0525
Standard Error Mean	1,996	1,5355	8,3274	1,6841

This informal observation is confirmed also by the results of the Paired T-test associated to H_{0Eff} (Table 5.13, column 2).

Because the *p-value obtained* is $0,2705 > 0.05$, from the considered sample there is no difference from the point of view of effectiveness between the two test strategies.

For sure, the low values of fault detection effectiveness obtained in this experiment are not encouraging for any kind of test strategy. However, from a deeper analysis we highlighted the two following main reasons.

First, the choice of XACML policies: these are real ones, therefore they contain the mostly used functionalities and constructs. They are not artificially developed for testing objective; therefore, they are not a complete representation of the XACML policy population. However, here the target is to asses the effectiveness of the test strategy on the few functions, data types, XACML elements that are currently used in the practice, so to focus as much as possible the testing activity on the most critical aspects.

The second reason concerns the mutation operators adopted. The operators implemented in the framework for the generation of mutated PDPs are the standard ones and are applicable to any kind of Java program. They do not consider the peculiarities of the XACML language and therefore could not be targeted by the XACML requests used as a tests input.

5.8.4 Results: Size (RQ 2)

As presented in section 5.8.1, the aim of the RQ 2 is to ***Analyze*** GROOT and XACMET Strategies ***for the purpose of*** evaluation ***with respect to their*** cost in terms of number of test cases generated ***from the point of view of the*** researcher ***in the context of*** budget programming.

An important aspect for the selection of a test cases generation strategy is its cost evaluated in terms of: the time for the test cases execution and the time necessary for the debugging activity. Supposing that each test case has potentially the same impact on the overall testing effort, the execution time becomes directly connected with the number of test cases executed: i.e., the size of a test suite represents also its cost.

The samples relative to the sizes of GROOT and XACMET test strategies for Null Hypothesis testing (H_{0Size}) are reported in Table 5.12 (columns 4 and 5). In particular, the upper part of the table reports the samples associated to each test cases generation strategies, while the lower part reports some statistical information of each sample.

Table 5.13: *Paired T-Test: RQ 1 (Effectiveness) and RQ 2(Size)*

RQs	RQ 1	RQ 2
Label	GROOT- XACMET	
t	-1,1838	2,7426
df	8	8
p-value (2-tailed)	**0,2705**	**0,0253**
Mean	-1,4896	22,6667
Standard Deviation	3,7749	24,7942
Standard Error Mean	1,2583	8,2647
CI (Lower Bound)	-4,3913	3,6082
CI (Upper Bound)	1,412	41,7251

Except for *Xacml-Nottingham-Policy-1* policy, XACMET strategy generates smaller test suites with respect to those generated by GROOT. Considering the size metric, the test suites of XACMET cost about 70% less than those of GROOT one, but reaching the same quality in terms of errors found or mutants killed.

This result is formally confirmed by the Paired T-test associated to H_{0Size} and illustrated in Table 5.13, column 3. The *p-value obtained* ($0,0253 < 0.05$) suggests, rejecting the Null Hypothesis H_{0Size}, that there is no difference from the point of view of the size between the two test strategies. Therefore, we can conclude that the XACMET strategy outperformed GROOT strategy in terms of cost.

5.8.5 Results: APFD (RQ 3)

As presented in section 5.8.1 the aim of RQ 3 is to ***Analyze*** GROOT and XACMET Strategies ***for the purpose of*** evaluation ***with respect to their*** effectiveness in terms of APFD ***from the point of view of the*** researcher ***in the context of*** interruption of XACML PDP testing activity.

For unexpected time constraints or budgets reduction reasons, the testing activity could be interrupted before its overall completion. In this case, not all the test case could be executed with a consequent risk for the final quality and efficiency of the product developed.

In this case, the standard metric adopted for measuring such a risk is the APFD metric, which is calculated by taking the weighted average of the percentage of faults detected during the execution of the test suite. It is formally defined as follows [88]:

$$APFD = \frac{\sum_{i=1}^{n-1} F_i}{n \times l} + \frac{1}{2n}$$

where, n is the number of test cases in the test suite T, l is the number of faults, and F_i is the number of faults detected by at least one test case among the first i test cases in T.

By construction, APFD values range from 0 to 1; higher values imply faster (better) fault detection rates. Thus, APFD is commonly used to evaluate prioritization techniques, because it is able to estimate the speed with which an ordered test suite can reach the maximum number of discovered faults. In other words, it is the measure of how quickly the faults are detected by a testing strategy. Improving the rate of fault detection can have an impact on the testing cost and effort: software engineers may locate and correcting faults earlier in advance and better evaluate the risk of test activity interruption [44, 88].

Since the APFD measure is connected to the concept of (ordered) test suites having the same cardinality, its application in this experiment requires that the number of XACML requests generated by the GROOT and XACMET test strategy is the same. For this, the following corrective actions have been applied:

Test suite size and requests selection the cardinality *N* of the test suite has been fixed to the minimum one between the values of the GROOT and XACMET test strategy and the *N* test cases randomly selected among those available.

Prioritize the test cases each reduced test suite has been ordered in terms of mutated PDPs killed, so to assure the optimal fault detection effectiveness of each set.

APFD calculation for each ordered test suite the APFD has been calculated.

Considering in detail the *Prioritize the test cases* action, the technique used, called mutation-based heuristic, is a greedy-optimal selection of test cases computed on the mutation coverage. The heuristic is a sub-optimal algorithm, since it orders the test cases according to the cumulative number of different mutants killed. For implementing the selected heuristic, the mutated PDPs killed by each test case have been organized into a matrix (request, mutated PDPs) where each cell(i, j) is equal to 1 if the i-th request kills the j-th mutated PDP, and 0 otherwise.

Informally, the algorithm for implementing the mutation-based heuristic works as follows:

1. it calculates the number of killed PDPs for each request;

2. it selects the request s.t. the number of killed PDPs is maximum;

3. it removes the request and the mutated PDPs killed by such request from the matrix;

4. it repeats the steps 1-3 until all requests are selected.

Table 5.14: *RQ 3: APFD*

Xacml Policies Subjects	APFD	
	GROOT	XACMET
2_73020419964_2	0,932588598	0,9159689
create-document-policy	0,843173312	0,876790123
demo-11	0,956210668	0,953673777
demo-26	0,931931707	0,917708333
demo-5	0,964039803	0,965003406
read-document-policy	0,874445036	0,905128205
read-informationunit-policy	0,779981164	0,849919094
read-patient-policy	0,85009274	0,893767705
Xacml-Nottingham-Policy-1	0,980102041	0,978416605
Samples Statistics		
N	9	9
Mean	0,9014	0,9174
Standard Deviation	0,0676	0,0422
Standard Error Mean	0,0225	0,0141

To avoid experimental bias, the above steps have been repeated ten times and the APFD computed.

The samples relative to the APFD of GROOT and XACMET test strategies for Null Hypothesis testing (H_{0APFD}) are reported in Table 5.14 (columns 2 and 3). In particular, the upper part of the table reports the samples associated to each test case generation strategies, while the lower part reports some statistical information of each sample. The

Table 5.15: *Paired T-Test: RQ 3(APFD)*

RQs	RQ 3
Label	XACMET - GROOT
t	1,6135
df	8
p-value (2-tailed)	**0,1453**
Mean	0,016
Standard Deviation	0,0297
Standard Error Mean	0,0099
Confidence Interval Probability	0.95
Confidence Interval of the Difference (Lower Bound)	-0,0069
Confidence Interval of the Difference (Upper Bound)	0,0388

results of the Paired T-test associated to H_{0APFD} are illustrated in Table 5.15. Because the *p-value obtained* is $0,1453 > 0.05$, from the sample considered there is no difference from the point of view of the APFD between the two test strategies, i.e., XACMET and GROOT have the same velocity in reaching their effectiveness.

5.8.6 Discussion

Automation and replication are two important aspects for the assessment of the effectiveness of different testing strategies and key factors for the overall testing process.

Thus, the target of the work: defining and executing a controlled experiment for the comparison of two test cases generation strategies, specifically GROOT and XACMET approach, in the context of access control.

According to the analysis and results in the presented experiment, the two strategies have the same effectiveness (RQ1) and behaviour in terms of APFD (RQ3), while they are significantly different in terms of cardinality of the test suites (RQ2), and therefore in their cost. This means that in case of budgets or effort constraints, because both of the test strategies provided the same performance in terms of fault detection, applying the XACMET approach could be a winning solution for reducing testing time and guaranteeing the same product quality.

A crucial aspect emerged during the comparative experiment analysis, that has never been considered before in the literature related to the assessment of XACML-based systems: during the test suite execution, there is the possibility that different sets of distinct Mutated PDPs can be executed by different test strategies. Ignoring such data evidence and comparing the test strategies just in terms of effectiveness or APFD could really produce invalid conclusions and therefore wrong test strategy selection. By leveraging XMF and the data collected during the experiment execution, the mitigation of such a risk was possible and a correct evaluation provided.

Concerning the validity of the experiment, i.e., the amount of confidence in the results, the important key factors are Subjects, Object and Parameters of the experiment. Here below, the strategies used to minimize the threats to validity are described.

With respect to ***confidence in the results***, the controlled experiment used data and measurements that satisfy the principles of independence, homogeneity and normality.

With respect to ***internal validity*** the crucial points are: the policy used, the PDP engine selected and the tool integrated for the generation of the mutants set. Concerning the Subject of the experiment, the policies included in the XMF framework are a good representative of real world XACML Policies, because they contain most of the constructs and functionalities actually used in the practice. However different policies, like for instance those of XACML conformance test suite, may produce different results. The Object of the experiment, i.e., the Sun PDP implementation integrated in the XMF framework, is one of the most adopted in access control systems and therefore its quality and performance are well established. However, different XACML-based implementations could be considered. Finally, considering the Parameter of the experiment, i.e., the tool integrated for the derivation of the mutants set, we included the already existing μJava tool because it is one of the most widely used in object oriented environment. Previous works guarantee that its performance can be comparable to others available, however it could be possible that other mutation tools may produce different results.

With respect to ***external validity***, we compared two test strategies: *GROOT* and XACMET, so as to have elements of comparable performance. Other strategies could have been considered and different results provided, but the purpose was only to show the use of controlled experiment for test strategy comparison and not to select the best one.

Part IV

GENERAL_D: Application Examples

CHAPTER 6

GENERAL_D & Legal Text

THE GDPR's sixth principle, *Integrity and Confidentiality*, dictates that personal data must be protected from unauthorised or unlawful processing. To this aim, we propose a systematic approach for authoring access control policies that are by-design aligned with the provisions of the GDPR. We exemplify it by considering realistic use cases.

This chapter refers to the Application Example 1 described in Chapter 3, Section 3.5.1. In order to support performing the required steps for authoring GDPR-based ACPs, we have customized our reference architecture GENERAL_D (see Figure 3.2)

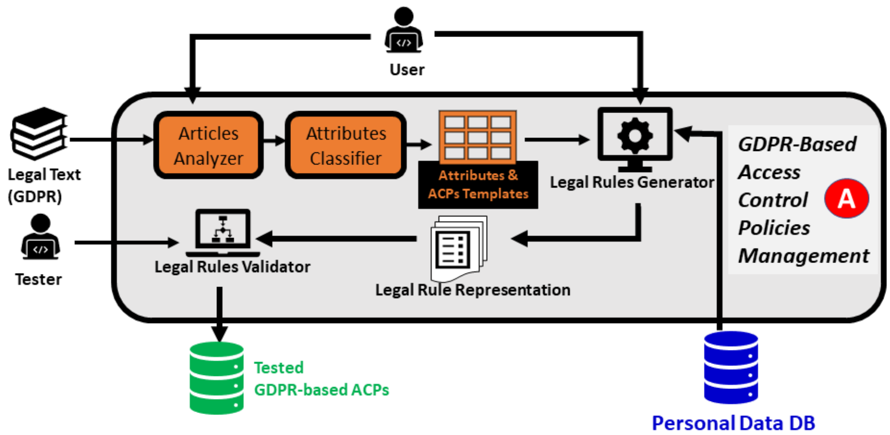


Figure 6.1: *GENERAL_D Customization for Handling Example 1.*

as depicted in Figure 6.1. More precisely, the customization involves Module (A) by specializing the component *Legal Text Analyzer* with the following components: 1. *Ar-*

ticles analyzer; 2. *Attributes Classifier*; 3. *Attributes & ACPs Templates*.

This chapter is based on the related publication:

- [32] Cesare Bartolini, **Said Daoudagh**, Gabriele Lenzini, Eda Marchetti: Towards a Lawful Authorized Access: A Preliminary GDPR-based Authorized Access. ICSOFT 2019: 331-338

6.1 Introduction

The new GDPR is changing how *Personal Data* should be processed. It states, in Art. 5.1(f), that "[data] should be processed in a manner that ensures appropriate security of the personal data [..] using appropriate technical or organisational measures (integrity and confidentiality)".

AC systems can be such a measure. AC is a mechanism used to restrict access to data or systems according to ACPs, i.e., a set of rules that specify who has access to which resources and under which circumstances [205]. By implementing them, one can gain compliance with the principle of Integrity and Confidentiality, but when enriched with policies elicited from the GDPR's provisions, we believe, AC systems can realize a compliance by-design to the GDPR's provisions expressed in the policies.

As an initial trivial mapping, according to the GDPR, Personal Data can be considered the resources, while the Controller, the Processor, or the Data Subject are the subjects requesting access to the resources. But, besides this simple mapping, it may be challenging for ACPs designers to *identify*, to *extract*, to *translate* and to *encode* the GDPR's provisions into enforceable ACPs [238]. Provisions can be ambiguous and can include implicit information. They are also unstructured and therefore not straightforwardly expressible in a formal policy.

All these issues call for a systematic process for the design of ACPs properly linked to the GDPR. Failures this task may have serious consequences: not only the AC system enforcing the ACPs can leave personal data unprotected, but the AC system may also become unlawful for the specific context of the GDPR.

The risk can be mitigated by promoting the adoption of AC systems enforcing policies systematically designed for expressing GDPR's provisions, consequently the scope or our research.

Recent literature provides partial solutions to this problem. In [92], for instance, the authors propose an approach to extract ACPs from the Data Protection Directive (Directive 95/46/EC), the document that before the GDPR was a reference point for the protection of personal data. In [53] the authors discuss an approach for implementing Attribute-Based AC policies tailored to the protection of resources in an industrial setting; although the proposal is an example of systematic implementation of policies, it does not consider any legal framework.

Our proposal is to leverage those results by combining them and by providing a unified framework able to design ACPs in reference to the legal framework of the GDPR. In particular, inspired by the principle of *Data Protection by-design*, we discuss how to develop such ACPs by gathering access control AC requirements from the GDPR.

The remainder of this chapter is organized as follows. We recall the Legal Ontologies in Section 6.2, where we also discuss the related work. In Section 6.3 we describe a simple scenario used as reference in the remaining sections. In Section 6.4 we describe our approach and in Section 6.5 we apply it. In Section 6.5.1 we conclude and point out the future work.

6.2 Background and Related Work

Legal Ontologies As stated in Chapter 4, designing ACPs in reference to the GDPR requires to refer, within a policy, to GDPR concepts and to relationships among them. It also demands for a consistent vocabulary along the whole lifecycle of the development of the ACPs. An help in this direction comes from semantic web technologies and in particular from the legal ontologies. Among the legal ontologies currently available, we select PrOnto "that aims to provide a legal knowledge modelling of the privacy agents, data types, processing operations, rights and obligations" [183]. However, in this chapter we refer to the RAccOnto ontology defined in Chapter 4, which leverages PrOnto.

Related Work In literature there are several works that use access control as main means of protecting personal data. For example, authors in [61] report an initial proposal for an automatically enforceable policy language for access and usage control of personal information, aiming at transparent and accountable data usage. A formal definition of the consent is introduced in [226], where the authors defined a privacy preference language explicitly designed to fulfill consent-related requirements and to suit constrained execution environments. Only some proposals take as an explicit reference a given data protection law. For instance, the Health Insurance Portability and Accountability Act (HIPAA) [8] was considered as a case study in [63]. Here authors evaluated XACML as a candidate specification language for HIPAA privacy rules. They based their evaluation on the set of features required to sufficiently express HIPAA, proposed in literature.

A work closer to ours is reported in [92]. Here the authors examined the feasibility of translating the articles related to access control of the directive, and also provided an implementation. In the industrial environment, authors in [53] proposed a systematic methodology for the implementation of ABAC [125] solutions in real contexts.

However, all the available proposals either focus only some aspects of the GDPR or do not provide implementations or are not specific for legal requirements. Differently from these works, this chapter aims at defining a systematic approach for gathering as many GDPR requirements as possible so as to comply with the regulation, and consequently to provide ACPs in line with the GDPR.

6.3 Running Example

In explaining our proposal, we refer to a simple scenario (see Figure 6.2). A customer *Alice* (the data subject) wants to purchase goods online from *ABC* (the controller), an e-commerce company which provides an online service for ordering and delivering goods. ABC follows two marketing strategies, both using customer personal data: (1) *Untargeted Marketing*: the customers' E-mail is used to advertise novelties, such new services or special sales; (2) *Location-based Targeted Marketing* (or *Geomarketing*): a customers' location is processed to customise the user experience of who is visiting the platform that provides the service.

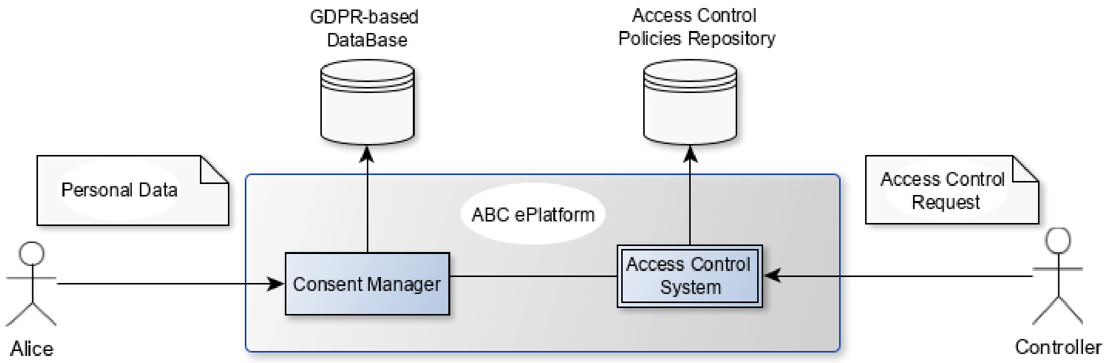


Figure 6.2: *E-Commerce Scenario.*

After the GDPR entered into force, ABC wants to adopt an AC systems to be compliant with the GDPR obligations. Its main objectives are: (1) to regulate the access to personal data; (2) to guarantee that its processing is lawful; (3) to facilitate data subjects in exercising their rights.

6.4 The Proposed Approach

Our approach has three phases: (1) *GDPR-based ACP Template Generation*; (2) *Legal Use Cases Definition*; (3) *Access Control Policies Authoring*.

6.4.1 Phase 1: GDPR-based ACP Template Generation

The GDPR text is analysed in search for provisions that spot a relation with AC so as to derive a meta-model (i.e., ACP template) for each of them.

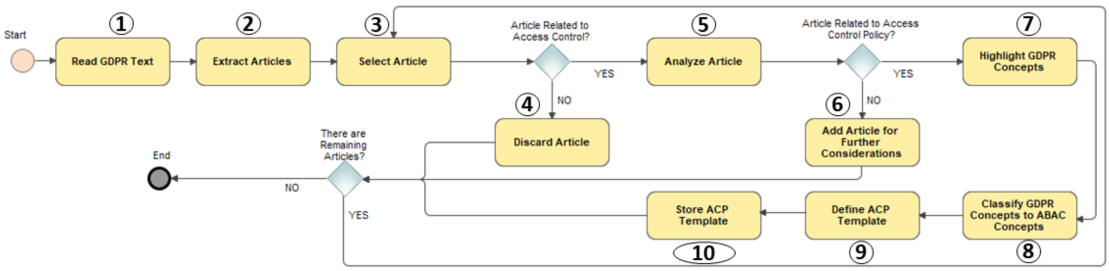


Figure 6.3: *GDPR Articles Selection and Templates Generation Process.*

This phase is organized in ten activities (see Figure 6.3). The first six (from 1 to 6) aim at selecting only the articles related to access control and discard all the remaining ones (activity 4).

Subsequently, the selected articles are then distinguish between articles related to ACPs (activity 5) and the ones related to AC mechanism (activity 6). The former group is used for the definition of meaningful ACPs. The latter is used to gather legal requirements from the architectural point of view, and it is out of the scope of the current work. Indeed, the collected functional and non-functional requirements will be used during the ACPs enforcement.

During the activity 7 all and only the attributes related with AC are identified. The selection of these attributes is driven by a conceptual model of the GDPR as they are represented in the RAccOnto ontology.[1]

For aim of completeness, we report the following sentence, as a simple example, where the identified GDPR-based attributes are highlighted:

Data Subject can **access** her/his **Personal Data**.

The next activity (8) is then aimed to classify the identified attributes into the commonly-used entities (or categories) in AC, namely, *Subject*, *Resource*, *Action* and *Environment*. Specifically, in ABAC terms **Data Subject** is classified as a *Subject*, **access** is classified into the *Action* category, and finally **Personal Data** is classified as a *Resource*. For instance, considering the above sentence, the identified attributes are classified as reported in the following:

Data Subject$_{[Subject]}$ can **access**$_{[Action]}$ her/his **Personal Data**$_{[Resource]}$

Finally, the last two activities (9 and 10) involve the definition of GDPR-based ACP templates, where the natural language statements are transformed in a machine-readable representation and the relations between attributes are identified.

Considering the previous example, we need to clarify the meaning of **her/his** and to define possible relations between the attributes *Data Subject* and *Personal Data*. Article 4(1) can be used for the purpose. Specifically, it states that: *'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, [. . .]*; this means that the Personal Data have the property of identifying a particular Data Subject. We can express this property as:

DataSubject = PersonalData.Owner

Consequently, a possible GDPR-based ACP template related the aforementioned sentence could be:

((Subject = Data Subject) $\wedge$ (Resource = Personal Data) $\wedge$ (Action = access)
$\wedge$ (Data Subject = PersonalData.Owner)) $\implies$ (Authorization = Permit)

6.4.2 Phase 2: Use Cases definition and ABAC attributes selections

Depending on the peculiarities of the specific application scenarios and the selected GDPR articles, the use cases are defined, customized and better specified for each user of the system, e.g., Data Subject or Controller to gather AC requirements in terms of concrete attributes. The second phase is then made up of two main steps.

Step 1. Legal Use Case Definition which includes the development of ACPs able to guarantee by design some of the Data Subject's rights, such as the right of access of personal data (Article 15) and the right to data portability (Article 20). This is in line with the Article 12.2, which is worded as follows: *"The controller shall facilitate the exercise of data subject rights under Articles 15 to 22. [. . .]"*. Indeed, on the basis of the template developed in the first phase (Section 6.4.1), the controller can

[1]Thus, we restricted our study to the concepts described within RAccOnto representation.

automatically, easily and promptly setting up customized ACPs as soon as the consent is obtained from a data subject and in line with the GDPR's provisions. This allows the controller to act *without hindrance* and *without undue delay* (pursuant to Article 20) to wishes of the data subject to exercise her/his rights. As a consequence, the data subject can exercise her/his rights as soon as the ACPs become enforceable from the access control system, i.e., when the policies are deployed in the ACPs repository.

To better explain the proposed methodology, we consider the running example from Section 6.3. We suppose that Alice, at registration time within the ePlatform, provided the ABC Company her name, her E-mail address, and the name of the city where she has the permanent address. We also assume that, at a later moment, Alice wanted to know which data she gave to the ABC Company during the registration, so as to exercise her right of access pursuant to Article 15.1.

Consequently, ABC defined the following authorization requirement:

ABC Req: Alice can read her name, E-mail, and her permanent city.

Of course, without the appropriate access control mechanisms, the specified authorization requirement could hardly be enforced. For this, the next activity.

Step 2. AC Attribute Identification and Classification We identify AC attributes directly from the *Legal Use Case* and for each of them:

1. the specific category is defined. This includes categories of data subjects, e.g., customer or employee and categories of personal data, e.g., biodata, financial data, health data or biometric data and so on;

2. the proper classification is identified. This include to classify the attributes according to the commonly used entities (or categories) of AC specification, i.e., Subject, Resource, Action and Environment.

By referring to the requirement **ABC Req**, the identified attributes are highlighted as follows:

ABC Req: Alice can **read** her **name**, **E-mail**, and her **permanent city**.

A possible classification of those attributes is then reported in Table 6.1, where (1) column *Identified Attribute* contains the identified attributes; (2) column *Attribute Category* shows a possible classification of those attributes into a specific category[2]; (3) while column *AC Category* illustrates the classification attributes into the commonly used entities in AC.

Table 6.1: *Attribute Classification Example.*

Identified Attribute	Attribute Category	AC Category
Alice	Customer	Subject
read		Action
name	Biodata	Resource
E-mail	Contact data	Resource
permanent city	Location data	Resource

[2]Note that classification refers only to the personal data and not to the processing operations such as *read* action.

6.4.3 Phase 3: Authoring and Assessing the GDPR-based ACPs

The first two phases provide the necessary building blocks for authoring and assessing concrete, meaningful and enforceable ACPs. Phase 3 is composed of three steps: (1) *Attribute Matching*; (2) *Authoring the GDPR-based ABAC Policy*; and (3) *Assessing the GDPR-based ABAC Policy*.

Step 1. Attributes Matching. The GDPR-based attributes identified in **Phase 1** are connected and instantiated with the concrete ABAC attributes identified in **Phase 2**.

The process we adopted for this aim is illustrated in Figure 6.4. Therefore, by referring to the requirement **ABC Req**, *Alice* is classified as *Data Subject*; the *read* action is connected to the *access* one; and finally, *Name*, *E-mail* and *Permanent City* attributes match *Personal Data* one.

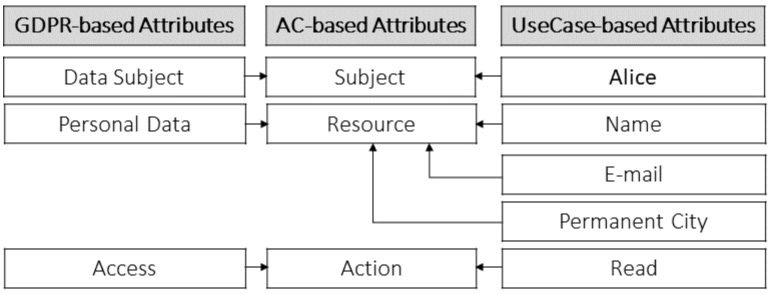


Figure 6.4: *Attributes Matching Example.*

Step 2. Authoring the GDPR-based ABAC Policy. The concrete and enforceable ACPs are obtained by performing two activities:

1. instantiate the ACP templates (see **Phase 1**) with actual attributes gathered from the legal use cases, as in **Phase 2** and
2. translate the resulting policies in a given formalism or language[3].

Consequently, by referring to the classification of the attributes of **ABC Req** defined in **Phase 2** and to the policy defined during the *Authoring Access Control Templates* activity of **Phase 1**, a possible abstract ACP looks like the following:

(Subject = Alice) $\wedge$ (((Resource = Name) $\wedge$ (Subject = Name.owner)) $\vee$ ((Resource = E-mail) $\wedge$ (Subject = E-mail.owner)) $\vee$ ((Resource = PermanentCity) $\wedge$ (Subject = PermanentCity.owner))) $\wedge$ (Action = read) $\Longrightarrow$ (Authorization = Permit)

The second activity involves the translation of the abstract ACP into a reference formalism or language. In this work we refer to the widely used XACML standard [180] to express the GDPR-based ABAC policies; but, one can choose any other implementation of ABAC model. An example of a concrete XACML policy is provided in Section 6.5.

Step 3: Assessing the GDPR-based ABAC Policy The last step is in charge of checking whether the authored GDPR-based policies conform with intended access rights, i.e., it verifies the correctness of the authored policies. In literature different proposals

[3]The approach aims at providing a generic ACP i.e., an independent representation from any formalism. This helps one to author ACPs in different languages that refer to different formalisms such as ABAC and RBAC. For aim of clarity, here, we illustrate how to encode actual ACPs by referring the ABAC model and its implementation XACML.

target the problem of policy assessment and are generally divided into: model-based testing [239], and combinatorial based testing [47]. We refer to the literature for more details.

6.5 Application Example

In this section we illustrate the application of our approach to the GDPR. To this regard, we first selected the articles related to AC and then we provided an ACP model for each of them (**Phase 1** of the approach). This allows the controller, or his/her delegate (e.g., an internal security administrator), to write ACPs in line with the GDPR according to the principle of data protection by design and by default. Consequently, the usage scenario and required attributes have been defined (**Phase 2**). Finally, the ACP templates have been instantiated so as to obtain enforceable ACPs (**Phase 3**).

Phase 1. From a procedural point of view, we firstly parsed the text of the GDPR (*Read GDPR Text* activity) and we selected only ninety-nine articles[4](*Extract Articles*). For each selected article (*Select Article*), we evaluated its adherence to the concept of access control: not pertinent articles have been consequently discarded (*Discard Article*). The remaining ones have been further analyzed (*Analyze Article*) to assess whether they could be related either with ACP concepts or with the AC mechanisms (*Add Article for Further Considerations*). As final results, among the ninety-nine selected articles, only forty-one have been considered as related to access control. Specifically: three of them were concerning only AC mechanisms; eight were referring only ACPs, and thirty articles related to both ACPs and AC mechanisms. Consequently, only thirty-eight articles have been used to derive GDPR-based ACP templates.

As an example, in the remaining of the section we illustrate the proposed approach only for one of the final selected articles, which is related to the management of both the purpose and the consent given by the data subject.

6.5.1 Lawfulness of Processing

For providing a lawful authorized access of personal data by the controller, the first step is to guarantee that all the accesses authorized by the AC system (or processing activities in general) are based on lawful basis. To this purpose, the Art. 6 lists as first basis the Consent: this is the most general concept and the most critical from a legal point of view[5]. Specifically, during the *Highlight GDPR Concepts* activity (see Figure 6.3), we refer to the sub-paragraph of the Art. 6.1(a) which words:

> Processing shall be lawful only if and to the extent that at least one of the following applies: (a) the data subject has given **consent** to the **processing** of his or her **personal data** for one or more specific **purposes**.

Consequently, by referring to the RAccOnto ontology we identified the following four GDPR concepts, as highlighted before: (1) consent, (2) processing, (3) personal data, and (4) purposes.

[4]In this proposal we focused on the articles that are mandatory.
[5]For more details we refer to the *Guidelines on consent under Regulation 2016/679* of the WP29.

During the activity *Classify GDPR Concepts into ABAC*, we classified personal data as *Resource* and processing as *Action*.

Concerning the purposes attribute, we distinguished the purpose for which the personal data is collected and the purpose for which the data is requested or accessed. For this, we referred to the XACML Privacy Policy Profile [181] provided by the XACML standard (see also Chapter 4). This specification describes a profile of XACML for expressing privacy policies and defines two attributes and one rule as in the following:

(1) the **resource:purpose** attribute "indicates the purpose for which the data resource was collected";

(2) the **action:purpose** attribute "indicates the purpose for which access to the data resource is requested"; whereas,

(3) the defined rule "stipulates that access shall be denied unless the purpose for which access is requested matches [...] the purpose for which the data resource was collected".

Therefore, since the purposes listed in Art. 6 refer to the purposes for which the personal data was collected, we classified the identified attribute as *Resource*, and more precisely as an attribute of personal data.

The same strategy has been adopted for the consent as well, i.e., we defined the consent as a special attribute of the specific purpose for which the personal data is collected. In case of personal data, we considered the consent as a BOOLEAN contextual attribute. This allows the controller to manage also the right of the data subject "to withdraw his or her consent at any time" pursuant the Art. 7 (*Conditions for consent*). As a consequence the consent attribute has been classified as Environment attribute with the following result:

> [...] (a) the data subject has given **consent**$_{[Environment]}$ to the **processing**$_{[Action]}$ of his or her **personal data**$_{[Resource]}$ for one or more specific **purposes**$_{[Resource]}$

During *GDPR-based ACP Template* activity, the following ACP template associated the Art. 6.1(a) has been derived:

> ((Resource = PersonalData) $\wedge$ (Action = processing) $\wedge$ (Action.purpose = PersonalData.purpose) $\wedge$ (PersonalData.purpose.consent = YES)) $\implies$ (Authorization = Permit)

Table 6.2: *Legal Use Case: Attribute Classification.*

Identified Attribute	Attribute Category	AC Category
	Req 1	
ABC company	Controller	Subject
Alice	Customer	Subject
send		Action
E-mail	Contact data	Resource
Consent		Environment
untarget marketing	purpose	Resource

Phase 2. A possible **Legal Use Case** aligned with Art. 6.1(a) concerns the registration phase within the ePlatform and the actions required by the controller to obtain the explicit consent from its customers.

More precisely, by referring the scenario in Section 6.3, to use the online service provided by the ABC company, Alice needs to create an account within the *ePlatform* (see Figure 6.2), and submits a set of personal data, i.e., Name, Surname, E-mail Address, Home Address, the Gender, and Birthdate.

Afterwards *Consent Manager* asks Alice her consent for processing some of her personal data for the purposes defined by Controller, i.e., Untarget and Location-based target marketing.

Consequently, for each requested consent Alice gives or denies her consent. In particular, we consider the specific situation in which she gives only the explicit consent of processing her E-mail Address for Untarget Marketing purpose, and withhold her consent for Geomarketing purpose.

Finally, the Consent Manager stores and sends the collected information to the AC System for authoring an ACP related to Alice. A possible authorization requirement related to Alice's consent is:

Req 1: **ABC company** (Controller) can **send** communications only for **untarget marketing purpose** using the **E-mail** of **Alice**, because of the **consent** given.

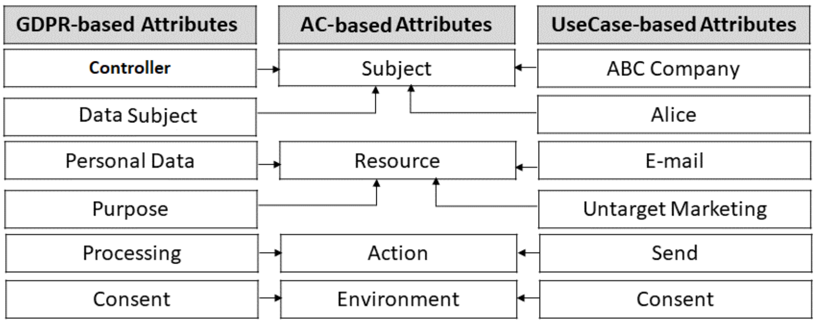


Figure 6.5: *Article 6.1(a): Attributes Matching.*

The next activity is the **Access Control Attribute Identification and Classification**, where the attributes based on the above requirements are identified and reported in the first column of Table 6.2. The table depicts the classification of the identified attributes into a commonly used access control categories as well.

