UNIVERSITEIT VAN AMSTERDAM

Master Thesis

An Analysis of Privacy Policy Languages and their Compliance with the General Data Protection Regulation (GDPR)

Author: Ilias Daia (2622922)

1st supervisor:	MSc. Milen G. Kebede
2nd supervisor:	Dr. Adam Belloum
2nd reader:	Dr. Tom van Engers

A thesis submitted in fulfilment of the requirements for the joint UvA-VU Master of Science degree in Computer Science

April 11, 2023

"I want to start this paper in the Name of Allah (God), The most Gracious, the most Mercifull.

Our beloved Prophet Muhammad (peace be upon him) said in a report, Indeed Allah loves one who does any kind of work, he does it with excellence (Itqaan)."

Prophet Muhammed (Salla Allahou 'Alayhi wa Sallam)

Abstract

by Ilias DAIA

In the early years of the internet, it was challenging to hold businesses accountable for improper data collection and protection practices. With the advent of data law and the regulation of online privacy, this has changed. The General Data Protection Regulations (GDPR) were implemented by the European Union a few years ago. Data security and user privacy are both protected in this way. According to GDPR, user agreements must be transparently stated and users cannot be duped into giving up their privacy rights. The privacy policy language then comes into play in order to convert such a regulation into a language that can be understood by computers. This thesis examines the state of the art for GDPR-compliant privacy policy languages, looking for shortcomings and problems as well as solutions. In order to identify the qualities and attributes that might make a language possibly helpful for an application with the GDPR data law profile.

Keywords: GPDR, Privacy, Data Law, policy languages, ODRL, XCAML

Acknowledgements

I'd like to thank Dr. Adam Belloum and Milen G. Kebede for allowing me to work on this project and ensuring its completion. Both Adam and Milen were extremely kind and supportive throughout this project, which greatly aided me as a student and made me feel at ease, allowing the collaboration to run smoothly.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Context

Big data is becoming more and more popular, which has increased concerns in the IT community about the issue of data privacy [1]. The emergence of big data has changed our attention away from the volume of data. The privacy and security of data are a pressing issue that require immediate attention. Data leakage is never a minor issue, and recently, the public has begun to pay more attention to data security. Due to increasingly stringent data protection laws, many businesses and organizations, including hospitals, are unable to legally disclose or expose the private information of their clients. The defence of data security and privacy is being strengthened not just by individuals, but also by groups and society. The goal of the General Data Protection Regulations (GDPR), which the European Union put into effect on May 25, 2018 is to safeguard users' personal information and data security. GDPR must clearly state the user agreements and cannot trick or persuade users to waive their privacy rights [2]. Websites of companies and other services that are open to the public are one way that personal data is processed. It is crucial to make sure that visitors to these websites are informed of how their personal information is handled, how its correctness is maintained, how its data integrity and confidentiality are safeguarded. These elements are critical because user data has become a valuable resource. To optimize their commercial services for the online consumer market, businesses analyze user data internally or share/sell it with advertisers and researchers. Consequently, it is expected

that all websites have a relevant privacy statement to ensure the legal, ethical, and transparent processing of user data. Machine-readable policy languages have been available for many years and enable the communication of an individual or organization's preference to grant access to a particular resource, thereby controlling the functioning of actual systems over real data. However, policy languages are insufficient to address the diverse responsibilities related to privacy and data protection that computers can aid in.
To specify concepts and rules in the domain, vocabularies and computer ontologies have emerged in recent years. These can be used to either simply record information as RDF or to operate ontology-based information systems. Although numerous privacy languages exist, they often adopt forms that are useful for their implementations, without necessarily utilizing the GDPR as their specific frame of reference. [3].

1.1.1 Natural Language Ambiguity

One of the biggest challenge of Natural language processing(NLP) is ambiguity. NLP ambiguity is challenging since when trying to understand the meaning of a word various factors should be considered such as context and how the word is used generally in the society. Natural language processing also brings about ambiguity since words change meaning over time and can also mean one thing in a specific domain and a different meaning in another domain. It is considered ambiguous when anything can be understood in two or more different ways or senses. Lexical ambiguity, which occurs in a single word, is referred to as such; structural ambiguity, which occurs in a sentence or clause. There are many instances of lexical ambiguity. A note can refer to either a musical tone or a brief written document. A statement you know to be untrue is a "lie," and the present tense of the verb "to lay" is to be or place oneself flat. We can also consider the word "ambiguity" by itself.

It can indicate ambiguity, the desire to convey multiple meanings, the likelihood that one or both interpretations were intended, and the actuality that a phrase has several interpretations. Ambiguity tends to grow with usage frequency. Ambiguity can occasionally mean something clever or dishonest in everyday discourse. According to Harry Rusche, ambiguity should include verbal nuances that allow different interpretations of the same language component. The compound term polysemy refers to a fundamental linguistic characteristic [4]. Another name for polysemy is multiplication or radiation. When a word develops a more extensive range of meanings, this occurs. Paper, for instance, is derived from Greek papyrus. It originally referred to writing materials created from Nile papyrus reeds, then to various writing materials. It refers to official documents, research findings, family letters, newspapers, or archives. When a single verb has many senses connected in some predictable fashion, this is referred to as complementary polysemy. According to the study, ambiguity and language evolution are related. Over countless years, language has developed into a far more complicated phenomenon from a collection of symbols subject to rules. It has become nearly hard to determine whether this is in our favour due to the sheer number of unclear situations that could arise. To acquire a more comprehensive understanding, we should consider the various ambiguities that frequently occur in natural languages [5].

1.1.2 Types of Ambiguity

Many different ambiguities exist:

Lexical Ambiguity: It is the word's inherent ambiguity. Regarding the syntactic class it belongs to, a term may be confusing, like study or book. Lexical category disambiguation, also known as parts-of-speech tagging, can be used to resolve lexical ambiguity. Numerous words could fall under more than one lexical group. Each word in a phrase is given a part-of-speech or lexical category, such as a noun, verb, pronoun, preposition, adverb, adjective, etc., through the process of part-of-speech tagging.

Lexical Semantic Ambiguity:lexical ambiguity occurs when a single term is connected to several senses. For instance: cricket, fast, bat, bank, etc. Using word sense disambiguation (WSD) techniques, lexical semantic ambiguity is resolved. WSD tries to automatically assign the word's meaning in the context in a computational manner [6] [7].

Syntactic Ambiguity: Syntactic ambiguities made up the structural ambiguities. There are two types of structural ambiguity: attachment ambiguity and scope ambiguity.

Scope Ambiguity: Operators and quantifiers are involved in scope ambiguity. Take this as an example: Older adults were transported to secure areas. The extent of the adjective, or the volume of material it qualifies, is unclear. Specifically, whether the building is for "old men and women" or "(old men and women)"? Quantifiers' scope is frequently unclear, which leads to ambiguity [8]. All men adore women. The interpretations include the idea that there is a lady for every guy, and that every man has a favorite woman.

Attachment Ambiguity: The sentence has attachment ambiguity if a constituent can fit in more than one position in a tree structure. Attachment ambiguity results from uncertainty about which part of a sentence to attach a phrase

or clause. Take this illustration: The man used the telescope to observe the girl. It's unclear if the man saw the girl holding the telescope in person, or if he observed her using his telescope [9]. Whether the preposition "with" is associated with the girl or the male affects the connotation. Take this as an example: Purchase books for kids. The preposition "for children" can be used adjectivally with the object noun books or adverbially with the verb "purchase".

Semantic Ambiguity: This happens when the words' meanings are open to interpretation. There are two ways to read the sentence, even after the syntax and word meanings have been sorted out. Take this as an example: Sita and Priya both adore their mothers. It might be interpreted that Priya adores Sita's mother or that Priya adores her mother. Semantic ambiguities are caused by the fact that, in general, a computer cannot tell what is logical from what is not logical. Consider the following scenario: A moving car struck a post. These interpretations include The vehicle struck the pole while it was moving, and the pole was also moving when the collision occurred. As a result of our ability to discriminate between what is logical (or conceivable) and what is not, the first interpretation is preferred to the second. It's not that simple to provide a computer with a world model. Take this as an example: We observed his duck. The word "duck" can refer to either the person's bird or a movement. When a speaker uses an ambiguous term or phrase, semantic ambiguity occurs [10].

Discourse Ambiguity:A shared world or body of information is required for discourse-level processing, and this context is used for interpretation. Discourse level refers to anaphoric ambiguity [11].

Anaphoric Ambiguity: Entities that have already been brought into the dialogue are anaphors. Take the horse running up the hill as an illustration. It was steep. It quickly grew weary. Both instances of the anaphoric reference "it" raise questions. Since steep refers to a surface, "it" can be a hill. A horse might be "it" because "tired" applies to all animate objects [12].

Pragmatic Ambiguity: One of the most challenging challenges in NLP is dealing with pragmatic ambiguity, which is when the context of a word allows for various interpretations. Processing user intention, mood, belief world, and other very complicated duties are part of the challenge [13].

Computational Linguistics and Ambiguity: Computational linguistics has two main goals: to enable computers to analyze and process natural language and to gain a better understanding of how humans process language by analogy with computers. One of the most significant issues in understanding natural language is the ambiguity problem. Humans can often resolve ambiguities using context and general knowledge, but computers struggle to effectively utilize context as they lack this information. This problem arises when computers search the internet for information regarding different meanings of search phrases and when translating text into another language using machine translation. Due to the limited capacity of computers to handle polysemy, attempts to comprehend human language solely using computers have been unsuccessful.

Efforts to address the ambiguity issue have focused on statistical and knowledge-based systems. Knowledge-based approaches require system engineers to encode a vast amount of knowledge about the outside world and create procedures for utilizing it to determine the meaning of text. In contrast, the statistical approach requires a significant amount of annotated data, which

software engineers use to create methods that compute the most likely solutions to the ambiguity based on words or grammatical forms and other straightforward conditions. However, there is currently no working computer system capable of determining the intended meaning of text used in dialogue. [14].

Nevertheless, the importance of resolving the polysemy issue will ensure that all efforts are made. The true essence of computing science, machine learning, will be within reach once we accomplish this aim, in my opinion. However, there is still a great deal of language context, in particular, that computers need to learn.

Ambiguity in Natural Language and Machine Translation: At the core of machine learning is the concept of transforming spoken language into a coherent sequence of commands with the highest level of precision and uniformity. However, due to the ambiguity present in human language, it is nearly impossible for machines to fully comprehend and process it. One of the main challenges in machine translation is to mechanically convert natural language from one language to another while maintaining the meaning of the input text and creating fluent content in the output language. This has been a topic of study in artificial intelligence for many years, and recently, large-scale empirical methodologies have been adopted, leading to significant improvements in translation quality.[15].

1.2 Motivation

Many different policy languages have been proposed and implemented by various types of institutions. Access control languages from the security domain, in addition to privacy-specific policy languages, can be used to express

the privacy controls that both organizations and users want to express. In this research, We identify policy languages that may be suitable for privacy policies by comparing the usability of policy languages in various settings. Our focus is on a few key elements of privacy policies that must be present in order for the policy to be compliant with data protection laws (e.g. GDPR).

1.3 Problem statement

The ubiquity of internet usage has led to a surge in online businesses, with a majority of companies transitioning their operations to the virtual world. With consumers sharing a wealth of personal information on these online platforms, they must demand high standards of online privacy. Businesses recognize the need to safeguard users' privacy and to adhere to privacy regulations. Privacy is a paramount concern for organizations, as a failure to provide adequate assurances of privacy may result in a loss of consumer trust, leading to negative repercussions for the company. Given the importance of protecting user privacy in the digital age, this study aims to investigate the various types of privacy policy languages that organizations can employ to ensure consumers' privacy.

Organizations and consumers alike need to be informed on the existing privacy policy languages. Some languages are designed to aid business organizations easily express their privacy policies in strategies that are compliant to policy enforcement. There are also privacy policy languages that are designed to help consumers understand and define their privacy requirements. Since each language has its own syntax, it is necessary for organization to understand this syntax and the available mechanism for implementation. Privacy policies must be effectively enforced through auditable mechanism to achieve the level of compliance the privacy legislation's and laws require. Privacy policies play a crucial role in achieving such goal.For that reason, in

this work, we aim to investigate selected state-of-the-art privacy policies to quantify to what extent these policies meet legal requirements in achieving compliance.

1.4 Research questions

To address the state-of-the-art of privacy policy languages, the following research questions were set up:

- **RQ1: What are the current gaps and challenges of state-of-the-art privacy policy languages in terms of specifying policies from legal provisions?**

  - Which common challenges do current privacy policy languages face?

  - Which current gaps and challenges exist in the privacy properties of policy languages?

- **RQ2: What features should be integrated into current privacy policy language to attain correct enforcement of privacy policies?**

  - What features are essential for all privacy policy languages?

  - What features are common in most privacy policy languages?

- **RQ3: Which privacy policy languages are most suiting for specifying the GDPR privacy policies?**

  - Current state-of-the-art privacy policy

  - Comparison between existing privacy policy languages and GDPR

  - Data privacy policy translation to machine language

1.4.1 Research Method

To gather a comprehensive collection of published privacy policy languages, a diverse range of search platforms were employed. Along with the search

tool provided by the university library, the study utilized several other platforms such as IEEE Explore, ScienceDirect, Scopus, Springer Link, Semantic Scholar, Google Scholar, and ACM.

1.4.2 Inclusion and exclusion criteria

To select articles that fit the research question, the following inclusion and exclusion criteria were defined for the literature part of the research:
The inclusion [I] and exclusion [E] criteria are listed as follows:

Inclusion and Exclusion criteria		
N.	**[I]**	**[E]**
1	Studies that concern privacy policy languages	Studies that are not complete or require a paid subscription.
2	Studies that are specific concerning GDPR	Studies of data law that do no mention the use of privacy policy
3	Researches about data privacy law	Studies that don't mention the use of data privacy law
4	Studies in the business application domain of privacy	Studies that focus on other aspects of policy languages and don't mention privacy
5	The article focuses on privacy-by-design and/or privacy-by-default for GDPR	

TABLE 1.1: Table of criteria

1.5 Structure of Thesis

The thesis is structured into seven distinct sections. In the second section, the background information is presented, and the available literature is searched for and compiled. The related work is then gathered and summarized. In the third section, the methodology used in this study is described in detail, including the pipeline used and the rationale behind the choices made. Chapter five presents the research carried out, with a detailed analysis of the

outcomes. The sixth section contains the conclusion and discussion, where the research questions are addressed again, and a reflection on limitations is provided. Finally, in the seventh section, the primary difficulties with each component of the data law and machine language are outlined in detail.

2 Background

This chapter discusses the current state-of-the-art tools, relevant technology, and the related academic papers. First, the relevant literature and work on privacy policy languages are investigated. Following that, a taxonomy of potential languages is presented.

2.1 Data privacy law

The most significant data privacy and protection regulation in many years is the General Data Protection Regulation (GDPR). Although the GDPR is a law of the European Union, it is applicable to any organization, regardless of where it is located, that collects or processes the data of EU citizens. The GDPR forced businesses throughout the world to make crucial decisions and changes in how they collect and process Personally Identifiable Information (PII) from their employees and customers due to the global nature of commerce and people's movements [16].

Privacy policies are now the de-facto means of explaining how a business or organization collects, shares, and uses personally identifiable information (PII), particularly with regard to its website. Posting privacy policies is required by numerous governments across the world (including the FTC in the US). Additionally, a lot of people work to protect the PII of consumers by implementing laws and rules about these practices [17].

A policy language can be used to ensure the confidentiality of personally identifiable information (PII), as well as context information and metadata

that could potentially reveal PII. Rules can be established and enforced to safeguard the privacy of specific objects. Depending on the scope of the policy language, it may consider the privacy of system users or data owners whose information is stored in the system. [18].