Phase 3. The result of **Attribute Matching** phase is reported in Figure 6.5. As depicted in the figure, for example, *ABC Company* is classified as *Controller*; the *send* action is connected to the *Processing* one; and, *E-mail* attributes match *Personal Data*.

Based on this mapping, the next activity **authoring policy** produces the enforceable ACP reported in Figure 6.6. It is composed of a *Target* element, stating that the policy is applicable to *ABC Company*, and two rules. The former, with effect Permit, states that *ABC Company* can access *E-mail* to perform *sendEmail* action (see the *Target* element of the rule) if and only if the owner of that E-mail is *Alice* and for *untarget Marketing* purpose accessing (see the Condition element of the rule). The latter is the default rule aiming at denying the access in case the first rule is not applicable.

```
Policy ............................ root element
│                                   PolicyId = Article6(1)(a)
│                                   rule-combining-algorithm:permit-overrides
├── Target ........................ Company can send communications only for un-target marketing purpose
│   │                               using the E-mail of Alice, because of the consent given
│   └── Subject ................... Subject-id = ABC Company
├── Rule .......................... RuleId = untargetMarketing, Effect = Permit
│   ├── Target .................... The sent e-mail actions
│   │   ├── Resource .............. subject-id = ABC Company
│   │   ├── Resource .............. resource-id = E-mail
│   │   └── Action ................ action-id = sendEmail
│   └── Condition ................. ABC Company.
│       └── And ................... And Operator
│           ├── string-equal ...... resource:owner = "Alice"
│           ├── string-equal ...... purpose = "untarget Marketing"
│           ├── string-equal ...... action:purpose = purpose.
│           ├── string-equal ...... resource:purpose = purpose.
│           └── boolean-equal ..... resource:purpose:consent = true.
└── Rule .......................... RuleId = defaultRule, Effect = Deny
                                    Default: deny all, which is not allowed explicitly.
```

Figure 6.6: *A Possible XACML Policy for Article 6.1(a).*

Remark. This chapter presented a systematic approach to gather access control requirements from the GDPR. This approach is the first step towards a definition of an access control solution based on the GDPR. Although grounded in a domain-related implementation (i.e., compliance to the GDPR), the approach yields a more general spectrum, since it can be applied to different data protection regulations and more in general to any legal text that implicitly contains, or suggests, data protection requirements. To the best of the authors' knowledge, the novelty of our work is the systematic approach to join and improve the current academic proposals for the extraction of legal by-design ACPs from the data protection regulation with the approaches currently used in industrial environment for implementing ABAC.

CHAPTER 7

GENERAL_D & User Stories

BECAUSE of the GDPR's principle of "data protection by design and by default", organizations who wish to stay lawful have to re-think their data practices. Access Control (AC) can be a technical solution for them to protect access to "personal data by design", and thus to gain legal compliance, but this requires to have Access Control Policies (ACPs) expressing requirements aligned with GDPR's provisions. Provisions are however pieces of law and are not written to be immediately interpreted as technical requirements; the task is thus not straightforward. The *Agile software development methodology* can help untangle the problem. It promotes detailed procedure and form for describing requirements such as the specification of *User Stories*. These are concise yet informal requirement descriptions telling who, what and why something is needed by users. Additionally User Stories are organized into prioritized lists, called *backlogs*. Inspired by the Agile development process this work advances the notion of *Data Protection Backlogs*, which are lists of User Stories about GDPR provisions described in terms of technical requirements. Thus for each User Story we provided its corresponding ACP, so as to make easier the design and implementation of GDPR compliant AC systems.

This chapter refers to the Application Example 2 described in Chapter 3, Section 3.5.2. In order to support performing the required steps for authoring GDPR-based ACPs, we have customized our reference architecture GENERAL_D (see Figure 3.2) as depicted in Figure 7.1. More precisely, the customization involves Module A by specializing the component *Legal Text Analyzer* with the following components: 1. *User Stories Tool*; 2. *User Stories*; and 3. *User Stories BD*.

This chapter is based on the related publication:

- Cesare Bartolini, **Said Daoudagh**, Gabriele Lenzini, Eda Marchetti: GDPR-Based

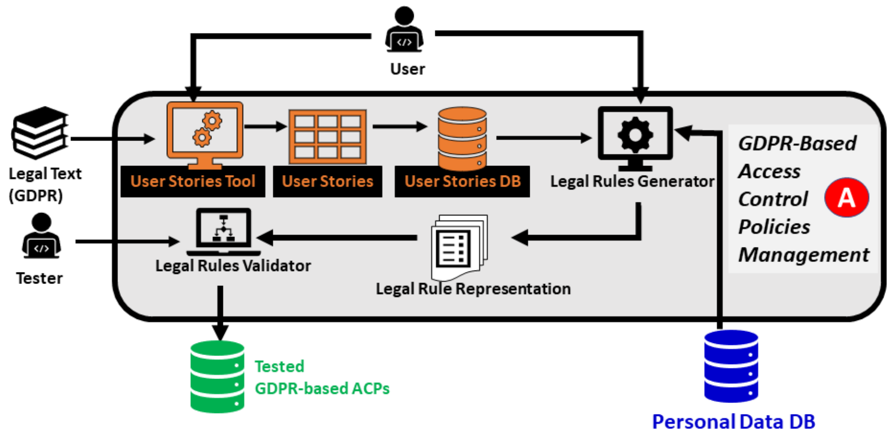


Figure 7.1: *GENERAL_D Customization for Handling Example 2.*

User Stories in the Access Control Perspective. QUATIC 2019: 3-17

7.1 Introduction

Nowadays, the Information Technology (IT) domain is moving towards systems with growing complexity, where digitalization, artificial intelligence, interconnection and mobility are some key factors. Indeed, in their multidisciplinary nature, they require an extensive deployment of advanced ICTs, as well as the adoption of effective measures for strengthening security, trust, dependability and privacy. These aspects have to be considered over the whole Software Development Life Cycle (SDLC), from the gathering of the requirements to the deployment and subsequent maintenance of the system.

Over the last decade, especially for small and medium enterprises, *Agile Software Development (ASD)*, first introduced in the Agile Manifesto [98], and its subsequent evolutions such as eXtreme Programming (XP) and Scrum [134] are becoming commonly-adopted software development processes. Basically, ASD is an iterative approach that focuses on incremental specification, design and implementation, while requiring a full integration of testing and development. In this development process, a common means of capturing the user's needs and describing the value that the user would get from a specific functionality of the system is the so-called *User Story* [10]. From a practical point of view, a User Story focuses on a requirement written according to a specific format (see paragrafo 7.2) and guidelines on how to implement it. Usually, depending on the granularity of the story, different names can be used for defining its contents: large ones may be known as *Epics*, and small ones as *Features*, *User Stories*, and *Tasks* [10].

In practice, many times small organisations and software development groups have not the required effort (in terms of budget and time) for exhaustively implementing the requirements elicitation phase so they rely on previously collected set of User Stories. Indeed, these are also able to speed up the subsequent implementation because technical details are already included. However, specific User Stories targeting privacy requirements are currently missing: organisations need to collect and develop their privacy requirement. In this case, lack of time and expertise could have the side effect to release software with high privacy risks [19].

With the entering into force of the GDPR (see sezione 2.2) this situation is not affordable anymore, because the regulation is changing how *Personal Data* should be processed. Consequently, as stated in Capitolo 3, privacy concerns have to be considered at early stage of development of services and systems that manage and process Personal Data following the Privacy-By-Design approach.

Within the Agile development the use of *security backlogs*, i.e., a prioritized features list describing the functionalities to be included in the final product [10], is already adopted. These backlog items are often provided in the form of User Stories [19]. The set of security backlogs is therefore a list of ready-made specifications of security items (requirements and task descriptions) useful for the implementation. An example of a security User Story related to access control is reported in paragrafo 7.2.

Therefore, following this tendency, the contribution of this chapter consists of three main parts (see Figure 7.2): i) introduce the concept of *Data Protection Backlog* that contains User Stories based on GDPR requirements; ii) map specific provisions of the GDPR to User Stories; and iii) provide, for each User Story, the corresponding specification/implementation as Access Control Policy so as to assure the GDPR compliance design.

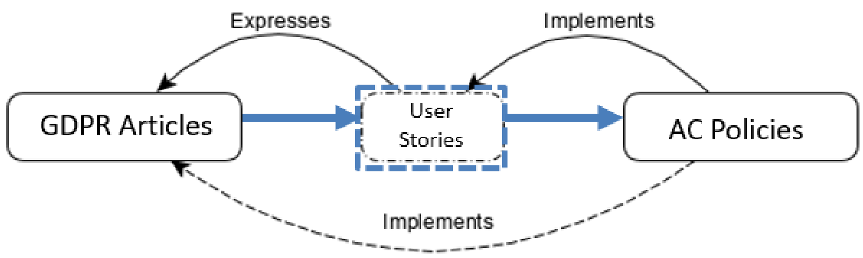


Figure 7.2: *Overview of the Proposal.*

To this purpose, in this chapter, we present our solution for a compliant implementation of the GDPR, by encoding User Stories, and consequently the GDPR provisions, as ACPs. In order to minimize errors and issues, a consolidated verified and predefined structure of ACPs [229] is provided. In line with this trend, for each identified User Story related to access control, this chapter provides a GDPR-based ACP template. Certainly, the templates represent meaningful, concrete and predefined building-blocks for ACP specification, that can be adopted and refined for the different scenarios, so as to overcome possible misinterpretations and reduce security and privacy risks.

The *Data Protection Backlog* containing the set of User Stories, each associated with a proper ACP template, would be a valid starting point for privacy requirements specifications, and a generic guidance for who are facing the GDPR implementation problem. Consequently, when a new development starts, the developer or access control policies architect could pick up the related predefined User Story and easily implement it.

The reminder of this chapter is organized as follows. We recall User Stories in sezione 9.2, where we also illustrate the related work. The proposed GDPR-based User Stories model is described in sezione 8.3, and in sezione 7.4 we show the process to the derive the *Data Protection Backlog* containing the User Stories and the associated ACPs. Finally, paragrafo 7.4 concludes the chapter.

7.2 Background and Related Work

In this section, we firstly recall User Stories, and then we provide an overview of the proposals dealing with the main topics of this chapter, i.e., representing security and privacy by means of User Stories, and how to put in relation AC environment and the GDPR.

User Stories. User Stories are an important part of an Agile development process because they represent a valid means to writing simple and understandable requirements [231]. Currently, their adoption is massively growing [130], and several definitions are available [153]. However, most of them agree that commonly a User Story is a short, yet a simple description of a feature from the perspective of an end-user or customer of the system. A User Story typically presents the following structure:

> As a *[end user]*, I want to achieve *[goal]* so that *[I realize the following benefit of]*.

An example of a security User Story related to ACP, reported in [10], is as follows:

> As *[an information security manager]* I want *[that it is clearly defined which user accounts are authorised for which activities]* so that *[the effectiveness and correctness of access controls can be checked]*.

One key factor of the widespread use of User Stories is that they can be written at different levels of detail. They can cover large amount of functionalities and in this case are generally known as *Epics*. However, an epic is generally too large for being easily implementable into a single Agile iteration. Thus, it usually split into multiple smaller User Stories before it is worked on. This is for instance the case of features, User Stories, and tasks [10]. In some cases, User Stories are detailed more, by adding conditions of satisfaction, i.e., a high-level description of what needs to be true after the Agile User Story is completed.

Although there is no specific customer's or user's role for writing User Stories, having a common set of *product backlog* of Agile User Stories is an essential factor for the successful development of a system. Indeed, the product backlog can be used to select and prioritize the list of the functionalities that have to be developed in different iterations of the Agile process.

Related Work. An important innovation for speeding up the development of software has been the introduction of Agile development and the Scrum methodology. Over the last years, literature has moved an important criticism to these kinds of approaches because they mostly ignore the security risk management activity [10, 20, 203, 230]. Thus the concepts of security should be considered during all stages of the software development Life Cycle. In Agile environment this commonly means integrating security principles in terms of security backlog [19, 20]. The security backlog is a set of ready-made User Stories that can be used to cover the security requirements [216]. This new backlog can be used to manage and mitigate the security risks associated with the software [203, 230].

The introduction of GDPR requirements in the secure software development adopted into the Agile processes for discovering and solving security threats is not sufficient anymore to guarantee the required privacy level, and few proposal are recently targeting this issue. Among these, in [203] the authors propose a Threat Poker method to exercise both security risks and privacy risks and evaluate the effort needed to remove the corresponding vulnerabilities in the software developed. However, the proposal is mainly focused on the estimation of the seriousness of security and/or privacy risks during software development. Similarly, in [165] the authors present an Agile process for the definition of security and privacy in terms of User Stories, in order to develop a framework to manage Personal Health Information. In particular, the authors highlight the need for suitable policies and procedures for data security and privacy management, so as to make the framework compliant with regulations.

This chapter is inspired by a proposal [218] that describes a semi-formalized, constrained natural language format for User Stories. The format uses variables to precisely correlate various parts of the story with a predefined format, to express strictly-defined operators in a (almost) natural language. The authors also showed a possible way to extract access control information for role-based access control from this format.

Differently, our proposal here is aiming at defining User Stories related to the technical requirements gathered from legal text (in our case the GDPR), and providing ACPs that are based on the ABAC model.

7.3 GDPR-Based User Stories Conceptual Model

In an attempt to comply with the principle of *data protection by design*, laid out by Art. 25.1 of the GDPR, we detail a methodology for defining privacy-based User Stories and gathering them to ACPs requirements directly from the GDPR. From a practical point of view, this means first extracting, in an Agile perspective, User Stories that represent atomic privacy or legal requirements to be implemented so as to comply by-design with the GDPR. Then, considering systems that enforce an AC, defining an actionable list of simple AC system specifications which address the core requirements demanded by the GDPR.

The proposal would like to contribute to: i) an incremental development of the AC system, by guaranteeing that, by-design, it maintains compliance with the GDPR; ii) the Data Protection Impact Assessment (DPIA) along the development of the system; iii) a mapping between the implemented functionalities and the corresponding GDPR provisions. This will help to create a traceability mechanism useful for demonstrating GDPR compliance, as required by the *Accountability* principle.

The User Stories are built taking into account the GDPR concepts of *Data Subject*, *Controller*, *Processor*, *DPO*, and *Personal Data*.

The conceptual model for User Stories, used for the derivation of the actionable list, is shown in Figura 7.3. It is composed of three sub-models: the GDPR Model, User Stories Model and AC Model. The sub-models are combined into the process followed for going from the definition of the User Stories to specification of AC policies.

The sub-models have been voluntarily kept separated to increase the possible generalization of our proposal. Indeed, the GDPR Model and AC Model could be replaced by any other legal regulation or legislation which is suited for automatic enforcement.

The remainder of this section provides specific details about these sub-models.

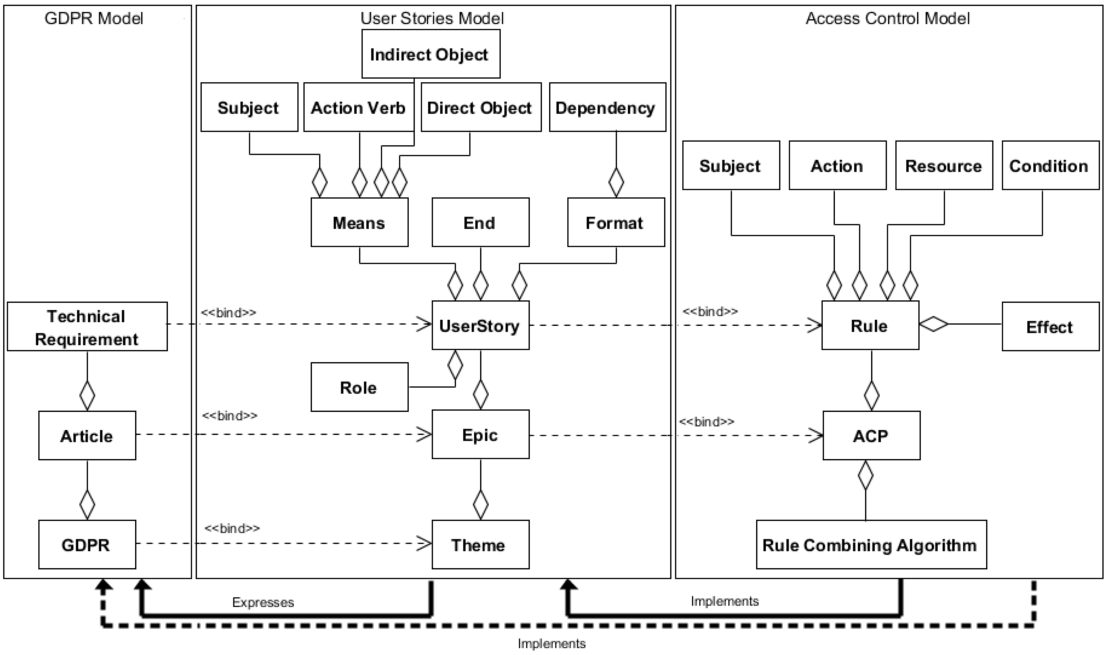


Figure 7.3: *The Conceptual Model of GDPR-focused User Stories.*

7.3.1 User Stories Model

The User Story used in our model is a modified version of model introduced in [152]. More precisely, we do not consider the *Clarification* and the *Quality* elements of the *End* component; we eliminate the *Adjective* element from the *Means* component; and finally, we introduce the *Theme* component as abstract level to better bind the User Stories to the GDPR.

As depicted in Figura 7.3, a User Story always includes one relevant *Role*, which is associated with the stakeholder or legal entity that expresses the need. Currently, due to the complexity of the GDPR text, the number of proposals trying to provide a conceptual model of the regulation [33, 185, 186] is increasing in literature. Among those available, and in order to relying on a formal base for the role specification, in this work we rely on the formalization provided by the PrOnto [183, 184] ontology, which we have already integrated in our RAccOnto ontology and described in Chapter 4. Consequently, the stakeholders that we are referring to are therefore *Controller*, *Processor*, *Data Protection Officer* (*DPO*), *Data Subject* and *Supervisory Authority*.

The *Format* of the User Story is a predefined template in which the role, means, and optional end(s) are specified. As described in Section 7.2 we refer to the most widespread format introduced in [64] which consist of:

As a [type of user], I want [goal], so that [some reason].

Differently, *Means* can have different structures that can be used to represent different types of requirements. Means have three common grammatical elements: i) a subject with an aim; ii) an action verb that expresses the action related to the feature being requested; and iii) a direct object (and optionally an indirect object) on which the subject executes the action.

The End of a User Story explains why the means are requested. However, User Stories often include other types of information, such as dependency on another functionality, i.e., implicit references to a functionality which is required for the means to be realized. This is useful in the context of the regulations since legal text often use the cross-reference mechanism between articles.

In the GDPR context, a possible User Story related to Art. 30.4 could be:

> As a *[Supervisory Authority]*, I want *[to access the record of processing activities]*, so that *[I can monitor those processing operations]*.

7.3.2 The GDPR Model

In this study we model the GDPR only from a structural point of view. As described in sezione 2.2, the mandatory part of the GDPR is composed of ninety-nine articles organized in chapters; some chapters are then broken in sections. The GDPR's articles present a structure that involves at least other two levels (paragraphs and letters). Consequently, each article may include one or more technical requirements.

In order to be aligned with the structure defined in User Stories model, we model the GDPR as an aggregation of articles. More precisely, we do not consider the recitals, and we collapsed all the aforementioned complex structure of the regulation in a more simple one that includes only three levels: *GDPR* → *Article* → *Technical Requirement*.

This simple structure helps in binding the GDPR core code with the concept of Theme in Agile terminology; then, the articles represent Epics which contain one or more small and manageable technical requirements, each expressed by means of a User Story.

7.3.3 The Access Control Model

As state in Chapter 2, Section 2.3.1, an ACP defines the AC requirements of a protected system, i.e., a set of AC *Rules* that specify who (e.g., Controller, Processor or Data Subject) has access to which resources (e.g., Personal Data) and under which circumstances [205]. The AC rule is often specified using Natural Language Access Control Policy (NLACP)), that presents the following structure: *[Subject] can [Action] [Resource] if [Condition]* [115].

The Access Control Model used in this proposal is a simplified version of the Policy Language Model provided by the XACML standard [180]. In Chapter 4, we have already leveraged that model in the context of semantic web, by proposing RAccOnto ontology expressing a GDPR profile for access control. Differently, here we leverage the same model in the Agile perspective by connecting the main access control concepts with User Stories ones. Even simple, the model captures all the essential concepts for the design of both simple and more complex ACPs.

For the aim of completeness and for the aim of having a self-contained chapter, we recall that the model consists of *Rule* class, which represents the most elementary unit of policy enforceable by an ACS (see Figura 7.3). The rule is composed of one single *Subject*[1], one single *Action*, one single *Resource* and one single optional *Condition*.

[1]Note that the Subject expressed in this model is different from the one defined in the User Stories Model: the Subject in that model represents a grammatical function in the formulation of the means; while Subject in the AC domain represents an active entity which covers a role. The Subject in this model is an entity that can semantically be correlated with the Role entity in the User Stories Model.

The *Effect* associated with the rule represents the rule-designer's intended consequence of a *True* evaluation for the rule. We recall the the usual two values allowed for the rule's effect are: *Permit* and *Deny*. As depicted in Figura 7.3, the rule represents an expression of an atomic technical requirement described by a User Story. The ACP class is a composition of rules and *Rule Combining Algorithm* which defines strategy by which the results of evaluating the rules are combined when the ACS evaluates the policy. As in Figura 7.3, the ACP in associated with Epic.

7.4 User Stories Related to Access Control

The process we used to define the set of User Stories, related to the provisions of the GDPR and the AC rules, is composed of three steps: (1) *GDPR Articles Selection*; (2) *User Stories Definition*; (3) *GDPR AC Rules Definition*. Figure 7.4 depicts an overview of proposed process, whereas in Figure 7.5 we detail more the main activities involved in each steps.

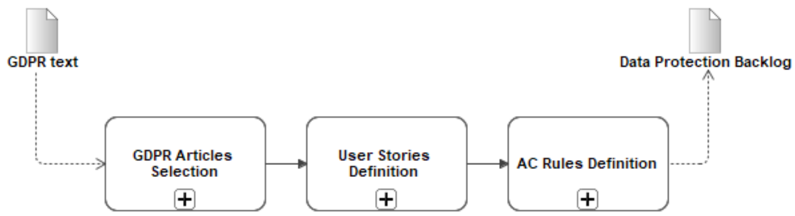


Figure 7.4: *GDPR-focused User Stories Definition Process.*

GDPR Articles Selection. The input of the process is the GDPR text. Firstly, we selected only the mandatory part of the GDPR which consists of ninety-nine articles; for each article (step 1(a) in Figure 7.5), we decided whether is related to AC concept, i.e., AC language or AC mechanism, and consequently we created an Epic associated to the current article (step 1(b) in Figure 7.5). The result of this step was the section of forty-one Epics (GDPR articles) related to AC. Specifically, three of them were concerning only AC mechanism; eight were referring only ACPs, and thirty articles related to both ACPs and AC mechanism. For more details about this step we refer to Chapter 6 and its related publication [32].

User Stories Definition For each article identified in the previous phase, we extracted one or more technical requirements and defined a specific User Story for each of them (step 2 in Figure 7.5). Thus, the User Stories were added to the Epic associated with the current article. In order to trace the covered GDPR's articles during the Agile development process, we defined a for each Epic an identifier (named EpicID [2]) able to find the GDPR's article the Epic is referring to. Similarly, we defined an identifier for each User Story (called UserStoryID [3]) with the purpose to the specific part of the GDPR's article the User Story related to (e.g., the paragraph or the letter of the article).

[2]The identifier EpicID has the following structure: GDPR.Epic.Article.[articleNumber].

[3]The identifier UserStoryID has the following structure: [EpicID].[ParagraphNumber].[letter].US.[progressiveNumber]

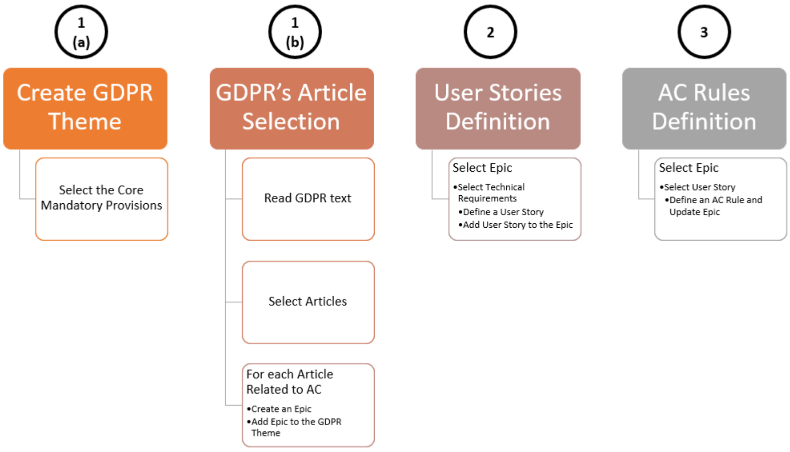


Figure 7.5: *Details of User Stories Definition.*

GDPR AC Rules Definition The final step (i.e., step 2 in Figure 7.5) deals with the translation of the technical requirements associated with the AC language, and consequently we defined an AC rule for each User Story conceived in the previous step. The procedure used in this step is the same defined in Chapter 6, where we have defined a systematic approach for deriving ACPs directly from the GDPR. We therefore refer reader to Chapter 6 for more details about this step. The only difference is that instead of starting from the GDPR's articles text, here this step starts from their representation as User Stories expressed in the format defined in Section 7.2. However, in literature there exist different proposals for the derivation of ACPs from the natural language [15, 238] or controlled natural language [92]. It is out of the scope of this chapter going into details about the different procedures of extracting ACPs.

As in Figura 7.4, the result of this process is a *Data Protection Backlog*, i.e., a Privacy Backlog containing a set of AC rules organized in User Stories, Epics and Theme. This is a ready-made solution to be used during the Agile development of an ACs system aligned with the GDPR requirements. Indeed, when an SME want to promote the GDPR initiative by using AC as main technical means to comply with the GDPR, developers or AC architects can pick-up one User Story per time and smoothly implement the corresponding the ACP. In this way, the SME can achieve the GDPR compliance incrementally, in Agile perspective.

In Tabella 7.1 we present an extract of the defined Data Protection Backlog related to the GDPR. The User Stories are reported from both the perspective of the Data Subject and the Controller, by considering Art. 6, Art. 7, and Art. 15.

The table is composed of three columns: the column **Article** (first column) contains the GDPR's articles. The column **User Story** contains the GDPR-based User Stories defined. Finally, the third column contains the AC rules related to the User Stories.

Remark. This chapter presents an Agile methodology to gather access control requirements from the GDPR by using the concept of User Stories. This methodology is a first step towards a formal definition of access control solutions addressing GDPR requirements in Agile environment. To the best of the our knowledge, an Agile methodology for the specification of User Stories, organized in Data Protection Backlog, i.e., Privacy

Table 7.1: *GDPR-focused User Stories: Controller and Data Subject Perspectives*

Article	User Story	AC Rule
Art. 6.1(a)	As a [Controller], I want [to process Personal Data only if Data Subject has given consent for one or more specific purpose], so that [the processing shall be lawful].	[Controller] can [Process] [Personal Data] If [PersonalData.purpose = Processing.purpose AND PersonalData.purpose.consent = TRUE]
Art. 7.3	As a [Data Subject], I want [to withdraw my consent], so that [I can exercise my right as stated in Art. 7.3]	[Data Subject] can [Withdraw] [PersonalData.purpose.consent] If [PersonalData.owner = DataSubject AND PersonalData.purpose.consent = TRUE]
Art. 15.1	As a [Data Subject], I want [to access my Personal Data and all the information], so that [I can be aware about my privacy]	[Data Subject] can [Action = access] [PersonalData] AND [Resource = PersonalData.purposes] AND [Resource = PersonalData.categories] if [PersonalData.owner = Data Subject]
Article 15.3	As a [Data Subject], I want [to download a copy of my Personal Data], so that [I can check their correctness]	[Data Subject] can [download] [Personal Data] If [PersonalData.owner = Data Subject]

Backlog, aimed at extracting legal ACPs from the GDPR is novel. Although grounded in a domain-related implementation (i.e., the GDPR), the Agile methodology yields a more general spectrum, since it can be applied to different data protection legislation that encodes ACPs specification.

In our case, the generation of a set of ACPs aligned with the GDPR was conceived in three phases: the selection of GDPR's articles related to access control; the definition of a Data Protection Backlog containing User Stories extracted from the selected GDPR's articles; and finally, the definition of access control rules, each related to a specific User Story. Having a User Story (and consequently an access control rule) related to a specific GDPR provision helps to detect the rules that need to be updated when the regulation changes.

Ongoing and future work. As future, we are planning to consider the GDPR requirements referring access control mechanisms, i.e., requirements from the architectural point of view. Furthermore, we are currently investigating a comprehensive Data Protection Impact Assessment (DPIA) methodology (which is one of the legal requirements in the GDPR (Art.35)) by leveraging the conceived Data Protection Backlog. Whereas, ongoing work includes the validation of the User Stories by different Agile development teams in the context of European projects that address key regulations such as the GDPR. For example, within the CyberSec4Europe project we are validating the obtained results in two project's demonstrators in the context of Smart-City, whereas in BIECO project we are using User Stories as specification of privacy claims to be satisfied during the implementation of system of system (SoS) that processes Personal Data.

CHAPTER 8

GENERAL_D & External Consent Manager

THIS chapter provides evidences of the flexibility of the i.e., GENERAL_D proposal in adapting and integrating pre-existing solutions. In particular, we consider the integration of an available Consent Manager (CM) and an Access Control (AC) to aid organizations to comply with the GDPR. The idea is to use GENERAL_D for converting the GDPR machine-readable format provided by an External CM into a set of enforceable Access Control Policies (ACPs). In this chapter the defined the layered architecture able to makes systems compliant by-design with the GDPR. To validate the feasibility of this proposal, we provide a proof-of-concept by integrating and AC Manager, i.e., GENERAL_D, and an External Consent Manager coming from an industrial context.

This chapter refers to the Application Example 3 described in Chapter 3, Section 3.5.3, and the customization of GENERAL_D is depicted in Figure 8.6.

The content reported here is based on the related publication:

- [80] **Said Daoudagh**, Eda Marchetti, Vincenzo Savarino, Roberto Di Bernardo, Marco Alessi: How to Improve the GDPR Compliance through Consent Management and Access Control. ICISSP 2021: 534-541

8.1 Introduction

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. As a consequence, assuring the GDPR compliance, and therefore avoid the related fines, becomes an important research challenge.

Currently, many businesses are struggling in the definition of appropriate procedures and technical solutions for their development process so as to enforce and demonstrate the GDPR compliance [138]. More precisely, they recognized as a pivotal factor in the availability of automated supports for specifying privacy requirements, controlling personal data, and processing them in compliance with the GDPR.

From a practical point of view, scientific communities, private companies, and European projects such as CyberSec4Europe (Cyber Security Network of Competence Centres for Europe) [1] are identifying in the *Consent* and *Security* services the successful elements for automatic specification and enforcing the data protection regulation [138].

Indeed, consent services may allow citizens and companies to manage and track personal data in a straightforward, user-centric, and user-friendly manner; whereas, security services, and specifically authorization systems (i.e., AC), can enforce the data protection regulations taking into account additional legal requirements, such as the data usage purpose, user consent, and the data retention period. Therefore, the joint work of the consent and security services may overcome the challenging and error-prone task of extracting legal and machine-readable policies directly from the GDPR's rules.

Currently, different research activities have been devoted to define and implement privacy knowledge and rules [210], but no generic solution is still available. Along these lines, under the hypothesis that the joint integration of access control systems and consent manager can enhance the controller's and processor's compliance with the regulation, this chapter wants to provide the basic architecture of a generic and practical solution to solve the GDPR compliance problem.

In presenting our idea, we focus on the following primary Research Question (RQ):

> *How a consent manager solution can be improved with access control for assuring compliance with the data protection or privacy regulation, such as the GDPR?*

In answering that RQ, we present a possible *privacy-by-design* architecture by integrating AC and consent management systems. Finally, an implementation of the proposed architecture by using real available solutions for the consent management and the ACM, coming from both industry and academia, is presented.

Outline: Section 8.2 presents the basic concepts used along the proposal and related works; Section 8.3 describes the proposed solution by answering our RQ; Section 8.4 shows the proof-of-concept we implemented by instantiating the proposed solution with real artifacts coming from both industrial and academic contexts; and finally, Section 9.4 concludes the chapter and illustrates future and ongoing work.

8.2 Background and Related Work

This section introduces the main concepts used along the present chapter: Smart ICT Systems and Consent Manager; and reports the related work.

8.2.1 Smart ICT System

Smart ICT Systems (or Services) are becoming increasingly important in almost all industries and areas of today's society [172]. They rely on the integration and implemen-

[1] https://cybersec4europe.eu/

tation of innovative tools and techniques that make a given system *smart* to strengthen economic needs [204].

Despite their increasing significance, a distinct definition of Smart ICT has not yet evolved in the scientific literature. Nevertheless, it is possible to identify a very high-level abstract architecture for a standard Smart ICT System, as depicted in Figure 8.1.

Commonly, it is composed of a Smart ICT Core System that offers the main functionalities to Smart Services in terms of both hardware and smart software (e.g., Cloud Computing, Internet of Things, and Big Data). Consequently, developers use these functionalities to conceive and implement Smart Services that end-users consume to achieve a given business or personal needs.

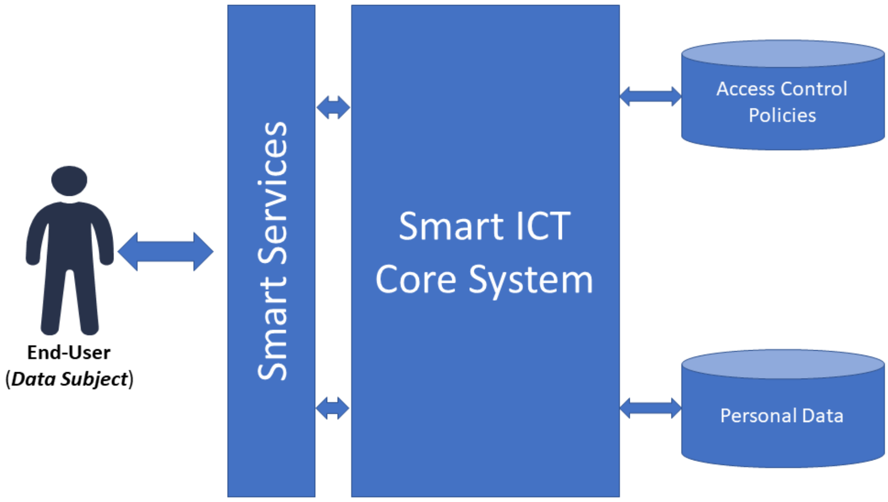


Figure 8.1: *A Smart ICT System.*

Smart ICT Core System is also in charge of managing the resource and data access by using either customized facilities or by relying on a specific Access Control System. To this purpose, in Figure 8.1 an Access Control Policies repository has been considered.