2.1.1 GDPR

It was challenging to hold companies responsible for improper data collecting and protection methods in the early days of the internet. This was partially due to the lack of legal development necessary to hold Data Collectors accountable for obligations that did not yet exist. As courts began to recognize the importance of data and its connection to privacy, legal tendencies began to change gradually. However, by hiding behind the complexities brought on by technology, Data Controllers and Processors were able to avoid accountability, or at the very least, to lessen it [19].

In 2016, the GDPR was approved to replace the 1995 Data Protection Directive. This regulation was the result of a complex negotiation process that spanned four years, during which multiple revisions were made to the legal text. The inconsistencies in data privacy laws across EU Member States were perceived as hindrances to the development of the EU's economy and sources of competitive distortion. Unlike the Data Protection Directive, the GDPR applies immediately to its recipients, without requiring any further implementation by EU Member States. The GDPR aims to enhance legal clarity by harmonizing data protection regulations, thereby eliminating any possible barriers to the free flow of personal data information. [16].

The introduction of GDPR has placed the responsibility of translating GDPR duties into software needs on software engineers. However, this task can be challenging, especially for developers who lack a fundamental understanding of the legal and security concepts stated in the law. To comply with

GDPR's rights and obligations, it's essential to determine a set of information flows that are connected to the information that needs to be shared between stakeholders. These stakeholders may be categorized as data subjects (DS), data controllers (DC), data processors (DP), recipients (Rp), supervisory authorities (SA), or data protection officers (DPO).[3].

Criteria Used By GDRP to Protect Data

Since the GDPR views the protection of personal data as a fundamental right of natural beings, it requires that personal data be:
(1) handled in a lawful, equitable, and open manner. (2) gathered for clearly defined and constrained goals. (3) sufficient, pertinent, and limited to what is required. (4) Reliable and current [20].
Therefore, when handling personal data, handlers are required to ensure they follow all the listed principles. Based on the first principle, GDRP states that any processing of personal data should be lawful and fair. Additionally, personal data should only be collected for clearly defined goals. The handlers should be able to explain why they are collecting personal data. The data should also be reliable and current, this means that the data collected must be accurate and processed in a manner that promotes integrity and security.

2.1.2 Requirements by law to Rights of the data subject

Right to be informed

An information flow is the transfer of information from one stakeholder to another necessary for the activation and grant of a right or obligation. For instance, if a data subject exercises his or her right to be forgotten, the controller is required to provide details about the reasons behind the request and

to share these details with any other controllers handling the same personal data [21].

Right of access

According to GDPR Article 15, the right to access is divided (legally) into two stages. According to GDPR Art. 15 Sec. 1, the data subject is entitled to a preliminary confirmation from the controller regarding the processing of their personal data. In the event that such processing takes place, the subject of the data should subsequently have access to both the processed personal data and the following details.

The first "substantive" right of the data subject is provided by GDPR Article 15. By granting them the right to request information and access from the Controller, the right enables the data subject to confirm the legitimacy of the processing. Access by data subjects is not always a simple entitlement to grant. In some circumstances, it could be necessary to limit the range of access due to competing interests. [2]

Data subjects have the right to access their personal data being processed, which includes the right to receive confirmation that their data is being processed. They also have the right to obtain a copy of the data in a commonly used electronic format. In addition, data subjects have the right to receive information about the reasons for processing, the categories of personal data concerned, the source of the data if it was not obtained directly from the subject, the recipients of the data, the storage period, and information about their rights, including the right to file a complaint with a Data Protection Authority (DPA). Finally, data subjects have the right to be informed about the existence of the DPA. [21]

Right to rectification

The right to request that the data controller update incomplete and inaccurate personal information belongs to the data subject.

Right to erasure or Right to be forgotten

Personal data should be erased when it is no longer necessary for the purposes for which it was collected. Additionally, if the data subject withdraws their consent, and there is no other legal basis for processing, the personal data must also be erased. Similarly, if the data subject objects to the processing of their data, the data must be erased. Personal data must also be erased if it is processed unlawfully or when there is a legal requirement to do so. Finally, personal data collected for the purpose of providing goods or services must be erased once that purpose has been fulfilled.

Right to restriction of processing

The data subject has the right to request that processing of their personal data be stopped in certain situations. These situations include when the accuracy of the data is disputed, when the processing is unlawful, but the data subject does not want the data to be erased, when the controller's stated purposes for processing are no longer valid, but the data subject needs the data for any pending legal claims, or when the data subject objects to the processing of their data.

Right to be notified

The data controller must inform both the data subject and any recipients to whom the data was disclosed about the processing of personal data. Additionally, the data controller must inform the data subject of the identity of these recipients.

Right to data portability

The data subject has the right to request that their personal data be transferred directly from one controller to another, without hindrance. They also have the right to receive their personal data in a commonly used and machine-readable format.

Right to object

A key data subject right under the GDPR is the right to object. The right to object to any sort of processing of personal data, including profiling, is granted to data subjects. Profiling is the automated processing of a person's personal information with the purpose of assessing specific facets of their personality. The right to object allows data subjects to object to processing that is done for direct marketing purposes, a controller's or a third party's legitimate interest, or both. The controller is required to stop processing personal data after a data subject objects, unless they can show compelling legitimate grounds for the processing that prevail over the data subject's interests, rights, and freedoms, or the processing is required for the establishment, exercise, or defense of legal claims.

Right to not be subjected to automated decision-making

the "right not to be subjected to automated decision-making, including profiling" [16]

2.2 Privacy policy specification languages

When drafting, evaluating, testing, approving, issuing, merging, analysing, amending, withdrawing, retrieving, and enforcing policies, privacy policy languages might be helpful. Languages for privacy policy were developed

to express the privacy restrictions that users and corporations wished to impose. Most privacy policy languages were developed with particular characteristics and attributes in mind. Most of these language design projects have been progressing for the past ten years [22].

There are several languages that can be used to convey privacy policies in a more accurate and computer-compatible manner. Some of these languages are designed to allow users to define their privacy preferences, while others are meant to help enterprises communicate their privacy policies in ways that are more conducive to policy enforcement.

Each language has its own syntax and implementation methods. However, there is no common metric for evaluating and contrasting these languages. Nonetheless, privacy policy languages should be straightforward and concise. As a result, they were created as lightweight XML markup languages. It's important to note that these privacy policy languages are not expected to carry out intricate flow controls or high-level mathematical computations. Instead, they are designed to facilitate clear communication of privacy policies and to enable users to make informed decisions [23].

In the context of this work, a policy is a set of guidelines that specify how to preserve a particular situation by deciding what to do. A set of syntax and semantics used to represent policies is known as a policy language. Rules that, when followed, protect the confidentiality, availability, and integrity of a specific system are created using a security policy language. The ability to create accountability rules is included in several security policy languages. When data is exchanged with third parties or stored on third-party systems, such as in cloud computing or for marketing, accountability is specifically taken into account.

2.2.1 Basic structure of a privacy policy

GDPR Structure

Before the introduction of GDRP, privacy policies were too complicated and could only be understood by well-educated people on privacy policies, such as attorneys. However, since the introduction of GDRP, it has become easy for everyone to understand privacy policies. One of the reasons the EU implemented GDPR was to better inform consumers and give them choice over how businesses acquire, use, share, safeguard, and treat their personal data. Making privacy rules clear and thorough became one of the new law's primary criteria. Noncompliance with this clause may result in substantial fines or perhaps prosecution. If you haven't already, it's time to go through the details of your privacy policy. In this piece, we'll go over how to create a GDPR-compliant privacy policy. The GDPR was developed based on seven principles: 1) legality, fairness, and openness; 2) purpose restriction; 3) data minimization; 4) accuracy; 5) storage restriction; 6) integrity and confidentiality (security); and 7) accountability. Accountability is a novel concept in data protection rules. In the United Kingdom, all other principles are equivalent to those in the 1998 Data Protection Act [24].

Lawfulness, Fairness and Transparency

According to the website of the Information Commissioner's Office, "data must be processed lawfully, fairly, and transparently." The overall concept underlying these ideas is simple. The intended use of data must be clearly and efficiently communicated so that the data subject understands exactly how their information is gathered and handled. Transparency in data sharing is created as a result, so no one involved is upset or unaware of how their data was processed.

Purpose Limitation

The International Organization for Standardization (ISO) has established that data must be collected for specific, explicit, and lawful purposes, and should not be processed in a manner that is inconsistent with those purposes. However, further processing of the data may be permissible if it is compatible with the original purposes and is carried out for public interest archiving, scientific or historical research, or statistical objectives.

This principle, known as the Purpose Limitation Principle, means that data cannot be held and repurposed for other purposes that were not initially disclosed to the data subject. This is related to the first principle of proper data collection, which requires transparency in reporting data usage. The Purpose Limitation Principle serves to prevent companies from deriving benefits from data through its sale or use for unknown future purposes.

Data Minimization

Data reduction is an important principle in data protection that emphasizes limiting the use of personal data to only what is necessary for the specific purposes for which it was collected. This principle is often reflected in the minimum necessary requirement in US data security laws, which mandates that only the minimum amount of data necessary to achieve a particular goal should be collected and processed.

According to the ICO, personal data should be "adequate, relevant, and limited to what is necessary in relation to the objectives for which they are processed." This means that companies and individuals must carefully consider what personal data they actually need to fulfil their intended purposes, and should avoid collecting or processing any more data than is necessary.

Data retention, processing, and distribution should also be limited and closely scrutinized before any data is collected from the data subject. This can help

to minimize the risk of data breaches, unauthorized access, and other security threats. By following the principle of data reduction, organizations can better protect personal data and maintain the trust of their customers and stakeholders [20].

Accuracy

To expand further, the accuracy principle emphasizes the importance of ensuring that personal data is up-to-date and correct. This means that organizations should take reasonable steps to verify the accuracy of the data they collect and ensure that it remains accurate over time. If an organization discovers that the data they hold is inaccurate, they must take steps to correct or erase it as soon as possible. Inaccurate data can cause significant harm to individuals, including reputational damage, financial loss, and discrimination. It can also impact the effectiveness of decision-making based on that data. Therefore, it is essential for organizations to implement procedures to regularly review and update the data they hold to maintain its accuracy. This includes verifying the data with the data subject, where necessary, and providing a means for the data subject to correct any inaccuracies. Overall, the accuracy principle is vital to ensure that personal data is used fairly and responsibly, and organizations must prioritize this principle in their data processing practices.

Storage Limitation

The principle of data retention is an essential part of GDPR compliance as it ensures that personal data is not kept for longer than necessary. According to the ICO, data should be retained in a way that allows individuals to be identified for no longer than is required for the purposes for which it was collected. However, there are exceptions where personal data can be held for longer periods if it is used for public interest archiving, scientific or historical

research, or statistical purposes. In such cases, it is essential to implement adequate technical and organizational measures to protect individuals' rights and freedoms.

To comply with this principle, organizations must be transparent with data subjects about how long their data will be retained and must ensure that it is destroyed appropriately once it is no longer needed. Non-compliance with data retention requirements can result in severe fines and penalties under GDPR. Therefore, it is critical for organizations to develop and maintain effective data retention policies and procedures to comply with this principle.

Integrity and Confidentiality

The principle of data confidentiality is crucial in maintaining consumer trust and preventing unnecessary data loss. It involves processing data on a need-to-know basis, similar to the principle of least privilege, where only authorized persons who require access to the information will be granted access. According to the ICO, data should be processed using adequate technical and organizational methods to ensure appropriate security, including protection against unauthorized or unlawful processing, accidental loss, destruction, or damage.

Data confidentiality requires that the privacy of customers' data is put at the forefront of business activities. Organizations should utilize data in a discrete and respectful manner that prioritizes the customers' information and privacy. Failure to comply with this principle can result in significant legal and financial consequences. Therefore, organizations must establish and maintain effective data security measures to ensure compliance with this principle.

Accountability

Finally, as stated on the ICO website, "the controller must be responsible for, and capable of demonstrating compliance." Anyone who handles data must be adequately trained and completely informed of what GDPR compliance entails. Finally, it is the controller's responsibility to guarantee that GDPR compliance is maintained and that client privacy is prioritized [20]. We believe that breaking down these principles can remove some of the uncertainty surrounding GDPR compliance and provide you with a clearer grasp of what GDPR compliance entails. Finally, GDPR compliance works to protect customers' privacy and ensure that everyone is informed of how their data is being used.

2.2.2 Different kind of policies

Policies can be of several forms, such as:

Constraint policies, which control the actions taken by the managed parties and define which actions are permitted, prohibited, and required. One prominent illustration of constraint rules is access control policies

Goal-based policies lay out objectives that the parties under their control must meet, such as trying to meet a task's due date or maintaining a minimum level of utilization.

Utility-based policies strive to get the greatest results in accordance with some value functions, such reducing energy use. [25]

2.3 Existing languages

For the languages with the greatest potential from the study, the following criteria were used: ([26]

Situation:

Languages have been developed to address privacy management in many contexts (for example, capturing internal corporate policies rather than user preferences), and the context directly affects the language's properties. We hypothesize that this quality is the most important quality to consider when selecting a language.

Representation:

Rules, rulesets, queries, and data are all represented in various ways by languages. In this study, the majority of the languages were studied using XML as their representation language. XML has been incorporated into some languages in a variety of ways to express their linguistic qualities. Additionally, there are differences in their respective vocabularies, underlying linguistic structures, and methods for representing data. We'll talk about the design aspects used in languages to express data, rules, rulesets, and queries in this attribute.

Evaluation:

For making judgments based on the rules, rulesets, queries, and data provided, different languages employ various strategies. The sequence of the various policy components, such as rules and rulesets, also affects how well a language is evaluated. With reference to the evaluation criteria, we intend to talk about the design elements of languages in this attribute.

Output Schema:

Depending on how the rules, rulesets, data, and queries are evaluated, different types of results (such as allow and deny) are produced by different

languages. We intend to talk about how output schema is implemented in the languages in this attribute.

Implementation:

Languages are used in the real world for a variety of reasons and deployments (such as the type of application the language can be used in, such as web or other apps). We intend to examine the specifics of language implementation in this characteristic.

Formalization:

Formalizing privacy policies is crucial for several reasons. First, it enables us to systematically evaluate policies to ensure that they comply with relevant laws and regulations. By making policies machine-readable, it becomes easier to enforce them and ensure that they are being followed. Second, formalizing privacy policies enables organizations to be more transparent about their data practices, which can build trust with customers and other stakeholders. Finally, formalizing privacy policies can improve consistency and reduce the risk of errors or omissions, which can help organizations avoid costly legal or reputational consequences. Overall, the formalization of privacy policies is an important step towards ensuring privacy and data protection for individuals and promoting responsible data practices by organizations.

2.4 State-of-the-art privacy policy languages

The literature is examined and presented in relation to a number of policy languages that are now considered to be state-of-the-art. We give a high-level overview of the policy languages under discussion in this section.

2.4.1 ODRL

IPR systems in Australia created the open standard known as ODRL to express machine-readable licenses for digital resources. In its background materials, ODRL mentions Erickson's work as well as that of the UK groups indecs7 and Editeur8. Each of these is an example of an industry effort to formally establish the sale of digital goods. ODRL is a collaborative effort that currently includes more than a dozen cooperating organizations, despite its origins in work at IPR systems. A secure encoding of ODRL statements and digital signatures are also included in the ODRL, which also consists of an expression language and a data dictionary with their own unique XML schemas. The abstractions that make up the expression language's vocabulary borrow their definitions from the data dictionary. [27]

DRM's presence and influence will start to be seen in more widespread desktop services and mobile devices as DRM systems transition from proprietary systems to open standards. "Rights Expression Languages" is a new domain created by the standardization of DRM (REL). The REL is an essential component of any DRM system since it offers data about the material, the owners of the rights, the usages, and the payments that must be safely generated, processed, and comprehended by all participants in the value chain. The Open Digital Rights Language is one of the most used RELs (ODRL).