8.2.2 Consent Management

One of the aims of the GDPR is to empower individuals and give them control over their personal data, and consent is the legal basis to support that control.

As stated in Chapter 2, Section 2.2, according to the GDPR, Art. 4.1, consent "means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her".

In general consent authorizes data sources to provision data to the data consumer and authorizes data requester to process that data. Consent has to refer to a data usage policy that can be linked to a consent formalization. Consent needs to be given in a clear manner so that the data controller can demonstrate that a valid consent has been given. In particular consent record should demonstrate: a) who consented b) when they consented c) what was consented d) how was consented c) whether a consent withdrawn occurred.

A Consent Manager has the goal to support the entire lifecycle of consent management; it enables the data subjects to trace and manage the given consents in a transparent manner, and controllers to share data among digital services according to GDPR requirements. To do this a consent manager has to address to specific requirements of user-centeredness, transparency, standardization and interoperability.

In the last years, several initiatives and working groups have contributed on a standard of a structured, simplified and machine readable formalization of a data processing consent. For instance, the Kantara Initiative has released the first consent receipt (CR) specification [2] with the purpose to decrease the reliance on privacy policies and enhance the ability for people to share and control personal information.

According to this specification, a consent receipt is a notice, created from a record of consents, given to an individual when he/she provides the agreement for the collection, use and sharing of his/her personal information. Kantara's CR specification provides a standard and machine-readable structure of the record so as to addressing the "proof of consent" principle of the GDPR.

Usually, Consent Manager has to manage consents in a distributed environment: several data sources and data requesters, belonging different organizations, can interact each others. Thus, it has to assure at the same time control and transparency to the individuals and to make clear how and for what purpose personal data is used. To this purpose, in [222] a privacy-centered architecture is proposed that integrates data security and semantic descriptions into a trust-query framework, enabling the provision of user consent as a service. This framework is based on the MyData approach [3] and provides tools for creating a service contract that honors rules for data exchange. The key concept of MyData is the " MyData operator" for service formalization and user centric consent management. Through the " MyData operator", individuals can control the use of their personal data across service, grant or deny access to data or define the service activities on their data.

8.2.3 Related Work

Over the last years, different solutions have been proposed for the enforcement of the GDPR compliance into the Smart ICT Systems. They can be roughly divided into the following categories:

- Solutions applicable at Business Processes level, i.e., mainly focused on the behavioral aspect [11, 28, 55, 219].
- Proposals providing supporting facilities for transforming the GDPR's text into executable access control policies. In this case, the policies are either systematically derived from the GDPR, e.g., [32, 83] or generated through intermediate formal structures [30, 58].
- Proposals easily enforceable into the Smart ICT Systems architecture. They can be roughly classified into: i) those using access control mechanisms for the protection of personal data within Smart ICT Systems perimeters [108]; ii) those using Smart ICT Systems users location information for authenticate the customer and manage

[2] https://kantarainitiative.org/download/7902/

[3] https://mydata.org/wp-content/uploads/sites/5/2020/08/mydata-white-paper-english-2020.pdf

his/her data [110]; and iii) those exploiting specific security attributes for assuring the GDPR compliance [25, 57, 123].

Our answer to the RQ wants to merge the best practices of the identified above research areas. Indeed, we are proposing the integration of the consent and access control management for enhancing the a generic ICT System with specific specific layer aiming at guaranteeing the the GDPR's demands.

8.3 A Privacy-By-Design Proposal for Smart ICT Systems

In this section, we answer the RQ presented in the introduction of this chapter by integrating consent and access control management for assuring compliance with a reference data protection legal framework, i.e., the GDPR.

In the remainder of this section, details about the proposed reference architecture are provided in abstract terms.Indeed, they are our positive answer to the RQ. Additionally, to remark the feasibility of the proposed solution, we also provide its possible instantiation by using GENERAL_D, the proposal of this thesis, and a real-world system coming from industrial context. Details of this integration are reported in Section 8.4.

8.3.1 A Privacy-By-Design Smart ICT System

In this section, we describe the *Privacy-By-Design Smart ICT System* layer, that provides features for interacting directly with smart services and end-users of the system to guaranteeing compliance with the European Data Protection regulation.

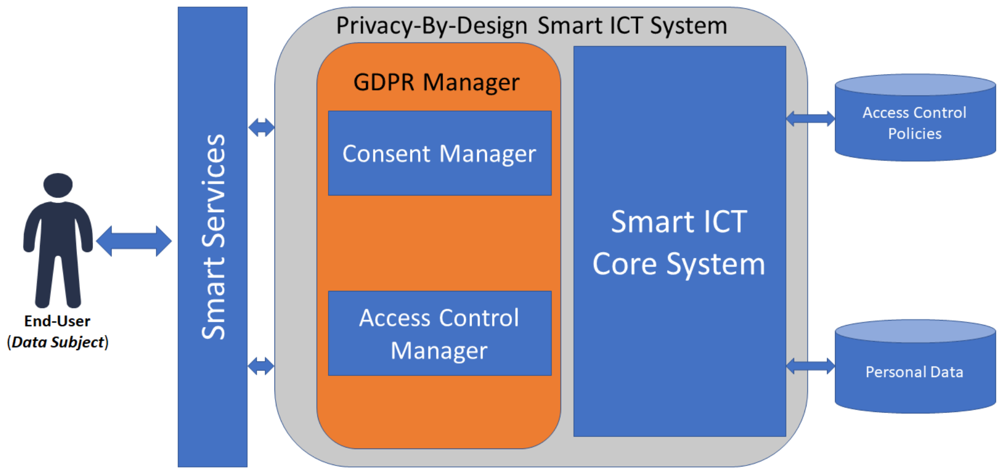


Figure 8.2: *A Privacy-By-Design Smart ICT System Proposal.*

By referring to Figure 8.1, the extended architecture is schematized in Figure 8.2, where the *Privacy-By-Design Smart ICT System* is represented by the external grey square. As in the figure, this layer has the responsibility to interact with end-users (in our case Data Subject and the Smart Services) of the Smart ICT system. It is also in charge of managing all the domain dependent activities that are necessary for the end-users interactions.

More precisely, the *Privacy-By-Design Smart ICT System* includes: (1) the components already part of the Smart ICT core systems (described in Section 8.2), and represented as blue square labeled *Smart ICT Core System*; and (2) a new layer called *GDPR Manager*, which is in charge of the translation and enforcement of executable access control policies. This is represented by the orange square labeled *GDPR Manager*, and detailed more in the remainder of this section.

8.3.2 GDPR Manager

The *GDPR Manager* includes two main components (see Figure 8.2): *Consent Manager* that provides a structured representation consent, and *Access Control Manager* that allows obtaining enforceable access control policies based on the collected consent.

Consent Manager. The aim of Consent Manager is to manage and control personal data during the interaction among Data Subjects and public and private services as Data Controller and Processors (e.g., PA, Social, IoT, B2C). It provides facilities for lawful data sharing processes, with the ability to grant and withdraw consent to third parties for accessing own personal data. Concerning Smart Services, the Consent Manager should allow them to define specific purposes for each operation (i.e., processing activities) and the data needed to accomplish the required tasks lawfully. Thus, the Consent Manager should include a consent-based, user-centric interface enabling: (1) the data subjects to manage, trace their own data and its associated consent; (2) the data controllers/processors to use consent to data sharing among digital services using personal data and meet the GDPR's requirements. Additionally, Consent Manager should guarantee by-design the compliance with the GDPR's demands, such as data minimization and purpose limitation principles.

Access Control Manager. Access Control Manager has the responsibility of creating ACPs that are compliant by-design with the GDPR. It works in collaboration with the Consent Manager by receiving, as input, the machine-readable specification of services definitions and the related Data Subjects' consents. More precisely, it uses: Personal Data related to Data Subject classified in categories as required by the GDPR; information about the Controller of each service and the defined purposes; the consent given by the Data Subject in terms of relation between Personal Data and Purposes. Based on that information, Access Control Manager is able to create specific Access Control Policies, each related to specific consent given by data subject so as to obtain an enforceable version. The peculiarities of the Access Control Manager are the possibility (a) to be integrated with different Consent Managers, and (b) to collaborating with different Access Control systems. This in order to guarantee the independence with specific input and output formats, and to be easily enhanced with standardized Access Control Systems, such as the one offered by the XACML standard, e.g., when this component is missed in the *Smart ICT Core System*.

8.4 Proof-of-Concept

In this section, we provide an instantiation of the architecture presented in the previous section by using real artefacts: CaPe [4](industrial open-source product) and GEN-

[4]https://www.cape-suite.eu/

ERAL_D (the proposal of this thesis). More precisely, we will show how they can collaborate and easily be integrated to achieve the GDPR compliance.

8.4.1 Use Case Scenario

To better explain the use of CaPe and GENERAL_D framework, we consider the following application example sets into a wellness environment. Alice, a Data Subject, wants to use a smart wellness application to monitor her daily activities to achieve a predefined training objective. The application is provided by the myWellness company (Controller). To meet Alice's needs, myWellness has so far defined different purposes, each related to a specific data set of Personal Data. At the time of subscribing to the myWellness application, Alice provided her personal data (i.e., Age, Gender, and Blood Cholesterol) and gave her consent for one purpose (i.e., MyCholesterol). Additionally, Alice gave her consent to share her personal data with a third-party company named zzz-HealthOrg company. In turn, myWellness gave to Alice controller's contacts that include: piiController, orgName, address, e-mail, and phone number.

8.4.2 Consent Manager: CaPe at Glance

CaPe provides an ICT suite for a consent-based, user-centric personal data management. It follows MyData [5] principles to exploit the potential of personal data, facilitates its control and new business opportunities in compliance with the GDPR. Thus, CaPe assures the following features: i) Consent authorizes Data Sources to provision data to Data Consumer and authorizes Data Requester to process that data; ii) Consent refers to a Data Usage Policy that can be linked to consent formalization; iii) Consent is given in a clear manner so as to let the data controller to demonstrate that a valid consent has been given; iv) Consent record clearly includes 1. Who consented; 2. When they consented; 3. What was consented; 4. How was consented; 5. Whether a consent withdrawn occurred.

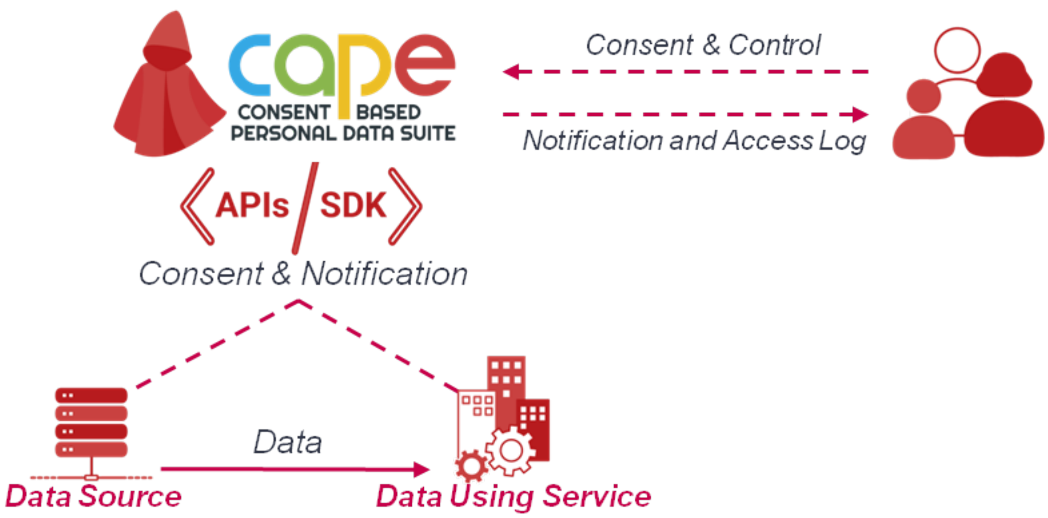


Figure 8.3: *Overview of the CaPe Consent Manager. Adopted from [80].*

[5] https://mydata.org/

Figure 8.3 shows an overview of the CaPe Consent Manager. As in the figure, CaPe acts as an intermediary for the communication between data subjects and data controllers, supporting the generation and management of dynamic consents.

As shown in Figure 8.4, the CaPe provides also two specific dashboards (the Data Controller Dashboard and the User Self-Service Dashboard) for let the overall management of the personal data management. Additionally, through these interfaces, CaPe provides specific features to **grant and withdraw consent** to third parties for **access** to data about oneself.

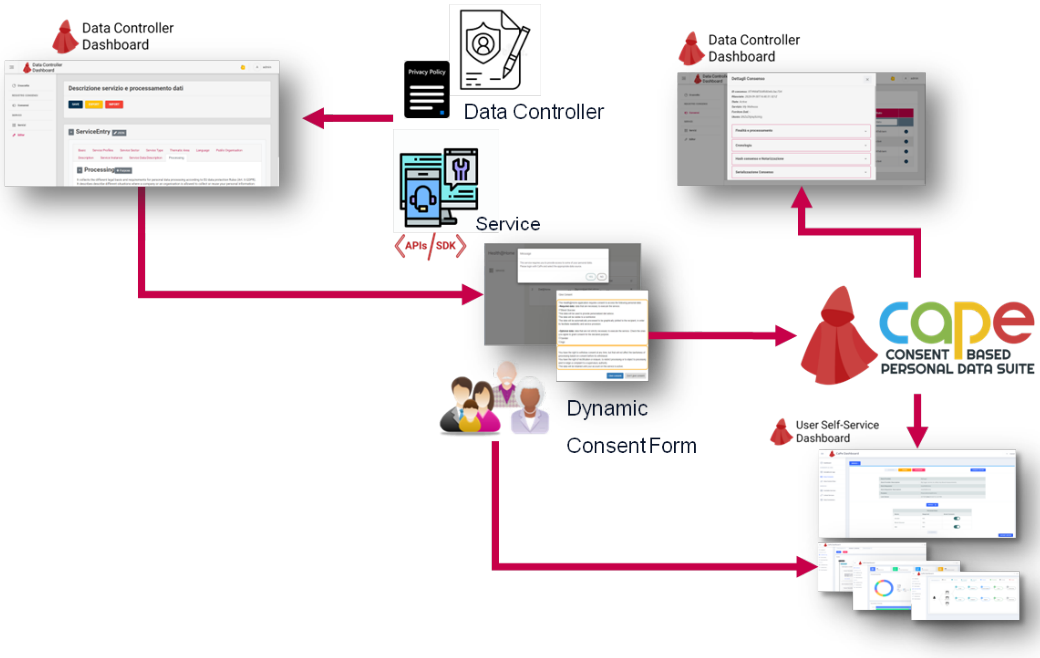


Figure 8.4: *How CaPe Works. Adopted from [80].*

A general use of CaPe is provided in Figure 8.4. In the depicted scenario, through the *Data Controller Dashboard* an organization (e.g., an SME) can model the legal basis for the processing of personal data: in a standardized manner; in accordance with the relevant information (i.e., purpose, processing, type of data and so on); and in line with the related privacy policy. According to the derived model, CaPe automatically generates the consent form that can be shown to the data subject. The two separated dashboards can let, on one side, the Data Controller to view and manage all the consents collected, on the other, the Data Subject, through the *User Self-Service Dashboard*, to check which data is used, how and for what purpose and to manage the related consents.

By referring to the reference Use Case Scenario previously introduced in Section 8.4.1, for confidential reasons we report in Figure 8.5 just an extract of the consent model derived by CaPe. As in the figure, the Consent is modeled as an entity having a unique ID identifying it and a status (Active, Non-Active).

A *Data Subject* is identified by its ID, and it is related to a set of *Personal Data*, each represented by a name/value pair. The Data Subject can give a *Consent* for processing his/her for a specific *Purpose* defined by the *Controller*. Each Purpose has a name and it is implemented by means a set of *Actions*. During the given consent phase, the Data

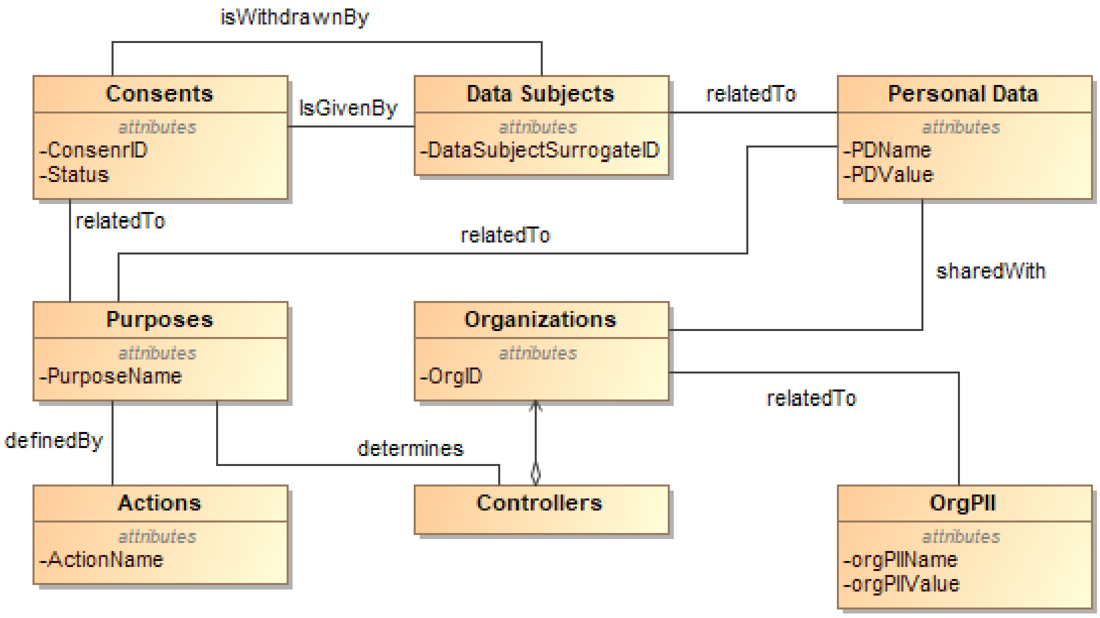


Figure 8.5: *Extract of the CaPe Consent Model.*

Subject can choose also to share his/her Personal Data with one or more *Organizations*, so as the controller can eventually achieve the defined purposes. As defined in Art. 7 of the GDPR, Data Subject can withdraw at any time the given consent. In the defined model, this is modeled as the *withdrawnBy* association between Data Subjects and Consent entities reported in Figure 8.5.

In the current implementation, CaPe encodes the instances of the defined model as Json files, and it then provides such files to the GENERAL_D for the aim of making the given consent directly enforceable by the Smart ICT Core System.

8.4.3 Access Control Manager: GENERAL_D

As reported in Chapter 3, GENERAL_D can be customized for different scenarios and consequently instantiated by using available tools presented in both academia and industry. In Figure 8.6, we illustrate its customization in the context of the proposal of this chapter. As in the figure, GENERAL_D instantiates the Access control Manager, and it is composed of three main components (see Figure 8.6): Json Manager; ACP Manager; and DBs Manager.

Json Manager has the responsibility to interact directly with CaPe described in the previous section. It receives the consent in Json format, and it parses that consent so as to extract the relevant information for the ACPs generation purpose. Such information includes, among others, the *Consent ID* and the *Consent Usage Rules* that contain the defined purposes of processing, the allowed operations, and Personal Data provided by the Data Subject.

ACP Manger is the core component of GENERAL_D framework. It has the responsibility of creating enforceable ACPs encoded in the XACML language. It interacts with: 1) Json Manager for retrieving the data to be processed (e.g., the Controller's

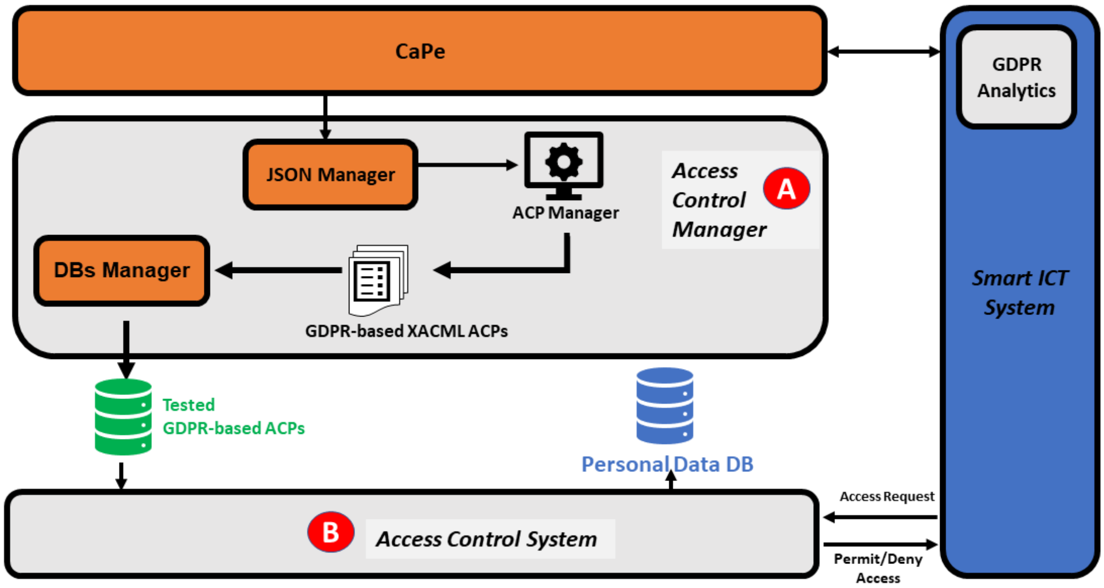


Figure 8.6: *Overview of GENERAL_D Access Control Manager.*

data, the defined purposes, the list of allowed third parties); 2) User Stories Manager for receiving the ACPs templates to be filled with those data. Therefore, ACP Manager combines the received data for deriving XACML policies, that it stores in the Access Control Policies repository.

DBs Manager offers databases supporting functionalities to the ACP Manager (e.g., create/modify/delete database, and insert/modify/delete specific entries in the available tables). In the considered implementation, DBs Manager relies on the MySQL Data Base Management System.

By referring to the Alice's activities in the myWellness' application scenario presented at the beginning of this section, and to CaPe's consent model shown in Figure 8.5, the algorithm implemented by GENERAL_D Framework, for deriving enforceable XACML policies, is reported in Algorithm 3.

From the behavioral point of view, the algorithm implements three scenarios:

1. Data Subject (Alice) gives his/her consent to Controller (myWellness company). In this case the XACML-based ACPs are generated from scratch and loaded into the database. These policies will be then made enforceable as soon as the Aclice's activities start, i.e., during the production phase.

2. Data Subject modifies his/her consent, e.g., Alice wants to modify her given consent so as to allow the management only to two of the three initially provided Personal Data. This involves the withdraw of the previously given consent and its substitution with a new one. In terms of the access control policies, this means to modify the related ACPs in DENY-ALL policies and create new ACPs for the modified consent. This behavior is in line with the accountability principle, because it lets the controller to demonstrate the compliance with the GDPR by showing the history of both the consent and the related ACPs modifications. Specifi-

cally, it refers to the transparency principle (Art. 5.1(a) “lawfulness, fairness and transparency”) and Art. 30 (“Records of processing activities”).

3. Data Subject (Alice) withdraws the given consent: i.e., prevent any access to Personal Data belonging the Data Subject. In terms of access control, this means to deny any access requests to those data. Practically, this can be enforced by the ACP Manager by setting the related ACPs to DENY-ALL.

From a procedural point, as shown in Algorithm 3, through the Json Manager component, GENERAL_D parses the Json file for retrieving the data of interest, i.e., Personal Data, Purposes and the third parties those data are shared with. Then, through the joint collaboration of ACP Manager and User Stories Manager, the ACPs templates can be instantiated for generating XACML-based policies that are compliant with the GDPR (GENERAL_D Framework’s outcomes).

In details considering the Algorithm 3:

- **line 1-4**. The algorithm takes as input the consent represented in Json format (`CJF`), and parses that file by obtaining its internal representation (`CJFAsPOJO`, line 4).
- **line 5-9** In case of active consent (Algorithm 3, line 6), the algorithm verifies whether the processed consent is a modification of an already given one. In case of modification, the related ACPs derived so far are modified to DENY-ALL policies (Algorithm 3, line 8).
- **line 10-13** For each consent related to a specific purpose, an XACML policy is generated (Algorithm 3.
- **line 14-15** In case of withdrawing the content, the received Json input contains the status *non-Active* (Algorithm 3, line 14), and in terms of AC, this means that no one is able to access Personal Data related Data Subject. This is reflected in denying all the incoming access requests, by triggering the default DENY-ALL policies modified in Algorithm 3, line 15.

Algorithm 3 GDPR-based ACP Derivation

1: **input:** CJF ▷ Consent as Json File
2: **output:** GAL ▷ GDPR-based ACP List of XACML policies
3: $GAL \leftarrow \{\}$
4: $CJFAsPOJO \leftarrow parse(CJF)$
5: $cID \leftarrow CJFAsPOJO.getCID()$
6: **if** $CJFAsPOJO.isActive()$ **then**
7:     **if** $isAlreadyGiven(cID)$ **then**
8:         $DenyAllPolicies(cID)$
9:     **end if**
10:     **Foreach** $c_j \in CJFAsPOJO$ **do**
11:         $ACP \leftarrow CreateACPS(c_i, cID)$
12:         $GAL.add(ACP)$
13:     **end for**
14: **else if** $!CJFAsPOJO.isActive()$ **then**
15:     $DenyAllPolicies(cID)$
16: **end if**
17: **return** GAL

By referring to the previously presented Use Case scenario (see Section 8.4.1), and by applying Algorithm 3, we illustrate in Figure 8.7 one of the obtained ACPs

```
Policy .......................................... PolicyId = alicePolicy
                                                  root element
                                                  rule-combining-algorithm:permit-overrides
├── Target ...................................... Lawfulness of Processing Sample Policy
│   ├── Subject ................................. Controller.orgName = myWellness
│   └── Subject ................................. ThirdParty.orgName = zzz-HealthOrg
├── Rule ........................................ RuleId = readRule, Effect = Permit
│   ├── Target ..................................................................................
│   │   ├── Resource ............................ Age
│   │   ├── Resource ............................ Gender
│   │   ├── Resource ............................ Blood Cholesterol
│   │   └── Action .............................. Action.Purpose=MyCholesterol
│   └── Condition ...............................................................................
│       └── And ................................. And Operator
│           ├── string-one-and-only ............. type-One-And-Only Function.
│           │                                     #Resource = 1
│           ├── string-equal .................... type-Equal Function.
│           │                                     Resource.owner = AliceID
│           └── string-is-in .................... Action ∈ {AGGREGATE,
│                                                 COLLECT, DERIVE, QUERY}
└── Rule ........................................ RuleId = defaultRule, Effect = Deny
                                                  default: deny all, which is not allowed explicitly.
```

Figure 8.7: *An XACML-like Policy authorizing Lawfulness of processing of Personal Data based of the Consent Given by the Data Subject (Art. 6.1(a)).*

in XACML-like format. This policy rules access to the Personal Data according to Art. 6.1(a) of the GDPR, i.e., when the processing activity is lawful based on the consent given by the Data Subject. As in the figure, the allowed subjects, to access Alice's Personal Data, are both myWellness (Controller) and zzz-HealthOrg (Third Party). This is expressed in the Target element of the XACML policy. As specified in the derived rule, the purpose of processing (i.e., MyCholesterol purpose) is achieved by allowing access to perform a specific set of Actions, i.e., AGGREGATE, COLLECT, DERIVE and QUERY.

Remark. Smart ICT Systems are gaining a certain amount of attention in the last years. They provide means for developing Smart Services in different domains such as Smart-Cities, Education, and Healthcare environments, to cite a few. However, with the entering into the force of the EU data protection regulation, i.e., the GDPR, proposed solutions for Smart ICT Systems lack appropriate supports to aid Controllers in developing Smart Services in compliance with the regulation. More precisely, the proposed solutions are not Privacy-By-Design conceived, i.e., the implemented services are not compliant with the GDPR from the early stage of their design. To overcome these difficulties, we have conceived a possible generic architecture that can be customized with real artifacts to accomplish the GDPR compliance. We have also provided a proof-of-concept consisting of the integration of two new tools coming from the industrial and academic sectors: CaPe (industrial product) and GENERAL_D Framework (an outcome of the current research). This integration has demonstrated the applicability and flexibility of our Privacy-By-Design solution for Smart ICT Systems.

Future work, we are planning to validate our approach by considering different Smart-Cities environments. the integration CaPe and GENERAL_D therefore will be validated in the currently available and emerging Smart-Cities platforms, such as the ones based on the market-ready open-source software FIWARE platform. In particular, currently, the integration is being validated within CyberSec4Europe project by considering a real Smart-Cities IoT platform provided by one of the partners of the project.

CHAPTER 9

GENERAL_D & Business Process

CURRENTLY, the scientific communities and private companies are actively working to provide theoretical and practical solutions for enforcing the adoption of the GDPR and its compliance problem. In line with the principle of data protection by design, the chapter proposes an approach for the automation and enforcement of GDPR requirements. The idea is to extend the currently adopted access control mechanisms so to leverage them to the enforcement of GDPR compliance during business activities of data management and analysis. From a practical point of view, this means to integrate into the existing business processes specific facilities for assisting in the design, development, maintenance, and verification of the GDPR requirements as well as to modify the language and architecture of the access control systems so as to let the management of GDPR principles and obligations. For this, the basic steps of the proposed approach are provided as well as an example used to clarify the integrated use of access control systems and business process models.

This chapter refers to the Application Example 4 described in Chapter 3, Section 3.5.4. In order to support performing the required steps for authoring GDPR-based ACPs, we have customized our reference architecture GENERAL_D (see Figure 3.2) as depicted in Figure 9.1. More precisely, the customization involves Module A by specializing the component *Legal Text Analyzer* with the *Business Process Analyzer* one.

This chapter is based on the related publication:

- Antonello Calabrò, **Said Daoudagh**, Eda Marchetti: Integrating Access Control and Business Process for GDPR Compliance: A Preliminary Study. ITASEC 2019

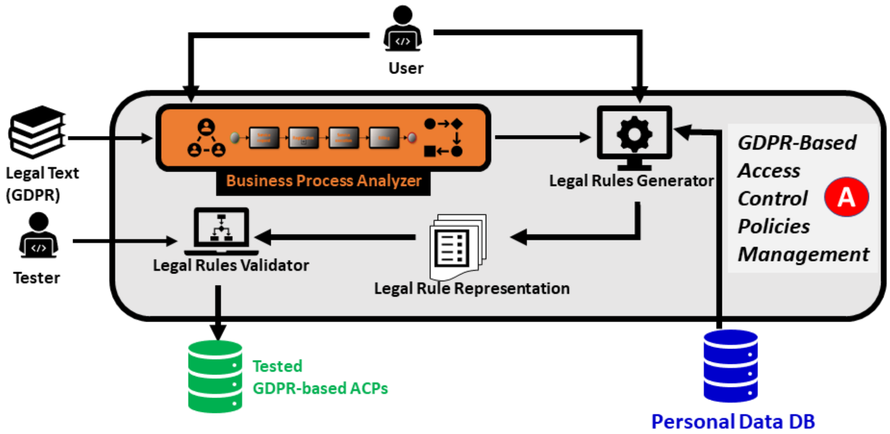


Figure 9.1: *GENERAL_D Customization for Handling Example 4.*

9.1 Introduction

The purpose of the GDPR is to harmonize the regulation of Data Protection across the EU member states and, at the same time, to enhance and to arise business opportunities within the Digital Single Market space. GDPR imposes several limitations of processing personal data and provides several provisions, defining responsibilities and fines in case of non-compliance.

In general, implementing GDPR requirements and demonstrating compliance, and therefore avoid the related penalties, is not a trivial problem. From a practical point of view, this issue can be reloaded as: making a given Data Management System (DMS) comply with the GDPR legal requirements, and providing the necessary information and evidences so that a supervisor authority could accept this as evidence of the compliance.

Currently, the scientific communities, as well as private companies, are actively working to provide theoretical and practical solutions for enforcing the adoption of the GDPR and its compliance problem. To facilitate this process and tackle the principle of *data protection by design*, contained in Art. 25.1 of the Regulation, an important step is the automation the enforcement of GDPR requirements. From this the idea of this work: improving the currently adopted security services and access control procedures so to leverage them to the enforcement of GDPR compliance during business activities of data management and analysis.

Indeed, the current trend of increasing automation and data exchange promoted by the Industry 4.0 is encouraging many industrial realities to the adoption of visual models, called Business Processes (BPs), to easily manage the assignment of tasks, the interactions between the different roles, and the changes in the organization or in the business activities [97]. Thus, in many enterprises (e.g., in SMEs), especially large ones, this means to integrate into their business processes specific facilities for authorization and access management so as to target the GDPR needs.

From a practical point to view, there are several possible ways to model a BP; perhaps the most popular and widespread adopted is Business Process Model and Notation (BPMN) [182], which provides a visual representation supported by a formal XML specification. Since BPMN is an extensible standard, it is possible to empower it to express activities related to data protection [27, 140, 233]. The main benefits of BPMN commonly rely on the possibility of having a clear and standard notation for creating a description of processes (in terms of participants and activities) and develop executable frameworks for the overall management of the process itself. Directly integrating, through the usage of security services, the GDPR requirements into the business process execution represents a key aspect both for privacy management and assurance [140, 233].

Following this idea, the solution presented in this chapter relies on two pillars:

- exploit the BPMN models for assisting in the design, development, maintenance, and verification of a system in order to comply with the GDPR requirements. This includes the detection of possible violations, with the objective to minimize the risk of sanctions from the supervisory authority;
- leverage the authorization systems, and in particular the access control ones, to tackle the problem of the GDPR compliance. This includes their integration into business process activities as external authorization service.

The BPMN leveraged with the appropriate mechanisms for the GDPR definition and compliance can provide a number of benefits:

1. it can be used by controllers processing personal data for having a clearer view of their duties with respect to data protection in the context of their business;
2. it can be used to check if the BPMN is compliant with the requirements imposed by the GDPR;
3. it can automatically suggest and perform the mandatory activities and obligations to be met to achieve the GDPR compliance;
4. it can be used to discover when specific GDPR obligations are not fulfilled at runtime; and finally,
5. it can supply auditors and supervisory authorities with a complete view of the process and the procedures adopted for data protection.

The chapter is organized as follows: Section 9.2 presents the background about the BP modeling and discusses related work by surveying existing literature concerning the integration of the GDPR's principles into BP; Section 9.3 introduces the basic steps of the propose approach; an example in Section 9.4 shows how the proposed approach can be used; finally, Section 9.4 gives a set of conclusions and the envisioned future work.