It is designed to be machine-actionable as a component of a system for enforcing digital rights. In accordance with its own open license, ODRL is made available to anyone who wants to use it entirely or in part in their own digital system for free. For those who are already familiar with XML, the documentation on the ODRL website, which includes graphic renderings of the XML schema, is fairly thorough and understandable [27].

The Open Digital Rights Language (ODRL) is a language specifically designed to express digital rights policies in a flexible and interoperable way.

ODRL provides an architecture, vocabulary, and encoding techniques for expressing claims about the use of digital materials and services. The language is designed to be expressive, allowing for complex policy statements to be made. The underlying ideas, entities, and connections that form the basis of ODRL policies are described in the ODRL Information Model, which provides a formal and precise definition of the language's syntax and semantics. Overall, ODRL provides a useful tool for organizations and individuals to formalize and enforce their digital rights policies [28]. The fundamental semantic paradigm for statements expressing permission, restriction, and duty for the use of content is defined by the ODRL Information Model. The fundamental concepts, entities, and relationships that serve as the basis for content consumption claims are covered by the information model. In order to make it simple for users to get this data, these machine-readable policies may be connected directly to the material they are associated with. A number of core elements and their interactions are involved in the Open Digital Rights Language (ODRL) [17], which is built on an expandable model for rights statements. Figure **??** depicts the general ODRL Model, which has the following three main components: (Assets, Rights and Parties).

Assets

The Assets are any tangible or digital items that can be uniquely identified, may include parts, and may come in a variety of formats. Assets can also be intangible representations of works or non-tangible manifestations of such works. For the purpose of enabling secure content dissemination, assets may also be encrypted. The rights include permissions, and permissions may come with restrictions, demands, and conditions. The actual uses or actions that are permitted in relation to the Assets are known as Permissions (for

example, "Playing" a video Asset). Limitations are restrictions on these permissions (for example, "maximum of 5 times" playback of the film). The obligations necessary to execute the permission are known as requirements (for instance, "Pay $5" each time you play the video). Conditions outline exceptions that, if they come to pass, revoke the permissions and may necessitate a new agreement (for instance, all permissions to play the film are withdrawn if your credit card expires). End users and Rights Holders are referred to as the Parties. Parties can be individuals, groups, or other roles with clear definitions. The asset consumers are often end users. The majority of the time, rights holders are organizations or individuals that have contributed in some way to the development, manufacture, or distribution of the asset and who are able to claim ownership of the asset and/or its permissions. Rights Holders may also receive royalties.

ODRL Rights Expression Model

There are two main components to the ODRL specification, as can be seen in 2.1. The "data dictionary" and the expression language model (as previously described). The terms for the real permissions, constraints, requirements, and conditions are gathered in the data dictionary. Examples of data dictionary terms are play, print, display, and execute. The two components are divided primarily to promote more semantic expansion and reuse. [23]

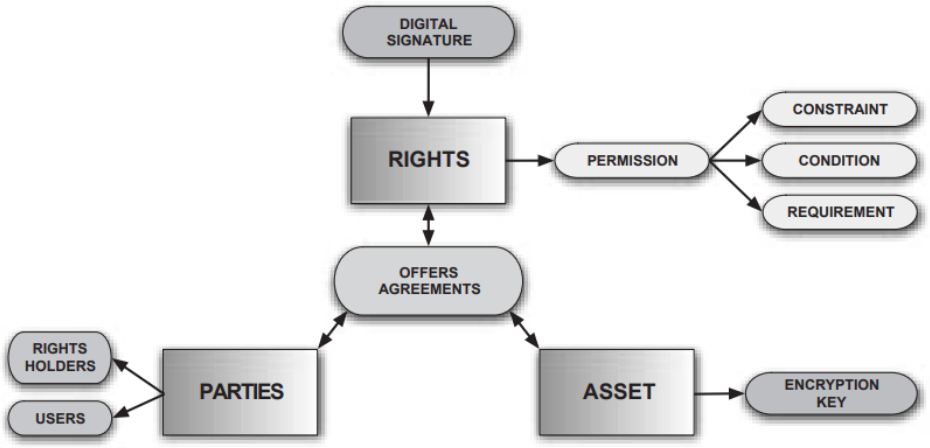

FIGURE 2.1: ODRL Rights Expression Model

2.4.2 P3P

P3P, which stands for Platform for Privacy Preferences, was developed by the World Wide Web Consortium (W3C) as a machine-readable language to help users settle disputes with providers regarding privacy practices and data requests. The development of P3P involved a consensus process that included international experts and representatives from more than a dozen W3C member organizations. When a user interacts with a website or application that uses P3P, the entity in charge of the service provides a machine-readable proposal that includes its identity and privacy policies. The proposal is identified by a Uniform Resource Identifier (URI) or group of URIs that relate to the domain. The proposal outlines the data elements that the service plans to collect and how each will be used, including who it may be shared with and whether it can be used to identify a specific individual. P3P uses a harmonized vocabulary, which is a set of information practice disclosures that describe the functionality of a service, not whether it complies with the law. The privacy proposal is presented in both English and P3P syntax, allowing for easy interpretation by both humans and machines [29].

2.4.3 LegalRuleML

LegalRuleML is a rule interchange language specifically designed for the legal domain. It was developed by the OASIS LegalRuleML Technical Committee and was recently adopted as an OASIS Standard in August 2021. LegalRuleML is based on the RuleML syntax and principles, which is an XML language that provides formal features for representing and reasoning about legal norms, rules, and regulations. One of the main features of LegalRuleML is the use of numerous semantic annotations that allow for the expression of distinct legal interpretations. LegalRuleML also includes the modeling of deontic operators, rule temporal management, rule authorial tracking, and a mapping to RDF triples. These features enable the metadata, context, and statements to be the basic aspects of a LegalRuleML document. The metadata section of a LegalRuleML document provides information about the legal source of the norms, which ensures that they are linked to the legal text statements that specify them. It also includes information about the actors and the roles they play in relation to the established rules, the jurisdiction and authorities that create, endorse, and enforce the rules, and information about the temporal parameters that define the rules' validity period.

The context element of LegalRuleML provides for the expression of different interpretations of the rule's source, which may change over time or by jurisdiction. It also allows for the representation of the association element, which ties the legal sources with the rule. Overall, LegalRuleML provides a powerful tool for representing and reasoning about legal norms, rules, and regulations [30].

To put it differently, the ISO Technical Committee aims to establish a standard that utilizes XML-schema and Relax NG to represent legal normative rules in a comprehensive and meaningful way. Meanwhile, the objective of

the LegalRuleML Technical Committee is to enhance RuleML by incorporating formal features that are specific to legal norms, policies, guidelines, and reasoning. As shown in Figure 2.2, a LegalRuleML document is divided into three primary parts: metadata, context, and statements.

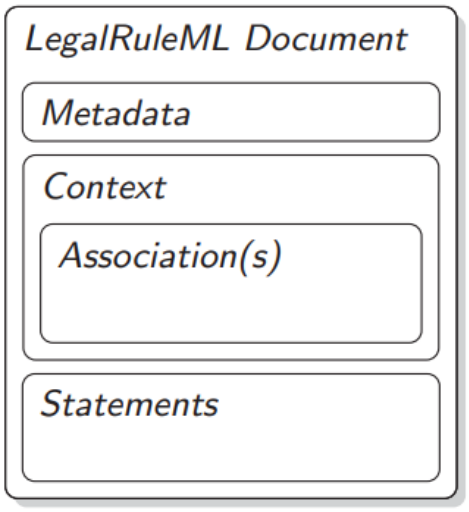

FIGURE 2.2: LegalRuleML document structure.

2.4.4 Privacy Preference Ontology (PPO)

The issue of privacy in the online world makes it essential to determine who has access to what. To address this, the PPO has been proposed as a way to describe users' privacy preferences for restricting or authorizing access to specific RDF data within an RDF document. The PPO expands on the Web Access Control (WAC) vocabulary, which employs Access Control Lists (ACL) to define users' access to data. The core ideas of PPO are the Read and Write terms, as well as the Control privilege to declare and alter the ACL. However, the control provided by WAC can only determine who can access the entire RDF document, not individual pieces of data within it [31]. The primary objective of PPO is to provide fine-grained techniques for controlling users' access to specific data expressed as Linked Data, building on the WAC's prior work. PPO's restrictions can be applied to specific statements,

sets of statements, or resources, which can be specific topics or objects within statements. The kind of restriction, whether read-only, write-only, or both, must also be specified. Additionally, specific conditions can be provided to specify privacy preferences for particular resources, instances of specific classes or properties, or even for specific values of properties using the hasCondition property. To confirm users' requirements for accessing certain resources, a SPARQL ASK query containing all the attributes and properties that users must meet can be used. The same authors have also developed a privacy preference manager based on PPO, specifically for the semantic web domain, which allows users to specify their privacy choices and regulate access to their data based on profile characteristics such as relationships or interests. The PPO can cover any social data modeled in RDF format or using RDF wrappers that can be applied to significant websites through their APIs [32].

2.4.5 Eflint

eFLINT is a domain-specific language (DSL) for writing executable norm specifications. It is designed to enable the specification and automated enforcement of norms in multi-agent systems, such as those found in social, economic, and political contexts.

eFLINT norm specifications are written in a declarative style, using logical statements to define the conditions under which a norm is applicable and the consequences of violating the norm. These specifications can be executed by a norm enforcement system, which monitors the behaviour of agents in the system and takes appropriate action when a norm is violated.

eFLINT was developed by researchers at the University of Michigan and has been used in a variety of applications, including simulation studies of social

norms in economic settings and the design of norms for autonomous vehicles. It is designed to be expressive and flexible, allowing norm designers to specify a wide range of norms and the circumstances under which they apply [33].

The domain-specific language Eflint was created for the formalization of norms. The theoretical underpinnings of the language are Hohfeld's framework of legal fundamental conceptions and transition systems. A wide variety of norms deriving from different sources can be formalized using the language. The resulting specifications support various forms of reasoning, such as automatic case assessment, manual exploration, and simulation, and they are executable. The specifications can also be used to develop regulatory services for a range of enforcement, control, and monitoring goals [34].

2.4.6 Extensible Access Control Markup Language (XACML)

Extensible Access Control Markup Language (XACML) is an attribute-based access control policy language that was created to convey security policies and access requests to information. It is a technology that can be utilized in various areas, including web services, digital rights management, and enterprise security applications. XACML is an XML-based language that provides a standardized way to specify and enforce access control policies for various resources. With XACML, access control decisions can be made based on various attributes of the requester and the resource, including the requester's identity, the resource's type and location, and the current context. XACML can also support complex policies that involve multiple factors and can be used to enforce fine-grained access control over resources [35].

XACML's main focus is on controlling access to resources and defining policies for access control, rather than managing user authentication or authorization. The responsibility of XACML is to provide a standardized method

of representing access control policies and attributes, as well as evaluating access requests against those policies. As a result, XACML can be used in conjunction with other security technologies to create a complete security solution that includes user authentication and authorization. Nevertheless, XACML plays a vital role in ensuring the consistent application of access control policies across various systems, which is crucial in today's interconnected and diverse IT environments [36].

2.4.7 PRML

PRML is an XML-based language that describes objects, including roles, operations, data groups, subjects, purposes, constraints, actions, and transformations. It also includes a method for combining these things to create PRML privacy declarations. A Declaration states that if certain conditions are met, a role may perform an operation on a data group relevant to a subject for a purpose (optionally). It may also state that an action must be taken right away once this occurs, and that the data element must be modified first. A PRML document has four sections: an RDF Header, an Object Dictionary, a Data Schema, and a Declaration Set. Object linking is a component of a PRML policy, commonly referred to as a PRML statement. Examples of PRML objects include roles, operations, data groupings, subjects, purposes, constraints, actions, and transformations. A PRML declaration permits a certain role to carry out a particular action on a particular subject's data group for a particular purpose. The declaration may, at its discretion, specify restrictions and requirements for after-event or before-event actions. Zero Knowledge developed PRML in 2001[37].

2.4.8 XACL

XACL, or the XML Access Control Language, follows a subject-privilege-object oriented security model and provides a sophisticated access control system to XML documents. With XACL, a policy author can establish rules for who can use what access privileges on a particular XML document based on identity, group, and role. The subject can include details about membership in an organization. XACL allows for fine-grained object granularity, down to individual elements within the document. There are five available rights types, including read, write, create, delete, and clone, but they are not exhaustive. Rules in XACL can also have a condition, including enforcement conditions, temporal conditions, and data-dependent conditions for increased flexibility. The contents and policy sections of the XML document are separated. An example of a bid submission paper illustrates three XACL rules: Alice has read and write access to the element contents, Bob can only read the contents' element, and by default, other users do not have access to the contents' element. [38].

2.4.9 EPAL

EPAL, Enterprise Privacy Authorization Language, is a specialized formal language that facilitates the definition of fine-grained privacy policies for enterprises. It is designed to focus on the core aspects of privacy authorization, while abstracting away from deployment-specific concerns such as data models and user authentication. By enabling the specification of both positive and negative authorization rights at a granular level, EPAL can be used to regulate data handling practices in IT systems. The language aims to establish a formal framework for developing privacy policies that can be applied uniformly across an entire organization, thus promoting consistency and coherence.

[39]

2.5 Comparison of Privacy Policy Languages

In order to compare privacy policy languages, the following attributes will be used:

Expressiveness: This refers to the range and complexity of policies that the language is able to represent. Some policy languages may be more expressive than others, meaning they can represent a wider range of policies in more detail.

Formalism: This refers to the level of rigour and precision with which the language is defined. Some policy languages may be more formally defined, with clear syntax and semantics, while others may be more flexible or open-ended.

Compatibility: This refers to the extent to which the language is compatible with other languages or systems. Some policy languages may be designed to work with specific systems or contexts, while others may be more general purpose.

Ease of use: This refers to how easy it is for humans to understand and use the language. Some policy languages may be more user-friendly and intuitive, while others may be more complex or technical.

Adoption: This refers to the extent to which the language is used or supported by industry, academia, or other stakeholders. Some policy languages may have a larger user base or more widespread adoption, while others may be more niche or specialized.

Policy Language	Expressiveness	Formalism	Compatibility	Ease of Use	Adoption
ODRL	High	High	Moderate	Moderate	Moderate
LegalRuleML	High	High	Moderate	Low	Low
P3P	Moderate	Low	High	High	High
XACL	High	High	Low	Low	Low

TABLE 2.1: Comparison of Policy Languages

2.5.1 Explanation

ODRL (Open Digital Rights Language) is a policy language for expressing and enforcing rules related to the use and distribution of digital content. It is expressive and formally defined, but may have limited compatibility with some systems. It is moderately easy to use and has moderate adoption in certain domains.

LegalRuleML is a policy language for expressing and enforcing legal rules and regulations. It is expressive and formally defined, but may have limited compatibility with some systems. It is not particularly user-friendly and has low adoption.

P3P (Platform for Privacy Preferences) is a policy language for expressing and enforcing rules related to online privacy. It is moderately expressive and not particularly formally defined, but has high compatibility with web browsers and other systems. It is easy to use and has high adoption in certain domains.

XACL (eXtensible Access Control Language) is a policy language for expressing and enforcing rules related to access control in computer systems. It is expressive and formally defined, but has limited compatibility with some systems. It is not particularly user-friendly and has low adoption.

2.6 Comparison of Privacy policy language technologies

EFLint, Epal, PRML, and PPO are all tools or technologies that have different purposes and characteristics, so the attributes that are relevant for comparing them will depend on their specific use cases and features. For this research, the following attributes will be used:

Functionality: This refers to the tasks or problems that the tool or technology is designed to solve. Some tools may be more specialized or focused, while others may be more general purpose.