9.2 Background and Related Work

The present work aims at use the authorization systems, and in particular the access control, into a business process so as to model the involved activities in compliance with the GDPR, and to enable the adoption of data protection by design and by default. There are many possible ways in which the two might be integrated, depending on the specific purpose; however, all the proposals are based on three building blocks: 1) the model of the business process; 2) the representation of the GDPR; and 3) the access control mechanisms for the enforcement of the GDPR requirements.

In the following basic concepts about the first topic is provided, while, and related work are discussed.

9.2.1 Business Processes

Business processes usually refer to any structured collection of related activities or tasks that are carried out to accomplish the intended objectives of an organization. The main focus is creating an abstract but meaningful representation of the real business domains and sharing a formalized definition, so as to improve expressiveness and to make easier the development of tools [124].

Usually, BPMN [182] is the formalism chosen to represent business models, which is the *de facto* standard for process modeling. It is indeed a rich and expressive language (but also a complex one) used for the tasks associated with process modeling [197].

In detail, a BPMN has four categories of graphical elements that can be used to build the diagrams:

1. *Flow Objects* are associated with the actions that can be performed in a business process and make up the behavior of the BP. They consist of Events, Activities, and Gateways;
2. *Connecting Objects* can be used to connect elements to each other in three different ways: Sequence Flows, Message Flows, and Associations;
3. *Swimlanes* give the capability of grouping the primary modeling elements. Swimlanes have two elements through which modelers can group other elements: Pools and Lanes;
4. *Artifacts* are used to provide additional information about the process that does not affect the flow.

In Figure 9.2 and Figure 9.5 examples of BPMN are provided.

To introduce the GDPR requirements in software business process modeling, an important step is provide mechanisms to extend the existing models so as to express legal provisions. Among the currently available proposals, in this work we refer to [89, 137, 174] that provides mechanism for manage content-oriented pattern and customized process views.

9.2.2 Related Work

In recent years, due to the complexity and the importance of the GDPR adoption and implementation, researchers and practitioners have devoted significant attention to the challenges raised from data protection principles policies and regulations interpretation [121]. At the same time, many supporting tools and applications have been developed to assist users in producing reports on the GDPR compliance [95].

Notwithstanding these important contributions, the integration of data protection rules into the commonly-used business processes is still an emerging challenge. In literature, a great deal of attention has been devoted either to include generic privacy aspects into the adopted business process [54, 84, 178, 219] or to assess of privacy and security analyses in all stages of system development [9, 49, 106, 220] or to verify the GDPR provisions [37, 87, 107].

This highlight the need of a standard methodology to perform an assessment of IT systems concerning privacy and security aspects especially targeting the GDPR requirements. In line with this field, the proposal of this work attempts to make easier the assessment of GDPR requirements, by explicitly integrating specified access control systems into the commonly adopted business process. In particular, the presented approach wants to integrate and extend the available proposal promoting business processes as a key solution for privacy issues [13, 37, 112, 195].

For this aim, an extension of the XACML reference architecture is promoted. In literature, there are several proposals aiming to satisfy the GDPR requirements through improvement of the reference XACML architecture. The main proposal are: [82] which focuses on authorization decision depending on the context as well as on the user's access privileges; [91] and [62] where authors designed a system that ensured the enforcement of multiple privacy policies within an organisation and throughout a distributed system; [189] which proposes a proof-of-concept implementation for the IoT environment where the security between the XACML reference architecture components was addressed; [189] where the proposed architecture is an integration oriented proposal aimed to make XACML easier to use by other systems.

Differently from the provided solutions, our idea is to decouple the authorization functionalities from the business logic. This lets to adapt and extend the XACML reference architecture with new features without modifying the business logic of the applications that use and consume Personal Data. Separation of concern from the architectural point of view should help one to propose scalable, manageable and extendible authorization solution.

9.3 Approach

There are different proposals addressing specific data protection principles by leveraging authorization systems (see for instance [185, 196]). However, currently only few are targeting the GDPR compliance problem and proposing access control systems as a key solution [37]. Indeed, as they are, the current access control mechanisms and techniques are not able to either satisfy the GDPR constrains or be easily integrated into the business process steps.

The proposal of the work presented in this chapter is to move a step ahead and provides a comprehensive methodology that combines, merges and integrates the access control system into the BP so as to address different aspects of the GDPR compliance problem. For aim of simplicity, here we restrict ourselves to the provisions directly related to access control.

Therefore, on the bases of the process presented in Chapter 3 and Example 3 of the same chapter, the approach adopted in this work for integrating the access control systems into the business process activities consists in the following steps:

GDPR-based use case definition (step ①) and Gather authorization requirements (step ②). This steps aim at analyzing the business process activities to establish a common basis to discuss with different stakeholders. This allows to identify only those activities that can be affected by the GDPR requirements, and for each of them the GDPR articles affecting it are detected. Therefore, the affected activities will be extended/substituted with sub-processes compliant with the GDPR specifications so as to enforce the GDPR provisions and make easier requirement reviews. To make easier this step a pre-defined set of sub-processes will be provided. These can include specific patterns that allows for example to obtain the consent, to withdraw the consent, or to transfer data to third-parties.

Identify required attributes (step 3). This step consists of (a) identifying the attributes involved in the (extended) activity by the GDPR, and (b) classifying them into the commonly used categories in AC (i.e., Subject, Resource, Action, and Environment).

Author authorization policies (step 4). This step aims at developing enforceable access control policies for each affected activity. To this purpose, we can leverage the result of Chapter 7 and Chapter 8 by using the ACPs templates. In particular, for each affected activity, ACP templates are identified, and for each of them a enforceable policy is generated by using the attributes and data involved in that activity,

Test ACPs & AC mechanisms (step 5). i.e., to ensure that the implemented XACML policies meet the GDPR requirements. State-of-the-art and specifically conceived testing techniques should be used according to the different purposes. This step involves also the evaluation of the adequacy of the current AC mechanisms in the context of the GDPR.

Deploy the architecture (step 6). This step aims at defining the contact point with existing systems. Since this step is usually business-dependent, a specific Policy Enforcement Point (PEP) will be defined for each application that interact with the authorization system. In particular, here we refer to implementation of those activities that are affected by the GDPR, and identified in Step 3. As stated in Chapter 2, Section 2.3.1, the flexibility of the XACML reference architecture allows to decouple the authorization functionalities from the business logic, i.e., separation of concern from the architectural point of view. This lets to adopt authorization systems as a services. The only effort a developer face is the implementation of the specific PEP based on the language used for executing business process, and more precisely, those activities affected by the GDPR.

Deploy the policies (step 7). This step requires the deploying the authored XACML policies according to the selected (production) environment, that realizes the Business Process Execution. Even this step is business-dependent, having the authorization as a service (see step 6) facilitates performing the deployment of the policies into the access policy repository, which is at this point depending on the authorization system used.

Run access reviews (step 8). This step aims at analysing the developed policies against a set of attributes to determine what these attributes grant.

In the next section more details about the proposed approach will be provided through a simplified, typical and realistic running example.

9.4 Application Example

We illustrate the proposed approach through a simple example related to a standard process for service provisions by a specialist/professional. Thus, the use case scenario considered is depicted in Figure 9.2, where the basic activities have been shown by means of a generic Business Process Model (BPM).

Figure 9.2: *Generic business process*

As in the figure, the BPM has four main phases (activities):

- *Service request*, in which the *Customer* and professional (*Seller*) establish the first contact and agree about the provisioning of a service;
- *Registration*, in which, for starting the collaboration, the (*Seller*) collects the *Customer* data (if he/she has not been already registered for the same service);
- *Service execution*, during this activity the requested service is provided;
- *Billing*, in which the collaboration ends with the production of an invoice.

As described in Section 9.3, during Step ② (*Gather authorization requirements*), the GDPR requirements relative to the activities of the BP are explicitly listed. It is part of this stage the identification of the data types affected by privacy constraints, the primary purpose of data collection as well as the optional purposes that could involve the data management. Additionally, the activities related to collecting, reading, storing, transmitting, or deleting personal data that have to be compliant with the definition provided in Art. 4.2 of the GDPR are also identified.

In Figure 9.3(a) the simplified version of the consent request form is provided. As in the figure, optional purposes can be also included, such as the usage of the customer's e-mail of physical address for sending: i) un-target news or advertisements (newsletters); ii) specific target marketing based on the customer's history.

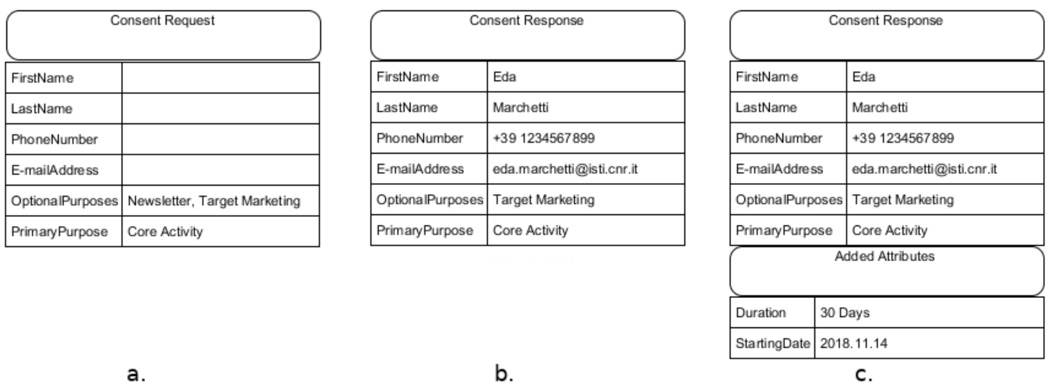


Figure 9.3: *a. Form request - b. Form response - c. Form response enriched*

Afterward, during Step ③ (*Identify required attributes*, the activities affected by the GDPR provisioning are highlighted in the BP model. These will be substituted or extended by specific activities or sub-processes in order to guarantee the GDPR compliance. In the considered, example only the *Registration* activity has been highlighted as critical from the GDPR point of view (see Figure 9.4), and therefore, improved with a set of compliant GDPR sub-tasks as shown in Figure 9.5.

Figure 9.4: *Enhanced Business Process Model*

In particular, Figure 9.5 details the new sub-process provided. Here, the *Customer* has been identified with the data subject and the *Seller* with the controller.

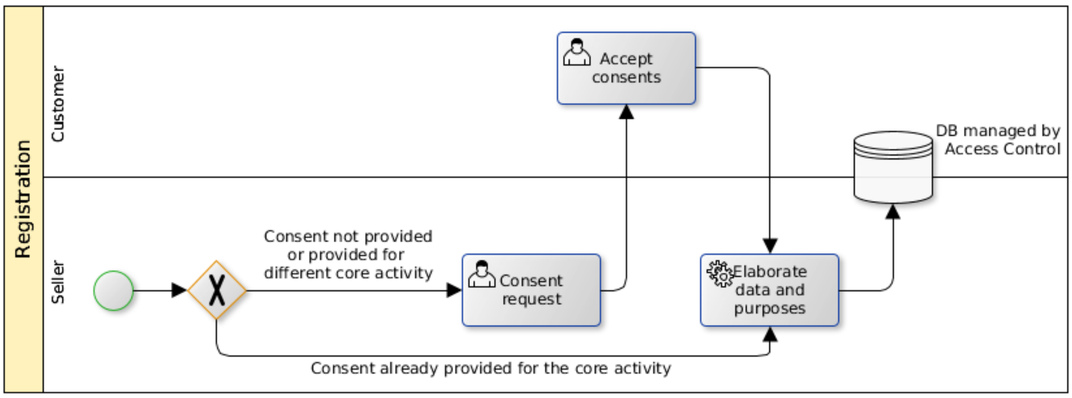


Figure 9.5: *Registration sub-process*

In this sub-process, the *Seller* checks if the *Customer* has already provided consent to the required service. If not, the *Seller* prepares the *Consent request* according to the form shown in Figure 9.3(a) and sends it to the *Customer*. This last fills the form with required data (see Figure 9.3(b)) and sends it back to the *Seller*.

According to the approach proposed in Section 9.3 the task `Elaborate data and purposes` implements the steps from Step ④ (*Author the authorization policy*) to Step ⑦ (*Deploy the policy*). It is in charge of converting the information collected into `XACML` policies/attributes encoding the GDPR principles and, setting up the access control mechanism in order to rule the data access through a common database.

Figure 9.3(c) shows an abstraction of attribute considered for policy specification. As in the figure, two additional attributes (Duration and StartingDate) are included in order to satisfy Art. 17 of the GDPR.

The procedure of how to extract access control policies directly from the GDPR's text is discussed in Chapter 6, and we refer to it for more details.

Figure 9.6 shows an extract of the policy derived using the data of Figure 9.3(c). The extended access control architecture will use the policy for ruling the access the database so as to guarantee the online GDPR conformance.

`Policy`	root element rule-combining-algorithm:permit-overrides
`└─ Target`	Sample Policy rules the access to the attributes of the *Customer*.
`   └─ AnyOf`	Or Operator
`      └─ AllOf`	resource-attributes = {Eda, Marchetti, +39 1234567899, eda.marchetti@isti.cnr.it}
`└─ Rule`	RuleId = rule1, Effect = Permit
`   └─ Target`	The *Seller* can perform any action for the *Core Activity* and / or *Target Marketing* purposes within 30 days starting from the *StartingDate*.
`      └─ AnyOf`	Or Operator
`         └─ AllOf`	PrimaryPurpose = Core Activity
`         └─ AllOf`	OptionalPurpose = Target Marketing
`   └─ Condition`	$30 \leq$ CurrentDate - StartingDate
`└─ Rule`	default: deny all, which is not allowed explicitly. RuleId = rule2, Effect = Deny

Figure 9.6: *An XACML policy using the data of Figure 9.3(c).*

According to Step 6 (*Test the policy*), before deploy the XACML policy on the access control system, an accurate testing activity is also performed so as to avoid possible security or privacy flaws. We refer reader to Chapter 5 for more details about how to test both access control policies and mechanisms.

Remark. Since the GDPR was about to be finalized, theoretical research and industry have started addressing the issue of compliance with the new Regulation. The idea of having integrated solutions that supports compliance throughout the various stages of the software development life cycle of data processing applications is in itself very simple, but its realization is far from that. This chapter moved a step ahead in the direction of *data protection by design* by improving the currently adopted security services and access control procedures. The target was to leverage them to the enforcement of GDPR compliance across the business activities related to data management and analysis.

From a technical point of view, the BPMN was selected as the target model to integrate in the business process the access control mechanism. Indeed, BPMN is a simple yet effective means of modelling a flow of activities (both man-made and automated). Of course, different modeling languages could have been considered. However, our proposal aimed to focus on the underlining idea than on the technical implementation details.

To exemplify the proposal presented herein, the basic steps of a feasible approach are provided. Moreover, an application example has been used to clarify the adoption of access control systems for protection of personal data during the BPMN modeling and execution.

As a future work we would like to prototype the proposed approach including the features for: extending the BPMN with specific patterns to perform modelling activities required to achieving compliance with GDPR; defining a comprehensive strategy for highlighting inconsistencies, with respect to the GDPR' obligations, during the Business Process execution.

CHAPTER 10

GENERAL_D & Indoor Localization Systems

NOWADAYS, availability of mobile devices have lead to an arising development of (indoor) location-based services that share a huge amount of (personal) information and data. However, without an accurate and verified management, they could become severe back-doors for security and privacy. Therefore, in this chapter we show, for the first time, how to integrate GENERAL_D within an indoor positioning infrastructure so as to internally guarantee by-design the enforcement of the GDPR's provisions. A prototype example is also provided for feasibility purposes.

This chapter refers to the Application Example 5 described in Chapter 3, Section 3.5.5, and the customization of GENERAL_D is depicted in Figure 10.2. In particular, it reports and extends the following related scientific contribution:

- Paolo Barsocchi, Antonello Calabrò, Antonino Crivello, **Said Daoudagh**, Francesco Furfari, Michele Girolami, Eda Marchetti: A Privacy-By-Design Architecture for Indoor Localization Systems. QUATIC 2020: 358-366

10.1 Introduction

Nowadays, the wide availability of mobile devices have lead to an arising development of (indoor/outdoor) location-based services for improving users daily life and works. In the outdoor environments, due to the Global Positioning System (GPS) diffusion, location-based services have already find their place in providing users with practical means for easily locate themselves, discover nearby facilities or being more productive and efficient in their business. In the indoor context, where GPS could not be efficient and effective, ILSs are increasing their presence and finding a very important role in daily life and business. Indeed, a lot of competitors are exploiting these kind of systems for providing commercial solutions, selling products, tracking facilities, social apps and services.

In this chapter, we refer explicitly to the indoor environment where the missing of a standardized reference architecture for ILSs is promoting a variety of solutions. They differ in the features provided and on the technical facilities which they rely and also on the security and specifically privacy enforcement they adopt. Indeed, by construction, location-based services are sharing a large amount of information and data. Consequently, special treatment should be devoted to personal data such as MAC or IP address, user localization with relative date and time, storage of the location visited, devices used and personal preferences.

Thus, hidden in the golden world of location-based services, fulls of interesting, useful, appealing features, there is an enormous amount of (personal) data that users are leaving in the different services and databases without being aware of the risk. Indeed, these data are the new gold for the commercial environment. Products are migrating from a *fee for a service* to a *free but not for free* service where the *not for free* part is paid by the personal data collected.

Beyond Snowden [100], people's sensitiveness about personal privacy, fortunately, has been increasing. The idea that someone could take personal data, images, position and behavior for developing the *architecture of oppression* stirs minds to start rebuilding a safe and healthy global software and hardware infrastructure. In parallel, as mentioned in Chapter 2, Section 2.2, the recent adoption of the GDPR [90] focuses the attention on the data protection principles: to strengthen the rights of individuals over their own data, and to make organizations more accountable regarding the previous directive. In this view, location-based services have to be built in line with GDPR provisions.

As we have discussed in the previous chapters, among security mechanisms presented in literature, one of emerging technical solution used for adequate fine-grained mechanisms that take into account legal requirements is to integrate in the systems architecture an AC solution [37, 195, 196]. However, notwithstanding the important role of AC systems, their integration inside the localization system architecture is still an emerging challenge, specifically considering the enforcement of the GDPR provisions.

Motivated by thereof, in this chapter we claim that the adoption of a specific consent manager (based for instance on the consent specification provided by Kantara initiative) and GDPR-based ACPs templates (e.g., those defined in Chapter 6 and Chapter 7) can help untangle the GDPR compliance in indoor environments. Indeed, this can be a starting point for conceiving a privacy-by-design architecture for ILSs in compliance with the GDPR's demands.

Therefore, the main contribution of this chapter is to provide a reference architecture for the indoor localization system which includes an AC system (i.e., GENERAL_D) able to guarantee the GDPR compliance.

As a result, we propose a privacy-by-design ILS architecture, where purposes of data management are explicitly defined, consents collected and the rights related to privacy and data protection correctly enforced.

To the best of our knowledge, the proposed architecture is the first attempt to integrate three separate research fields:

(1) the design and implementation of smart and easy-to-use indoor localization sys-

tems;

(2) the use of access control systems for resource and data management inside localization environment; and

(3) the enforcement of the GDPR provisions inside the localization systems.

In the next section, an overview of background and related work is presented. Then, in Section 10.3 we describe the proposed privacy-by-design solution, whereas in Section 10.4 we present an application example. Finally, Section 10.4.2 concludes the chapter.

10.2 Background and Related Work

In this section, we describe the main concepts related with indoor localization, consent manager and the integrated usage of AC inside the indoor localization systems. Additionally, their related works are also presented.

10.2.1 Indoor localization systems and location-based services

Several Indoor Localization Systems (ILSs) have been proposed in the last decade showing differences, in terms of methods and data sources, but a generally accepted cross-domain solution is still lacking and only high customized systems are starting to hit the market [192]. The main indoor localization systems and frameworks already available on the market are IndoorAtlas [1] (e.g., deployed into the Mumbai Airport for assisting travellers), Indoor Google Maps (i.e., it provides routes, places, and point of interest after a survey of the indoor area), and Anyplace [101] (i.e., it allows to define Points of Interest and to manage indoor maps). These solutions exploit the most promising ILS data sources based on the opportunistic exploitation of radio communication systems (e.g., BLE and Wi-Fi), on the usage of MEMS (e.g., inertial, pressure and magnetic sensors) generally available in mobile devices and, recently, using camera information.

Besides the positioning and localization functionalities, ILSs lead to the development of accurate Location-Based Services (LBSs). In fact, using LBS is already quite common for mobile users but LBS generally exploits WLAN infrastructures for indoor environments to determine a user's location. When the people location is known, the system can provide location related contextual information such as events, places, and point of interests or navigation for the mobile users. As a consequence, the tracking data are available to a wide number of LBS posing a risk of location privacy violation and, more in general, through these information a third party can infer personal behaviors and habits. In [136] authors show strong performances in protecting users from location tracking by the localization service based only on Wi-Fi measurements and without hindering the provisioning of location update.

10.2.2 Access Control Systems and location privacy inside the ILS

Notwithstanding its relevance, the literature is currently dedicating little attention to the problem of enhancing the privacy of indoor localization systems, explicitly considering

[1] https://www.indooratlas.com/

the enforcement of the GDPR provisions.

We refer to [113] for a systematic review of privacy in indoor positioning systems. As emerged by this survey, recent proposals evidenced how location and topology-aware are becoming security-relevant characteristics [109]. However, most of the research has been focused either on:

1. using technology Wi-Fi and the fingerprinting methods combined with cryptography solutions [129, 177, 245];
2. using access control mechanisms for (physical) protection within virtual perimeters [108];
3. using location information for automatically authenticate customer [110];
4. specific security attributes that do not fully cover the GDPR requirements [25, 57, 123].

To this purpose, location privacy is a crucial aspect for guaranteeing anonymity while using location-based services [148]. Usually, existing Location Privacy-Preserving Mechanisms (LPPMs) mainly rely on obfuscation mechanisms and anonymization mechanisms [12], or cryptography and shared information reduction mechanisms [252].

Without the pretend to be exhaustive, the analysis of the current state-of-the-art evidences that an integrated use of location information, in reference to Indoor Localization System, and access control mechanisms, able to manage the security and privacy enforcement of data managements in reference to the GDPR, still needs to be proposed.

This chapter, therefore, wants to enhance the current research by proposing for the first time a reference architecture for the indoor localization system, which includes a location and topology-aware access control system able to guarantee the compliance with the GDPR's provisions.

10.3 A Privacy-By-Design Solution

In this section, we describe the reference architecture for the indoor localization system, which includes an AC system for managing access to personal data in compliance with the GDPR. The proposal extends a solution presented in [99] and integrates the proposal of this thesis, i.e., GENERAL_D described in Chapter 3 Section 3.4. In the remainder of this section, the architecture and the behavior of the proposed solution are provided in more details.

10.3.1 Architecture

As schematize in Figure 10.1, the main components of the proposed reference architecture are: **User Agent (UA)** and **Localization Infrastructure (LI)**.

User Agent (UA). It lets the interaction between the users and the LI. It is included in the device (smartphone, tablet or smartwatch) used for the LI connection and is dependent by the nature of the available systems. As shown in Figure 10.1, the UA is in charge of managing the user interaction for:

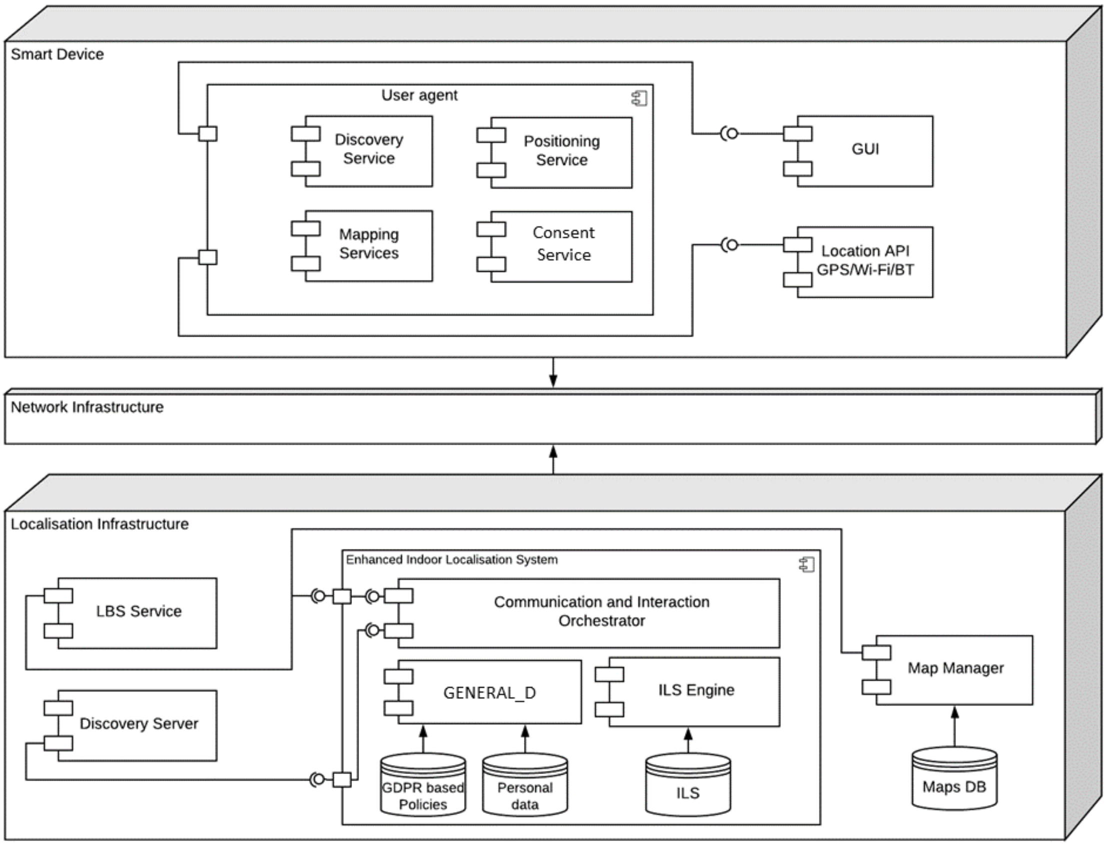


Figure 10.1: *Software architecture. Adapted from [23].*

- automatically detecting the existence of an Indoor Localization Server (ILS) (through the Discovery Service);
- allowing the localization of the device (through the Positioning Service);
- visualize the position on the indoor map (through the Mapping Service);
- accepting/denying the consents and sending/receiving access requests/responses (through the device GUI and Consent Service).

Localization Infrastructure (LI). It is a distributed infrastructure which computes the user location. It relies on the WiFi signals received by the UA and implements different features through three specific components:

1. *Map Manager:* this is a (federated) component managing the updating and storage of the internal maps.
2. *Discovery Server:* this component is in charge of sending the URL of the available ILSs to the different UAs that are currently listening, among a specific distance range. This component does not manage personal data therefore it is not under the control of GENERAL_D.
3. *Enhanced Indoor Localization System (E-ILS):* it is the core component of the LI. It has three main sub-components (Communication and Interaction Orchestrator,

the GENERAL_D framework and the ILS Engine). These components implement the main features during the execution of the service and rely on two database for collecting the required information and (personal) data. More precisely:

- the *Communication and Interaction Orchestrator* is the component in charge of managing the communication to and from the E-ILS;
- *GENERAL_D* which rules the resources and data access according to specific security policy and enforces the GDPR provisions;
- the *ILS Engine*, is in charge of estimating the User Agent's location. In turn, the *ILS Engine* returns back the User Agent its timestamped coordinates according to the map reference system (e.g., latitude and longitude as WGS84 reference system) [191].

In the next section more details about the *GENERAL_D*, a sub-component of E-ILS, are provided.

GENERAL_D in Indoor Localization Systems

Because different people (the data owner, the administrators, the guardians, the supervisors, and so on) or services (booking services, advertisement services, navigation services and so on) may ask the data access in different moments or situations, the *ILS Engine* and *GENERAL_D* have to work in strict collaboration. Indeed, this last component is in charge of evaluating each single data access request and allowing or denying the access according to the consent collected, the data validity period, the specific users/service rights and the access control policies established inside the overall Localization Infrastructure.

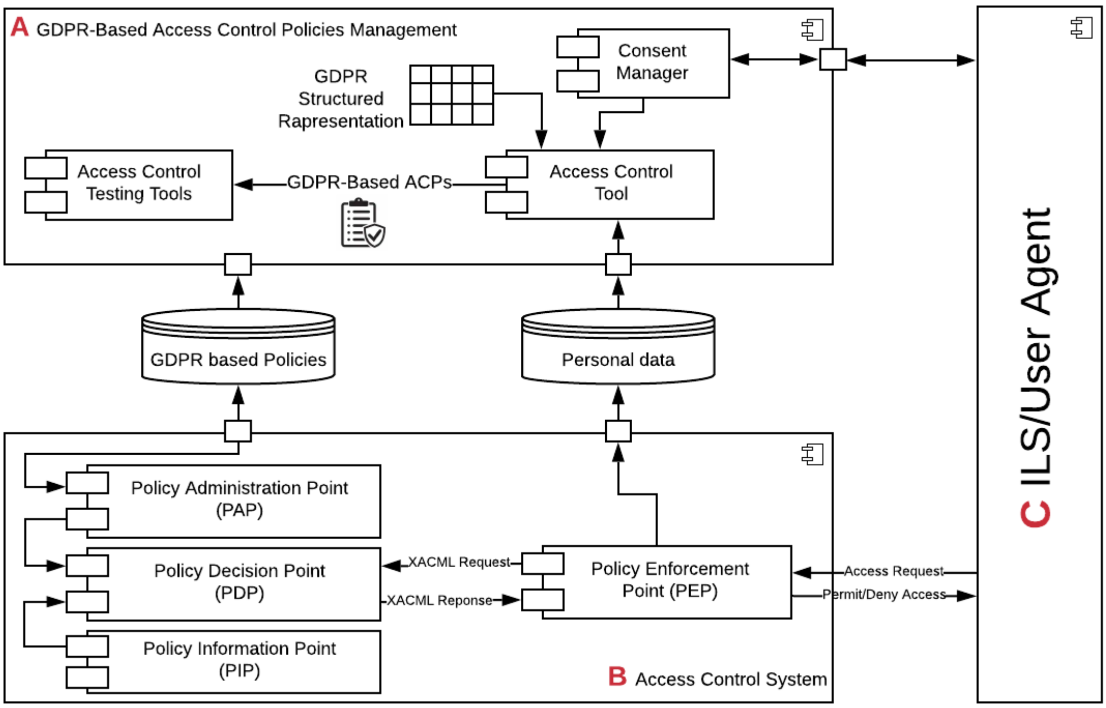


Figure 10.2: *Customization of GENERAL_D in the Context of ILSs. Adopted from [23].*

GENERAL_D can be customized for different scenarios and consequently instantiated by using available tools presented in both academia and industry. In Figure 10.2, we illustrate its customization in the context of the proposal of Indoor Localization, which provides facilities for performing the following steps, in accordance with the Life Cycle described in Example 5 of Chapter 3 (see Section 3.5.5):

1. Gather privacy requirements from collected consents (Step 1);
2. Identify privacy attributes (Step 2);
3. Author the GDPR-based policies (Step 3);
4. Test GDPR-based policies (Step 4);
5. Deploy GDPR-based policies on the E-ILS (Step 5); and
6. Manage the data access (Step 6).

As in the Figure 10.2, GENERAL_D is composed of the two modules: GDPR-Based Access Control Policies Management (module A) and Access Control System (module B).

GDPR-Based Access Control Policies Management. Module A provides facilities to perform steps from 1 to 4. It has the responsibility of interacting with User Agent's sub-component Consent Service, by receiving the collected consents from the end users of the ILS system. As in Figure 10.2, it is composed of three main components: 1) Internal Consent Manager; 2) ACP Manager; and 3) Access Control Testing Tools.

1. **Internal Consent Manager.** This component is in charge of performing steps 1 and 2. It interacts with Consent Service component by preparing the consents to be subscribed by the users. Therefore, after receiving the Consent Record representing the given consent by the end-user of the system (i.e., the Data Subject), *Consent Manager* extracts the useful personal data from the signed consents and stores them into the *Personal data* DB. It also collaborates ACP Manager component by providing it an internal representation of the processed consent by identifying all the required attributes for ACPs authoring purpose.

2. **ACP Manager.** This component performs Step 3. It receives an internal representation of the consent. The Structured Representations are then translated by ACP Manager component into enforceable GDPR-based ACPs, written in the XACML standard, by taking into account the data collected in the users consents. More precisely, this component leverages the result of Chapter 6 by using the obtained GDPR-based ACPs Templates as predefined ACPs structures [2] (i.e., *GDPR Structure Representation* of Figure 10.2). Therefore, by filling them with actual data contained in the consent provided by Internal Consent Manager, it is able to generate enforceable ACPs, each related to a specific GDPR's article that spots relation with access control. We refer to Chapter 6 for more details about the GDPR-based ACPs Templates generation.

[2] Note that, in case of adopting Agile methodology, we can refer and leverage the template associated to User Stories as discussed in Chapter 7.

3. **Access Control Testing Tools.** In Module A, the *Access Control Testing Tools* component integrates also different testing tools for validating the derived policies before to store them into the *GDPR based Policies* DB. We refer to Chapter 5 for more details about the currently available testing facilities included in GENERAL_D.

Access Control System. In this proposal, GENERAL_D is also in charge of managing the access to the personal data during the online use of indoor localization system. In this case, the access is regulated by Module B in accordance with the deployed ACPs, and consequently it performs step 6 by adapting and extending the current version of the XACML reference architecture. More details of the behavior of the GENERAL_D are provided through the application example described in section 10.4.2.

10.3.2 Behavioural Specification

The interaction among the components is reported in Figure 10.3. The indoor infrastructure periodically advertises its presence broadcasting an URI. The URI points to the meta-information of the ILS. When the end-user enables the indoor positioning on her/his device, the User Agent starts listening for such announces (`Listen for URI` activity).

Discovery. The Discovery phase ends when the User Agent accesses the URI in order to obtain the description of all the resources that are part of the infrastructure. The structure of the information obtained during the discovery phase represents a key-point. We report a schematic example of such information in Figure 10.4. The information reported can be represented following different formats, such as JSON or XML text-based format.

Access. The Access stage can now start. During this stage, the end-user grants or denies some consents required by the ILS to work properly (`Consent evaluation`). This process can also involve a more fine-grained assessment of the consent (`specialized consent acceptance` activity), depending on the kind of services to be used. If consents are accepted, the subscription data as well as the collected consents are used by the *GENERAL_D* for example:

- setting up the user specific GDPR-access control policies in order to guarantee that all the information collected and exchanged between the services are managed according to the GDPR's demands (`GDPR policies set-up` activity);
- preparing the required database infrastructure and security procedures so as to guarantee for instance: the isolation of the data from the point of view of storage; the data anonymization; the data deletion according with the consents collected (`Enabling procedures and repositories` activity).