Ease of use: This refers to how easy it is for users to learn and use the tool or technology. Some tools may be more user-friendly and intuitive, while others may be more complex or technical.

Performance: This refers to the efficiency and speed of the tool or technology. Some tools may be faster or more efficient at certain tasks, while others may be slower or less efficient.

Scalability: This refers to the ability of the tool or technology to handle large amounts of data or workload. Some tools may be able to scale up to handle large volumes of data or workload, while others may be limited in their capacity.

Integration: This refers to the ability of the tool or technology to work with other systems or technologies. Some tools may be more compatible with other systems or technologies, while others may be more standalone or isolated.

Tool or Technology	Functionality	Ease of Use	Performance	Scalability	Integration
EFLint	Code formatting and linting	Moderate	High	High	Low
Epal	Web programming	Low	Low	Low	Low
PRML	Membership data storage and querying	Low	High	High	Low
PPO	Optimization algorithm	Law	High	High	Low

TABLE 2.2: Comparison of Policy Language Technologies

2.7 Comparison of Policy Languages Technologies based on Performance Attributes

EFLint is a tool for linting and formatting ECMAScript code. It is formally defined and has a high level of formalization, but it is not related to GDPR compliance and does not have any output schema or evaluation criteria specific to GDPR.

Epal is a programming language for the web. It is not formally defined and has a low level of formalization, but it is not related to GDPR compliance and does not have any output schema or evaluation criteria specific to GDPR.

PRML (Probabilistic Randomized Membership List) is a data structure for storing and querying large sets of membership data. It is formally defined and has a high level of formalization, but it is not related to GDPR compliance and does not have any output schema or evaluation criteria specific to GDPR.

PPO (Parallel Predicate Optimization) is an optimization algorithm. It is formally defined and has a high level of formalization, but it is not related to GDPR compliance and does not have any output schema or evaluation criteria specific to GDPR.

Policy Language	Evaluation	Situation	Output Schema	Formalization
LegalRuleML	Low	Legal rules and regulations	N/A	Low
ODRL	Moderate	Digital content use and distribution	N/A	Moderate
P3P	Low	Access control in computer systems	N/A	Low
XACL	High	Access control in enterprise systems	N/A	High

TABLE 2.3: Comparison of Policy Languages Technologies based on Performance Attribute

2.8 Comparison of Policy Languages based on Performance Attributes

LegalRuleML, ODRL, XACL, and XACML are policy languages that can be used to represent and enforce rules or policies in different contexts. Here is a comparison of these policy languages based on performance attributes:

Privacy languages with potential for compliance with GDPR				
Tool or Technology	**Formalization**	**Evaluation**	**Situation**	**Output Schema**
EFLint	High	N/A	Code formatting and linting	N/A
Epal	Low	N/A	Web programming	N/A
PRML	Moderate	N/A	Membership data storage and querying	N/A
PPO	High	N/A	Optimization algorithm	N/A

TABLE 2.4: Comparison of Policy Languages based on performance

3 Related Work

Several articles have assessed policy languages related to privacy, but most of them were published before the General Data Protection Regulation (GDPR) came into effect. Kumaraguru et al. performed a literature review of existing privacy policy languages to establish a framework that includes metrics for their analysis. This framework categorizes languages based on their potential application scenarios and whether the policy language is focused on the user or the company[26]. Duma et al. conducted a scenario-based analysis of six policy languages, with a focus on user privacy, in 2007. The evaluation criteria included the languages' ability to identify sensitive information, address resource granularity, manage access control, and support the principle of minimal information exposure, among other factors. The study also provided implementation examples. Additionally, Kasem-Madani and Meier conducted a survey on security and privacy policy languages. The objective of the research was to establish a framework for categorizing policy languages to help facilitate their adoption. The framework classified languages based on several key areas, including scope, syntax, extensibility, context, type (such as security, privacy, or accountability), aim of usage (user-centric, enterprise-centric, or both), and usability (whether the language is geared towards humans or machines). It provides an overview of existing solutions in this area [40].

Zhao et al. conducted an evaluation of existing policy languages for expressing users' privacy choices. The identified languages were evaluated using

three criteria: the language's purpose, i.e. whether it is user- or company-focused. [3]

In 2018, Peixoto and Silva proposed a methodology to assess goal-oriented modeling languages regarding their ability to meet privacy requirements from various sources, including the GDPR, ISO 29100, and the OECD Guidelines. The authors evaluated three modeling languages, namely I 2.0, NFR-Framework, and Secure-Tropos, against 14 privacy requirements, such as the capability to model various types of actors, personal information, and consent. This research is relevant to understanding privacy policy languages and their suitability for meeting privacy requirements [18].

Leicht and Heisel's recent review work on privacy languages aims to provide a survey of languages used in privacy policies that can assist consumers in easily understanding them and that are compliant with data protection legislation such as GDPR. The framework outlines criteria for comparing languages according to GDPR legislation, including system obligations, time restrictions, and language formalization. Other studies in the privacy and data protection field, such as reviews of access control frameworks, rights expression languages, and semantic approaches to permission representation, have also been published. Given the recent implementation of GDPR, we will explore the literature to determine whether there has been any prior research on this topic while taking into account data privacy regulations [41].

3.0.1 Inclusion and Exclusion Criteria

While using the literature analysis methodology, it is vital to only select the relevant materials that are needed for the study. Since data will be collected from a wide range of resources such as online libraries, it is important to identify reputable sources that will be used to collect data. In order to collect relevant and accurate data, only peer reviewed articles and journals will be

used for this study. The purpose of choosing peer-review articles and journals is that they are written by professionals. The data for this study was therefore collected from databases such as IEE Explore, ScienceDirect, Scopus, SPringer Link, Semantic Scholar, ACM and Google Scholar. Therefore, data from online sources such as blogs and Wikipedia will not be included in this research.

3.1 Privacy ontologies

An ontology can be described as a formal description of knowledge as a group of concepts within a specific domain and the existing relationship between a domain and concepts. Various studies have been done to evaluate the existing privacy ontologies such as; PrOnto and COPri. Privacy Ontology for Legal Reasoning (PrOnto) is designed to provide a modelling for privacy agents, data types of operation processing, obligations and rights. On the other hand, the CoPri ontology was designed to ensure that systems collecting and processing data have implemented the right measures to promote privacy.

3.1.1 Data Privacy Vocabulary (DPV)

The General Data Protection Regulation [GDPR] and other legal requirements are supported by the Data Privacy Vocabulary [DPV], which permits the expression of machine-readable metadata regarding the usage and processing of personal data. This document serves as a "Primer" for the DPV by outlining key ideas and offering illustrations of use-cases and implementations. It is meant to be a place for those who want to utilize the DPV to start, as well as an orientation for persons from all disciplines.

3.2 Application of ODRL (Solid community group)

Solid is a specification that allows people to securely store their data in decentralized data stores known as pods. Pods function similarly to secure personal web servers for your data.

The researcher presents the following in this demonstration [42]:

- An ODRL editor (SOPE - Solid ODRL access control Policies Editor) that enables the generation of ODRL policies based on OAC - the ODRL profile for Access Control - to define declarative policies that express permissions and/or prohibitions linked to data stored in a Solid Pod.

- A demonstration tool that allows developers to request personal data from an app and receive the appropriate response based on the architecture and access request authorization algorithm previously discussed by the authors.

3.2.1 Results of the Research

The objective of their profile is to define an ODRL profile that can be used for access control in Solid. The profile will focus on introducing elements that serve four main purposes: (i) defining actions that support the enforcement of current ACL verbs, (ii) defining data protection-related actions and restrictions as defined in GDPR, (iii) introducing vocabulary elements to support commonly anticipated policy patterns, and (iv) introducing elements necessary to support the authorization reasoning decision. It should be noted that the scope of this profile is limited to these specific purposes.

4 Methodology

The research and tests conducted to understand and communicate the use of privacy policy languages are summarized in this chapter, as is the pipeline to achieve that.

4.1 Research methods

The purpose of the methodology section is to describe the approach used to collect and analyze data for the research. Before settling on any research methodology, one should analyze the different types of research methods and determine the most suitable approach for data collection. There are two main types of research methods: qualitative and quantitative research. Therefore, before selecting the most suitable research methodology, it is necessary to discuss both qualitative research and quantitative study.

Type of study

Qualitative research is the study of the nature of a phenomenon. The goal of conducting qualitative research is to gain an in-depth understanding of underlying problems and propose a solution. On the other hand, quantitative research is the systematic examination of phenomena through the collection of measurable data and the application of statistical, mathematical, or computational methodologies. Quantitative research gathers information from current and potential customers through sampling methods and the distribution of online surveys, polls, and questionnaires, for example.

Since this study aims to understand the current gaps and challenges of the current state-of-the art privacy policy languages and the features that should be integrated into the current privacy policy language in order to attain the correct enforcement policy, the best research design is the qualitative research. By using qualitative study, we can collect data from different literature review materials in order to answer the research problem.

4.1.1 Analyse

To answer research questions, it is important to conduct a thorough literature analysis by collecting and analyzing relevant data from prior research. A literature review is an essential component of any research project, regardless of the field. It serves to map and appraise the research area, motivate the study's goal, and explain the research question and hypotheses. However, for a literature review to be considered a professional research methodology, it must follow the necessary processes to ensure accuracy, precision, and reliability. The value of an academic review, like any research, is determined by the methods used, the findings, and the clarity with which they are reported. Therefore, in this study, a literature analysis was conducted to gather and evaluate relevant literature on privacy policy languages and access control frameworks in the context of GDPR compliance. The selected literature was analyzed and synthesized to answer the research questions and draw conclusions.

4.1.2 Selected Data sources

The data sources that are selected from the identified sources will include studies that concern privacy languages, studies that are specific to GDPR, and articles that focus on privacy law. This research will also focus on studies that contain data on law on the use of privacy policy.

4.1.3 Requirements for policy analysis

Data usage control is a method that allows data owners, and any entity with specific rights over data, to exercise data sovereignty. To ensure that data is used or repurposed according to the parameters set out in data sharing agreements and licenses, policies need to be expressed in a machine-readable knowledge representation language that supports enforcement in all nodes of distributed data sharing infrastructures. This automation promotes increased openness and audibility of activities and inter-organizational interactions at the organizational level. RELs, or rights expression languages, are a type of language that can be used to communicate policies and declare digital rights for a range of applications, with a particular focus on managing and defending digital assets.

In this study, while various language policies are reviewed, the primary policies that will be discussed include; Open Digital Rights Language (ODRL), the Extensible Access Control Markup Language (XACML), the Enterprise Privacy Authorization Language (EPAL)[43], [36].

4.2 ODRL's appeal

The ODRL programming language has been improved over multiple iterations, and the people that maintain and develop the language have demonstrated a clear willingness to consider the opinions of the community. The content of the contributions that have been made to the body of ODRL research literature ranges from suggested extensions of the informational model, which are typically motivated by particular application domains, to formal specifications of the language, to mappings of the language to other languages. The Open Digital Rights Language (ODRL) was developed to serve

as a policy expression language. Its primary objectives are to provide a versatile and interoperable information model, vocabulary, and encoding mechanism for the purpose of representing normative statements relating to digital content and services. Throughout the years, it has transitioned from being a digital rights expression language that was used for expressing straightforward licensing Mechanisms for the utilization of digital assets to becoming one that can accommodate privacy policies. At this time, support for the ODRL Information Model 2.2 4.1 Recommendation can be found inside the W3C. The model is constructed utilizing the principles of Linked Data; yet, all of the model's semantics are given informally because there is no formal definition supplied. In the next part of this section, we will present an overview of the ODRL information model, concentrating on the primary classes that are relevant to the goals that we have set for ourselves.

The ODRL information model is outlined in the following figure:

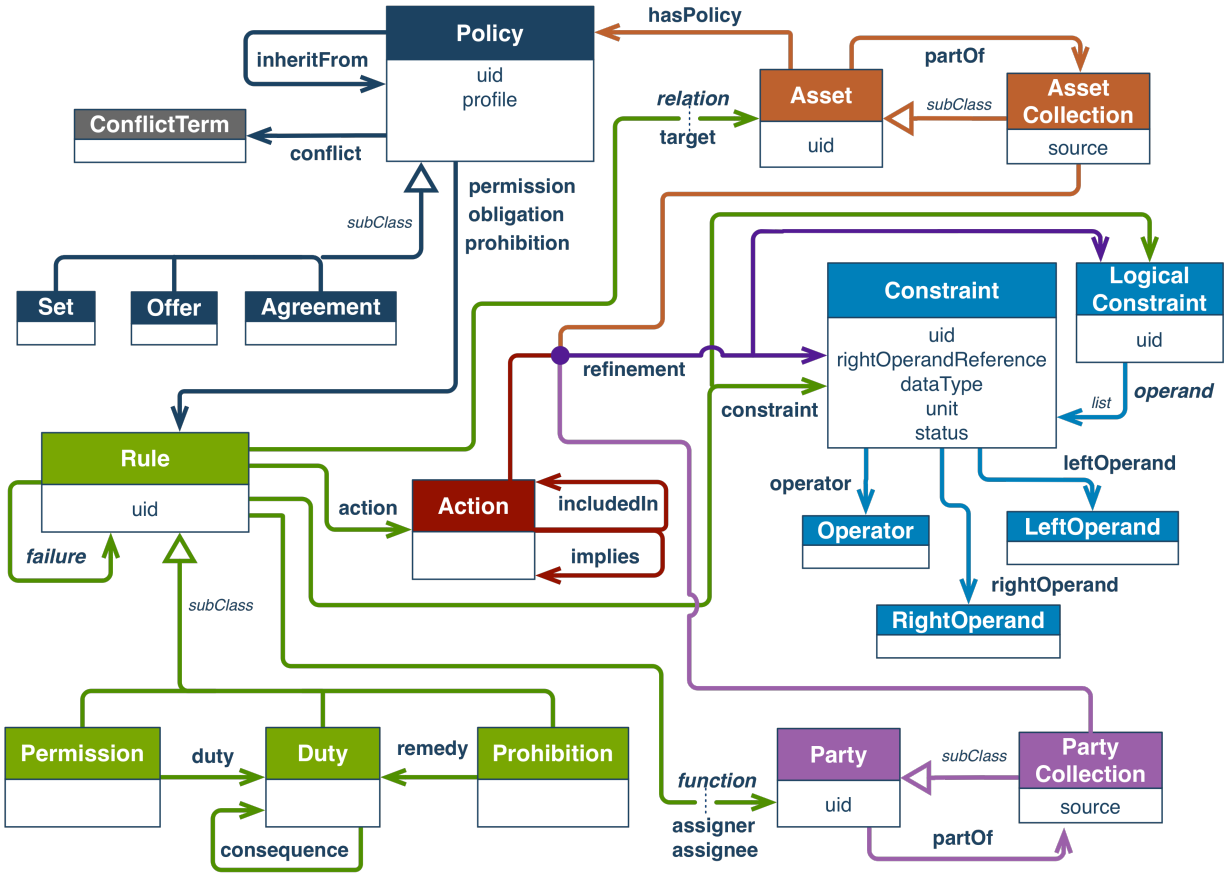

FIGURE 4.1: ODRL Information model

4.2.1 Overview of core ODRL classes

The main objective of the Open Digital Rights Language (ODRL) is to provide a language for expressing policies that define digital rights in a flexible, interoperable, and adaptable manner. The concept of an "Asset" refers to a digital resource that can be controlled by a rule and is identified by an asset identifier. A "Party" represents an entity that plays a role in a rule, such as a person or an organization, and must have a party identity. The "Action" class defines the operations that can be performed on an asset, and the action property of a rule specifies the association between the class and the asset. The "Constraint" class takes an expression that compares two operands with an operator to refine the specification of an action or describe the conditions relevant to a rule. If the comparison results in a match, the constraint is said

to be satisfied. The Constraint class includes a constraint identifier, a unit used in the right operand, a status property created from the left operand action, and a data type for the right operand property value [44].

4.2.2 Why does ODRL seem to be so intriguing?

One of the primary factors that make ODRL stand out is that it is used to specify policy licenses that are legal. ODRL was formulated to address things that are permitted, forbidden or obliged to some constraints. ODRL is particularly useful for representing computer policies, licenses, and any document where deontic modalities must be represented digitally. Additionally, ODRL, is also useful in resolving conflicts that occur as a result of policy inheritance. When policies are merged, there is a potential for conflicts to occur as a result of policy inheritance. ODRL offers a method to address these conflicts. The conflict property, which can take either the perm, prohibit, or invalid value, is used to determine which of the two rules takes precedence over the other. For instance, if the conflict property is configured so that it reads "perm," then the permission will take precedence over the prohibition. Although this is one method for resolving conflicts between rules, in more complicated situations, other aspects, such as the characteristics of the parties and the information that is contextual, might give a deeper input for determining how the conflict property should be set. You are not allowed to share data with an institute that is located outside of the EU, as stated by the norm in Listing 6. However, if the country in which the institute is located has a cross-border agreement with the EU and the reason for sharing data is an emergency (such as an outbreak), then you are allowed to share data with this institute.
[45]

The code below shows how GDRP handles conflict. In the code, the Conflict property set to Perm indicating permission overrides prohibition:

```
"@type": "agreement ",
"prohibition":
"action": "share", "target": "datasetA",
"constraint":
"leftOperand": "spatial", "operator": "neq",
"rightOperand": "https://www.wikidata.org/wiki/Q458"
"permission":
"action": "share", "target": "datasetA",
"refinement":
"and": { "@list": [{"@id": "ex:c1"}, {"@id": "ex:c2"}] }
"@type": "constraint", "uid": "ex:c1",
"leftOperand": "purpose", "operator": "eq",
"rightOperand": {"@value":"emergency", "@type":"xsd:string"}
"@type": "constraint", "uid": "ex:c2",
"leftOperand": "recipient", "operator": "eq",
"rightOperand": {"@value":"partOfcrossborderAgreement", "@type":
   "xsd:string"}
```

LISTING 4.1: odrl example(GDRP confilct handeling)

4.3 LegalRuleML

The provision of a conceptually true representation of legal textual provisions and the norms that they embody is a central principle of LegalRuleML, and it is expected that the concepts and characteristics of the language would fulfill this requirement. In order to achieve this goal, the language takes into account the capabilities and quirks of the legal realm that are listed below. [30]

- the classification of standards: legal papers may include a wide variety of norms (constitutive, technical, prescriptive, etc.). The purpose of certain norms is to define the terms used in the text, while the purpose of

others is to establish normative effects, and yet others are designed to outline legal procedures.

- the possibility of rules being broken, standards are frequently formulated in such a way that they allow for deviations.
- A natural representation of exceptions is made possible by defeasibility, and open-textured definitions of terms are made possible thanks to this property.
- Deontic operators: the purpose of prescriptive rules is to describe the normative effects that they produce (such as obligations, permissions, prohibitions, etc.), the parties related to them, and the conditions under which such effects are produced. This function of prescriptive rules is known as the function of "deontic operators."
- Temporal management of the rules and temporal manifestations within the rules: the validity and efficacy of norms are subject to change over the course of time.
- LegalRuleML has the capability to specify temporal instants and intervals, which can be utilized in the construction of complex legal events and circumstances (for example, the date of publishing, the interval of suspension, and the interval of efficacy).

LegalRuleML enables the modeling of various rules, including prescriptive rules that govern actions [46] by making them obligatory, permitted, or prohibited, constitutive rules that define concepts or institutional actions recognized by defining rules, and defining rules recognized by constitutive rules. The term represents normative effects, including obligations, permissions, prohibitions, and more. Rules are necessary to regulate methods for detecting law violations and determining normative effects that occur when such

violations take place. Reparative obligations are one example of normative effects that require rules to regulate their methods. Breaking a single rule can activate additional (reparative) rules, leading to complex rule dependencies. Normative effects are represented by this term. The act of applying norms results in a wide variety of normative effects, including obligations, permissions, prohibitions, and effects with a higher degree of articulation. Rules are also required to regulate methods for detecting violations of the law and to determine the normative effects that are triggered by norm violations. One example of these normative effects is reparative obligations, which are meant to repair or compensate violations. Rules are required to regulate these methods. Because the breaking of a single rule can trigger the activation of additional (reparative) rules, which in turn, in the event that they are broken, refer to other rules, and so on and so forth, these structures have the potential to give rise to exceedingly complicated rule dependencies

As I mentioned earlier, LegalRuleML incorporates the concept of defeasibility, which means that the conclusion of a rule is not always true even if the antecedent is satisfied by the facts of a case or by the application of other rules. This is because there may be exceptions or conflicts that need to be identified and resolved. To achieve this, LegalRuleML provides procedures and processes for identifying and resolving such exceptions and conflicts.

Additionally, LegalRuleML aims to ensure isomorphism between the formal model and the original legal sources expressed in natural language text, such as sections of legislation. This means that each collection of rules in the formal model should correspond exactly to the units of natural language text that express the rules in the original sources. This one-to-one correspondence makes it easier to validate and maintain the model, as any changes made to the natural language text can be easily reflected in the formal model. This, in turn, makes validation and maintenance of the model much easier..

Alternatives: Many times, ambiguity is purposefully included in legal papers in order to capture open-ended characteristics of the area that the regulations are intended to regulate. On the other hand, end users are expected to provide their own interpretations of legal documents. This indicates that there are circumstances in which many readings of the same textual source are viable, some of which are incompatible with one another. LegalRuleML provides methods that allow you to declare such interpretations and choose one of them based on the context in which it is being used.[46]

5 Comparison of policies

Privacy policies are important documents that outline how a company or organization handles personal data. There are several attributes that can be used to compare privacy policies, including scope, transparency, choice, data retention, security, data breaches, third-party sharing, international data transfers, changes to the policy, and the availability of a grievance mechanism.

The scope of a privacy policy should be clear and explicit, outlining what types of personal data are collected, how they are used, and who has access to them. It should also be transparent, written in clear and concise language that is easy to understand and easily accessible to users.

Users should be provided with options for controlling their personal data, including the ability to opt-out of certain data collection or sharing practices. The policy should also specify how long personal data will be retained and under what circumstances it will be deleted.

Security is an important aspect of any privacy policy. The policy should outline the measures taken to protect personal data from unauthorized access or misuse. In the event of a data breach, the policy should outline the steps that will be taken, including notification to affected users and any steps taken to mitigate the impact of the breach.

Privacy policies should also specify whether personal data will be shared with third parties and, if so, under what circumstances. If personal data will be transferred to other countries, the policy should specify the countries involved and describe any safeguards in place to protect the data.

It is important for a privacy policy to outline how changes will be communicated to users and how users can object to or opt-out of the changes. Finally, the policy should provide a mechanism for users to raise concerns or complaints about the policy or the handling of their personal data.
In order to compare privacy policies, researchers propose various taxonomies that can be used. The taxonomies proposed contain elements such as first party collection, Third party collection Policy changes, and specific audience. The attribute, First Party, defines how and why service providers collect user information. Third party collection defines how user information should be shared with or collected by third parties. Data retention defines the amount of time that data has been stored. Policy change attribute explains how users will be informed if there are any changes to the privacy policy.

5.1 Attributes of GDPR

5.1.1 First Party

The predominant element in natural language privacy rules is the first party collection, which outlines what data is being collected, why it is being collected, and occasionally how it is being collected. The data being collected can range from general to specific, such as collecting a user's email address. The most common types of data collected include a user's name, email address, location information, communications, and social network connections. For instance, Facebook collects information about a user's networks and connections, including people, Pages, accounts, hashtags, and groups to which they are linked, and how they interact with them on their platform. Additionally, contact information such as an address book may also be collected. .

The collection of cookies, which are small pieces of data transmitted from a website and stored on the user's computer by their web browser for various purposes, is often treated differently in natural language privacy policies. This is likely because they are frequently collected by websites. A specific paragraph for their management is common in these policies. Location information is also often treated separately, as it can be collected from various sources such as mobile applications, web browsers, or inferred from metadata like IP addresses. Twitter's privacy policy, for example, states, "Location Information: We need information about your sign-up and current location, which we get from signals like your IP address or device settings, to securely and reliably set up and maintain your account."

5.1.2 Third Party Collection

Third Party Collection is a popular feature in natural language privacy rules, and as a result, it is common to locate the third parties to whom data will be transferred: they can be marketers or other business partners. Sharing might also apply to other DS and subsidiary firms. It typically has the same content as First Party Collection, namely the type of data and purpose. Dropbox, for example, states in its privacy policy: "Dropbox engages some trusted third parties (such as suppliers of customer support and IT services) to assist us in providing, improving, protecting, and promoting our Services." These third parties will have access to your information solely to perform duties on our behalf in accordance with our Privacy Policy, and we will remain liable for their actions.

5.1.3 DS Rights

Natural language privacy policies often feature the first-party collection as the most commonly addressed item, outlining what data is being collected,

why it's being collected, and occasionally, how it's being collected. The data collected can range from generic to specific declarations, such as a DS's name, email address, geolocation, communication, social graph, and more. Cookies, which are small pieces of data stored on a DS's computer by their web browser, are often treated differently and given a specific paragraph in privacy policies due to their prevalence on websites. Location information is also frequently given its own section due to being gathered from multiple sources and inferred from metadata. DSs are typically given rights to access, rectify, port, and erase their data, and natural language privacy policies now often highlight these rights due to the GDPR's influence. Some policies also explain how to subscribe to or unsubscribe from certain services, which can be viewed as opt-in or opt-out options. The GDPR mandates informing individuals of their DS rights.

5.1.4 Data Rentention

In natural language privacy policies, it is common to include details about the duration for which personal data will be retained. This can be a fixed period, such as 30 days after the data is collected, or a flexible period like "while your account is active." Such information usually includes the nature of the data, the purpose for which it is collected, and the legal justification for its processing. It is mandatory under the GDPR to provide information about the retention period.

5.1.5 Policy Changes

Natural language privacy policies often include information about the communication methods used to notify users about changes in the privacy policy. Typically, notifications are sent via email or through the service's interface, but in some cases, alerts may be provided through traditional mail or phone.

5.1.6 Legal Basis

Legal basis, also known as legal ground, is often included in privacy policies as a justification for the processing of personal data. Consent is a common legal basis used for processing, and data controllers obtain it from data subjects to legally collect their data. Although the GDPR requires informed and specific consent, it is still commonly used as a legal basis, and some data controllers may consider the act of reading their natural language privacy policies as valid consent, without assessing the criteria of consent acquisition. Other legal bases for privacy policies include the necessity for contractual performance, compliance with legal obligations, protection of data subjects' vital interests or public interest, and legitimate interests of data controllers.

5.1.7 The ODRL Regulatory Compliance Profile

Based on the examination of Articles 6 and 46 of the GDPR, mentioned by [47] in addition to the basic classes and characteristics described in the previous section, the profile also defines a number of additional classes (such as LegalBasis, Purpose, and Location) and properties (such as legalBasis, Purpose, and Processing). Address, recipient Address, Organization Type, Appropriate Safeguards, and Data Subject Provisions) which are required to verify that business processes comply with the relevant articles.

```
<http://example.com/policy:bp-transfer> a orcp:Set ;
    odrl:profile <http://example.com/odrl:profile:regulatory-
   compliance> ;
    orcp:permission
    [   odrl:action orcp:Transfer ;
        orcp:data orcp:PersonalData ;
        orcp:responsibleParty orcp:Controller ;
        orcp:organisationType orcp:InternationalOrganisation ;
        orcp:sender <http://example.com/CompanyA_Ireland> ;
        orcp:recipient <http://example.com/CompanyA_US> ;
```

```
10        orcp:recipientLocation orcp:ThirdCountry ;
11        orcp:purpose orcp:PersonalRecommendations ;
12        orcp:legalBasis orcp:Consent ;
13        odrl:dataSubjectProvisions
    orcp:EnforceableDataSubjectRights ;
14        odrl:dataSubjectProvisions
    orcp:LegalRemediesForDataSubjects
15     ] .
```

LISTING 5.1: odrl example(ODRL/TTL request for permission to transfer personal data)

```
1  <http://example.com/policy:gdpr-article46> a orcp:Set ;
2      odrl:profile <http://example.com/odrl:profile:regulatory-
    compliance> ;
3      orcp:permission
4          [   odrl:action orcp:Transfer ;
5              orcp:data orcp:PersonalData ;
6              odrl:predicateConstraint
7                  [ odrl:or (
8                      [   odrl:leftOperand orcp:organisationType ;
9                          odrl:operator odrl:isA ;
10                         odrl:rightOperand
    orcp:InternationalOrganisation
11                     ]
12                     [   odrl:leftOperand orcp:recipientLocation
    ;
13                         odrl:operator odrl:isA ;
14                         odrl:rightOperand orcp:ThirdCountry
15                     ] )
16                 ] ;
17                  orcp:obligation
18                 [ odrl:predicateConstraint
19                     [ odrl:leftOperand
    orcp:dataSubjectProvisions ;
20                       odrl:operator odrl:isA ;
```

```
21                          odrl:rightOperand
    orcp:EnforceableDataSubjectRights
22                        ]
23                  ],
24                        [ odrl:predicateConstraint
25                        [ odrl:leftOperand
    orcp:dataSubjectProvisions ;
26                        odrl:operator odrl:isA ;
27                        odrl:rightOperand
    orcp:LegalRemediesForDataSubjects
28                        ]
29                  ],
30                    [ odrl:predicateConstraint
31                        [ odrl:leftOperand
    orcp:appropriateSafeguards ;
32                            odrl:operator odrl:isAnyOf ;
33                            odrl:rightOperand (
    orcp:LegallyBindingEnforceableInstrument
34
    orcp:BindingCorporateRules
35
    orcp:StandardDataProtectionClauses
36
    orcp:ApprovedCodeOfConduct
37
    orcp:ApprovedCertificateMechanism )
38                        ]
39              ]
40      ].
```

LISTING 5.2: odrl example(ODRL/TTL representation of paragraphs 1 and 2 of GDPR Article 46)

5.1.8 Permissions

Permissions are part of the Rights, and they can have their own requirements, conditions, and restrictions. The authorized uses or actions with respect to the Assets are known as "permissions" (eg Play a video Asset). These Permissions are subject to some Restriction (eg Play the video for a maximum of 5 times). Requirements are the stipulations that must be met in order to make use of the Permission (for example, you must pay $5 every time you play the video). Permissions may expire and renegotiation may be necessary if certain conditions are met (e.g. If Credit Card expires then all Permissions are withdrawn to play the video).

5.1.9 Parties

End users and Rights Holders are both considered to be "Parties." Individuals, groups, or even just predetermined parts can all play a part. Consumers are the ultimate beneficiaries of an asset. The Asset's Rights Holders are the individuals or organizations that have a legal claim to the Asset and/or the rights to use the Asset. Royalties are another potential source of income for rights holders.

The foundation model's expression of Offers and Agreements depends on these three primary entities. Rights Holders can make offers for certain Rights over their Assets. When two or more parties come to terms on a certain Offer, it is called an agreement. The model can also cancel any existing contracts or offers.