On the other side, the User Agent needs to save the policy accepted by the person and the identification code that will be used to communicate with the infrastructure resources.

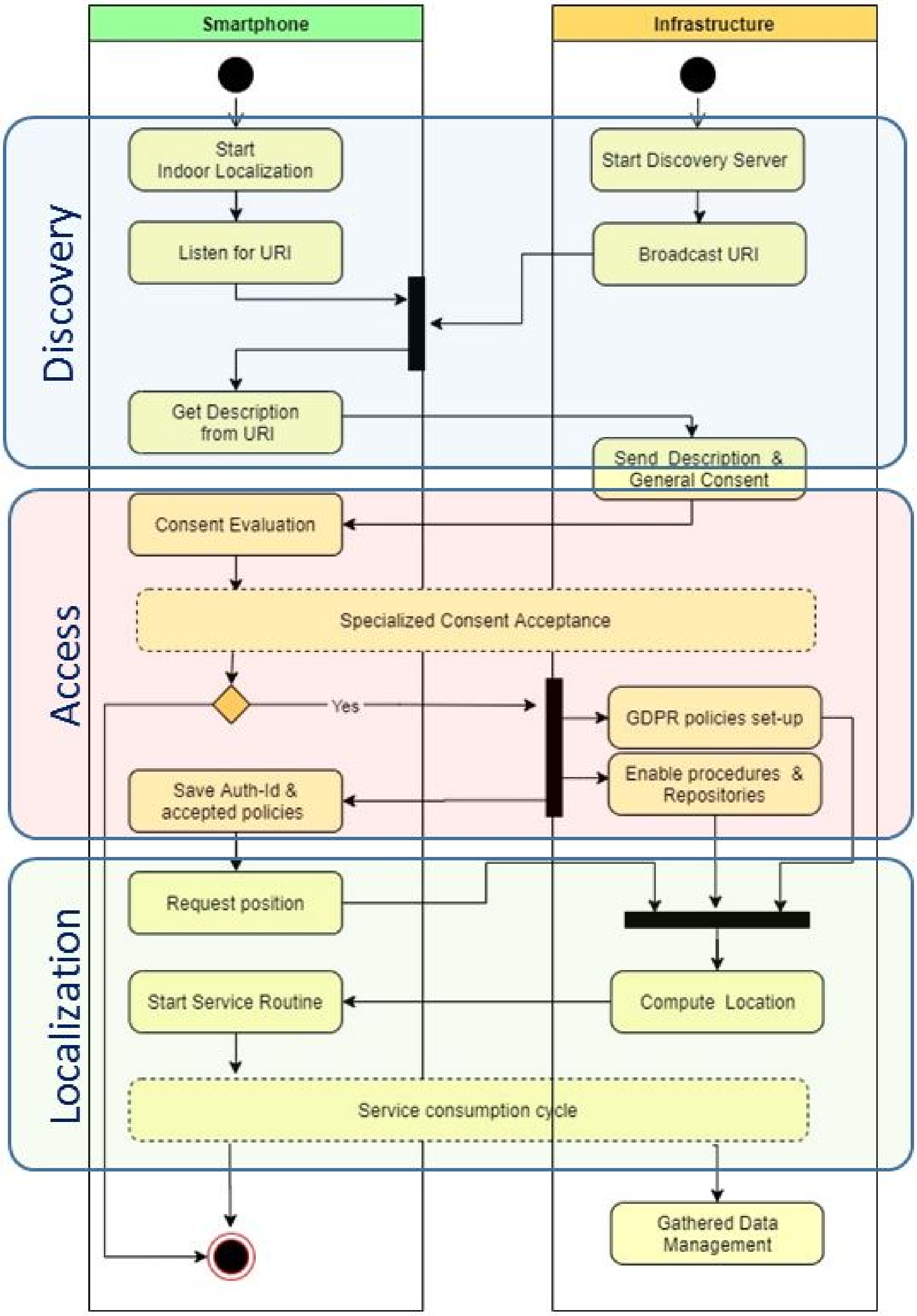


Figure 10.3: *Activity diagram for the system components. Adapted from [23].*

Localization. Once the access phase has been finalized, the `Service consumption cycle` starts, that is the Localization phase. At this point, the user interaction and (personal) data flow is realized taking into account the GDPR provisions such as: assure data treatment in line with purposes specified in the collected consents (Art. 5.1(a) of the GDPR), assure only authorized data transfer to third parties in accordance with the general principle demanded in Art. 44, and the enforcement of the Right to erasure ('right to be forgotten') defined in Art. 17 (`Gathered Data Management` phase).

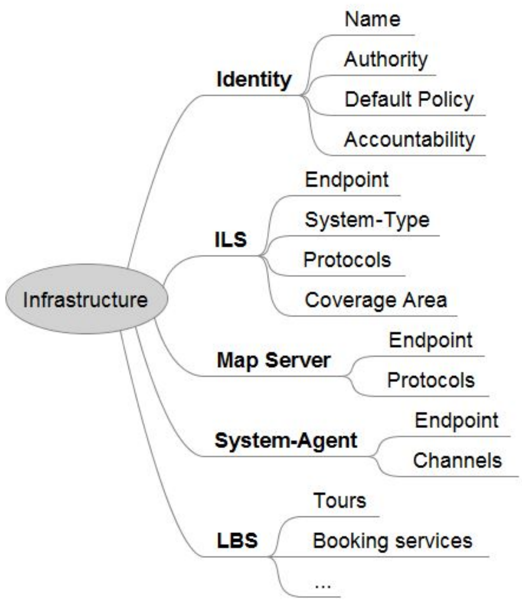


Figure 10.4: *Information describing an ILS through a Discovery Process. Adopted from [23].*

The process ends in case of consent revocation (in this case we are considering the right to withdraw as stated in Art. 7.3) or when the user requires to terminate a specific service.

10.4 Application Example

To better clarify the role of the GENERAL_D inside an Indoor Localization Infrastructure, we present here a simple typical realistic reference scenario, considering a proximity marketing service inside a mall: *Alice* (Customer, i.e., the Data Subject) provides (i.e., gives her explicit consent for processing) the ID of her smart device, the GPS data, the Wi-Fi signal data, and on-board sensors data (Personal Data), to the *proximity marketing service* (the Controller) for *advertising notifications* (i.e., the purpose of processing) when her position is near to a shop.

The infrastructure, by having the Alice's Personal Data, may offer several features such as: a navigation service for optimizing the path for completing a shopping list; a check-out management service notifying a user when to check-out; discount notifications so that to advertise a user when he/she is in proximity of a shop. This simple example highlights two important aspects:

1. from the commercial point of view, the collected localization data may represent an important source for improving their profitability by optimizing the placement of products, for measuring the effects of changing the shop layouts, for sending targeted advertisements, or for analyzing user behavior;

2. from the user point of view, the appealing facilities of the indoor positioning can make available a set of personal data that can be misused and exploited in a way different than expected.

10.4.1 Indoor Localization Enforcement

Considering the architecture presented in Figure 10.1, in this example the User Agent has been implemented by extending the Telegram instance as shown in [99]. The process is initiated by the user when selecting a Telegram menu to look for services available in the mall. Then, the User Agent, on behalf of the user, starts a discovery protocol in order to retrieve information about the available localization infrastructures, if existing. In particular, the Discovery service performs a periodic Bluetooth / WiFi scan so that to retrieve information encoded in the payload of advertising messages of such technologies, i.e., URL encoded in EddyStone beacon. In this case, the information available is a URL that enables the User Agent to download full description of the environment infrastructures and the available services.

The core components of the infrastructure is the ILS engine, that implements the localization algorithm and the Map Server in charge of provide details of the indoor environment to the User Agent. In the proposed implementation, the localization algorithm leverages on the WiFi signals received by the User Agent: it periodically scans WiFi beacons emitted by the WiFi Access Points in the nearby. Such signals (RSSI) are then transferred to the ILS which is in charge of estimating the User Agent's location. In turn, the ILS returns back the User Agent its timestamped coordinates according to the map managed by the Map Server component.

In the proposed implementation, a user receives also information usage of personal data and their purposes through the specific consent associated to the selected service. The target of this phase is to make users aware of the privacy risks connected with the subscription to an indoor position system. For example, who is the controller, how the data will be processed and for which purposes or the time of detention.

By referring to the use case scenario introduced at the beginning of this section, a possible consent record reporting the given consent provided by Alice to the proximity marketing service is reported in Figure 10.5.

In current implementation, among the consent format available in both industry and academia, we rely on "Consent Receipt Specification" proposed by Kantara initiative, and more precisely, we refer to the its draft GDPR extension version named "GDPR Explicit Consent Record & Receipt Extension for Kantara CISWG: Consent Receipt", which is under active development. The specification proposes a JSON schema for a consent receipt and it contains all the required GDPR's concepts useful for authoring ACPs in compliance with the regulation.

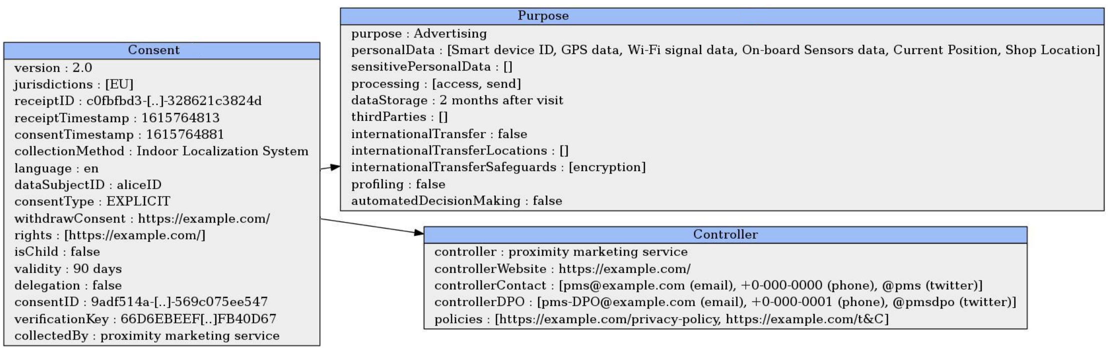


Figure 10.5: *Example of the Consent Record.*

In particular, as depicted in Figure 10.5, the receipt contains:

- Consent element, which reports information about:
  - the Data Subject, e.g., dataSubjectID, isChild attributes. In our example, Alice is an adult person identified by ID *aliceID*;
  - the consent itself, e.g., the consentTimestamp, the validity of the consent, the collection method and who performed it (i.e., the *proximity marketing service* in our case).
- Purpose element, which is related to the explicit given consent, contains the following attributes among others:
  - the purpose name and the allowed actions (i.e., processing) to achieve it;
  - the set of personal data the data subject is given to controller and their storage validity;
  - the involved third-parties in the processing, and whether personal data are using for profiling or for automated decision making.
- Controller element, instead, contains both its contact information and the DPO's contact.

10.4.2 GDPR-based Access Control Enforcement

According to the GDPR demands, the user personal data, the user's positions and their associated date and time, the information related to the services selected by the user as well as the ACPs are kept on an exclusive database ruled by GENERAL_D. Moreover, the collected data will be stored till the established GDPR specified deadline or until the specific deletion request from the user.

By referring the behavioral specification presented in Figure 10.3 and the components description in Figure 10.2, through the User Agent the user, i.e., the data subject, can interact with the Indoor Localization infrastructure, evaluates the consent related to the proximity marketing service and, if accepted starts the subscription to the selected services. From the Indoor Infrastructure point of view, the consent acceptance triggers its translation into specific access control policies by considering the procedural steps described in Chapter 3, Section 3.5.5 and briefly summarized previously in Section 10.3.

At this regard, although the consent receipt is mainly used as a proof of notice requirement for an explicit consent, in this work we leverage that receipt for ACPs derivation purpose. More precisely, the structure information contained is used for performing principally step 2, i.e., it is used as a means for identifying the concrete privacy attributes and their classification in terms of GDPR's concept. Concerning step 4, i.e., authoring access control policies in compliance with the GDPR, as stated previously, we recall that in this work we are relying on the result of Chapter 6 by using and leveraging the obtained GDPR-based ACPs Templates as predefined ACPs structures.

Algorithm 4 GDPR-based ACPs Authoring

1: **input:** CJF ▷ Consent as Json File
2: **output:** GAL ▷ GDPR-based ACP List of XACML policies
3: $GAL \leftarrow \{\}$
4: $CJFAsPOJO \leftarrow parse(CJF)$
5: $cID \leftarrow CJFAsPOJO.getCID()$
6: **if** $isAlreadyGiven(cID)$ **then**
7: $\quad DenyAllPolicies(cID)$
8: **end if**
9: $GATL \leftarrow loadGdprAcpsTempaltes()$
10: **Foreach** $gat_i \in GATL$ **do**
11: $\quad ACP \leftarrow CreateACPS(gat_i, CJFAsPOJO, cID)$
12: $\quad GAL.add(ACP)$
13: **end for**
14: **return** GAL

Consequently, for deriving enforceable XACML policies, by referring to consent receipt shown in Figure 10.5, the algorithm implemented in the current customization of GENERAL_D within the ILS is reported in Algorithm 4.

In details, the Algorithm 4 (line 1) takes as input the consent receipt represented in Json format (`CJF`), and returns a set of ACPs (`CJF`, line 2). It parses the file `CJF` by obtaining its internal representation (`CJFAsPOJO`, line 4). Then, it checks whether the consent was already given (Algorithm 4, line 6) and, in case, the algorithm modifies all the ACPs associated to the current consent into DENY-ALL policies (Algorithm 4, line 7). This is in line with the fact that we allow Data Subject to modify a given consent at any time, for example by modifying the purpose of processing or by withholding a given consent for some of his/her Personal Data.

```
Policy .................................. PolicyId = alicePolicy
                                          root element
                                          rule-combining-algorithm:permit-overrides
  Target ................................ Sample Policy
    Subject ............................. Subject = Alice
  Rule .................................. RuleId = readRule, Effect = Permit
    Target ..............................................................
      Resource .......................... Resource = Name
      Resource .......................... Smart device ID
      Resource .......................... GPS data
      Resource .......................... Wi-Fi signal data
      Resource .......................... On-board Sensors data
      Resource .......................... Current Position
      Resource .......................... Shop Location
      Resource .......................... Advertising
      Action ............................ Action = read
    Condition ...........................................................
      And ............................... And Operator
        string-one-and-only ............. type-One-And-Only Function.
                                          #Resource = 1
        string-equal .................... type-Equal Function.
                                          Resource.owner = Subject
  Rule .................................. RuleId = defaultRule, Effect = Deny
                                          default: deny all, which is not allowed explicitly.
```

Figure 10.6: *An XACML-like Policy.*

Afterwards, it retrieves the ACPs templates from a specif repository (`GATL`) (Algorithm 4, line 9); and consequently, for each of them it creates an enforceable policy and updates the set of the current list of policies (Algorithm 4, line 10-13). Finally, it returns the set of the derived access control policies that are built considering a given consent.

Figure 10.6 reports the derived GDPR-based ACP written in XACML-like language related to Art. 15.1 of the GDPR (Right of access by the data subject). Specifically, the policy is applicable to the subject *Alice* and contains two rules: (1) the first rule, with RuleId equal to *readRule*, represents the AC rule associated with Art. 15.1 and guarantees that Alice can read her provided personal information; (2) the second rule, called *defaultRule*, represents a standard default rule that denies all which is not allowed explicitly.

Remark. Data relative to paths executed by a user moving around the mall and its shops or the time spent in front of a showcase could be attractive sources of information from a marketing point of view. Commercial services could infer, from this data, users' habits and attitudes and exploit them for making business. In this light, having a system that by-design prevents the unauthorized usage of such data is extremely important for increasing the trustworthiness of end-users. This, considering the increasing attention focused by GDPR to the personal data-crumbs, is an issue we face in this work. We present in this chapter a reference architecture of an indoor localization system able to guarantee the GDPR compliance through the integration of specialized access control system enforcing the GDPR's provisions.

CHAPTER 11

GENERAL_D & COVID-19

BECAUSE ILSs know your position and consequently could potentially know who is near to you, in this chapter we leverage the privacy-by-design ILS architecture proposed in Chapter 10 to lawfully measure the distance between people. This allows to address, in a privacy preserving way, the new simple yet disruptive requirement imposed by countries as countermeasures to fight COVID-19 pandemic: the so-called social distancing. Indeed, in this chapter we take the opportunity to show the flexibility and applicability of the thesis proposal in a new emerging and not fully explored context. More precisely, we explore the possibility of adopting the indoor localization technologies to measure the distance among users in indoor environments. We discuss how information about people's contacts collected can be exploited during three stages: before, during, and after people access a service. By enhancing the reference architecture for an Indoor Localization System (ILS), presented in Chapter 10, we illustrate three representative use-cases: Visiting a Museum, Airport Access, and Shopping Assistant. We derive some architectural requirements, and we discuss some issues that concretely cope with the real installation of an ILS in real-world settings. Therefore, we explore the privacy and trust reputation of an ILS, the discovery phase, and the deployment of the ILS in real-world settings. We finally present an evaluation framework for assessing the performance of the architecture proposed.

This chapter is indented to extend and complement Chapter 10, by specializing Application Example 5 introduced in Chapter 3 section 3.5.3 in the context of COVID-19, and it reports the finding of the following scientific contribution:

- Paolo Barsocchi, Antonello Calabrò, Antonino Crivello, **Said Daoudagh**, Francesco Furfari, Michele Girolami, Eda Marchetti: COVID-19 & privacy: Enhancing of indoor localization architectures towards effective social distancing. Array, Vol-

ume 9, 2021, 100051, ISSN 2590-0056

11.1 Introduction

The recent COVID-19 pandemic has been imposing profound changes in our daily life. Most of the affected countries adopted different countermeasures in order to reduce the contagious rate. Among them, an effective action is the so-called *social distancing*. The idea is simple as disruptive at the same time: citizens are invited to maintain a certain physical distance from others. This recommendation applies when we interact with people out of our personal spaces, namely a restricted community of contacts.

Social distancing has become a new requirement in the way we access and provide *services*. In the context of the COVID-19 pandemic, policymakers have to re-think the way we visit a supermarket, we catch a bus, or we interact with colleagues at work. We consider two possible ways of guaranteeing such a requirement: manually or automatically. The manual approach is commonly adopted in our cities, such as in a shopping mall. In this case, an operator observes the scene acting to limit and prevent close contacts among people; for example by managing the waiting queue, verbally distancing customers, or by optimizing the displacements of goods so that to reduce involuntary contacts. Although such approach is relatively easy to implement, we argue that a complementary solution needs to be adopted on the long period; we refer to it as automatic social distancing.

In this work, we explore the possibility of automatically guarantying the social distancing indoor with the adoption of a privacy-preserving Indoor Localization System (ILS). We focus on those services that are generally available indoor, such as a museum, airport facilities, or a supermarket. In these representative use cases, users roam through a sequence of points of interest such as galleries of a museum, check-in desks, or aisles of a supermarket. Our approach consists of estimating the current location of people with the ILS and to compute the personal distance among the subjects involved. Knowledge on the existence of crowds can be exploited by suggesting to the customers an alternative path able to minimize gatherings with others.

In the last decade, ILSs have been widely adopted [249] in different scenarios; they are based on very different technologies, ranging from WiFi fingerprinting [213] to solutions based on ultra-wide band radio waves [228]. We argue that the accuracy obtained from the most advanced systems is now sufficient for the purpose of the social distancing [99]. As a meaningful example, we refer to class of ultra-wide band systems able to constrain the localization error in the range of centimeters while tracking moving objects [167].

In this chapter, therefore, we firstly leverage the privacy-by-design indoor positioning architecture, previously introduced (see Chapter10), for social distancing. Indeed, the proposed architecture is able to guarantee real-time user's location and privacy of the data collected, where purposes of data processing are explicitly defined, consents collected and the rights related to privacy and data protection correctly enforced according to the GDPR.

We then discuss three use-cases in which the architecture proposed can be adopted, namely visits a museum, airport access, and shopping in a supermarket. For each of the use-cases, we present the requirements for guaranteeing social distancing indoor. We also introduce some barriers currently preventing a massive adoption of ILS-based tracing systems and we conclude the chapter with an evaluation framework aimed at assessing the performance.

The innovation we propose with our work mainly consists of four aspects:

- we frame a reference architecture for social distancing based on an Indoor Localization System generalizing three common use cases;
- by leveraging GENERAL_D in the context of ILS, we realize a privacy-by-design ILS for social distancing grounded on the European GDPR framework;
- we summarize three typical use cases in which the proposed architecture can be adopted, by highlighting the intrinsic challenges of the social distancing;
- finally, we discuss four main barriers to overcome for an effective adoption of such technologies in real-world settings.

As recently reported by M. Zissman (MIT Lincoln Laboratory) in a recent article from J. Hsu [114], "[. . .] In a perfect world, something like this would have taken a couple years to implement. There just isn't the time [...]". We agree with such vision, and we consider that a great effort has to be spent for the integration of different technologies enabling the proximity detection of people both indoor and outdoor. Such effort will determine the success in fighting against the next pandemic.

Outline. Section 11.2 covers the background and related work in the field of Indoor Localization Technologies and Social distancing. Section 11.3 describes our reference architecture for an ILS. Section 11.4 reports three reference use cases, namely visiting a museum, airport terminal access and shopping assistance. In Section 11.5, we discuss some issues that we consider challenging for a real-world installation of an ILS and, finally, Section 11.6 describes our evaluation framework.

11.2 Background and Related Work

In the following, we focus on the main aspects that equally contributes to the proposed architecture: the related indoor localization technologies and their specific characteristics (Section 11.2.1); and the Indoor Localization Apps (Section 11.2.2). We discuss in particular the mobile application proposal able to deal with social distancing, exploring their main strengths and weaknesses in terms of authorization to access to the mobile resources requested to the user and their impact on user's privacy.

11.2.1 Indoor Localization Technologies

Several localization technologies have emerged in the last years to address the demanding of location-based services. We review two categories: Radio Frequency-based (RF) and non-RF based.

Among the RF technologies there exist systems based on the analysis of Wi-Fi [191], Bluetooth Low Energy (BLE) [164,188], LTE [141], and Ultra-Wide Band (UWB) [145, 200] signals.

Wi-Fi-based solutions have the advantage of exploiting the ubiquity of Wi-Fi Access Points. The most performing Wi-Fi solutions obtain high performance in terms of localization accuracy with reduced cost of maintenance and installation [111].

In the last few years, the Bluetooth Low Energy (BLE) standard has been adopted as cheap and viable technology for indoor navigation and localization. BLE tags are cheap and easy to deploy, moreover their battery life time spans from few months to years [1]. Indoor localization systems based on BLE often implement a range-based technique, according to which the moving target is localized in proximity of the BLE tags with the highest Received Signal Strength of the beacons emitted ([104] explores this technique for the purpose of tracing social interaction). However, more advanced solutions based on the beacon's angle or arrival and time of flight are also available with a very high accuracy level, as done with Quuppa[2] and the recent Bluetooth 5.1 stack.

Finally, the UWB network interface represents a recent and promising solution. Its accuracy can reach the centimeter-level with specific deployments. Its adoption has been increasing as Apple decided to provision the iPhone 11 with the U1 chip-set. As a result, we expect that in the near future other vendors will include such technology with Android-based smartphones. Some remarkable examples of UWB-based indoor solution are the Pozyx [22] and some recent works [50, 243, 248].

Non-RF based technologies for indoor localization rely on visual/camera [227], Visible Light Communication (VLC) [253], Inertial Measurement Unit (IMU) [214], and Magnetic Field Sensor (i.e., MEMS) [133].

The visual based systems exploit images captured by surveillance camera already deployed. The performance ranges in the centimetre scale but, in wide and public environments, the privacy regulations might limit their adoption on the large scale. Differently, if the user/target is equipped with a camera sensor, a visual-based system can reach accuracy performance around a meter of error. Furthermore, the end-user is required to keep the camera in a fixed position with the side-effect of influencing its natural way of moving.

The Visible Light Communication is an emerging optical technology for high-speed data transfer which uses visible light modulated and emitted by Light Emitting Diodes (LEDs). Indoor positioning systems based on VLC use light sensors (e.g., camera sensor) to measure the position and direction of the LED emitters but they generally require line of sight between emitters and receivers [253].

Systems based on Micro Electro-Mechanical Systems (MEMS) exploit the distortion of the Earth's magnetic field mainly due to structural steel elements (e.g., steel fire doors) and furniture. As an example, these distortions can be a discriminating factor in environments comprised by corridors, rooms and small areas. The performances of these systems generally drop in wide and open space because the distortion are considered less meaningful [212].

IMU based systems utilize tri-axial accelerometers and gyroscopes for sensing the motion. The combination of gyroscope and accelerometer is used to evaluate the heading direction [151]. Unfortunately, accelerometers are error prone due to random movements of human motion and, the gyroscope is susceptible of magnetic fields distortion. As a consequence, IMU-based systems generally reach low accuracy and require a complex calibration process to detect, for example, users' step length and the motion speed [215].

We finally report on Table 11.1 a comparison of RF and non-RF based techniques,

[1] BLE Tags can be configured in safe-mode with low power of emission and low advertisement rate.
[2] https://quuppa.com/

Table 11.1: *A comparison between indoor localization technologies. Adopted from [24].*

Technology	Strengths	Weaknesses	Accuracy	Scalability
Camera	High accuracy, low maintenance	Requires dedicated hardware, difficult user identification	0.5-1m	Medium
VLC	Potentially high accuracy, easy to install	Requires line of sight, requires additional hardware	0.5-1m	High
IMU	No infrastructure required	Requires high customization, error prone to drift problems	1-5m	N.A.
MEMS	Ubiquity of the signal, no infrastructure required	Error prone to interference, costly calibration process	2-5m	N.A.
Wi-Fi	Easy to implement, cost efficient	Medium accuracy, generally requires modifications to the APs	2-4m	High
BLE	Low energy consumption, low cost	Error prone to noise, medium accuracy in wide environment	2-4m	High
UWB	High accuracy in small environment	Requires dedicated hardware, high costs	0.5-1m	Low

with a summary of their weaknesses and strengths.

We finally survey some architectures for indoor positioning. Such architectures provide features for a quick integration, such as an SDK or APIs for third-party developers. Authors of [142] introduced a middleware architecture for fusing multiple sources of information, showing how a data fusion approach leads to improve performances in the same indoor environment. In [221], authors describe an extensible framework for exploring location data's multifaceted representations and exposing a query layer. Lastly, Anyplace [250] shows a similar idea to the other architectures above mentioned by releasing an open-source architecture in order to easily deploy indoor localization functionalities in new environments.

11.2.2 Indoor Localization Apps and Privacy

The most diffused technical solutions for guaranteeing social distancing are based on mobile applications. Apps enable an easy-to-use user interface and, at the same time, a massive diffusion through the well-known app stores. Currently, there exist several applications whose features span from tracing contacts, e.g., `Immuni` [2] (the application built by Italian Health Minister), to the possibility of managing a waiting queue, e.g., `ufirst` [5].

Depending on their features, the apps require access to several entities and purposes. From a privacy-preserving point of view, unfortunately, not all the apps expose neither a clear claim about the usage of the collected personal information nor a clear description concerning the usage of such data. Consequently, we argue that the end-user might remain skeptic in daily using such apps.

In order to provide a first outlook about the existing apps for social distancing or for detecting in social interactions [104], we report in Table 11.2 a selection of apps available on the Android Play Store and tested on commercial smartphones. The table reports the analysis of some features and the authorizations required. In particular, the group of columns labelled *Location*, *Others*, *Disk*, *Camera*, report the name of the permissions' classification provided by the Android Play Store for grouping the different features. Finally, for each group, we report some details concerning the permissions of each app based on the description provided by the developers. In particular, we report the following information:

1. *Approx location (column 2)*: if the app can localize the device within a wide area;
2. *Precise location GPS & net (column 3)*: if the app can accurately localize the device;
3. *Receive data from Internet (column 4)*: receive data form internet;
4. *View network conn. (column 5)*: if the app can check the networks to which the device has access;
5. *Full network access (column 6)*: if the app can access to any of the networks the device is connected with;
6. *Run at startup (column 7)*: if the app can automatic restart;
7. *Prevent device sleeping (column 8)*: if the app can prevent the device from switching in sleeping mode;
8. *Pair BT devices (column 9)*: if the app can pair with a Bluetooth device;
9. *Access BT settings (column 10)*: if the app can initiate the device discovery or modify the Bluetooth settings;
10. *Control vibration (column 11)*: if the app can control the device vibration;
11. *CRUD contents (column 12)*: if the app can perform CRUD operation;
12. *Take pictures or videos (column 13)*: if the app can take photos or record videos.

Taking a glimpse as a generic user to the installation and usage of the apps analysed in Table 11.2, the consent forms are very generic and sometimes do not intuitively declare the purposes of data collection, the duration of the data retention or the possibility of future exploitation of the data collected. For example, one application (2M Social distance checker) requests permission to access to the Call Log and Address Book without specifying how the data will be used, i.e., to enable the sharing of user experience with his/her contacts. From a technical point of view, in case of open-source applications [3] specific information about the real usage of sensors data or the procedures for managing them can be retrieved by accurate analysis of the source code. However, this operation is not feasible by common users without a computer science background and it is not allowed for proprietary application [4, 6].

Additionally, rarely there is a clear claim on where and how the collected data will physically be stored or distributed. Indeed, depending on the country where the DB is, the rules for accessing its information could be compliant with a privacy standard different from that required by the application country. The situation could be even worse in case the application is used by users belonging to different countries having not the same privacy rules. Consequently, there could be the risk of a personal data

Table 11.2: *Features of Social Distancing Mobile Apps. Adopted from [24].*

	LOCATION		OTHERS								DISK	CAMERA
App name	**Approx location**	**Precise location GPS & net**	**Receive data from Internet**	**View network conn**	**Full network access**	**Run at startup**	**Prevent device sleeping**	**Pair BT devices**	**Access BT settings**	**Control vibration**	**CRUD contents**	**Take pictures or videos**
Social Distancing Project (Su-Raksha)	✓							✓	✓	✓		
The Social Distancing App	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Social Distance	✓	✓	✓	✓	✓	✓	✓	✓	✓			
REXdistance Social Distancing					✓					✓	✓	✓
Social Distance Alarm		✓	✓	✓	✓	✓	✓	✓	✓			
Social Distancing					✓						✓	✓
1point5	✓	✓			✓		✓	✓	✓	✓		
Give Me Space The Best Social [...]		✓					✓	✓	✓	✓		
Social distance		✓			✓		✓	✓	✓	✓		
Social Distancer				✓	✓		✓					✓
Distancing App	✓		✓	✓	✓	✓	✓	✓	✓	✓		
Social Distance							✓	✓	✓			
Social Distancing App	✓	✓						✓	✓	✓		
Pistis.io Social Distancing App	✓	✓		✓	✓		✓	✓	✓	✓		
Distancing alarm	✓	✓			✓							
2M Social distance checker	✓	✓										
Social Distancing App - Wearable	✓	✓					✓	✓	✓	✓		
Keep Distance	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Immuni				✓	✓	✓	✓	✓				

management not completely compliant with the consents signed by the app users. For instance, in case of `Immuni` [2], developers clearly claim that the DB will be physically positioned in Italy and managed by the Italian Health Minister in compliance with the GDPR.

Social distance can also be implemented with ad-hoc hardware components like people counter and smart bracelets [1, 173]. These solutions have the benefit of guaranteeing reliability, since they do not depend on the user's device. The features offered are limited to tracing the contacts with others or alerting when a user gets too close to another. More advanced features can also be implemented with data analysis techniques but, at the current stage, we were not able to find remarkable examples in the current literature.

Concerning security (i.e., access control) and privacy reference to ILS, a recent systematic review of privacy in indoor positioning systems [113] pointed out that current proposals in literature show how location and topology-aware are becoming feature for the security [109]. However, most of the research has been focused either on: i) using technology Wi-Fi and the fingerprinting methods combined with cryptography solutions [129, 177, 245]; ii) using access control mechanisms for (physical) protection within virtual perimeters [108]; iii) using location information to automatically authenticate customer [110]; and iv) specific security attributes that do not fully cover the GDPR's requirements [25, 57, 123].

11.3 Overview of the Integrated Architecture

In the following, we detail the reference architecture based on an Indoor Localization System (ILS) for guaranteeing social distancing. The architecture relies on two layers: the smart device and the localization infrastructure. As detailed in Figure 11.1 these layers include components such as: a User Agent for managing the interactions on behalf of the end-user, the Indoor localization system including GDPR-based Access

Control subsystem and the Map server. The reference architecture extends the solution presented in [99], and it integrates the GDPR-based access control described in [32,78], i.e., GENERAL_D.

11.3.1 Aim and Scope

The approach we followed with this work is to firstly framing a reference architecture to be adopted in very different scenarios. To this purpose, our effort has been mainly focused on generalizing a common architectural design of a remote ILS, based on three main building blocks: a map server, the ILS engine, and a discovery agent to broadcast its existence. We then focus on the client side in the form of a smart device, as the primary interface to interact with the end-users. The architecture we propose has been deliberately designed without constraining to any of the common scenarios we daily experience. Differently, we tried to provide the community with a modular architecture to be customized. Furthermore, the current literature concerning the ILS does not identify a standard de-facto for indoor localization, rather multiple and heterogeneous solutions are available. To the best of our knowledge, this work introduces a privacy-by-design solution, mainly inspired by the European GDPR framework, as one of the most advanced regulations about privacy in force since the last decade.

11.3.2 Architecture Requirements

Indoor localization systems are based on very different technologies. A standardization process of these systems is therefore the first objective that should be pursued in order to increase the spread and the usability of location-based services.

We argue that standardized programming interfaces for the design of an ILS have a twofold benefit:

- to provide inter-operable location-based services to the end-users;
- to integrate in a seamless way outdoor and indoor localization systems.

Concerning the first benefit, its adoption can be used not only to locate and track people, but also to measure their physical distance. Its adoption can be considered an effective counter-measure to track, prevent and analyze how close people are in indoor environments. We refer to Section 11.4 for a in-depth description of three use-cases in which we describe the adoption of an ILS in real-world scenarios. Moreover, the standardization will increase the possibility for a user-agent to discover and to bind to any of the ILS available indoor. Such aspect is crucial for an open market, since it breaks the silos of custom and vertical solutions available so far.

Concerning the second benefit, we consider that the current user experience for outdoor localization systems (e.g., GPS) must be preserved also indoor. Under this context, the standardization could improve the design of systems enabling the hand-off between outdoor and indoor areas. We imagine a smooth transition from an outdoor map (e.g. provided by OpenStreepMap or Google Maps) to detailed indoor maps provided by an ILS.

As first step towards such standardization process is the definition of architectural requirements to be considered. In the following, we report a list of four requirements that we consider mandatory:

- *To discover the available Indoor Localization System dynamically* (**R1**).