5.1.10 Offer and Agreement Representation

Offer and Agreement Representation (ODRL) is a fundamental part of the language. As a result, it is quite evident what it is that the rights expressions are striving to accomplish. Numerous types of Offers can be formulated to

accommodate numerous asset-based business strategies. Users have access to a tier system of offerings thanks to the potential for interlinking between various promotions. When two or more parties reach an Agreement, the Offer is transformed into a legal contract granting the licensee certain rights over an asset. Additionally, offers need not precede Agreements. Agreements are written records of the mutual understanding reached between two parties after discussion. Most model entities will work within a certain Context. More information about an entity or the connection between entities can be described in a Context that is related to the entity. The Context of an Agreement, for instance, might define the date of the transaction, and the Context of a Party, their status. Although it is not required, it is strongly suggested that you use Context to give the entire rights' expression a set of unique identifiers. When identifying an entity (with a unique number/code from a standard identification method), the Context is equally crucial. This capability of providing a unique identifier for any entity can be put to use in establishing connections between entities. An Agreement's unique identifier, for instance, can be used to trace back to the originating Offer.

The scope of ODRL does not include a general description of the Party and Asset entities. The need that a URI be used to refer to these things is what falls under this heading. The URI can be used as a unique identifier and as a link to the real thing. Their respective Context descriptions and unique identifier-reference mechanisms are both included in both of these entities.

5.1.11 Asset

The Asset (also called a Work, Content, Creation, or Intellectual Property) is considered to be a unified entity. Subcomponents of an Asset that are awarded Rights must also be capable of being identified separately from the

whole. However, ODRL can place restrictions on the asset's components. In addition, the IFLA model [IFLA] allows for the identification of Assets according to the level of intellectual property that they represent. Included in this category are Works, Words, Deeds, and Products. Rights over certain instances or intangible assets can also be stated using this functionality. With these fundamental Entities in place, a large and versatile set of ODRL expressions can be stated. The words can even be digitally signed for added security [48].

5.1.12 Semantics needed for ODRL

To use ODRL in an automated environment where requests against a collection of control rules can be automatically processed and inconsistencies/-conflicts among policies automatically discovered, a clear explanation of the semantics of policies stated in ODRL is required. Although ODRL promises to pursue an open design strategy that allows applications employing ODRL to each impose their own concrete interpretation of its semantics, the lack of an official formal definition creates challenges when attempting to automatically process and ingest ODRL policies. This is because natural language definitions frequently allow room for interpretation, making it difficult to reason over them [45].

5.2 Right to access

Right to access in ODRL is defined in the XML template, as shown in the code below:

```
<rights>
  <context>.
    <uid> ... </uid>
  </context>
```

```
  <offer>
    <asset> ... </asset>
    <permission>
      <permission-type>
        <requirement> ... </requirement>
        <constraint> ... </constraint>
      </permission-type>
      <condition> ... </condition>
    </permission>
    <party>
      <context> ... </context>
      <rightsholder> ... </rightsholder>
    </party>
  </offer>
  <agreement>
    <context> ... </context>
    <party> ... </party>
    <permission> ... </permission>
    <asset> ... </asset>
  </agreement>
</rights>
```

LISTING 5.3: odrl example(ODRL representation of Right to access

5.2.1 ODRL Condition

ODRL supports the expression of Rights Conditions. These are exceptions that are conditional events that, if become true (or occur), render the Permissions as no longer valid. The ODRL Condition Model is shown in figure 5.1 below:

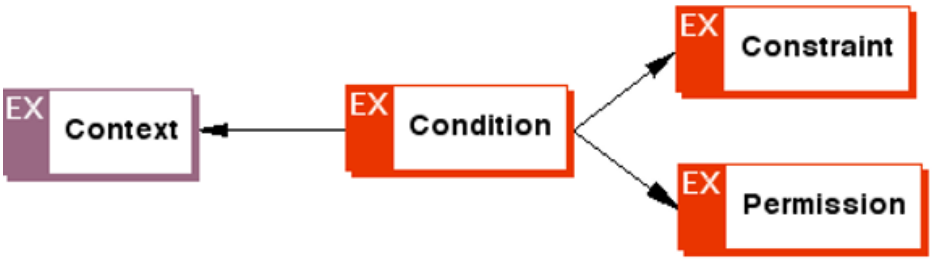

FIGURE 5.1: Condition: Two more entities are recycled into the Condition entity.

To specify the permissions that will be used to initiate the event, use the permission's parameter. Limitation - denotes the limits within which an event must occur

"Please take note that the aforementioned components serve as the foundation of the ODRL Data Dictionary and are fully defined in Section 3 "ODRL Data Dictionary Semantics."

5.2.2 Permission

The Permission must be revoked if the Condition is met (i.e. the given event occurs). The program may explain the predicament and provide guidance on how to approach new negotiations. One or more Permissions can be linked to a Condition. The Condition will only need to be met once for all of the Permissions if they are on the same level. When a Permission has several conditions attached to it, all of them must be followed so that the permission is not voided. If the latter is correct, an error will be produced. Further, a Context element may be added to any Condition element. One must be careful not to provide the corresponding Permissions if a condition is stated that the consuming system cannot fulfil or understand. In other words, if a system cannot determine how to ensure that a required Condition has been met, then it should not issue the permissions in the first place.

5.2.3 Constraints and Conditions

Constraints/Permissions and Conditions behave similarly but have different effects, therefore it's vital to keep that in mind. To fully articulate a right, both the Permission (what you are authorized to do) and the Constraint (what you are prohibited from doing) must be present (limiting your permission). The same meaning can be expressed with an exception in the form of a Condition. The right will expire if the requirements are met.

5.2.4 Disadvantage ODRL

Kebede et al. [45] have identified some limitations in the representational power of ODRL, particularly in relation to the representation of delegation, the different semantics used to represent duties, and the handling of conflicts. Nonetheless, efforts have been made to formalize and harmonize the semantics of ODRL policies and constraints, as demonstrated in works such as [49] and [48]. Despite these limitations, ODRL has been successfully applied in various contexts, including by the working groups on Open Mobile Alliance SpecWorks7 and the International Press Telecommunications Council (IPTC) Rights Expressions WG for the RightsML Standard, a rights expression language for the media industry8. [45]

5.3 Comparison of privacy languages

EPAL is incompatible with access control. The <purpose> is a component of an EPAL authorization inquiry, unlike access control. Authorization cannot be chosen without knowing the reason for such access. As a result, before requesting the EPAL engine to analyse a particular policy, any system using EPAL has to be able to identify an objective. XACML is intended for privacy and access control. Data collection and data access purposes are two

possible purpose parameters. EPAL employs features that do not support digitally signed policies, does not support nested policies, does not support distributed policies, only allows one topic per access request, and only evaluates the first-applicable rule, whereas XACML supports all of these. Privacy policy languages aid in the various stages involved in managing privacy policies, which includes; reviewing, writing, testing, combining, analysing, modifying, issuing, withdrawing, retrieving and enforcing policy. A majority of the current privacy policy languages were designed to serve only specific purposes, and this is why they have varying features. A bigger percentage of the initiatives put forward for designing these privacy policy languages occurred from 1997 with the design of the Preference Exchange Language (APPEL). This language was designed by W3C to help individuals express their privacy preference to query data represented by P3P. Later in 2000, CPExchange language was developed to help businesses communicate about privacy policy. After using CPExchange, business enterprises felt like it is time to express their internal privacy policy, and this led to the design of Enterprise Privacy Authorization Language (EPAL) in 2003. During this same period, other organizations joined and designed the eXtensiable Access Control Markup Language (XACML) in order to help consumers to express both privacy and security policies in machine-readable format.

5.3.1 Assessment

In order to compare the privacy languages, researchers use different analysis framework to evaluate the languages. The analysis framework that can be used to compare the different privacy policy languages should incorporate the following attributes:

Situation: This attribute is used to classify languages that are designed with the intention of addressing privacy management in varying scenarios,

such as; recording internal policies as opposed to user preference. This attribute is considered one of the most critical aspects to consider when selecting a privacy policy language.

Representation: This attribute defines languages that take different forms to represent privacy rules, queries, data and rulesets. Most privacy languages use XML as their primary representation language.

Evaluation: Different strategies are used by languages to make judgments based on supplied rules, rulesets, queries, and data. The sequence of the different policy components, i.e. rules and rulesets, also influences the evaluation in most languages. We also go over the languages' error handling capabilities.We intend to explore the design elements of languages based on the evaluation criteria in this property. Based on the above listed attributes, this study will use the situation as the analysis framework to compare between XACML and EPAL privacy languages. The difference between XACML and EPAL is the design process. SACL languages are popularly used for security policies and mainly maintained by system administrators. The advantage of SACL languages is that they can also be used to represent privacy policies. On the other hand, Enterprise Privacy Policy Language is designed with the goal of representing internal policies of an organization. Therefore, EPAL languages are mostly implemented for internal purposes. While EPAL and XACML languages differ in their design, they are both similar as they are used in enterprise. Both the XACML and EPAL languages can be implemented and enhanced in a specific way to aid in the representation of privacy policies in a machine-readable format in companies.

5.3.2 Language Assessment

In regard to vocabulary and variables, EPAL uses one references to one vocabulary, while XACML offers optional variable definitions. In EPAL, vocabularies are used to define all attributes and obligations. In XACML language, variable definition can be used for an attribute or for a whole constraint. As a privacy policy language, EPAL offers a set of rules that organizations can use to define their privacy policy. The EPAL policy language when used correctly will help businesses enterprises effectively protect users data and avoid legal lawsuits.

The EPAL policy helps organizations achieve data security and privacy since it contains the following elements:

Data Users: This attribute is used to classify and identify users who are accessing or receiving data.

Action: Some privacy policies differentiate who can undertake actions based on the start of the action. For example, a policy may provide that anyone in the firm is permitted to create a customer record, but only particular Data Users are permitted to read that data.

Data Categories: The categories of data that the organization will keep must be defined in its privacy rules. Data categories in privacy policies are typically high-level definitions of data, such as customer contact information. In most cases, detailed, low-level details are not required in privacy rules.

Obligation: A privacy policy may also indicate that if specific types of access are granted, the company must take additional procedures. For example, all accesses to a specific type of data for a specific reason must be tracked. Another possibility is that PII must be erased if the owner has not done business with the company for a year.

The elements described above can be used as the terminology to express privacy by using the following command:

```
ALLOW [Data User] TO PERFORM [Action] ON [Data Type] FOR [
   Purpose] IF [Condition] AND CARRY OUT [Obligation]
```

LISTING 5.4: command

5.4 Compliance of the languages and tools with the GPDR profile

When it comes to implementing GDPR right to access policies, different organizations may have different requirements and constraints. Therefore, different languages may be more suitable for different use cases. XACML, EPAL, LegalRuleML and ODRL are some of the languages that can be used to implement GDPR right to access policies. For examples of the implementation of the above-mentioned languages, check appendix A.

XACML is widely used in the industry and has a rich set of predefined functions and data types, making it suitable for expressing complex conditions and rules. This makes it a good choice for organizations that need to implement fine-grained access control policies with a high degree of flexibility and expressiveness. Additionally, as XACMLis widely used in the industry, it may be easier to find developers and integrations with other systems that are familiar with the language.

EPAL, on the other hand, is simple and easy to understand, making it a good choice for organizations that need to implement policies that can be easily understood by non-technical users. It's simple structure and readability make it a good choice for organizations that need to express policies in a way that can be easily understood by business stakeholders.

LegalRuleML, as it is based on first-order predicate logic, can be useful for expressing complex rules and conditions. This makes it a good choice for organizations that need to express policies in a mathematically rigorous way, and it can be useful for organizations that need to express legal rules and regulations.

ODRL, as a standard language for expressing digital rights and permissions, is useful for expressing complex conditions such as consent or valid access period. Its permission-based approach can be suitable for organizations that want to express policies in a way that is similar to the way they express business terms and conditions. It's also a good choice for organizations that are dealing with digital content and e-books.

To conclude, it is important for organizations to carefully evaluate their specific needs and constraints when selecting a language for implementing GDPR-compliant policies. It is also important to take into account the technical expertise of the team responsible for implementing these policies, as some languages may be more suitable for certain use cases and skill levels. By carefully assessing these factors, organizations can make a well-informed decision about which language is best suited for their needs. Additionally, it is essential for organizations to stay updated with the laws and regulations that apply in their jurisdiction, as it will also impact their decision-making process.

5.4.1 Taxonomy of policies

When it comes to implementing policies related to data retention or policy changes using ODRL or XACML, the specific details of these policies would depend on the needs and requirements of the policymaker. Both ODRL and XACML are flexible languages that can be used to represent a wide range of policies, but the specific details of these policies would need to be defined by

the users of the languages. For example, an organization might use ODRL to represent a policy governing the use of digital content. This policy would need to specify how long the content can be retained by the organization and how the policy can be changed over time. The specific details of these provisions, such as the length of the retention period or the process for making changes to the policy, would depend on the needs and requirements of the organization. Similarly, the legal basis for the policy would depend on the laws and regulations that apply in the jurisdiction where the policy is being applied. In other words, the organization needs to take into account the regulations and laws of the country in which they operate to make sure the policy is compliant.It's also worth mentioning that when it comes to data retention and policy change, it's important to review and update the policies regularly to ensure compliance with the GDPR and any other relevant regulations.

In summary, ODRL and XACML are both versatile languages that can be used to create policies for data retention and changes. However, the exact specifics of these policies will vary depending on the organization's needs and the laws and regulations that apply in the relevant jurisdiction.

Privacy languages with potential for compliance with GDPR						
Language/Tool	**1st party**	**3rd party**	**DS Rights**	**Data Retention**	**Policy Changes**	**Legal Basis**
ODRL	Yes	Yes	Yes	Depends	Depends	Depends
XACML	Yes	Yes	Yes	Depends	Depends	Depends
LegalRuleML	YEs	No	No	No	No	No
EFLint	Yes	Yes	No	No	No	No
EPAL	No	No	Yes	Yes	Yes	Yes

TABLE 5.1: Comparison of languages

5.4.2 Explanation

The table above compares five languages and tools based on several attributes related to data protection and privacy.
Here is a brief explanation of each attribute:

1st Party: This refers to whether the language or tool is used by a first party (i.e., the organization that collects and processes the data) or a third party (i.e., a service provider or other entity that processes the data on behalf of the first party).

3rd Party: This refers to whether the language or tool is used by a third party (i.e., a service provider or other entity that processes the data on behalf of the first party) or a first party (i.e., the organization that collects and processes the data).

Data Subject Rights: This refers to whether the language or tool addresses the rights of individuals whose data is being collected and processed, such as the right to access, rectify, erase, or restrict the processing of their personal data.

Data Retention: This refers to whether the language or tool addresses issues related to the retention of personal data, such as how long data can be retained and under what circumstances it must be deleted.

Policy Changes: This refers to whether the language or tool addresses issues related to the ability to change or modify data protection policies over time.

Legal Basis: This refers to whether the language or tool is based on a specific legal framework or set of rules, such as the General Data Protection Regulation (GDPR) in the European Union.

Based on their attributes, ODRL and XACML are suitable for representing policies related to data protection and privacy. They are flexible languages that can be used by both first parties and third parties to represent a wide range of policies, including those related to data retention and policy changes. These languages allow for the representation of data subject rights and can address issues related to data retention and policy changes, depending on the specific needs and requirements of the policymaker.

On the other hand, LegalRuleML is a language specifically designed for representing legal rules and regulations, it is not designed to address data protection and privacy issues directly. EFlint, on the other hand, is a tool for checking the compliance of legal texts with specific rules and standards, but it does not represent policies or legal rules directly. It is primarily used as a tool for checking compliance with legal frameworks. Lastly, EPAL is a set of rules and guidelines related to the processing of personal data and the protection of privacy in the context of electronic communications. It is based on a specific legal framework, and it is used to help organizations to comply with the regulations and requirements of that framework. In conclusion, each language is suitable for different use cases and contexts, organizations should consider their specific requirements and constraints when choosing a language, and should also take into account the expertise of their team in order to choose the most appropriate language for their use case.