  A discovery process should be defined to enable a person to look up for services available in a specific environment [103]. The process can be triggered through the Web or based on short-range network infrastructures. We refer to the first approach as *global search*, since the user queries the Web looking for an ILS available, e.g., in a supermarket. In this case, the user fetches the meta-information of an ILS via the HTTP or similar protocols. Differently, the second approach is referred to as *local search* since the user looks up for nearby ILSs, by exploiting network interfaces such as Wi-Fi Direct, Bluetooth or LTE-Direct and the upcoming 5G. Such interfaces allow to look up for surrounding services in the range of few meters.

- *Indoor localization systems must self-describe their features to ensure interoperability with heterogeneous systems* (**R2**).

  We expect the definition of a common language for describing the features provided by an ILS. More specifically, ILS has to advertise some core information, such as: the localization technology adopted, the privacy requirements, the location of the indoor map and any other resource required for a device to discover, connect and access to the ILS. The benefit of such language is the possibility of replicating the user-experience for outdoor navigation (e.g., through Google Maps or similar) also indoor.

- *Privacy must be guaranteed and the service policies must be well defined and verifiable* (**R3**).

  One of the most critical aspects for the location-based services is the possible loss of control of the personal data collected. As a meaningful example, we recall the contact tracing apps also exploiting the device localization and designed for the purpose of mitigating the COVID-19 pandemic. In these cases, end-users are worried about non-expect usage of the data collected for commercial purposes. Some example of data that we consider critical are: the timestamp of the contacts, the IDs of the contacts, any information about the device used and, in some cases, the GPS location of the users.

  Localization systems suffer intrinsically from this problem, therefore independently from the contingency period, explicit mechanisms for accepting policies, together with the ability to verify and manage the data collected, must be designed and implemented in accordance with the various national laws.

- *Indoor localization systems must be easy to use, intuitive and interactive* (**R4**).

  The interaction between the ILS and the end-user needs to specifically tackled. We argue that their success also depends on the way a user interacts with it. Most of the people already interacts fluently with GPS navigation systems (e.g., Google Maps, Garmin or TomTom charts). In particular, users search, discover and navigate toward a specific location that can always be represented as a pair of coordinates in the space (e.g., lat, long as WGS84 coordinates). The same user-experience should be replicated also for indoor environment, even if a higher number of challenges are present. To this purpose, we consider mandatory to design intuitive work-flows. As for example, the end-user should be able not only

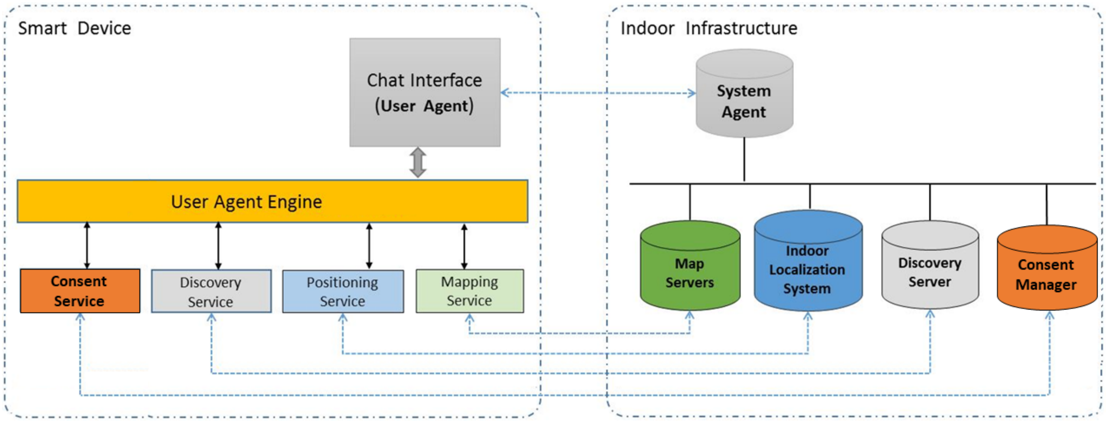


Figure 11.1: *Functional Components of the Integrated Architecture. Adapted from [24].*

to search for a location, but also to navigate toward person, to pick up a list of products in a specific order or to meet a moving target indoor. A further level of interaction is also represented by the possibility of asking to an ILS context information describing the environment, such as the existence of a crowd or the waiting time before accessing a service. Such level of interactivity can be obtained by designing multi-modal interfaces, such as the Instant Messaging (IM) paradigm. Indeed, the IM paradigm implements the best metaphor for managing the exchange of information between the end-users and a system in a intuitive way. The user can chat with the ILS in order to get the position of its target, to receive suggestions, to be notified proactively and/or to be guided step-by-step toward its final destination.

11.3.3 Architectural Components

The requirements **R1** to **R4** are grounding for our architectural design. We describe in this section, several functional components to be deployed in two distinct layers: those present on the user device and those made available by the indoor infrastructure. This distinction easily recalls the two methods through which the position of a user is estimated: self-positioning processed by the smartphone and remote positioning processed at the local infrastructure. We also consider the possibility of having hybrid solutions. We report in Figure 11.1 an overview of the components described. For the sake of brevity, we do not consider here the possibility of other solutions that could make use of the Cloud. For example, internal maps could be downloaded from any server on the Cloud (e.g., Google Maps), as well as a route to a target calculated by a navigation service available on the cloud. We mainly focus on abstract functionalities common to all the architectures and how they are to be described. Therefore, many concrete architectures can be derived by combining the abstract components we report in Section 11.4.4. We report in Figure 11.2 an overview of the main components.

In the first group of components, the most important is the **User Agent**. It can be described as an intelligent software component that operates on behalf of the end-user. The main functions it provides are: global and local discovery (R1), to manage the privacy for the end-users (R3), to interact with the local infrastructure to estimate position of the end-user (R2), to interact with the end-user (R4). Other functional components that could be installed on the device are: the **Navigator** and a **Translator**; the former to manage navigation or determination of the shortest routes; the latter is increasingly adopted to facilitate vocal interaction, as done with commercial vocal assistants (Amazon Alexa, Apple Siri, Google Home) (R4).

The second group of components concern the infrastructure. In particular, the **Indoor Location System** and the **Map Server** (R2). The first consists of hardware and software artifacts deployed in the environment that are functional to the estimation of the user's position and the data protection, as detailed in Section 11.3.4. The second one provides the indoor maps and features of the indoor environment useful for navigating. Other components we foresee are the **Discovery Server** which provides the description of the resources available by the infrastructure (R1), **Consent Manager** which is the counter-part of **Consent Service** for managing the lawfulness of processing of personal data (R3), and the **System Agent** which is the counter-part with which the User Agent can communicate. In particular, the System Agent can be seen as a regular chat user to which send requests for assistance or information (R4), it can be implemented by an Instant Messaging bot. Components deployed in the infrastructure are interfaced by respective services orchestrated by the User Agent. A person trough the User Agent interface can interact with the System-Agent of the indoor infrastructure.

11.3.4 Indoor Localization System and Data Protection

In this section, for the aim of completeness, here we recall briefly the internal structure of a generic ILS, which we have already presented in Chapter 10. We consider it provides three main sub-components, as detailed in Figure 11.2: the **ILS Engine**, the **GENERAL_D** framework and the **Communication and Interaction Orchestrator**. Such components implement the main features of an ILS. Moreover, we expect that they rely two distinct database for collecting the information.

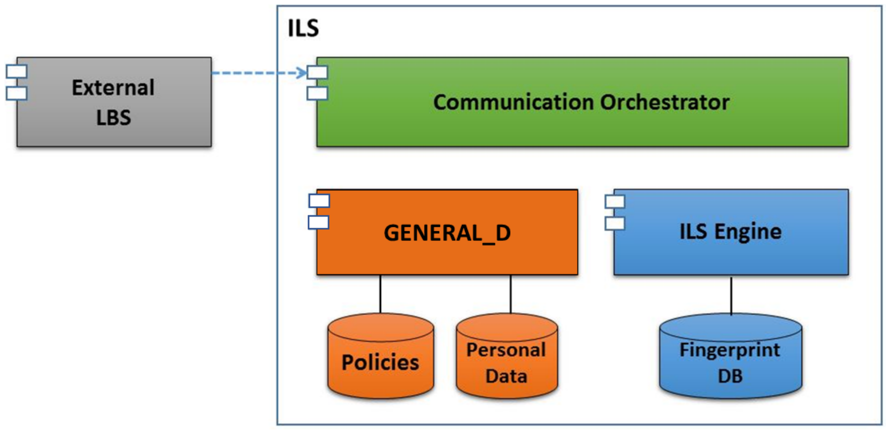


Figure 11.2: *ILS and Data protection components. Adapted from [24].*

The *ILS Engine* implements the core functionality of the localization algorithm: it returns back to the User Agent the timestamped coordinates according to the map reference system (e.g., latitude and longitude as WGS84 reference system) [191, 193].

Data provided by an ILS can be simultaneously accessed by multiple actors. More specifically, the end-user, the system administrator or a generic supervisor might require

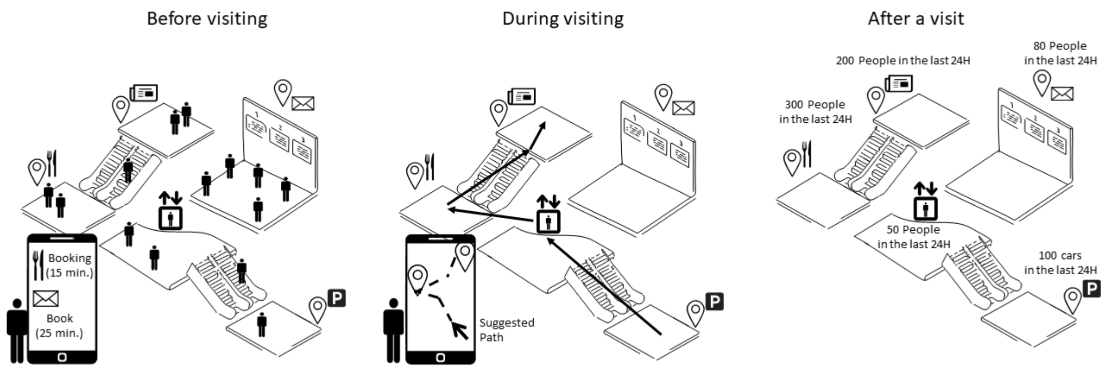


Figure 11.3: *How information about social distance can be used before, during and after visiting a generic indoor environment. Adopted from [24].*

access to specific data. In order to manage the different grants for the actor, we consider that the ILS and GENERAL_D components have to cooperate (see Figure 11.2). In particular, GENERAL_D is in charge of evaluating each single data access request and to allow or deny the access according to several factors. They are: the consents collected, the data validity period, the specific users/service rights and the access control policies established inside the overall Localization Infrastructure.

Finally, the *Communication and Interaction Orchestrator* is the component in charge of managing the communication to and from the ILS. This component is exploiting publish-subscribe design pattern through extensible events. It is in charge to instantiate communication channels and manage flows of notifications and events data. Those events can be structured adopting several asynchronous messaging technologies, such as *Java Messaging Service (JMS)*, *Advanced Message Queueing Protocol (AMQP)*, *Message Queueing Telemetry Transport (MQTT)*, in order to decoupling not only the locations of the publishers and subscribers, but also decouple them temporally.

11.4 Designing Indoor Social Distancing

We now discuss how to guarantee social distancing of users in three representative use-cases, namely visiting a museum, accessing an airport and shopping assistant. We describe for each of them the overall user-experience, the requirements to be guaranteed and some enabling technologies that can be used for adopting an indoor localization system.

As a general observation, information about the physical distance among people can be exploited in three different stages: *before*, *during* and *after* the end-user visits a location. More specifically, knowing how people dispose indoor during the such stages, can increase the user-experience and improve the effectiveness of countermeasures to the diffusion of diseases (see Section 11.2). We report in Figure 11.3 an example of how the information concerning the social distancing can be used during during visiting a generic indoor environment. More specifically:

- *before* visiting an environment, the end-user can plan her/hist visit so that to avoid crowded time slots. Planning the visit allows to minimize the probability of involuntary contacts and it can reduce queue to access to specific services;

- *during* a visit, the end-user can optimize the way she/he moves indoor, so that to reduce the contact probability with others. As a representative example, we refer to the possibility of planning the order of products to pick while shopping. The path selection can be achieved by prioritizing those areas of a shopping mall scarcely visited by customers;
- information about the social distance can also be used *after* the visit. More specifically, knowledge about how many people visited a specific location, can be used to plan an efficient cleaning schedule, to better dispose products and, more generally, to allow the service provider to re-think the way services are accessed. Data collected after the visits can be used for statistical purpose as well, in order to measure if the actions taken prevent the existence of crowd.

11.4.1 Use Case 1: Visiting a Museum

Visiting a museum is a challenging scenario: it is difficult to guarantee a proper social distance among visitors. A museum is generally organized with a visiting path designed to guide visitors through the artworks. Users can decide to follow the path and to move across the rooms in a specific order. Small museums do not recommend any specific path to follow, while others have multiple paths. Users following a path do not have any time restriction during their visits, they are free to move across the rooms and to rest. Moreover, the number of visits in a museum changes dramatically according to several factors, such as the day of the week, the hour, the scheduled holidays as well as the whether conditions. The combination of such factors determines the existence of burst of visits, as discussed in [119] that are challenging to predict.

The requirements that need to be preserved during a visit are:

- users have to be able to respect easily the distance from other visitors;
- the total number of visitors has to be managed;
- design multiple visiting paths so that to reduce the encounter probability among visitors;
- the user-experience needs to be guaranteed during a visit;
- users can wear a wristband or install a specific app before their visit.

We consider that the adoption of an indoor localization system in a museum can support the adoption of effective countermeasures for limiting crowded areas. More specifically, the knowledge of the position of the visitors can be used for 1) observing the way visitors access the museum and 2) to manage in real-time the flow of visits. Concerning the first goal, we argue that it is highly important to measure quantitatively the way visitors access a museum. More specifically, it is possible to measure the total amount of visits, the visits during a specific time interval, the visiting time for each room and artwork, the existence of preferential paths during a visit and other metrics useful to describe the social attitude of the users. Such information, can be in turn used to meet the second goal, namely to plan the visits according to the requirements previously reported.

Users can be localized by adopting proximity-based technologies such as Bluetooth Low Energy (BLE) or the UltraWide frequencies. Such technologies are becoming more and more popular. In particular, BLE is already available in most of the commercial smart devices, while the UltraWide technology is expect to diffuse in the near future. It is worth to notice that the iPhone 11 already provides the U1 chip-set. Bluetooth and UltraWide band allow to detect proximity not only between users but also between users and points of interests such as artworks or furniture. Moreover, such technologies can be easily integrated with personal devices such as smartphones or an audio guides without the need of a complex network infrastructure.

Some remarkable works already addressed the problem of localizing people in a museum, we refer to [14, 236, 237] for further details.

11.4.2 Use Case 2: Airport Access

The layout of an airport is generally a combination of indoor/outdoor multi-floor environments with restricted areas. Except for shops, generally the airports offer multi-storey buildings within wide and open spaces in which users freely roam. An airport terminal provides several services for the end-users. Some of them are mandatory to all the passengers, others are optional and provided only for the entertainment purpose. The number of users accessing to such services changes dramatically according to the seasons and according to external factors, such as weather conditions, strikes and delays of the flights.

The requirements that need to be preserved for travelers in order to guarantee appropriate social distance are:

- the airport facilities have to be accessed with a pre-determined order. More specifically, users have to check-in, to pass through the security clearances and, finally, to step toward the destination gate. The order and the time to complete the previous steps should be orchestrated so that to consider: the amount of users, the existence of crowds and any situation leading to involuntary contacts;
- users have to be able to respect easily the distance from other users;
- the user-experience must be preserved as much as possible.

Also for this use-case, we consider beneficial the adoption of a modular indoor localization system tracking the distance among users. In particular, the indoor localization system can be used for detecting the existence of crowds in a specific location, e.g., check-in desks, and to prevent other passengers to stack in queue. Furthermore, such localization system can be also exploited for security tasks, such as identification and tracking of *target* subjects.

The majority of the terminals are already equipped with Wi-Fi networks available also for traveller. Some works already address the problem of localizing users in an airport. Authors of [168] propose a multi-modal solutions based on Wi-Fi and BLE tags, through the availability of a precise map of the environment and an accurate survey of the environment. The end-user can benefit of services based on positioning information on their own commercial smartphone. In [102] authors propose an interesting approach, based on BLE technology in an airport scenario, using a combination of Received Signal Strength Indication (RSSI) and Time-of-flight.

11.4.3 Use Case 3: Shopping Assistant

A shopping mall is generally organized with product-specific aisles. Customers are free to decide the order of the products to pick. Malls also have some locations for specific fresh products, such as bakery or fresh-fish. In these locations, the self-service is generally not allowed and customers interact with an operator. Some small/mid-size malls guide customers through a suggested walking path, while in large-scale malls, customers can freely move within the areas.

Similarly to the other use-cases, the number of users can change along the time. The burst of visits can be roughly predicted since the working schedule forces many people to shop during the evening or week-ends. In this context, the requirements for respecting the social distance between customers are:

- specific products need to be booked in advance so that to reduce the number of customers waiting in the same location;
- users should be able to easily respect the distance from other users;
- the user-experience must be preserved as much as possible.

A shopping mall equipped with an indoor localization system can provide several services for customers. We foresee the possibility of optimizing the path to follow in order to completing the shopping list. Moreover, the supermarket can provide a queue management service that notifies the customers when to approach to a specific desk. Finally, a supermarket can provide services for personalized advertising to the end-users. In fact, indoor positioning systems shall make available a set of personal data which can be exploit to promote sales products or to promote temporary offers.

In the last decade, supermarkets have been equipped with internet access, throughout Wi-Fi Access Points deployed in the environment. These APs can be exploited also to provide an indoor localization services. The more promising technique in this environment is the fingerprinting technique, where the RSSI previously collected together with the position of the user is leveraged to infer the current user position [199]. In fact, the accuracy of this technique ranges in the order of few meters and, exploiting also inertial sensors of the smartphone, ILSs are generally able to localize users accurately.

A distributed ILS can provide meaningful information before, during and after shopping. For example, before shopping, users can use the aggregated information about the number of current buyers to plan the purchases or not. During shopping, the user is reassured about the use of the ILS which can provide a "safe route" as described before. After shopping, information related to all the routes followed by customers can be used to thoroughly sanitize the most frequented spaces.

11.4.4 A reference architecture for different use cases

Although our goal is not to define a reference implementation of the architecture described in Section 11.3, we consider that some of the components in Figure 11.2 can be implemented with existing software artefacts available in the current literature. We report in this section some meaningful examples both for the Indoor Infrastructure and for the Smart Device.

Concerning the Indoor Infrastructure, the Map Server is responsible for managing the indoor map. In particular, it provides the base maps or a tile set covering a specific area. The Map Server couples with the client side, in charge or downloading (possibly, with parallel connections) and rendering the map, e.g., on a 5-inch screen. Both modules are available in literature and can be re-used as third-party black boxes. As for example, the open source map-view solutions, open layers and leaflet are available. According to the specific needs, it is also possible to adopt different Map Server such as mapbox, Google Maps and AcrGis [3].

Concerning the ILS engine the literature also offers some interesting and open source solutions that can be deployed as off-the-shelf products, among them we refer to Anyplace as a complete framework for indoor localization comprising API, Viewer, Navigator and Logger components [4]. We finally mention some existing discovery protocols that can be embedded with the Indoor Infrastructure to discover the server in a seamless way. In particular, the SLP, UPnP, ZeroConf and WS-Discovery are old-but-robust valuables candidates for discovering networked resources [166]. Moreover, if the goal is to implement a local discovery then the Bluetooth beaconing and the Wi-Fi probing also represent two interesting protocols that can be used to broadcast small chunk of information.

Finally, in relation to the Smart Device we found several client interfaces that can be customized. Among them, we consider that Telegram app [5] is a valuable alternative since it offers the possibility of customizing the popular chat-based application by reusing most of features available. Such choice allows to include specific features enabling the localization, the discovery and the map rendering in a chat box. We finally remark that guidelines for choosing the proper technical solution are out of the scope for this work, but, it is worth to remark that these design decisions strongly depend on the considered use cases.

11.5 Towards Social Distancing through ILS

We now discuss some issues related to the concrete possibility of adopting an indoor localization system for the purpose of measuring the distance among users. This section covers different aspects of its adoption. In particular, in subsection 11.5.1 we discuss the impact on the privacy and and trust reputation of the ILS. Subsection 11.5.2 focuses on the discovery phase of ILS. Subsection 11.5.3 presents two alternatives for the social distancing, namely a manual and automatic approach and, lastly, subsection 11.5.4 concludes with a description of some challenges of the deployment phase of an ILS in real-world settings.

11.5.1 Privacy and Trust Reputation

Our first consideration faces with the problem of how to guarantee privacy of data collected by an ILS. We refer to [7] for complete survey also covering the following issues.

Privacy by design encompasses seven principles that should be followed [59]: proactive privacy protection instead of remedial action after privacy violations have hap-

[3] https://openlayers.org, https://leafletjs.com/
[4] https://www.indoorlocation.io/
[5] https://github.com/DrKLO/Telegram

pened; privacy as the default setting; privacy embedded into the design; full functionality with full privacy protection; privacy protection through the entire life cycle of the data; visibility and transparency; and respect for user privacy. Solutions for incorporating these principles in the design of an ILS are necessary. In parallel, data minimization approaches should be considered as a best practice for privacy by design adoption.

Furthermore, we argue that information sharing, active defense and automation methods should be integrated with an ILS. Thus, we consider mandatory to develop efficient methods to create, disseminate, and consume threat intelligence in a standardized and admissible way. It is also necessary to adopt defense mechanisms able to increasing the cyber adversary's cost by decreasing their overall efficiency of the active cyber operation. In parallel, in order to make the solutions effective, automation should be considered and solutions integrated into business workflow, governance, and structure control.

We also consider an orthogonal aspect of the privacy, namely the trust reputation of the ILS. Since the architecture described in Figure 11.1 involves a variety of components, it is required to implement different protections policies and to ensure that there are no privacy leaks at any of the stages we modelled in Figure 10.3. Additionally, the architecture should be deployable across different systems and environments maintaining the required level of trust.

Another aspect linked with the management of the trust of the system, is how to guarantee trust for third-party components that an ILS can integrate. As for example: multiples Map servers and different implementations of ILSs can coexist with the design presented in Section 11.3. To this purpose, we foresee some possible solutions: to provide interoperability recommendations and specifications; to define specific governance; to provide on-line verification and validation tools in order to identify the security risks. In parallel, data should be encrypted both at rest and in transit.

11.5.2 Discovering an ILS with Local and Global Interfaces

The capability of discovering an ILS automatically is a central aspect. We consider two possible approaches for the discovery phase: local and global. The local discovery is based on the analysis of local signals when entering a new environment. In this case, the user exploits short-range network interfaces looking up for nearby signals. However, we consider that a global search is required as well. In this last case, a standard search through a web-browser allows to query and to connect with the ILS. We recall the well-known user experience though which users look for services on a search engine. The search engine summarizes to the user a box with key information about the service, such as the street address, the opening hours, the popularity of the service (e.g., Google Popular Times). We expect to extend such list, by also reporting the information of the Indoor Localization System, e.g., showing an URL with the meta-information reported in Figure 10.4.

Mobility in multiple indoor environments increases privacy issues. Continuing on the example of the outdoor navigation services offered by Google, we know that the people who activate the history of their positions are tracked by Google, which, through the user account, allows you to view your movements and possibly eliminate them entirely. In the case of indoor navigation, this information will be collected by multiple subjects who must make it accessible to the owners of the data both for consultation and for modification. The task of the User Agent in this case becomes essential, because it must be capable of maintaining a history of the indoor sites visited. In particular, it must keep track of the policies and consents signed by the user, as well as links to the various interfaces to access the consultation and modification services of personal data. Nevertheless, much of this information must be conveyed during the Discovery process. Privacy management in general is more complex than the use case presented here, depending on whether the localization techniques used are Self-positioning or Remote positioning based. Systems that intrinsically guarantee privacy should be favored, in which the position is estimated by the User Agent (self-positioning) and is not known by other subjects, such systems are also more scalable. However, with respect to social distancing, you must in any case give up your rights and reveal your position even if used only anonymously, therefore defining an access control based on GDPR is always an indispensable step.

11.5.3 A Dichotomy of Manual and Automatic Social Distancing

Another crucial aspect is the safety distance among people (usually fixed in the range of 1-2 meters) which is normally perceivable on sight. People in favor of using automatic tools to support social distancing are already well prepared to keep the right distance from others. We observe two conflicting requirements: firstly, service providers (e.g., a shopping center) aim to increase the number of customers while, secondly, customers are interested to access a service scarcely populated. Therefore, a service obeying to the current prescriptions will grant the access to the maximum number of admitted customers. Such situation is generally perceived by the final users as potentially unhealthy, even if customers stay 1-2 meters away from others. Such consideration is predominantly of psychological nature. However, we argue that also the adoption of apps for preserving the social distance do not resolve the dichotomy between number of customers and distance among them. In fact, the false positive/negative alerts of such app, combined with the privacy issues previously mentioned, discourage their use in the daily basis.

Under this respect, the technology adopted by the apps is determinant for their successful adoption on the large scale. More specifically, range-based applications (i.e., based on Wi-Fi or Bluetooth signal strength) often fail in crowded scenarios or in those environments characterized by barriers and/or stances. Differently, the adoption of indoor localization system based on the data-fusion techniques are more reliable in such circumstances. Data-fusion allows to gather and to combine heterogeneous sensing and context information. Although more complexity with respect to a range-based approach, fusing data together allows to overcome issues such as body attenuation, indoor reflections and multi-path fading. The side-effect of an Indoor Localization System is the mostly represented by its installation costs.

The current trend is to adopt solutions for preserving the social distance that are based on apps for smartphones. We consider that such approach might fail on the large-scale and on the long-term. We consider necessary to understand those practicable alternatives and how to gradually move from the use of apps to the use of infrastructures, such as an Indoor Localization System.

If we consider that people are well predisposed for social distancing through the use of sight, a first discriminating factor is the type of environment. In open spaces, such as a supermarket, people will have greater ease of self-determination if a situation is risky or not. Differently, in indoor and constrained environments people need to be supported with automatic tools.

The transition from manual to automatic systems for social distancing requires bridging technologies able to reduce the deployment costs. As a remarkable example, we mention those systems designed to count the number of people in each room. Once a certain density has been reached, the system warns incoming people, in order to limit the access to such places. In any case, even if a precise localization system is not used, common interfaces must be studied through which to communicate to all end-users. Other aspect to consider is that the turnout of people could be estimated from the reservations that are made to visit a certain environment. This practice is currently used by the most visited museums, where you can buy tickets online and avoid long queues to buy tickets. In other environments such as airports, by integrating the various information systems of the airline companies, the number of people at a certain time can be determined on the basis of the scheduling of flights departing and arriving. Obviously this is an alternative to preparing new infrastructures for localization, but it is an estimate that can be affected by various random factors, lost reservations, flight delays, random congestion. But even in this case, an interface to people who access the environment or system is necessary to allow checking the crowding status and possibly receive notifications.

11.5.4 Deploying an ILS in Real-World Environments

We now discuss some deployment issues of an ILS at realistic conditions.

Deploying an ILS requires to accomplish at least the following two steps: survey of the environment and hardware installation and system calibration. Such steps are required for all the use-cases we detailed in Section 11.4.

The first step requires to visit the environment where the ILS is supposed to be deployed, with the goal of considering features of potential impact to the performance of the system. Some examples are: the building-material of the environment, the dimension of the area to be covered and the existence of outdoor/indoor areas. The building material of the environment has a great impact to the propagation of radio signals. As for example, concrete-based walls heavily attenuate 802.11 signals modulating at 2.4GHz, with respect to wooden or drywall. Moreover, the shape of the environment is another feature that influences the signal propagation. Wireless signals, generally, propagate more easily in open spaces due to the limited presence of obstacles. Finally, the existence of outdoor areas to be covered also influences the overall performance.

The previous step leads to the installation of the hardware required by an ILS. This step, usually, requires to find places where to deploy anchor nodes enabling the localization of the users, such as Wi-Fi Access Points, Bluetooth tags or UltraWide band boards. The hardware to be deployed often requires a power supply source in the nearby, the absence of surrounding obstacles and a safety distance from the end-users. The combination of such requirements makes the deployment a challenging task in places not designed for such purpose.

The last step copes with the configuration of the ILS. With the term configuration, we refer to all the settings depending on environmental settings. As a meaningful examples, we refer to the fingerprint-based techniques (see subsection 11.2.1). In this case, the localization system requires a database mapping the quality of the radio signals (e.g., Received Signal Strength Indicator) with a number of locations. Such database is generally built only after the hardware installation and it can be obtained with a data collection campaign often achieved manually by an expert. Another representative example of configuration is represented by all the algorithm settings of the ILS it-self. Such settings, very often, model features of the environment and they can be tuned only after the installation of the system in the target environment. Nevertheless, the configuration of an ILS is not one-shot task. Rather, real-world localization systems configured and re-configured multiple times during their life cycle. Some factors that require a new round of configuration are: environmental changes such as new obstacles or a new layout of the environment, new areas to be covered or modifications due to hardware replacement.

11.6 Measuring the Performance of the Integrated Architecture

We finally focus on the assessment of the performance of the integrated architecture as a crucial part of applicability of the solution we propose in this work. Our goal is to frame a reference architecture based on localization techniques for the purpose of measuring quantitatively the distance among people roaming in an indoor environment. In this picture, both the user experience and the hardware/software components can be measured to understand the effectiveness and its real applicability in real-world scenarios. To this purpose, we consider a set of measurable KPIs addressed to the four main players: the End-Users, the Smart Device, the Indoor Infrastructure and the Service Providers. We detail the motivation behind the such choices, how to measure the KPIs, the unit of measurement and any critical issue arising from the KPI. Table 11.3 summarizes the KPIs we propose.

Remark. Computing the inter-personal distance among people in real-time represents a challenging task. However, the recent COVID-19 pandemic imposes such requirement to the way people interacts and to the way people access services in indoor environments.

Countries affected by such pandemic reacted to the emergency in different ways by adopting counter-measures that, in some circumstances, might be not effective after the lock-down phase. In particular, we focus on exploitable technologies for guaranteeing social distance among people that are generally employed in the field of indoor localization. In this work, we describe the adoption of an Indoor Localization System (ILS) with a twofold goal. On one hand, the ILS can be adopted to localize people and, on the other hand, for measuring the in-between physical distance. We first present some functional requirements for an ILS and a reference architecture. Then, we present three significant use-cases where an ILS can be adopted for measuring distance among users. We discuss how information describing the distance among people can be used during three stages: before, during and after accessing a service. We also discuss some issues and new possible lines of investigation concerning the design of an ILS for the purpose of the social distance. In particular, our attention moves towards the design of discovery protocol able to identify available ILSs indoor and to the adoption of privacy mechanisms for the treatment of sensitive information collected about end-users. The

Table 11.3: *Evaluation framework of the reference architecture. Adopted from [24].*

	KPI	Objective	Measuring tool	Unit	Issues
End-user	Personalized feedbacks	To measure the overall impression of the final users (feel of protection, motivation is using the app)	questionnaire, survey	Statistics with the reported answers	1. Share the questionnaire 2. Bias of the answers caused by frustration and anxiety emotional states
Smart-Device	User acceptance	The success of ILS depends on the way the user interacts with it	users feedback: average number of scores received	Statistics with the reported answers	none
	Energy consumption	To measure the impact of the app to the battery life-cycle	Reporting APIs provided by iOS and Android SDKs	(Milli) Watts consumed by the app	none
	App usage	To estimate the usage of the app and the voluntary/involuntary stops of thew app	Reporting APIs provided by iOS and Android SDKs	1. Average usage time 2. Number of crash 3. Number of stops of the app	To manage appropriately any sensitive information collected
	Discovery and Access latency	To measure the time required by the app to discovery and access to the Indoor infrastructure	Profiling APIs available for iOS and Android SDK	Milliseconds	none
	Initial localization	To measure the time required by the app to compute the initial localization of the device	Custom reporting APIs profiling.	Milliseconds	none
	Maps Data transferred	To measure the amount of data transferred for rendering indoor maps	Profiling APIs available for iOS and Android SDK or Custom reporting APIs profiling	#byte	none
Indoor infrastructure	location latency	To measure the time required by the infrastructure to localize the smart device.	Custom profiling API server-side	Milliseconds	none
	ILS load	To measure the computational load of the ILS to estimate the position of all the devices connected	Performance profiling tools (e.g. Java JMX, Python DataDog client, Visual Studio profiler)	1. CPU load 2. RAM allocated 3. Data structure inspectio	Overhead of the profiling tool
	Map Server load	To measure the computational load of the Map Server to provide maps to the clients	Performance profiling tools (e.g. Java JMX interface, Python DataDog client, Visual Studio)	1. CPU load 2. RAM allocated 3. Data structure inspection 4. Data transfer rate	Overhead of the profiling tool
	performance of the proximity detection	To measure the correct detection of devices in proximity	Custom reporting API	Confusion matrix from which extract: Accuracy, Precision, F1, k-Statistics etc. metrics	To compare the results obtained with a reliable ground-truth to build the confusion matrix
	update location frequency	ToI measures the system's capacity to re-compute the user's locations seamlessly	Custom reporting API	Ratio between the number of received samples and the number of expected samples (for instance one every second).	none
Service-provider	Installation complexity	To measure the technical issues behind a correct deployment the indoor infrastructure	Custom reporting tool maintenance effort	Time of installation and of configuration,	none

letter point is, in our opinion, one of the most important barrier to the adoption and diffusion location-based services. We argue that a more transparent approach for the data treatment would benefit the adoption of such location-based services offered by ILSs.

Part V

Conclusion and Discussion

CHAPTER 12

Concluding Remarks

Pairing up Data Privacy and Data Security is becoming pivotal for promoting trustworthiness in services and products managing personal data, and for guaranteeing the data subject's rights. By defining the Integrity and Confidentiality principle (Art. 6.1(f)), the European legislator poses security at the heart of the GDPR. It dictates that personal data must be protected from unauthorised or unlawful processing. One of the cornerstone of security is the access control, which is ruled by access control policies specifying who is allowed to access Personal Data. However, the security of processing is not an isolated obligation, but comes together with the GDPR's Accountability principle (Art. 6.2). Indeed, according to this principle, security measures are at the same time an obligation and a technical means to implement other data protection obligations. Additionally, the GDPR imposes to the controllers and processors to adopt the Data Protection by Design and by Default (Art. 25), highlighting the necessity of engineering solutions for enforcing data privacy requirements into ICT services.