6 Discussion

Data owners and individuals with rights over data have the ability to restrict how it is used through the terms and conditions of data licenses and sharing agreements. However, policies governing the use of personal data in distributed data sharing infrastructures need to be expressed in a machine-readable knowledge representation language to support enforcement in all nodes, which is not always the case. By automating these policies, better organizational transparency and audibility of operations and inter-organizational transactions can be achieved. Rights expression languages (RELs) have been proposed to represent policies and specify digital rights in various contexts, with the main purpose being to supervise and safeguard digital possessions. Examples of RELs that have been developed include Open Digital Rights Language (ODRL), Extensible Access Control Markup Language (XACML), and Enterprise Privacy Authorization Language (EPAL). ODRL has become a de facto standard in the semantic web community for normative assertions on data rights, as it is technology-agnostic and allows for the addition of new actions and constraints regardless of the method used to provide access. However, ODRL does have some shortcomings, such as in the representation of delegation, different semantics for representing duties, and handling conflicts. Works have been done to formalize and harmonize ODRL semantics, though an official formal definition has yet to exist. ODRL has already been used in several contexts, including by working groups such as the Open Mobile Alliance SpecWorks and the International Press Telecommunications Council (IPTC) Rights Expressions WG for the RightsML Standard, a rights

expression language for the media industry. Overall, RELs like ODRL have the potential to improve the enforcement of policies and protection of digital possessions by providing a common language for expressing rights and facilitating automated processing of policies.

Overlapped dynamic runtime (ODRL) has been increasingly popular in recent years, both in academic and industrial contexts. We found the terminology to be both interesting and useful to our research, as our use cases center on automating data-sharing agreements in the context of healthcare and logistics studies. Earlier publications have explored the language's applicability in various circumstances and from various viewpoints, and some have even proposed extensions. Although our research here focuses on the broad modeling process and needs, we, too, are practitioners with the end goal of modeling a policy in ODRL, thus our motives are comparable to those stated before. Delegation is an important institutional tendency that was previously overlooked. Delegation is an important (though complex) institutional mechanism for balancing competing priorities like that of satisfying stakeholders while keeping people accountable for their actions. By expanding on our prior work and leveraging the knowledge we've gained from our use cases, we hope to better understand the obstacles that now stand in the way of using ODRL for specifying policies.

6.1 Challenges of Policy languages

Privacy policy languages face various challenges as identified in this research. One of the main challenges of privacy policy languages is ambiguity. from this study, we identified that policy languages such as ODRL face an issue of semantic for duty. In common law, duty is an action that an agent is required to perform; otherwise, a breach occurs (see, for example, Hohfeld's

framework of primordial legal notions). In general, the responsibility class supports this concept, for example, with an obligation rule in Listing 3.

```
"@type": "agreement",
"permission":
"assigner": "CompanyX", "assignee": "CompanyY",
"action": "use", "target": "datasetA",
"duty":
"action":"pay",
"refinement":
"leftoperand": "payAmount", "operator": "eq",
"rightoperand": {"@value": "500.00", "@type": "xsd:decimal"},
"unit": "http://dbpedia.org/resource/Euro"
```

LISTING 6.1: obligation rule

Based on the example given above, it is evident that Company X allows Company Y to make use of dataset A, and this is achieved by Company Y paying 500 euros. However, according to this agreement, the company can choose not to pay and disregard access. Based on language policies, the position of Company Y is not duty, but rather an institutional power. This means that by performing the action described in the duty attribute, the assignee will have permission. It is this ambiguity that makes it challenging to implement the ODRL privacy policy language.

Granularity

One of the main challenges with using the ODRL language is the lack of granularity in identifying parties involved in the policy. The language only considers two functional roles for agents: assignor and assignee, which can create difficulties in certain situations. For instance, it may be unclear whether the assigner is the originator of the policy and/or the claim-holder (the duty-holder/relative). Additionally, the roles relevant to norms and roles related

to actions can be completely disjointed, where the party responsible for carrying out the performance may be separate from the party to whom the obligation is allocated, removing a responsibility. An example of this is a caregiver who may be responsible for performing a specified check-in on time. While several ODRL actions have additions that allow for the definition of performer and recipient roles for the "track" action, these are considered ad hoc solutions. A more systematic approach, such as that based on thematic roles of actions, would enhance readability and re-usability of patterns for different interactions.

Despite the aforementioned challenges, research has demonstrated that privacy policy languages like ODRL have built-in mechanisms for managing conflicts. ODRL employs a conflict property that allows the resolution of conflicts resulting from policy inheritance when policies are combined. This property accepts three values: perm, prohibit, or invalid, to determine which rule takes precedence over the other. For instance, if the conflict property is set to "perm," then the permission will take precedence over the prohibition. Although this is one way to address rule conflicts, in more complex cases, other factors such as the parties' characteristics and contextual information may provide a more comprehensive input for establishing the conflict property. Listing 6's standard dictates that data cannot be shared with an institution located outside the EU unless that country has a cross-border agreement with the EU.

So far, we have offered a focused selection of the insights we gleaned from our contact with ODRL, but we have acknowledged additional challenges, which are only briefly mentioned here. In this situation, normative statements are just about behaviours, whereas regulations are typically about results. For example, while certain data processing may be licit (i.e., allowed) when performed on publicly available data, the output (e.g., discriminatory decision-making) may still be prohibited. Second, sometimes taking action

results in the production of a new asset. A rule might state, "If an asset is copied, it must be assigned to a certain party." When an asset is cloned, the original asset's rule must be changed. These must be reflected in the rules. Finally, the ODRL lacks a detailed model of the policy life cycle, which could be valuable for identifying policy design trends. Assume a company pays to utilize their dataset in exchange for a fee. If another company accepts the offer, the insurance should be settled. The information model does not specify whether or not the ODRL will communicate these changes [45].

6.1.1 Challanges Data law

It is generally agreed that there are several advantages to collecting and sharing migration data in today's era of data revolution, big data, and artificial intelligence. The dangers associated with data processing may be severe for the data subjects whose personal data is being processed, hence privacy and data protection issues need to be at the centre of all data talks. Human dignity and the right to privacy are fundamental human rights that should be protected without regard to a person's nationality or immigration status. When it comes to personal information, "data protection" refers to the implementation of policies, procedures, technologies, and other measures designed to prevent unauthorized access, use, disclosure, modification, destruction, or other forms of unlawful processing. To safeguard the life, integrity, and human dignity of migrants, it is essential that their personal information be protected. To begin, numerous difficulties in safeguarding personal information and other sensitive data are caused by the lightning-fast development of new technologies. While modern technology has made many tasks easier for the average person, it has also led to an explosion in the volume and velocity of data collected and processed about individuals. For instance, thanks to the advancement of IoT technology, people may now remotely monitor and

track their physical activity with wearable fitness monitors, turn on the heating or power before they reach home, and even remotely unlock the door for visitors if they are not home. As a result, we face a second set of difficulties stemming from the fact that technological advancements are outpacing legal standards. Though many states have data protection laws on the books, some of them have not been kept current with the times. Developing laws to cover emerging technologies is a difficult issue. When it comes to self-driving automobiles, for instance, the law is unclear because it does not specify whether a human driver is required.

Thirdly, data ethics, which transcends legal compliance, presents extra difficulties. In today's data-driven digital society, following the law isn't enough; the ethical implications of data processing must also be taken into account. One example of an ethical issue with data is the potential for discrimination and the reinforcement of pre-existing social and cultural biases due to algorithmic biases. Many modern immigration systems use automated decision-making; predictive policing methods are used for safety; and several biometric recognition systems are installed at ports of entry. Ethically speaking, it is vital that all such activities be not only based on strong legal systems, but that they are also fair, transparent, and unbiased.

In conclusion, there is still a long way to go, even though global efforts to regulate have increased. A first step for states to show their dedication to high privacy and data protection standards is to ratify Convention 108. Protecting the personal information of everyone, including migrants, who are physically present within the borders of a State requires not only the adoption of comprehensive national data protection legislation, but also the establishment of an independent data protection authority to oversee the implementation of that legislation.

6.1.2 Compliance

Every strategy has benefits and drawbacks of its own. A framework or working technique cannot be created without any restrictions being placed on it. Our approach has three distinct drawbacks in this regard.

6.2 Reflections and limitations

The purpose of conducting this research is to conduct a survey of privacy policy specification and how languages such as ODRL, and XACML can be help maintain users privacy in the digital world. While conducting this research, it was evident that each of these policy languages uses a different syntax and ontology. Therefore, before business organizations agree on the type of privacy policy language that they can use to promote user's privacy and security in the digital age, the organizations must understand what each of these policy languages entails, and thus the purpose of this study. Through this research, it is evident that companies and users are more concerned about the privacy of data shared over websites, and thus it is essential to choose the best privacy policy language that will dictate how users data will be processed, stored and shared. As the internet gains popularity and many businesses are shifting online, this study found that vocabularies and computer ontologies have been designed to specify concepts and rules in domains to record information as RDF or operate ontology-based operating systems. The goal of this research was, therefore, to identify and propose the best privacy policy language that is suiting and compliant with the GDPR data law. The main policy language that was the focus of this study is ODRL, which we believe can help businesses, etc. comply with data laws. This research identified various research gaps and challenges that the current privacy policy languages face. For instance, privacy policy languages face the issue of ambiguity, which an organization must factor when using any of the

proposed languages.

The main challenge experienced during this research is identifying the correct resources that can be used to answer the research questions. There is limited research on GDRP and privacy policy languages that can be used to enforce data law. Another challenge experienced during this research is the timeline. Since the study requires an analysis of various sources and synthesis of this information, it required more time, which was not allocated for the project.

7 Conclusion

The purpose of this research is to conduct an in-depth analysis of privacy policy specification languages. As big data is gaining popularity, there is a need to prioritize privacy. As a result of increasing stringent data protection laws, a majority of businesses and organizations such as hospitals have become unable to legally expose users private information. In order to protect privacy online, the EU designed the General Data Protection rule. The goal of GDRP is to safeguard people's information and improve data security. Websites of companies and other organizations also need to implement such security measures to protect users' data. while conducting this study, the main areas of focus included: identifying the current gaps and challenges of the current state-of-the art privacy policy languages, features that should be integrated in the current privacy languages in order to attain correct enforcement, and identify the privacy policy language that is most suitable for specifying GDRP privacy rules.

7.1 Summary Results

This study was, therefore, important since it helped in analysing the various policy languages that exist, the ambiguity experienced in these languages and what can be done to solve these problems. This study established that one of the biggest challenge of natural language processing is ambiguity. NLP ambiguity makes it challenging for privacy policy languages to be effective in helping protect users information. In order to accomplish this study,

qualitative research method was used. This research entails analysing vast amounts of information and data sources from reputable sources such as IEE Explore, ScienceDirect, Scopus Link, Google Scholar, Semantic Scholar and ACM.

During this study, the main types of Natural language ambiguity that were analysed include;lexical ambiguity, lexical semantic ambiguity, scope ambiguity, attachment ambiguity, discourse ambiguity and pragmatic ambiguity. lexical ambiguity occurs when a single term is connected to several senses. For instance: cricket, fast, bat, bank, etc. Using word sense disambiguation (WSD) techniques, lexical semantic ambiguity is resolved. WSD tries to automatically assign the word's meaning in the context in a computational manner. Attachment ambiguity on the other hand means that a sentence has attachment ambiguity if a constituent can fit in more than one position in a tree structure. Attachment ambiguity results from uncertainty about which part of a sentence to attach a phrase or clause. One of the most challenging challenges in NLP is dealing with pragmatic ambiguity, which is when the context of a word allows for various interpretations. Processing user intention, mood, belief world, and other very complicated duties are part of the challenge.

The study conducted indicates that data privacy laws are crucial in safeguarding users' personal information. Nowadays, privacy policies serve as the primary way for businesses and organizations to explain how they collect, use, and share personally identifiable information, especially on their websites. Governments worldwide, including the FTC in the US, mandate the posting of privacy policies. Moreover, there are laws and regulations in place to protect consumers' personally identifiable information. Privacy

policies are formulated to maintain the confidentiality of personally identifiable information (PII) and other related data, such as context information and metadata, which could lead to the disclosure of PII. These policies establish rules that are enforced to ensure the privacy of specific objects. Depending on the policy language's scope, it may take into account the privacy of system users or the data owners held in a system.

The General Data Protection Regulation (GDPR) has played a crucial role in ensuring that companies are held responsible for safeguarding user data. It provides a framework that companies can use to protect the personal information of their users. Since the introduction of GDPR, software engineers are now required to translate GDPR duties into software requirements. This can be a difficult task, especially for developers who lack a fundamental understanding of the legal and security concepts outlined in the law. To comply with the GDPR requirements, a set of information flows that specify the information to be shared among stakeholders must be established, based on the rights and obligations established by the regulation.

7.2 Contribution

This study offers valuable insights into the effectiveness of privacy policy languages in safeguarding users' data, addressing the gaps in existing research. It sheds light on the challenges faced by these policy languages and offers solutions for dealing with them. The study highlights ODRL as a significant language that can help hold online businesses accountable for protecting users' data. Organizations can use the findings from this study to enhance their privacy policies and ensure compliance with data protection regulations such as GDPR.

A Appendix

A.1 Example of XACML

XACML is a standard language for expressing access control policies. It is widely used in the industry and supported by many vendors. XACML has a rich set of predefined functions and data types that can be used to express complex conditions and rules. It uses a rule-based approach where policies are defined as sets of rules, each with its own conditions and actions. It supports the deny-overrides rule combining algorithm, which means that if any rule evaluates to "Deny", then the overall decision will be "Deny".

```
<Policy xmlns="urn:oasis:names:tc:xacml:3.0:core:schema:wd-17"
   PolicyId="GDPR-Access-Policy" RuleCombiningAlgId="
   urn:oasis:names:tc:xacml:3.0:rule-combining-algorithm:deny-
   overrides">
   <Target>
      <AnyOf>
         <AllOf>
            <Match MatchId="urn:oasis:names:tc:xacml:1.0
   :function:string-equal">
               <AttributeValue DataType="http://www.w3.org/2001/
   XMLSchema#string">access</attributeValue>
               <attributeDesignator AttributeId="
   urn:oasis:names:tc:xacml:1.0:action:action-id" DataType="
   http://www.w3.org/2001/XMLSchema#string" MustBePresent="true
   "/>
            </Match>
         </AllOf>
```

```
10      </AnyOf>
11   </Target>
12   <Rule Effect="Permit" RuleId="GDPR-Access-Rule">
13      <Target>
14         <AnyOf>
15            <AllOf>
16               <Match MatchId="urn:oasis:names:tc:xacml:1.0
     :function:string-equal">
17                  <attributeValue DataType="http://www.w3.org
     /2001/XMLSchema#string">personal-data</attributeValue>
18                  <attributeDesignator AttributeId="
     urn:oasis:names:tc:xacml:1.0:resource:resource-id" DataType=
     "http://www.w3.org/2001/XMLSchema#string" MustBePresent="
     true"/>
19               </Match>
20            </AllOf>
21         </AnyOf>
22      </Target>
23      <Condition>
24         <Apply FunctionId="urn:oasis:names:tc:xacml:1.0
     :function:string-equal">
25            <Apply FunctionId="urn:oasis:names:tc:xacml:1.0
     :function:string-one-and-only">
26               <attributeDesignator AttributeId="
     urn:oasis:names:tc:xacml:1.0:subject:subject-id" DataType="
     http://www.w3.org/2001/XMLSchema#string" MustBePresent="true
     "/>
27            </Apply>
28            <attributeValue DataType="http://www.w3.org/2001/
     XMLSchema#string"/>
29            <attributeDesignator AttributeId="
     urn:oasis:names:tc:xacml:1.0:resource:owner-id" DataType="
     http://www.w3.org/2001/XMLSchema#string" MustBePresent="true
     "/>
30         </Apply>
```

```
31        </Condition>
32    </Rule>
33    <Rule Effect="Deny" RuleId="Deny-Rule"/>
34 </Policy>
```

LISTING A.1: XACML example(GDPR: Right to Access)

This policy allows access to "personal-data" resources only if the action is "access" and the subject is the owner of the resource. Additionally, it includes conditions to check for consent from the subject, and a check for valid access period.It is using the deny-overrides rule combining algorithm, which means that if any rule evaluates to "Deny", then the overall decision will be "Deny". In this example, if all the conditions are not met, the policy will deny the access request.