The solution promoted in this thesis was based on the definition of AC systems that guarantees compliance with the GDPR. This required the definition of Access Control Policies (ACPs) able to express requirements aligned with GDPR's provisions and features for automatically translating the natural language requirements of the law into technical ones. The promoted solution elicited different research activities including the *identification*, *extraction*, *translation* and *encoding* of the GDPR's requirements. In particular, the standardized and enforceable ACPs structure has been selected as final representation of the personal data in compliance with the GDPR.

According to the main goal of this thesis, i.e., *To leverage AC systems, the de facto mechanisms used to restrict data access, as a technical means for protecting "personal data by-design", and gaining legal compliance with the GDPR*, the research activity contributed to:

1. define a GDPR-based Life Cycle for authorization systems and a reference architecture, enabling data protection by-design;

2. leverage the state-of-the-art about legal ontology by defining a GDPR-based AC ontology useful for building ACPs in reference to the GDPR;

3. define a GDPR profile for a standardised AC language;

4. define a systematic approach for gathering and developing ACPs compliant-by-design with the regulation;

5. advance the notion of Data Protection Backlogs by introducing specific User Stories focused on GDPR provisions and their technical requirements;

6. enable an Agile development of ACSs;

7. define a comprehensive testing framework for validating both GDPR-based and traditional ACSs;

8. promote the application of ACSs in different contexts.

The contribution of this thesis has been developed in order to answer different research questions. For aim of completeness in the following we recall each of them, discussing the challenges and activities performed for gaining their final reply.

(RQ 1) How can authorization systems, and in particular AC, be used for guaranteeing compliance with the GDPR?
Inspired by Data Protection by Design obligation, we defined: i) a by-design GDPR-based Life Cycle for developing the access control system in compliance with the GDPR; ii) a reference architecture for its (semi)-automation which included different available artifacts (see Chapter 3). The Life Cycle was composed of eight main steps: from defining GDPR-based use case, developing, testing, deploying and reviewing both ACPs and ACMs. We voluntary conceived the Life Cycle abstract so to make it flexible and applicable in different contexts. Finally, five possible realistic scenarios, and their customization, have been presented demonstrating the feasibility of using authorization systems, and in particular ACs for guaranteeing compliance with the GDPR.

(RQ 2) To what extent can the GDPR's obligations be represented and enforced using Access Control Technologies?
To answer this research question, we have defined a systematic approach to gather access control requirements from the GDPR (see Chapter 6). In particular, we focused on improving and joining academic proposals with approaches adopted in industrial environment.

Enforcing the obligations using Access Control Technologies included three phases: the translation of the most suitable GDPR articles into a GDPR-based ACP templates; the definition of customized legal use case for each GDPR article related to ACP; and finally, the generation of enforceable ACPs in a given language. In this proposal, we referred to the ABAC model and to its standardized implementation XACML. Although grounded in a domain-related implementation (i.e., the GDPR), the proposed approach opened the path of a more general spectrum of researches that include: to adapt the methodology to other AC models (i.e., RBAC) and other AC languages, and to represent through AC Technologies any legal text that encodes data protection specification.

(RQ 3) Is it possible to gather technical requirements from the legal specifications defined in the GDPR?
To answer this question, we have proposed an Agile methodology to gather AC requirements from the GDPR by using the concept of User Stories (see Chapter 7). For this, a conceptual model of GDPR-based User Stories has been introduced. It unfolds the GDPR's structure of the mandatory articles into basic and concrete elements and includes three sub-models (i.e., the GDPR Model, User Stories Model and AC Model) useful for automatically translate the User Stories into AC policies.

Additionally, to demonstrate how to derive technical requirements from the legal specifications defined in the GDPR, example of application has been provided. It included: the selection of GDPR's articles related to access control; the definition of a Data Protection Backlog containing User Stories extracted from the selected GDPR's articles; and finally, the definition of access control rules, each related to a specific User Story.

Although grounded in a domain-related implementation (i.e., the GDPR), the Agile methodology yields a more general spectrum, since it can be applied to different data protection legislation that encodes ACPs specification.

(RQ 4) For accomplishing compliance with the GDPR, are there other supporting technologies that can be integrated with AC?
Thanks to the peculiarity of the AC, other supporting facilities to perform specific functionalities can be easily integrated. For accomplishing compliance with the GDPR, two of them have been identified: Semantic Web (in particular legal ontologies), and Consent Management.

We used the former to express GDPR concepts and relationships among them. In particular, among the legal ontologies we selected and leveraged PrOnto one. Thus, we presented the RAccOnto ontology (see Chapter 4) specifically conceived in the context of AC. We have also defined an XACML GDPR Policy Profile, that provided a set of standard attributes to be used within the policy to identify the GDPR's concepts.

Considering the Consent Management, we explored two possible ways of managing the consent: in Chapter 8, we explored the possibility to integrating an industrial Consent Management application; whereas in Chapters 10 and 11 we showed the use of the Kantara GDPR Explicit Consent Record as a reference format for collecting, managing and classifying the GDPR's concepts.

To speed up the integration of these technologies, a generic architecture has been also presented. Finally, demonstration of the feasibility of the proposed solution have been performed through specific use cases.

(RQ 5) In which application domains can Access Control Technologies be employed to achieve the GDPR compliance?
The widespread adoption of ACSs in ICTs made them ideal candidates for being adopted in different application domains. To demonstrate every application domain could use the ACSs to regulate access to Personal Data, three specific yet representative case studies have been proposed: Smart ICT Systems, Business Processes and ILSs.

In particular, in Smart ICT Systems we introduced appropriate supports to aid controllers in developing Privacy-By-Design Smart Services. Thus, we enhanced the generic architecture Smart ICT Systems with a new layer called *GDPR Manager* (see Chapter 8) which lets: i) the user-friendly interaction with end-users of the Smart ICT system (i.e., Data Subject and the Smart Services); ii) the managing of the domain dependent activities; and iii) the automatic derivation of ACPs in accordance with the collected consents.

Considering the Business Processes, in Chapter 9, we leveraged them to automatically enforce the GDPR provisions during the activities related to data management and analysis. Therefore, we integrated in the business process the use of a GDPR-based access control mechanism. For this, we have customized the proposed Life Cycle (see Chapter 3) and exemplified its use in: i) the identification of the target BPMN activities, and ii) the definition and adoption of the access control systems able to protect personal data during the BPMN modeling and execution.

Finally, considering the ILSs, we presented a reference architecture able to guarantee the GDPR compliance through the integration of specialized access control system enforcing the GDPR provisions (see Chapter 10). Additionally, we also promoted the adoption of GDPR-aware ILSs for the purpose of the social distancing by providing them with opportunely defined privacy mechanisms for the treatment of sensitive information collected about end-users (see Chapter 11).

(RQ 6) Is it possible to realize test environments for the validation of (GDPR-aware) access control systems?
In order to reply this research question (see Chapter 5), we defined a testing framework capable to formally validate both ACPs and ACMs, by enabling to conduct CEs in the context of AC. In particular, the testing process can be used for: test strategy selection and derivation, test case execution and results evaluation, and finally Oracle definition. Additionally, for assessing GDPR-based test cases generation strategies, we defined a generic methodology based on mutation analysis and we illustrated the development of a CE experiment in the context of AC.

Gantt Chart: RQs and Related Scientific Contributions. For the aim of completeness, in Figure 12.1, we report a graphical representation of the lifetime of each RQ.

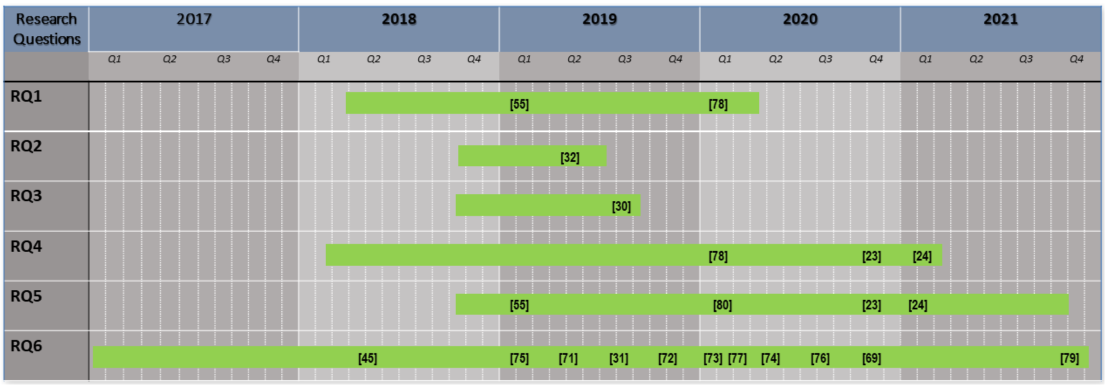


Figure 12.1: *Gantt Chart: RQs and Related Scientific Contributions.*

More precisely, we report a Gantt chart showing how the six RQs have been addressed over time. Specifically, each green line represents the starting and the end date of each research question. Additionally, the numbers in each green line refer to the related scientific contributions. In particular, the numbers position highlights either journal publication time or the conference (workshop) presentation time. As reported in the figure, there are cases in which the same publication targets two or more RQs (for instance n. 23 or n. 78).

12.1 Future Works and Open Problems

This thesis covers different research topics and technologies. Despite the accuracy devoted in investigating the main goal and all its research questions, future works are still possible and open problems still a challenge. In this section, we report a list of the main activities that could be performed as future works.

Standardization of the XACML GDPR Policy Profile. This research gave us the opportunity to study the limitations of currently proposed XACML Privacy Profile. However, in its original structure, the profile is not sufficiently adequate for representing all the GDPR's requirements: indeed, it targets just the concept of purpose. In this thesis, we have presented a possible extension of the attributes of XACML Privacy Profile for encoding GDPR's concepts in XACML policies. However, investigations about its adequacy and its impact on the scientific community should be still performed. We consider the standardization of our XACML GDPR Policy Profile a possible and achievable future work.

Validation of the Results Through Real Case Studies. In writing this thesis, we provided several examples of application of the GENERAL_D proposal to realistic applications or case studies. However, the complete validation into a real context is still missing. Indeed, COVID-19 introduced delay in the collaboration with the partners involved in two EU projects (namely BIECO[1] and CyberSec4Europe[2]) and prevented us for the collection of results from the real application field. We are confident that in the next (hopefully) few months we move quick to a positive resolution, and real validation data could be collected.

Discussions with Legal Experts. The proposals presented in this thesis were guided, and sometimes developed together, with data protection legal experts. We therefore consider the thesis solutions and results internally validated. However, external validation is still necessary. Independent legal experts should be put in the loop, so as to validate whether the developed ACPs can capture and express the legal meaning of the related GDPR's provisions. This will also quantify the completeness and the correctness of the translation of the norms. In the future, we want to adopt the methodology proposed by [34]. Here authors conceived an interdisciplinary agile methodology that put in loop a pool of legal experts for validating their proposal, which encodes the GDPR's articles in LegalRuleML. This kind of investigation could shed the light on the effectiveness of our approach within legal communities.

Methodology to Verify and Demonstrate the Compliance with the GDPR. The accountability principle dictates that "controller shall be responsible for, and be able to demonstrate compliance with" the other principles of the regulation. In this thesis, we extensively

[1] Building Trust in Ecosystems and Ecosystem Components (BIECO) Programme H2020 Grant Agreement No 952702. https://www.bieco.org/

[2] CyberSec4Europe H2020 Programme Grant Agreement No. 830929. https://cybersec4europe.eu

explored the compliance, but future works still remain for the demonstration. As future work, we would like to investigate the proposals of: [156] for using blockchain technologies for auditability and accountability purpose; [81] for addressing specific challenge within the XACML context or [61] for extending the language. All these solutions help guaranteeing the accountability-by-design (ex-ante). In contrast, the work in [17] proposes an ex-post solution by defining a specific language to validate accesses for demonstrating the compliance with the regulation starting from log analysis.

Release the Reference Architecture. During this thesis, different implementations of the GENERAL_D architecture have been provided. It is part of our future works to release a standardized reference architecture that could be easily customized for different applications.

User Stories Templates in Other Contexts. The defined User Stories could be specialized considering other contexts. For this, we are currently investigating a comprehensive Data Protection Impact Assessment (DPIA) methodology (which is one of the legal requirements in the GDPR (Art. 35)) for leveraging the conceived Data Protection Backlog. It is part of the ongoing work also the validation of the provided User Stories by different partners in the context of European projects that address key regulations such as the GDPR. Specifically, within the CyberSec4Europe project, we are validating the obtained results in two project's demonstrators in the context of Smart-City; whereas in BIECO project we are using User Stories as specification of privacy claims to be satisfied during the implementation of System-of-System (SoS) that processes Personal Data.

GENERAL_D for other Legal Frameworks. Applying and adapting our methodology for other legal requirements seems to be interesting future development. Indeed, although grounded in a domain-related implementation (i.e., compliance with the GDPR), the approaches we have conceived throughout this research yields a more general spectrum, since it can be applied to different data protection regulations and, more in general, to any legal text that implicitly contains or suggests data protection requirements. This paves the way for investigating the adoption of our approach to the new coming ePrivacy regulation as well as to the eIDAS regulation.

Bibliography

[1] ifeel-you bracelet. https://www.iit.it/iit-vs-covid-19/ifeel-you-bracelet.

[2] Immuni, uno strumento in più contro l'epidemia. `https://www.immuni.italia.it/`.

[3] Kunai. `https://github.com/kunai-consulting/OpenTrace`.

[4] Skyook. https://syook.com/the-social-distancing-app/.

[5] ufirst, risparmia tempo con ufirst. https://www.ufirst.com/.

[6] Who has access to your smartphone data? https://cacm.acm.org/magazines/2020/10/247585-who-has-access-to-your-smartphone-data/fulltext.

[7] Deliverable D4.3: Research and Development Roadmap. `http:\cybersec4europe.eu\wp-content\uploads\2020\02\D4.3.Research-and-Development-Roadmap-1-Submitted.pdf`, 2020.

[8] Accountability Act. Health insurance portability and accountability act of 1996. *Public law*, 104:191, 1996.

[9] Amir Shayan Ahmadian, Daniel Strüber, Volker Riediger, and Jan Jürjens. Supporting privacy impact assessment by model-based privacy analysis. In *Proceedings of the The 33rd ACM/SIGAPP Symposium On Applied Computing (SAC)*. ACM, April 2018.

[10] J Ahola, C Frühwirth, M Helenius, L Kutvonen, J Myllylahti, T Nyberg, A Pietikäinen, P Pietikäinen, J Röning, S Ruohomaa, et al. Handbook of the secure agile software development life cycle. *University of Oulu*, 2014.

[11] Agnes Åkerlund and Christine Große. Integration of data envelopment analysis in business process models: A novel approach to measure information security. In *ICISSP*, pages 281–288, 2020.

[12] Raed Saeed Al-Dhubhani, Jonathan Cazalas, Rashid Mehmood, Iyad Katib, and Faisal Saeed. A framework for preserving location privacy for continuous queries. In *International Conference of Reliable Information and Communication Technology*, pages 819–832. Springer, 2019.

[13] Khalid Alissa, Jason Reid, Ed Dawson, and Farzad Salim. Bp-xacml: An authorisation policy language for business processes. In *Information Security and Privacy: 20th Australasian Conference, ACISP 2015, Proceedings [Lecture Notes in Computer Science, Volume 9144]*, pages 307–325. Springer, 2015.

[14] S. Alletto, R. Cucchiara, G. Del Fiore, L. Mainetti, V. Mighali, L. Patrono, and G. Serra. An indoor location-aware system for an iot-based smart museum. *IEEE Internet of Things Journal*, 3(2):244–253, 2016.

[15] Manar Alohaly, Hassan Takabi, and Eduardo Blanco. Automated extraction of attributes from natural language attribute-based access control (ABAC) policies. *Cybersecurity*, 2(1):2, 2019.

[16] Krzysztof Apt. *Principles of constraint programming*. Cambridge university press, 2003.

[17] Emma Arfelt, David Basin, and Søren Debois. Monitoring the gdpr. In Kazue Sako, Steve Schneider, and Peter Y. A. Ryan, editors, *Computer Security – ESORICS 2019*, pages 681–699, Cham, 2019. Springer International Publishing.

[18] Paul Ashton and Linamaria Pintor-Escobar. Five ways to tackle phd research anxieties triggered by covid-19 lockdowns. Nature, June 2020.

[19] Vishal Asthana, Izar Tarandach, Niall ODonoghue, Bryan Sullivan, and Mikko Saario. Practical security stories and security tasks for agile development environments. *Online, July*, 2012.

[20] Zulkarnain Azham, Imran Ghani, and Norafida Ithnin. Security backlog in scrum security practices. In *2011 Malaysian Conference in Software Engineering*, pages 414–417. IEEE, 2011.

[21] Earl T Barr, Mark Harman, Phil McMinn, Muzammil Shahbaz, and Shin Yoo. The oracle problem in software testing: A survey. *IEEE transactions on software engineering*, 41(5):507–525, 2015.

[22] Valentín Barral, Pedro Suárez-Casal, Carlos J. Escudero, and José A. García-Naya. Multi-sensor accurate forklift location and tracking simulation in industrial indoor environments. *Electronics*, 8(10):1152, Oct 2019.

[23] Paolo Barsocchi, Antonello Calabrò, Antonino Crivello, Said Daoudagh, Francesco Furfari, Michele Girolami, and Eda Marchetti. A privacy-by-design architecture for indoor localization systems. In Martin J. Shepperd, Fernando Brito e Abreu, Alberto Rodrigues da Silva, and Ricardo Pérez-Castillo, editors, *Quality of Information and Communications Technology - 13th International Conference, QUATIC 2020, Faro, Portugal, September 9-11, 2020, Proceedings*, volume 1266 of *Communications in Computer and Information Science*, pages 358–366. Springer, 2020.

[24] Paolo Barsocchi, Antonello Calabrò, Antonino Crivello, Said Daoudagh, Francesco Furfari, Michele Girolami, and Eda Marchetti. COVID-19 & privacy: Enhancing of indoor localization architectures towards effective social distancing. *Array*, 9:100051, 2021.

[25] Paolo Barsocchi, Antonello Calabrò, Erina Ferro, Claudio Gennaro, Eda Marchetti, and Claudio Vairo. Boosting a low-cost smart home environment with usage and access control rules. *Sensors*, 18(6):1886, 2018.

[26] Cesare Bartolini. Software testing techniques revisited for owl ontologies. In *International Conference on Model-Driven Engineering and Software Development*, pages 132–153. Springer, 2016.

[27] Cesare Bartolini, Antonello Calabró, and Eda Marchetti. Integrating gdpr in business process modeling. In *Technical Report*, 2018.

[28] Cesare Bartolini, Antonello Calabrò, and Eda Marchetti. Enhancing business process modelling with data protection compliance: An ontology-based proposal. In *Proceedings of the 5th International Conference on Information Systems Security and Privacy, ICISSP 2019, Prague, Czech Republic, February 23-25, 2019.*, pages 421–428, 2019.

[29] Cesare Bartolini, Antonello Calabrò, and Eda Marchetti. GDPR and business processes: an effective solution. In *Proceedings of the 2nd International Conference on Applications of Intelligent Systems, APPIS 2019, Las Palmas de Gran Canaria, Spain, January 07-09, 2019*, pages 7:1–7:5, 2019.

[30] Cesare Bartolini, Said Daoudagh, Gabriele Lenzini, and Eda Marchetti. Gdpr-based user stories in the access control perspective. In *Quality of Information and Communications Technology - 12th International Conference, QUATIC 2019, Ciudad Real, Spain, September 11-13, 2019, Proceedings*, pages 3–17, 2019.

[31] Cesare Bartolini, Said Daoudagh, Gabriele Lenzini, and Eda Marchetti. Testing of GDPR-based access control policies. In *Poster Session of ESORICS 2019, Luxembourg, Luxembourg, September 23-27, 2019*, 2019.

[32] Cesare Bartolini, Said Daoudagh, Gabriele Lenzini, and Eda Marchetti. Towards a lawful authorized access: A preliminary gdpr-based authorized access. In *14th 14th International Conference on Software Technologies (ICSOFT 2019), Prague, Czech Republic, July 26-28, 2019.*, pages 331–338, 2019.

[33] Cesare Bartolini, Andra Giurgiu, Gabriele Lenzini, and Livio Robaldo. Towards legal compliance by correlating standards and laws with a semi-automated methodology. In *Benelux Conference on Artificial Intelligence*, pages 47–62. Springer, 2016.

[34] Cesare Bartolini, Gabriele Lenzini, and Cristiana Santos. An agile approach to validate a formal representation of the gdpr. In Kazuhiro Kojima, Maki Sakamoto, Koji Mineshima, and Ken Satoh, editors, *New Frontiers in Artificial Intelligence*, pages 160–176, Cham, 2019. Springer International Publishing.

[35] V. R. Basili and H. D. Rombach. The tame project: towards improvement-oriented software environments. *IEEE Transactions on Software Engineering*, 14(6):758–773, Jun 1988.

[36] Victor R. Basili, Gianluigi Caldiera, and H. Dieter Rombach. The goal question metric approach. In *Encyclopedia of Software Engineering*. Wiley, 1994.

[37] 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.

[38] Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *Scientific American*, 284(5):28–37, 2001.

[39] Elisa Bertino, Piero A. Bonatti, and Elena Ferrari. TRBAC: A temporal role-based access control model. *ACM Trans. Inf. Syst. Secur.*, 4(3):191–233, 2001.

[40] Elisa Bertino, Gabriel Ghinita, and Ashish Kamra. Access control for databases: Concepts and systems. *Foundations and Trends® in Databases*, 3(1–2):1–148, 2011.

[41] A. Bertolino, S. Daoudagh, F. Lonetti, and E. Marchetti. Automatic XACML Requests Generation for Policy Testing. In *Proc. of ICST*, pages 842–849, April 2012.

[42] A. Bertolino, S. Daoudagh, F. Lonetti, and E. Marchetti. Xacmut: Xacml 2.0 mutants generator. In *Proc. of 8th International Workshop on Mutation Analysis*, pages 28–33, 2013.

[43] A. Bertolino, F. Lonetti, and E. Marchetti. Systematic XACML Request Generation for Testing Purposes. In *Proc. of SEAA*, pages 3–11, 2010.

[44] Antonia Bertolino, Said Daoudagh, Donia El Kateb, Christopher Henard, Yves Le Traon, Francesca Lonetti, Eda Marchetti, Tejeddine Mouelhi, and Mike Papadakis. Similarity testing for access control. *Information and Software Technology*, 58:355 – 372, 2015.

[45] Antonia Bertolino, Said Daoudagh, Francesca Lonetti, and Eda Marchetti. An automated model-based test oracle for access control systems. In *Proceedings of the 13th International Workshop on Automation of Software Test, AST@ICSE 2018, Gothenburg, Sweden, May 28-29, 2018*, pages 2–8, 2018.

[46] Antonia Bertolino, Said Daoudagh, Francesca Lonetti, Eda Marchetti, Fabio Martinelli, and Paolo Mori. Testing of polpa-based usage control systems. *Software Quality Journal*, 22(2):241–271, 2014.

[47] Antonia Bertolino, Said Daoudagh, Francesca Lonetti, Eda Marchetti, and Louis Schilders. Automated testing of extensible access control markup language-based access control systems. *IET Software*, 7(4):203–212, 2013.

[48] Antonia Bertolino, Yves Le Traon, Francesca Lonetti, Eda Marchetti, and Tejeddine Mouelhi. Coverage-based test cases selection for xacml policies. In *Proceedings of ICST Workshops*, pages 12–21, 2014.

[49] Felix Bieker, Nicholas Martin, Michael Friedewald, and Marit Hansen. Data protection impact assessment. In Marit Hansen, Eleni Kosta, Igor Nai-Fovino, and Simone Fischer-Hübner, editors, *Privacy and Identity Management*, volume 526 of *IFIP Advances in Information and Communication Technology*, pages 207–220. Springer, 2018.

[50] K. Bregar and M. Mohorčič. Improving indoor localization using convolutional neural networks on computationally restricted devices. *IEEE Access*, 6:17429–17441, 2018.

[51] Lionel Briand and Yvan Labiche. Empirical studies of software testing techniques: Challenges, practical strategies, and future research. *SIGSOFT Softw. Eng. Notes*, 29(5):1–3, September 2004.

[52] Dan Brickley and R.V. Guha. RDF Schema 1.1. W3C Recommendation, World Wide Web Consortium, February 2014. `https://www.w3.org/TR/rdf-schema/`.

[53] 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.

[54] Erik Buchmann and Jürgen Anke. Privacy patterns in business processes. In Maximilian Eibl and Martin Gaedke, editors, *Proceedings of the 47. Jahrestagung der Gesellschaft für Informatik (INFORMATIK)*, pages 793–798. Gesellschaft für Informatik, September 2017.

[55] Antonello Calabrò, Said Daoudagh, and Eda Marchetti. Integrating access control and business process for GDPR compliance: A preliminary study. In *Proceedings of the Third Italian Conference on Cyber Security, Pisa, Italy, February 13-15, 2019.*, 2019.

[56] Antonello Calabrò, Francesca Lonetti, and Eda Marchetti. Access control policy coverage assessment through monitoring. In *Proc. of TELERISE*, pages 373–383, 2017.

[57] Antonello Calabrò, Eda Marchetti, Davide Moroni, and Gabriele Pieri. A dynamic and scalable solution for improving daily life safety. In *Proceedings of the 2nd International Conference on Applications of Intelligent Systems*, pages 1–6, 2019.

[58] Renato Carauta Ribeiro and Edna Dias Canedo. Using mcda for selecting criteria of lgpd compliant personal data security. In *The 21st Annual International Conference on Digital Government Research*, dg.o '20, page 175–184, New York, NY, USA, 2020. Association for Computing Machinery.

[59] Ann Cavoukian. *Privacy by Design: Leadership, Methods, and Results*, pages 175–202. Springer Netherlands, Dordrecht, 2013.

[60] Ann Cavoukian et al. Privacy by design: The 7 foundational principles. *Information and privacy commissioner of Ontario, Canada*, 5:12, 2009.

[61] 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.

[62] David W Chadwick and Kaniz Fatema. An advanced policy based authorisation infrastructure. In *Proceedings of the 5th ACM workshop on Digital identity management*, pages 81–84. ACM, 2009.

[63] Omar Chowdhury, Haining Chen, Jianwei Niu, Ninghui Li, and Elisa Bertino. On xacml's adequacy to specify and to enforce hipaa. In *Proceedings of the 3rd USENIX Conference on Health Security and Privacy*, HealthSec'12, pages 11–11, Berkeley, CA, USA, 2012. USENIX Association.

[64] Mike Cohn. *User stories applied: For agile software development*. Addison-Wesley Professional, 2004.

[65] Maria Luisa Damiani, Elisa Bertino, Barbara Catania, and Paolo Perlasca. GEO-RBAC: A spatially aware RBAC. *ACM Trans. Inf. Syst. Secur.*, 10(1):2, 2007.

[66] S. Daoudagh, F. Lonetti, and E. Marchetti. XACMET: XACML Testing & Modeling. *Software Quality Journal*, 2019.

[67] Said Daoudagh. A Data Warehouse and a Framework for the Validation and Testing of Access Control Systems. Master's thesis, Department of Computer Science, University of Pisa, Italy, 2017.

[68] Said Daoudagh, Donia El Kateb, Francesca Lonetti, Eda Marchetti, and Tejeddine Mouelhi. A toolchain for model-based design and testing of access control systems. In *Proc.of MODELSWARD*, pages 411–418. IEEE, 2015.

[69] Said Daoudagh, Francesca Lonetti, and Eda Marchetti. An automated framework for continuous development and testing of access control systems. *Journal of Software: Evolution and Process*, n/a(n/a):e2306. e2306 smr.2306.

[70] Said Daoudagh, Francesca Lonetti, and Eda Marchetti. Assessment of access control systems using mutation testing. In *Proceedings of the First International Workshop on TEchnical and LEgal aspects of data pRIvacy*, pages 8–13. IEEE Press, 2015.

[71] Said Daoudagh, Francesca Lonetti, and Eda Marchetti. A decentralized solution for combinatorial testing of access control engine. In *Proceedings of the 5th International Conference on Information Systems Security and Privacy*, ICISSP'19, 2019.

[72] Said Daoudagh, Francesca Lonetti, and Eda Marchetti. A framework for the validation of access control systems. In Andrea Saracino and Paolo Mori, editors, *Proceedings of the 2nd International Workshop on Emerging Technologies for Authorization and Authentication*, 2019.

[73] Said Daoudagh, Francesca Lonetti, and Eda Marchetti. Assessing testing strategies for access control systems: A controlled experiment. In *Proceedings of ICISSP 2020, Valletta, Malta, February 25-27, 2020*, 2020.

[74] Said Daoudagh, Francesca Lonetti, and Eda Marchetti. Continuous development and testing of access and usage control: A systematic literature review. In *ESSE 2020: 2020 European Symposium on Software Engineering, Rome, Italy, November 6-8, 2020*, pages 51–59. ACM, 2020.

[75] Said Daoudagh, Francesca Lonetti, and Eda Marchetti. A general framework for decentralized combinatorial testing of access control engine: Examples of application. In Paolo Mori, Steven Furnell, and Olivier Camp, editors, *Information Systems Security and Privacy*, pages 207–229, Cham, 2020. Springer International Publishing.

[76] Said Daoudagh, Francesca Lonetti, and Eda Marchetti. XACMET: XACML testing & modeling. *Softw. Qual. J.*, 28(1):249–282, 2020.

[77] Said Daoudagh and Eda Marchetti. Defining controlled experiments inside the access control environment. In Slimane Hammoudi, Luís Ferreira Pires, and Bran Selic, editors, *Proceedings of the 8th International Conference on Model-Driven Engineering and Software Development, MODELSWARD 2020, Valletta, Malta, February 25-27, 2020*, pages 167–176. SCITEPRESS, 2020.

[78] Said Daoudagh and Eda Marchetti. A life cycle for authorization systems development in the GDPR perspective. In Michele Loreti and Luca Spalazzi, editors, *Proceedings of the Fourth Italian Conference on Cyber Security, Ancona, Italy, February 4th to 7th, 2020*, volume 2597 of *CEUR Workshop Proceedings*, pages 128–140. CEUR-WS.org, 2020.

[79] Said Daoudagh and Eda Marchetti. Graduation: A gdpr-based mutation methodology. In *Quality of Information and Communications Technology - 14th International Conference, QUATIC 2021, Faro, Portugal, September 8-11, 2021, Proceedings*, pages –, 2021.

[80] Said Daoudagh, Eda Marchetti, Vincenzo Savarino, Roberto Di Bernardo, and Marco Alessi. How to improve the GDPR compliance through consent management and access control. In Paolo Mori, Gabriele Lenzini, and Steven Furnell, editors, *Proceedings of the 7th International Conference on Information Systems Security and Privacy, ICISSP 2021, Online Streaming, February 11-13, 2021*, pages 534–541. SCITEPRESS, 2021.

[81] M. Davari and E. Bertino. Access control model extensions to support data privacy protection based on gdpr. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 4017–4024, Dec 2019.

[82] Maryam Davari and Elisa Bertino. Reactive access control systems. In *Proceedings of the 23Nd ACM on Symposium on Access Control Models and Technologies*, SACMAT '18, pages 205–207, New York, NY, USA, 2018. ACM.

[83] Farah Dernaika, Nora Cuppens-Boulahia, Frédéric Cuppens, and Olivier Raynaud. Accountability in the A posteriori access control: A requirement and a mechanism. In *Quality of Information and Communications Technology - 13th International Conference, QUATIC 2020, Faro, Portugal, September 9-11, 2020, Proceedings*, volume 1266 of *Communications in Computer and Information Science*, pages 332–342. Springer, 2020.

[84] Vasiliki Diamantopoulou, Nikolaos Argyropoulos, Christos Kalloniatis, and Stefanos Gritzalis. Supporting the design of privacy-aware business processes via privacy process patterns. In *Proceedings of the 11th International Conference on Research Challenges in Information Science (RCIS)*. IEEE, May 2017.

[85] Hyunsook Do, Sebastian Elbaum, and Gregg Rothermel. Infrastructure support for controlled experimentation with software testing and regression testing techniques. In *Empirical Software Engineering, 2004. ISESE'04. Proceedings. 2004 International Symposium on*, pages 60–70. IEEE, 2004.

[86] Hyunsook Do, Sebastian Elbaum, and Gregg Rothermel. Supporting controlled experimentation with testing techniques: An infrastructure and its potential impact. *Empirical Software Engineering*, 10(4):405–435, 2005.

[87] Bob Duncan. Can EU general data protection regulation compliance be achieved when using cloud computing? In *Proceedings of the Ninth International Conference on Cloud Computing, GRIDs, and Virtualization (CLOUD COMPUTING)*, pages 1–6. IARIA, February 2018.

[88] Sebastian Elbaum, Alexey G Malishevsky, and Gregg Rothermel. Test case prioritization: A family of empirical studies. *IEEE Transactions on Software Engineering*, 28(2):159–182, 2002.

[89] Rik Eshuis and Paul W. P. J. Grefen. Constructing customized process views. *Data Knowl. Eng.*, 64(2):419–438, 2008.

[90] 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*, L119:1–88, May 2016.

[91] Kaniz Fatema, David W. Chadwick, and Stijn Lievens. A multi-privacy policy enforcement system. In *Privacy and Identity Management for Life*, pages 297–310, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[92] Kaniz Fatema, Christophe Debruyne, Dave Lewis, Declan OSullivan, 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.

[93] Fedora. Fedora Commons Repository Software. `http://fedora-commons.org/`.

[94] David F Ferraiolo, Ravi Sandhu, Serban Gavrila, D Richard Kuhn, and Ramaswamy Chandramouli. Proposed nist standard for role-based access control. *ACM Transactions on Information and System Security (TISSEC)*, 4(3):224–274, 2001.

[95] 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.

[96] K. Fisler, S. Krishnamurthi, L.A. Meyerovich, and M.C. Tschantz. Verification and change-impact analysis of access-control policies. In *Proc. of ICSE*, pages 196–205, 2005.

[97] Elena Fleacă, Bogdan Fleacă, and Sanda Maiduc. Process modeling as key technique for embedding the practices of business process management in organization. In *International Conference on Exploring Services Science*, pages 89–99. Springer, 2016.

[98] Martin Fowler, Jim Highsmith, et al. The agile manifesto. *Software Development*, 9(8):28–35, 2001.

[99] Francesco Furfari, Antonino Crivello, Paolo Barsocchi, Filippo Palumbo, and Francesco Potortì. What is next for indoor localisation? taxonomy, protocols, and patterns for advanced location based services. In *2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–8. IEEE.

[100] Barton Gellman and Jerry Markon Gribbons. Edward snowden says motive behind leaks was to expose surveillance state. 2013.

[101] Kyriakos Georgiou, Timotheos Constambeys, Christos Laoudias, Lambros Petrou, Georgios Chatzimilioudis, and Demetrios Zeinalipour-Yazti. Anyplace: A crowdsourced indoor information service. In *2015 16th IEEE International Conference on Mobile Data Management*, volume 1, pages 291–294. IEEE, 2015.

[102] Davide Giovanelli and Elisabetta Farella. Rssi or time-of-flight for bluetooth low energy based localization? an experimental evaluation. In *2018 11th IFIP Wireless and Mobile Networking Conference (WMNC)*, pages 1–8. IEEE, 2018.

[103] M. Girolami, P. Barsocchi, S. Chessa, and F. Furfari. A social-based service discovery protocol for mobile ad hoc networks. In *2013 12th Annual Mediterranean Ad Hoc Networking Workshop (MED-HOC-NET)*, pages 103–110, 2013.

[104] M. Girolami, F. Mavilia, F. Delmastro, and E. Distefano. Detecting social interactions through commercial mobile devices. In *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 125–130, 2018.

[105] M. Golfarelli and S. Rizzi. *From Star Schemas to Big Data: 20+ Years of Data Warehouse Research*, pages 93–107. Springer International Publishing, Cham, 2018.

[106] António Gonçalves, Anacleto Correia, and Luis Cavique. Data protection risk modeling into business process analysis. In Osvaldo Gervasi, Beniamino Murgante, Sanjay Misra, Giuseppe Borruso, Carmelo M. TorreAna, Maria A.C. Rocha, David Taniar, Bernady O. Apduhan, Elena Stankova, and Alfredo Cuzzocrea, editors, *Computational Science and Its Applications – ICCSA 2017*, volume 10404 of *Lecture Notes in Computer Science*, pages 667–676. Springer, 2017.

[107] Duarte Gonçalves-Ferreira, Mariana Leite, Cátia Santos-Pereira, Manuel E. Correia, Luis Antunes, and Ricardo Cruz-Correia. HS.Register. In Adrien Ugon, Daniel Karlsson, Gunnar O. Klein, and Anne Moen, editors, *Building Continents of Knowledge in Oceans of Data*, volume 247 of *Studies in Health Technology and Informatics*, pages 81–85. IOS Press, 2018.

[108] Brian Greaves, Marijke Coetzee, and Wai Sze Leung. Access control requirements for physical spaces protected by virtual perimeters. In Steven Furnell, Haralambos Mouratidis, and Günther Pernul, editors, *Trust, Privacy and Security in Digital Business*, pages 182–197, Cham, 2018. Springer International Publishing.

[109] Brian Greaves, Marijke Coetzee, and Wai Sze Leung. A comparison of indoor positioning systems for access control using virtual perimeters. In Xin-She Yang, Simon Sherratt, Nilanjan Dey, and Amit Joshi, editors, *Fourth International Congress on Information and Communication Technology*, pages 293–302, Singapore, 2020. Springer Singapore.

[110] J. Haofeng and G. Xiaorui. Wi-fi secure access control system based on geo-fence. In *2019 IEEE Symposium on Computers and Communications (ISCC)*, pages 1–6, 2019.

[111] Suining He and S-H Gary Chan. Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons. *IEEE Communications Surveys & Tutorials*, 18(1):466–490, 2015.

[112] Emil Heuck, Thomas T Hildebrandt, Rasmus Kiærulff Lerche, Morten Marquard, Håkon Normann, Rasmus Iven Strømsted, and Barbara Weber. Digitalising the general data protection regulation with dynamic condition response graphs. In *Proceedings of the 15th International Conference on Business Process Management (BPM)*, pages 124–134, September 2017.

[113] S. Holcer, J. Torres-Sospedra, M. Gould, and I. Remolar. Privacy in indoor positioning systems: A systematic review. In *2020 International Conference on Localization and GNSS (ICL-GNSS)*, pages 1–6, 2020.

[114] J. Hsu. The dilemma of contact-tracing apps: Can this crucial technology be both effective and private? *IEEE Spectrum*, 57(10):56–59, 2020.

[115] Chung Tong Hu, David F Ferraiolo, David R Kuhn, Adam Schnitzer, Kenneth Sandlin, Robert Miller, and Karen Scarfone. Guide to attribute based access control (abac) definition and considerations [includes updates as of 02-25-2019]. Technical report, 2019.

[116] Vincent C Hu, Rick Kuhn, and Dylan Yaga. Verification and test methods for access control policies/models. *NIST Special Publication*, 800:192, 2017.

[117] JeeHyun Hwang, Evan Martin, Tao Xie, and Vincent C. Hu. Policy-based testing. In *Encyclopedia of Software Engineering*, pages 673–683. Taylor & Francis, 2011.

[118] JeeHyun Hwang, Tao Xie, Vincent Hu, and Mine Altunay. Acpt: A tool for modeling and verifying access control policies. In *Proc. of International Symposium on Policies for Distributed Systems and Networks (POLICY)*, pages 40–43, 2010.

[119] Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter [Van den Broeck]. What's in a crowd? analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, 271(1):166 – 180, 2011.

[120] S. Islam and P. Falcarin. Measuring security requirements for software security. In *2011 IEEE 10th International Conference on Cybernetic Intelligent Systems (CIS)*, pages 70–75, Sep. 2011.

[121] IT Governance Privacy Team. *EU General Data Protection Regulation (GDPR)*. IT Governance Publishing, second edition, 2017.

[122] Andreas Jedlitschka and Dietmar Pfahl. Reporting guidelines for controlled experiments in software engineering. In *Empirical Software Engineering, 2005. 2005 International Symposium on*, pages 10–pp. IEEE, 2005.

[123] Christian Damsgaard Jensen, Kristine Geneser, and Ida C. Willemoes-Wissing. Sensor enhanced access control: Extending traditional access control models with context-awareness. In Carmen Fernández-Gago, Fabio Martinelli, Siani Pearson, and Isaac Agudo, editors, *Trust Management VII*, pages 177–192, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[124] John Jeston and Johan Nelis. *Business Process Management*. Routledge, 3^{rd} edition, 2014.

[125] Xin Jin, Ram Krishnan, and Ravi Sandhu. A unified attribute-based access control model covering dac, mac and rbac. In Nora Cuppens-Boulahia, Frédéric Cuppens, and Joaquin Garcia-Alfaro, editors, *Data and Applications Security and Privacy XXVI*, pages 41–55, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[126] Charles M Judd, Eliot R Smith, and Louise H Kidder. Research methods in social relations, fort worth: Holt, rinehart and winston, 1991.

[127] René Just. The major mutation framework: Efficient and scalable mutation analysis for java. In *Proceedings of the 2014 international symposium on software testing and analysis*, pages 433–436. ACM, 2014.

[128] Natalia Juristo Juzgado and Ana María Moreno. *Basics of software engineering experimentation*. Kluwer, 2001.

[129] K. Järvinen, H. Leppäkoski, E. Lohan, P. Richter, T. Schneider, O. Tkachenko, and Z. Yang. PILOT: Practical privacy-preserving indoor localization using outsourcing. In *2019 IEEE European Symposium on Security and Privacy (EuroS P)*, pages 448–463, 2019.

[130] M. Kassab. The changing landscape of requirements engineering practices over the past decade. In *2015 IEEE EmpiRE*, pages 1–8, Aug 2015.

[131] Meryem Kassou and Laila Kjiri. A goal question metric approach for evaluating security in a service oriented architecture context. *CoRR*, abs/1304.0589, 2013.

[132] A. S. M. Kayes, Jun Han, and Alan W. Colman. An ontological framework for situation-aware access control of software services. *Inf. Syst.*, 53:253–277, 2015.

[133] Seong-Eun Kim, Yong Kim, Jihyun Yoon, and Eung Sun Kim. Indoor positioning system using geomagnetic anomalies for smartphones. In *2012 International conference on indoor positioning and indoor navigation (IPIN)*, pages 1–5. IEEE, 2012.

[134] Henrik Kniberg. *Scrum and XP from the Trenches*. Lulu. com, 2015.

[135] Andrew J. Ko, Thomas D. Latoza, and Margaret M. Burnett. A practical guide to controlled experiments of software engineering tools with human participants. *Empirical Softw. Engg.*, 20(1):110–141, February 2015.

[136] Andreas Konstantinidis, Georgios Chatzimilioudis, Demetrios Zeinalipour-Yazti, Paschalis Mpeis, Nikos Pelekis, and Yannis Theodoridis. Privacy-preserving indoor localization on smartphones. *IEEE Transactions on Knowledge and Data Engineering*, 27(11):3042–3055, 2015.

[137] Agnes Koschmider and Hajo A. Reijers. Improving the process of process modelling by the use of domain process patterns. *Enterprise IS*, 9(1):29–57, 2015.

[138] Krenn S. et al. Deliverable D3.2: Cross Sectoral Cybersecurity Building Blocks. `https://cybersec4europe.eu/wp-content/uploads/2020/06/D3.2-Cross_sectoral_cybersecurity-building-blocks-v2.0.pdf`, 2020.

[139] Krzysztof Kuchcinski and Radoslaw Szymanek. Jacop-java constraint programming solver. In *CP Solvers: Modeling, Applications, Integration, and Standardization, co-located with the 19th International Conference on Principles and Practice of Constraint Programming*, 2013.

[140] Wadha Labda, Nikolay Mehandjiev, and Pedro Sampaio. Modeling of privacy-aware business processes in BPMN to protect personal data. In *Symposium on Applied Computing, SAC 2014, Gyeongju, Republic of Korea - March 24 - 28, 2014*, pages 1399–1405, 2014.

[141] Jung Ho Lee, Beomju Shin, DongHyun Shin, Jaehun Kim, Jinwoo Park, and Taikjin Lee. Precise indoor localization: Rapidly-converging 2d surface correlation-based fingerprinting technology using LTE signal. *IEEE Access*, 8:172829–172838, 2020.

[142] Filip Lemic, Vlado Handziski, Nitesh Mor, Jan Rabaey, John Wawrzynek, and Adam Wolisz. Toward standardized localization service. In *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–8. IEEE, 2016.

[143] Ang Li, Qinghua Li, Vincent C Hu, and Jia Di. Evaluating the capability and performance of access control policy verification tools. In *Proc. of MILCOM*, pages 366–371, 2015.

[144] Bixin Li, Xiaobing Sun, Hareton Leung, and Sai Zhang. A survey of code-based change impact analysis techniques. *Software Testing, Verification and Reliability*, 23(8):613–646, 2013.

[145] Peihao Li, Xu Yang, Yuqing Yin, Shouwan Gao, and Qiang Niu. Smartphone-based indoor localization with integrated fingerprint signal. *IEEE Access*, 8:33178–33187, 2020.

[146] Yongchao Li, You Li, Linzhang Wang, and Guanling Chen. Automatic xacml requests generation for testing access control policies. In *SEKE*, pages 217–222, 2014.

[147] Tomer Libal and Alexander Steen. Towards an executable methodology for the formalization of legal texts. In *International Conference on Logic and Argumentation*, pages 151–165. Springer, 2020.

[148] Bo Liu, Wanlei Zhou, Tianqing Zhu, Yong Xiang, and Kun Wang. Location privacy-preserving mechanisms. In *Location Privacy in Mobile Applications*, pages 17–31. Springer, 2018.

[149] Francesca Lonetti and Eda Marchetti. Chapter three - emerging software testing technologies. *Adv. Comput.*, 108:91–143, 2018.

[150] Francesca Lonetti and Eda Marchetti. On-line tracing of xacml-based policy coverage criteria. *IET Software*, 2018.

[151] Chuanhua Lu, Hideaki Uchiyama, Diego Thomas, Atsushi Shimada, and Rin-ichiro Taniguchi. Indoor positioning system based on chest-mounted imu. *Sensors*, 19(2):420, 2019.

[152] Garm Lucassen, Fabiano Dalpiaz, Jan Martijn E. M. van der Werf, and Sjaak Brinkkemper. Improving agile requirements: the quality user story framework and tool. *Requirements Engineering*, 21(3):383–403, Sep 2016.

[153] Garm Lucassen, Fabiano Dalpiaz, Jan Martijn E. M. van der Werf, and Sjaak Brinkkemper. The use and effectiveness of user stories in practice. In Maya Daneva and Oscar Pastor, editors, *Requirements Engineering: Foundation for Software Quality*, pages 205–222, Cham, 2016. Springer International Publishing.

[154] Yu-Seung Ma, Jeff Offutt, and Yong-Rae Kwon. MuJava: a mutation system for Java. In *Proceedings of the ICSE*, pages 827–830, 2006.

[155] L. Madeyski and N. Radyk. Judy - a mutation testing tool for java. *IET Software*, 4(1):32–42, Feb 2010.

[156] Damiano Di Francesco Maesa, Paolo Mori, and Laura Ricci. A blockchain based approach for the definition of auditable access control systems. *Comput. Secur.*, 84:93–119, 2019.

[157] Frank Manola and Eric Miller. RDF Primer. W3C Recommendation, WWW Consortium, 2004. `http://www.w3.org/TR/rdf-primer/`.

[158] Mantelero A. et al. Deliverable D4.2: Legal Framework . `https://cybersec4europe.eu/wp-content/uploads/2020/09/CS4E-D4.2-Legal-Framework_post-rev_20200914_v1.1.pdf`, 2020.

[159] Markatos E. et al. Deliverable D4.4: Research andDevelopment Roadmap 2. `https://cybersec4europe.eu/wp-content/uploads/2021/02/D4.4-Research-and-Development-Roadmap-2-v3.0-submitted.pdf`, 2020.

[160] Evan Martin and Tao Xie. Automated Test Generation for Access Control Policies. In *Supplemental Proc. of 17th International Symposium on Software Reliability Engineering (ISSRE)*, November 2006.

[161] Evan Martin and Tao Xie. Automated test generation for access control policies via change-impact analysis. In *Proc. of SESS*, pages 5–11, May 2007.

[162] Evan Martin and Tao Xie. A fault model and mutation testing of access control policies. In *Proc. of WWW*, pages 667–676, May 2007.

[163] Evan Martin, Tao Xie, and Ting Yu. Defining and measuring policy coverage in testing access control policies. In *Proc. of ICICS*, pages 139–158, December 2006.

[164] Tania Martin, Georgios Karopoulos, José Luis Hernández Ramos, Georgios Kambourakis, and Igor Nai Fovino. Demystifying COVID-19 digital contact tracing: A survey on frameworks and mobile apps. *CoRR*, abs/2007.11687, 2020.

[165] Fergal McCaffery, Özden Özcan-Top, Ceara Treacy, Pangkaj Paul, John Loane, Jennifer Crilly, and Arthur Mc Mahon. A process framework combining safety and security in practice. In *Systems, Software and Services Process Improvement*, pages 173–180, Cham, 2018. Springer International Publishing.

[166] Elena Meshkova, Janne Riihijärvi, Marina Petrova, and Petri Mähönen. A survey on resource discovery mechanisms, peer-to-peer and service discovery frameworks. *Computer Networks*, 52(11):2097 – 2128, 2008.

[167] Daniel Minoli and Benedict Occhiogrosso. Ultrawideband (uwb) technology for smart cities iot applications. In *2018 IEEE International Smart Cities Conference (ISC2)*, pages 1–8. IEEE, 2018.

[168] Benjamin Molina, Eneko Olivares, Carlos Enrique Palau, and Manuel Esteve. A multimodal fingerprint-based indoor positioning system for airports. *IEEE Access*, 6:10092–10106, 2018.

[169] Boris Motik, Peter F. Patel-Schneider, and Bijan Parsia. OWL 2 Web Ontology Language structural specification and functional-style syntax (second edition). W3C recommendation, World Wide Web Consortium, 2012.

[170] T. Mouelhi, F. Fleurey, and B. Baudry. A generic metamodel for security policies mutation. In *Proc. of ICSTW*, pages 278–286, 2008.

[171] Mirco Nanni, Gennady Andrienko, Albert-László Barabási, Chiara Boldrini, Francesco Bonchi, Ciro Cattuto, Francesca Chiaromonte, Giovanni Comandé, Marco Conti, Mark Coté, et al. Give more data, awareness and control to individual citizens, and they will help covid-19 containment. *Ethics and Information Technology*, pages 1–6, 2021.

[172] Jens Neuhüttler, Rudolf Fischer, Walter Ganz, and Florian Urmetzer. Perceived quality of artificial intelligence in smart service systems: A structured approach. In Martin Shepperd, Fernando Brito e Abreu, Alberto Rodrigues da Silva, and Ricardo Pérez-Castillo, editors, *Quality of Information and Communications Technology*, pages 3–16, Cham, 2020. Springer International Publishing.

[173] Quang Huy Nguyen, Princy Johnson, Trung Thanh Nguyen, and Martin Randles. A novel architecture using ibeacons for localization and tracking of people within healthcare environment. In *2019 Global IoT Summit (GIoTS)*, pages 1–6. IEEE, 2019.

[174] T. H. H. Nguyen, T. P. Hong, and N. Le Thanh. An ontological approach for organizing a knowledge base to share and reuse business workflow templates. In *2017 Seventh International Conference on Information Science and Technology (ICIST)*, pages 271–277, April 2017.

[175] Qun Ni, Elisa Bertino, Jorge Lobo, Carolyn Brodie, Clare-Marie Karat, John Karat, and Alberto Trombetta. Privacy-aware role-based access control. *ACM Trans. Inf. Syst. Secur.*, 13(3):24:1–24:31, 2010.

[176] Changhai Nie and Hareton Leung. A survey of combinatorial testing. *ACM Computing Surveys (CSUR)*, 43(2):1–29, 2011.

[177] R. Nieminen and K. Jrvinen. Practical privacy-preserving indoor localization based on secure two-party computation. *IEEE Transactions on Mobile Computing*, pages 1–1, 2020.

[178] Nicolás Notario, Eleonora Ciceri, Alberto Crespo, Eduardo González Real, Ilio Catallo, and Sauro Vicini. Orchestrating privacy enhancing technologies and services with BPM tools. In *Proceedings of the 12th International Conference on Availability, Reliability and Security (ARES)*. ACM, August–September 2017.

[179] OASIS. eXtensible Access Control Markup Language (XACML) Version 2.0. `http://docs.oasis-open.org/xacml/2.0/access_control-xacml-2.0-core-spec-os.pdf`, February 2005.

[180] OASIS. eXtensible Access Control Markup Language (XACML) Version 3.0. `http://docs.oasis-open.org/xacml/3.0/xacml-3.0-core-spec-os-en.html`, January 2013.

[181] OASIS. XACML v3.0 Privacy Policy Profile Version 1.0. http:docs.oasis-open.orgxacml3.0privacyv1.0xacml-3.0-privacy-v1.0.html, January 2015.

[182] Object Management Group. Business process model and notation, January 2011.

[183] Monica Palmirani, Michele Martoni, Arianna Rossi, Cesare Bartolini, and Livio Robaldo. Legal ontology for modelling GDPR concepts and norms. In *Legal Knowledge and Information Systems - JURIX 2018: The Thirty-first Annual Conference, Groningen, The Netherlands, 12-14 December 2018.*, pages 91–100, 2018.

[184] Monica Palmirani, Michele Martoni, Arianna Rossi, Cesare Bartolini, and Livio Robaldo. Pronto: Privacy ontology for legal reasoning. In *Electronic Government and the Information Systems Perspective - 7th International Conference, EGOVIS 2018, Regensburg, Germany, September 3-5, 2018, Proceedings*, pages 139–152, 2018.

[185] Harshvardhan J. Pandit, Kaniz Fatema, Declan O'Sullivan, and Dave Lewis. Gdprtext - gdpr as a linked data resource. In *The Semantic Web*, pages 481–495, Cham, 2018. Springer International Publishing.

[186] Harshvardhan J Pandit and Dave Lewis. Modelling provenance for gdpr compliance using linked open data vocabularies. In *PrivOn@ ISWC*, 2017.

[187] Mike Papadakis, Marinos Kintis, Jie Zhang, Yue Jia, Yves Le Traon, and Mark Harman. Chapter six - mutation testing advances: An analysis and survey. volume 112 of *Advances in Computers*, pages 275 – 378. Elsevier, 2019.

[188] Jan Pelant, Zdenek Tlamsa, Vlastimil Benes, Ladislav Polak, Ondrej Kaller, Libor Bolecek, Jan Kufa, Jiri Sebesta, and Tomas Kratochvil. Ble device indoor localization based on rss fingerprinting mapped by propagation modes. In *2017 27th International Conference Radioelektronika (RADIOELEKTRONIKA)*, pages 1–5. IEEE, 2017.

[189] Óscar Mortágua Pereira, Vedran Semenski, Diogo Domingues Regateiro, and Rui L. Aguiar. The XACML standard - addressing architectural and security aspects. In *IoTBDS*, pages 189–197. SciTePress, 2017.

[190] Santiago Pina Ros, Mario Lischka, and Félix Gómez Mármol. Graph-based xacml evaluation. In *Proc. of the 17th ACM symposium on Access Control Models and Technologies*, pages 83–92, 2012.

[191] Francesco Potortì, Antonino Crivello, Michele Girolami, Paolo Barsocchi, and Emilia Traficante. Localising crowds through wi-fi probes. *Ad Hoc Networks*, 75:87–97, 2018.

[192] Francesco Potortì, Sangjoon Park, Antonio Ramon Jimenez Ruiz, Paolo Barsocchi, Michele Girolami, Antonino Crivello, So Yeon Lee, Jae Hyun Lim, Joaquín Torres-Sospedra, Fernando Seco, et al. Comparing the performance of indoor localization systems through the evaal framework. *Sensors*, 17(10):2327, 2017.

[193] F. Potortì, A. Crivello, M. Girolami, E. Traficante, and P. Barsocchi. Wi-fi probes as digital crumbs for crowd localisation. In *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–8, 2016.

[194] Alexander Pretschner, Tejeddine Mouelhi, and Yves Le Traon. Model-based tests for access control policies. In *Proc. of First International Conference on Software Testing, Verification (ICST)*, pages 338–347, 2008.

[195] Qusai Ramadan, Mattia Salnitriy, Daniel Strüber, Jan Jürjens, and Paolo Giorgini. From secure business process modeling to design-level security verification. In *Proceedings of the ACM/IEEE 20^{th} International Conference on Model Driven Engineering Languages and Systems (MODELS)*, pages 123–133. IEEE, September 2017.

[196] Silvio Ranise and Hari Siswantoro. Automated legal compliance checking by security policy analysis. In Stefano Tonetta, Erwin Schoitsch, and Friedemann Bitsch, editors, *Computer Safety, Reliability, and Security*. Springer International Publishing, 2017.

[197] Jan Recker. Opportunities and constraints. *Business Process Management Journal*, 16(1):181–201, 2010.

[198] Kantara Initiative Technical Specification Recommendation. Consent receipt specification 1.1.0. kantara initiative consent & information sharing work group. In David Turner Mark Lizar, editor, *Kantara Initiative Technical Specification Recommendation*, 2018.

[199] Valerie Renaudin, Miguel Ortiz, Johan Perul, Joaquin Torres-Sospedra, Antonio Ramón Jiménez, Antoni Pérez-Navarro, Germán Martín Mendoza-Silva, Fernando Seco, Yael Landau, Revital Marbel, et al. Evaluating indoor positioning systems in a shopping mall: The lessons learned from the ipin 2018 competition. *IEEE Access*, 7:148594–148628, 2019.

[200] Matteo Ridolfi, Samuel Van de Velde, Heidi Steendam, and Eli De Poorter. Analysis of the scalability of uwb indoor localization solutions for high user densities. *Sensors*, 18(6):1875, 2018.

[201] Livio Robaldo, Cesare Bartolini, Monica Palmirani, Arianna Rossi, Michele Martoni, and Gabriele Lenzini. Formalizing gdpr provisions in reified i/o logic: the dapreco knowledge base. *Journal of Logic, Language and Information*, 29(4):401–449, 2020.

[202] Per Runeson and Martin Höst. Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, 14(2):131, Dec 2008.

[203] Hanne Rygge and Audun Jøsang. Threat poker: Solving security and privacy threats in agile software development. In *Nordic Conference on Secure IT Systems*, pages 468–483. Springer, 2018.

[204] Nader Samir Labib, Chao Liu, Saharnaz Esmaeilzadeh Dilmaghani, Matthias Brust, Grégoire Danoy, and Pascal Bouvry. White paper: Data protection and privacy in smart ict-scientific research and technical standardization. Technical report, ILNAS, 2018.

[205] R. S. Sandhu and P. Samarati. Access control: principle and practice. *IEEE Communications Magazine*, 32(9):40–48, Sep. 1994.

[206] Ravi S. Sandhu. Lattice-based access control models. *Computer*, 26(11):9–19, 1993.

[207] Ravi S Sandhu. Role-based access control. In *Advances in computers*, volume 46, pages 237–286. Elsevier, 1998.

[208] David Schuler and Andreas Zeller. Javalanche: Efficient mutation testing for java. In *Proceedings of ES-EC/FSE*, pages 297–298, New York, NY, USA, 2009. ACM.

[209] Yu seung Ma, Jeff Offutt, and Yong Rae Kwon. Mujava : An automated class mutation system. *Journal of Software Testing, Verification and Reliability*, 15:97–133, 2005.

[210] Sforzin A. et al. Deliverable D3.11: Definition of Privacy by Design and Privacy Preserving Enablers. `https://cybersec4europe.eu/publications/deliverables/`, 2020.

[211] Muhammad Shahid, Suhaimi Ibrahim, and Mohd Naz'ri Mahrin. A study on test coverage in software testing. *Advanced Informatics School (AIS), Universiti Teknologi Malaysia, International Campus, Jalan Semarak, Kuala Lumpur, Malaysia*, 2011.

[212] Wenhua Shao, Haiyong Luo, Fang Zhao, and Antonino Crivello. Toward improving indoor magnetic field–based positioning system using pedestrian motion models. *International Journal of Distributed Sensor Networks*, 14(9):1550147718803072, 2018.

[213] Wenhua Shao, Haiyong Luo, Fang Zhao, Yan Ma, Zhongliang Zhao, and Antonino Crivello. Indoor positioning based on fingerprint-image and deep learning. *IEEE Access*, 6:74699–74712, 2018.

[214] Wenhua Shao, Haiyong Luo, Fang Zhao, Cong Wang, Antonino Crivello, and Muhammad Zahid Tunio. Depedo: Anti periodic negative-step movement pedometer with deep convolutional neural networks. In *2018 IEEE international conference on communications (ICC)*, pages 1–6. IEEE, 2018.

[215] Wenhua Shao, Haiyong Luo, Fang Zhao, Cong Wang, Antonino Crivello, and Muhammad Zahid Tunio. Mass-centered weight update scheme for particle filter based indoor pedestrian positioning. In *2018 IEEE wireless communications and networking conference (WCNC)*, pages 1–6. IEEE, 2018.

[216] Tuuli Siiskonen, Camillo Särs, A Vähä-Sipilä, and A Pietikääinen. Generic security user stories. *Handbook of the Secure Agile Software Development Life Cycle, Pietikinen Pekka and Rning Juha (Eds.). University of Oulu, Oulu*, 2014.

[217] Dag IK Sjøberg, Jo Erskine Hannay, Ove Hansen, Vigdis By Kampenes, Amela Karahasanovic, N-K Liborg, and Anette C Rekdal. A survey of controlled experiments in software engineering. *IEEE transactions on software engineering*, 31(9):733–753, 2005.

[218] Ścibor Sobieski and Bartosz Zieliński. User stories and parameterized role based access control. In *Model and Data Engineering*, pages 311–319. Springer, 2015.

[219] Ana Sokolovska and Ljupco Kocarev. Integrating technical and legal concepts of privacy. *IEEE Access*, 6:26543–26557, May 2018.

[220] Uros Stevanovic, David Groep, Ian Neilson, Stefan Paetow, and Wolfgang Pempe. Data protection impact assessment-an initial guide for communities, April 2018.

[221] Graeme Stevenson, Juan Ye, Simon Dobson, and Paddy Nixon. Loc8: a location model and extensible framework for programming with location. *IEEE Pervasive Computing*, 9(1):28–37, 2009.

[222] X. Su, J. Hyysalo, M. Rautiainen, J. Riekki, J. Sauvola, A. Maarala, H. Hirvonsalo, P. Li, and H. Honko. Privacy as a service: Protecting the individual in healthcare data processing. *Computer*, 49(11):49–59, nov 2016.

[223] Sun Microsystems. Sun's XACML Implementation. `http://sunxacml.sourceforge.net/`, 2006.

[224] TAS3 Project. Trusted Architecture for Securely Shared Services. `http://www.tas3.eu/`.

[225] Fatih Turkmen, Jerry den Hartog, Silvio Ranise, and Nicola Zannone. Analysis of xacml policies with smt. In *Proc. of International Conference on Principles of Security and Trust*, pages 115–134. Springer, 2015.

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

[227] Dominik Van Opdenbosch, Georg Schroth, Robert Huitl, Sebastian Hilsenbeck, Adrian Garcea, and Eckehard Steinbach. Camera-based indoor positioning using scalable streaming of compressed binary image signatures. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 2804–2808. IEEE, 2014.

[228] Wipassorn Vinicchayakul, Sathaporn Promwong, and Pichaya Supanakoon. Study of uwb indoor localization using fingerprinting technique with different number of antennas. In *2016 International Computer Science and Engineering Conference (ICSEC)*, pages 1–4. IEEE, 2016.

[229] Sandra Wachter. Normative challenges of identification in the internet of things: Privacy, profiling, discrimination, and the gdpr. *Computer law & security review*, 34(3):436–449, 2018.

[230] W. Wang, A. Gupta, and N. Niu. Mining security requirements from common vulnerabilities and exposures for agile projects. In *2018 IEEE 1st International Workshop on Quality Requirements in Agile Projects (QuaRAP)*, pages 6–9, Aug 2018.

[231] Xinyu Wang, Liping Zhao, Ye Wang, and Jie Sun. The role of requirements engineering practices in agile development: an empirical study. In *Requirements Engineering*, pages 195–209. Springer, 2014.

[232] Goitom Kahsay Weldehawaryat and Basel Katt. Towards a quantitative approach for security assurance metrics. In *The Twelfth International Conference on Emerging Security Information, Systems and Technologies; SECURWARE 2018 September 16, 2018 to September 20, 2018-Venice, Italy*. International Academy, Research and Industry Association (IARIA), 2018.

[233] Sören Witt, Sven Feja, Andreas Speck, and Christian Prietz. Integrated privacy modeling and validation for business process models. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops, Berlin, Germany, March 30, 2012*, pages 196–205, 2012.

[234] Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, and Björn Regnell. *Experimentation in Software Engineering*. Springer, 2012.

[235] Claes Wohlin, Per Runeson, Martin Höst, Magnus C Ohlsson, Björn Regnell, and Anders Wesslén. *Experimentation in software engineering*. Springer Science & Business Media, 2012.

[236] X. Wu, R. Shen, L. Fu, X. Tian, P. Liu, and X. Wang. ibill: Using ibeacon and inertial sensors for accurate indoor localization in large open areas. *IEEE Access*, 5:14589–14599, 2017.

[237] H. Xia, J. Zuo, S. Liu, and Y. Qiao. Indoor localization on smartphones using built-in sensors and map constraints. *IEEE Transactions on Instrumentation and Measurement*, 68(4):1189–1198, 2019.

[238] Xusheng Xiao, Amit Paradkar, Suresh Thummalapenta, and Tao Xie. Automated extraction of security policies from natural-language software documents. In *Proceedings of the ACM SIGSOFT FSE '12*, FSE '12, pages 12:1–12:11, New York, NY, USA, 2012. ACM.

[239] Dianxiang Xu, Michael Kent, Lijo Thomas, Tejeddine Mouelhi, and Yves Le Traon. Automated model-based testing of role-based access control using predicate/transition nets. *IEEE Transactions on Computers*, 64(9):2490–2505, 2015.

[240] Dianxiang Xu, Roshan Shrestha, and Ning Shen. Automated coverage-based testing of xacml policies. In *Proceedings of the 23nd ACM on Symposium on Access Control Models and Technologies*, pages 3–14. ACM, 2018.

[241] Dianxiang Xu, Zhenyu Wang, Shuai Peng, and Ning Shen. Automated fault localization of xacml policies. In *Proceedings of the 21st ACM on Symposium on Access Control Models and Technologies*, SACMAT '16, pages 137–147, New York, NY, USA, 2016. ACM.

[242] Dianxiang Xu and Yunpeng Zhang. Specification and analysis of attribute-based access control policies: An overview. In *Proc. of Eighth International Conference on Software Security and Reliability-Companion (SERE-C)*, pages 41–49. IEEE, 2014.

[243] Y. Xu, Y. S. Shmaliy, Y. Li, and X. Chen. Uwb-based indoor human localization with time-delayed data using efir filtering. *IEEE Access*, 5:16676–16683, 2017.

[244] Fara Yahya, Robert J Walters, and Gary B Wills. Using goal-question-metric (gqm) approach to assess security in cloud storage. In *International Workshop on Enterprise Security*, pages 223–240. Springer, 2015.

[245] Zheng Yang and Kimmo Järvinen. The death and rebirth of privacy-preserving wifi fingerprint localization with paillier encryption. *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, pages 1223–1231, 2018.

[246] Stephen S. Yau and Junwei Liu. A situation-aware access control based privacy-preserving service matchmaking approach for service-oriented architecture. In *2007 IEEE International Conference on Web Services (ICWS 2007), July 9-13, 2007, Salt Lake City, Utah, USA*, pages 1056–1063. IEEE Computer Society, 2007.

[247] S. Yoo and M. Harman. Regression testing minimization, selection and prioritization: A survey. *Softw. Test. Verif. Reliab.*, 22(2):67–120, March 2012.

[248] W. You, F. Li, L. Liao, and M. Huang. Data fusion of uwb and imu based on unscented kalman filter for indoor localization of quadrotor uav. *IEEE Access*, 8:64971–64981, 2020.

[249] Faheem Zafari, Athanasios Gkelias, and Kin K Leung. A survey of indoor localization systems and technologies. *IEEE Communications Surveys & Tutorials*, 21(3):2568–2599, 2019.

[250] Demetrios Zeinalipour-Yazti and Christos Laoudias. The anatomy of the anyplace indoor navigation service. *SIGSPATIAL Special*, 9(2):3–10, October 2017.

[251] Nan Zhang, Mark Ryan, and Dimitar Guelev. Evaluating access control policies through model checking. In *Information Security*, volume 3650 of *Lecture Notes in Computer Science*, pages 446–460. 2005.

[252] Tao Zhang, Sherman SM Chow, Zhe Zhou, and Ming Li. Privacy-preserving wi-fi fingerprinting indoor localization. In *International Workshop on Security*, pages 215–233. Springer, 2016.

[253] W Zhang and Mohsen Kavehrad. Comparison of vlc-based indoor positioning techniques. In *Broadband access communication technologies VII*, volume 8645, page 86450M. International Society for Optics and Photonics, 2013.

[254] Y. Zhang and B. Zhang. A new testing method for xacml 3.0 policy based on abac and data flow. In *2017 13th IEEE International Conference on Control Automation (ICCA)*, pages 160–164, July 2017.