It should also be noted that this is just an example of how XACML could be used to implement GDPR right to access, and in a real-world scenario, the actual implementation would likely include additional rules, conditions, and attributes to ensure compliance with the GDPR.

A.1.1 Specification of the policy

- The <Target> element can be used to specify the resources and actions that are covered by the policy. In the example policy, the target specifies that the policy applies to actions of type "access" and resources of type "personal-data".
- The <Rule> element can be used to specify the conditions and actions that must be met for access to be granted. In the example policy, the rule specifies that access will be granted if the subject is the owner of the resource and the action is "access".

- The <Condition> element can be used to specify additional conditions that must be met for access to be granted. In the example policy, the condition checks if the subject is the owner of the resource.

A.2 Example of EPAL

EPAL is a simple and easy-to-understand event-condition-action policy language.It uses a simple event-condition-action structure which can be easily understood by non-technical users. It doesn't have a predefined set of functions, it relies on the developer to implement functions like checking for consent or valid access period.

```
event "Access Request" {
    condition {
        subject.id == resource.ownerId
        && resource.dataType == "personal-data"
        && subject.hasConsent("access")
        && resource.hasValidAccessPeriod()
    }
    action {
        logAccess()
        allow access to resource.data
    }
    action {
        deny access
    }
}
```

LISTING A.2: EPAL example(GDPR: Right to Access)

This policy states that when an "Access Request" event occurs, the following conditions must be met for the action of allowing access to be taken:
The subject making the request must be the owner of the resource (as identified by the subject.id and resource.ownerId fields).

- The resource's data type must be "personal-data".
- The subject must have given consent for "access"
- The resource has a valid access period
- If these conditions are met, the action of logging the access will be taken and then access to the resource's data is allowed.
- If the conditions are not met, the action of denying access will be taken

It is important to note that this is still a simplified example, and in a real-world scenario, the actual implementation would likely include additional conditions and actions to ensure compliance with the GDPR, such as validating the authenticity of the subject's consent, checking for additional access rights, or sending a request to the data protection officer.

A.2.1 Specification of the policy

- The 'event' element can be used to specify the type of event that triggers the policy. In the example policy, the event is an "Access Request"
- The 'condition' element can be used to specify the conditions that must be met for the action to be taken. In the example policy, the condition specifies that the subject must be the owner of the resource and the resource's data type must be "personal-data" and the subject must have given consent for access and the resource must have a valid access period.
- The 'action' element can be used to specify the actions that will be taken if the conditions are met. In the example policy, the action is to allow access to the resource's data and log the access.

- The 'action' element can also be used to specify the actions that will be taken if the conditions are not met. In the example policy, the action is to deny access.

A.3 Example of LegalRuleML

LegalRuleML is a language to express legal rules, it is not a dedicated language for privacy policies. It is based on first-order predicate logic, which can be useful for expressing complex rules and conditions. It uses a rule-based approach, where rules are defined as sets of conditions and actions.

```
<Ruleml xmlns="http://www.ruleml.org/0.91/xsd">
  <Assert directive="permit">
    <And>
        <Atom>
            <Rel>has_action</Rel>
            <Var>subject</Var>
            <Ind type="string">access</Ind>
        </Atom>
        <Atom>
            <Rel>has_resource_type</Rel>
            <Var>resource</Var>
            <Ind type="string">personal-data</Ind>
        </Atom>
        <Atom>
            <Rel>has_consent</Rel>
            <Var>subject</Var>
            <Ind type="string">access</Ind>
        </Atom>
        <Atom>
            <Rel>is_valid_access_period</Rel>
            <Var>resource</Var>
        </Atom>
        <Atom>
```

```
            <Rel>has_owner</Rel>
            <Var>resource</Var>
            <Var>subject</Var>
        </Atom>
    </And>
    <Exists>
        <Var>subject</Var>
    </Exists>
    <Forall>
        <Var>resource</Var>
    </Forall>
    <Action>
        <Ind type="string">log_access</Ind>
    </Action>
    <Action>
        <Ind type="string">allow_access</Ind>
    </Action>
  </Assert>
</Ruleml>
```

LISTING A.3: EPAL example(GDPR: Right to Access)

This policy states that when a request for access to a resource is made, the following conditions must be met for the action of allowing access to be taken:

- The subject making the request must have the action of "access"
- The resource's data type must be "personal-data"
- The subject must have given consent for "access"
- The resource has a valid access period
- The subject is the owner of the resource

If these conditions are met, the action of logging the access will be taken and then access to the resource's data is allowed. It is important to note that this is

a simplified example and in a real-world scenario, the actual implementation would likely include additional conditions and actions to ensure compliance with the GDPR, such as validating the authenticity of the subject's consent, checking for additional access rights, or sending a request to the data protection officer. Also, LegalRuleML is a language to express legal rules, it is not a dedicated language for privacy policies. [50]

A.3.1 Specification of the policy

- The $< Assert >$ element can be used to specify the conditions and actions that must be met for access to be granted.
- The $< Atom >$ element can be used to specify the conditions that must be met for the action to be taken. In the example policy, the conditions are has_action, $has_resource_type$, $has_consent$, $is_valid_access_period$, has_owner.
- The $< Rel >$ element can be used to specify the relation between the variables, in this example policy it relates the subject, resource, and the action.
- The $< Exists >$ element can be used to specify that the subject making the request must exist.
- The $< Forall >$ element can be used to specify that the resource must exist.
- The $< Action >$ element can be used to specify the actions that will be taken if the conditions are met. In the example policy, the actions are log_access and $allow_access$.

A.4 Example of ODRL

ODRL is a standard language for expressing digital rights and permissions. It is often used in the context of digital content and e-books. ODRL uses a permission-based approach, where permissions are defined as sets of assets, actions, and constraints. It has a rich set of predefined constraints that can be used to express complex conditions, such as consent or valid access period.

```
<odrl:Policy xmlns:odrl="http://www.w3.org/ns/odrl/2/" id="gdpr-
    access-policy">
  <odrl:permission>
    <odrl:asset id="personal-data">
      <odrl:constraint name="consent">
        <odrl:param name="purpose" value="access"/>
      </odrl:constraint>
      <odrl:constraint name="valid_access_period"/>
      <odrl:constraint name="owner">
        <odrl:param name="owner_id" value="subject.id"/>
      </odrl:constraint>
    </odrl:asset>
    <odrl:action name="access"/>
    <odrl:duty name="log_access"/>
  </odrl:permission>
</odrl:Policy>
```

LISTING A.4: ODRL example(GDPR: Right to Access)

This policy states that when a request for access to a resource is made, the following conditions must be met for the action of allowing access to be taken:

- The subject making the request must have given consent for the purpose of "access"
- The resource has a valid access period
- The subject is the owner of the resource, identified by the "subject.id" parameter

- The action of "access" must be performed

If these conditions are met, the action of logging the access will be taken and then access to the resource is allowed.
It's important to note that this is a simplified example and in a real-world scenario, the actual implementation would likely include additional conditions and actions to ensure compliance with the GDPR, such as validating the authenticity of the subject's consent, checking for additional access rights, or sending a request to the data protection officer.

A.4.1 Specification of the policy

- The <odrl:Policy> element can be used to define the overall policy.
- The <odrl:permission> element can be used to specify the permissions that are granted by the policy.
- The <odrl:asset> element can be used to specify the resources that are covered by the policy. In the example policy, the asset is "personal-data"
- The $< odrl : constraint >$ element can be used to specify the conditions that must be met for the permission to be granted. In the example policy, the constraints are consent, $valid_access_period$, and owner.
- The $< odrl : action >$ element can be used to specify the actions that are covered by the policy. In the example policy, the action is "access".
- The $< odrl : duty >$ element can be used to specify actions that must be taken in addition to granting the permission. In the example policy, the duty is "log_access".

A.5 Comparing the policies

When it comes to implementing policies related to the GDPR's right to access [51], it's important to note that the example policies provided are simplified and do not cover all aspects of GDPR compliance. In a real-world scenario, the actual implementation would likely need to include additional elements, attributes, and conditions to ensure compliance with the GDPR. One key aspect of GDPR compliance is ensuring that the subject's consent is valid. The implementation should ensure that the consent was given freely, specifically for the purpose of access, and that the subject is aware of their rights under the GDPR [16]. This can be achieved by implementing a consent management system that captures and records consent, and by implementing a mechanism for revoking consent. Another important aspect is checking for additional access rights. The implementation should ensure that the subject has the right to access the data and has not exercised their right to erasure or their right to object to processing. Additionally, organizations should have a process in place for sending a request to the data protection officer (DPO) in case of uncertainty or disputes regarding data access rights, or if the subject has exercised their right to access and the organization is unable to comply with the request [52].

To ensure GDPR compliance, organizations should also conduct regular risk assessments and data protection impact assessments (DPIAs) to identify and mitigate any potential risks to the rights and freedoms of individuals. Organizations should also implement robust data protection and security measures to protect personal data from unauthorized access and breaches [50].
In conclusion, implementing GDPR right to access policies is a complex task that requires a combination of technical and legal expertise. Organizations should consider their specific requirements and constraints when choosing

a policy language, and should also take into account the expertise of their team. Regularly reviewing and updating policies and procedures, as well as conducting risk assessments and data protection impact assessments, can help organizations ensure compliance with the GDPR and protect the rights of their customers and users. It is always a good idea to consult the GDPR regulation and to take advice from legal experts or a Data Protection Officer (DPO) to ensure that your organization's policies and procedures are compliant with the GDPR [53].

Bibliography

[1] C. Zhang, Y. Xie, H. Bai, B. Yu, W. Li, and Y. Gao, "A survey on federated learning," *Knowledge-Based Systems*, vol. 216, p. 106 775, 2021.

[2] F. Alizadeh, T. Jakobi, J. Boldt, and G. Stevens, "Gdpr-reality check on the right to access data: Claiming and investigating personally identifiable data from companies," in *Proceedings of Mensch und Computer 2019*, 2019, pp. 811–814.

[3] B. Esteves and V. Rodrıguez-Doncel, "Analysis of ontologies and policy languages to represent information flows in gdpr," *Semantic Web*, no. Preprint, pp. 1–35, 2022.

[4] P. Jackson, "Understanding understanding and ambiguity in natural language," *Procedia Computer Science*, vol. 169, pp. 209–225, 2020.

[5] D. M. Berry, "Ambiguity in natural language requirements documents," in *Monterey Workshop*, Springer, 2007, pp. 1–7.

[6] R. A. Kadir, R. A. Yauri, and A. Azman, "Semantic ambiguous query formulation using statistical linguistics technique," *Malaysian Journal of Computer Science*, pp. 48–56, 2018.

[7] R. Krovetz and W. B. Croft, "Lexical ambiguity and information retrieval," *ACM Transactions on Information Systems (TOIS)*, vol. 10, no. 2, pp. 115–141, 1992.

[8] S. Y. Chen and B. van Tiel, "Every ambiguity isn't syntactic in nature: Testing the rational speech act model of scope ambiguity," in *Proceedings of the Society for Computation in Linguistics 2021*, 2021, pp. 254–263.

[9] J. Stetina and M. Nagao, "Corpus based pp attachment ambiguity resolution with a semantic dictionary," in *Fifth Workshop on Very Large Corpora*, 1997.

[10] S. Reisz, R. Duschinsky, and D. J. Siegel, "Disorganized attachment and defense: Exploring john bowlby's unpublished reflections," *Attachment & Human Development*, vol. 20, no. 2, pp. 107–134, 2018.

[11] A. Alabduljabbar, A. Abusnaina, Meteriz-Yildiran, and D. Mohaisen, "Tldr: Deep learning-based automated privacy policy annotation with key policy highlights," in *Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society*, 2021, pp. 103–118.

[12] S. Ezzini, S. Abualhaija, C. Arora, and M. Sabetzadeh, "Automated handling of anaphoric ambiguity in requirements: A multi-solution study," in *In Proceedings of the 44th International Conference on Software Engineering (ICSE'22), Pittsburgh, PA, USA 22-27 May 2022*, 2022.

[13] F. Macagno and S. Bigi, "Types of dialogue and pragmatic ambiguity," in *Argumentation and Language—Linguistic, Cognitive and Discursive Explorations*, Springer, 2018, pp. 191–218.

[14] A. Cohen *et al.*, "Why ambiguity?" *Between*, vol. 40, 2006.

[15] E. Kamsties and B. Peach, "Taming ambiguity in natural language requirements," in *Proceedings of the Thirteenth international conference on Software and Systems Engineering and Applications*, vol. 1315, 2000.

[16] P. Voigt and A. Von dem Bussche, "The eu general data protection regulation (gdpr)," *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, vol. 10, no. 3152676, pp. 10–5555, 2017.

[17] J.-M. Kim and H.-S. Chung, "Odrl ontology extention model and prototype design for the specification of the rights to use digital contents,"

Journal of Convergence for Information Technology, vol. 10, no. 1, pp. 13–21, 2020.

[18] M. M. Peixoto and C. Silva, "Specifying privacy requirements with goal-oriented modeling languages," in *Proceedings of the XXXII Brazilian symposium on software engineering*, 2018, pp. 112–121.

[19] N. Mousavi Nejad, P. Jabat, R. Nedelchev, S. Scerri, and D. Graux, "Establishing a strong baseline for privacy policy classification," in *IFIP International Conference on ICT Systems Security and Privacy Protection*, Springer, 2020, pp. 370–383.

[20] A. Goldsteen, G. Ezov, R. Shmelkin, M. Moffie, and A. Farkash, "Data minimization for gdpr compliance in machine learning models," *AI and Ethics*, vol. 2, no. 3, pp. 477–491, 2022.

[21] M. L. Rustad and T. H. Koenig, "Towards a global data privacy standard," *Fla. L. Rev.*, vol. 71, p. 365, 2019.

[22] H. Wang, L. Sun, and E. Bertino, "Building access control policy model for privacy preserving and testing policy conflicting problems," *Journal of Computer and System Sciences*, vol. 80, no. 8, pp. 1493–1503, 2014.

[23] R. Iannella, "The open digital rights language: Xml for digital rights management," *Information Security Technical Report*, vol. 9, no. 3, pp. 47–55, 2004.

[24] A. W. Chickering and S. C. Ehrmann, "Implementing the seven principles: Technology as lever," *AAHE bulletin*, vol. 49, pp. 3–6, 1996.

[25] H. Proper, S. Hoppenbrouwers, *et al.*, "Towards utility–based selection of architecture–modelling concepts," 2005.

[26] P. Kumaraguru, L. Cranor, J. Lobo, and S. Calo, "A survey of privacy policy languages," in *Workshop on Usable IT Security Management (USM*

07): Proceedings of the 3rd Symposium on Usable Privacy and Security, ACM, 2007.

[27] R. Garcıa and J. Delgado, "An ontological approach for the management of rights data dictionaries.," in *JURIX*, 2005, pp. 137–146.

[28] S. Steyskal and A. Polleres, "Defining expressive access policies for linked data using the odrl ontology 2.0," in *Proceedings of the 10th International Conference on Semantic Systems*, 2014, pp. 20–23.

[29] W. W. W. Consortium *et al.*, "The platform for privacy preferences 1.0 (p3p1. 0) specification," 2002.

[30] T. Athan, G. Governatori, M. Palmirani, A. Paschke, and A. Wyner, "Legalruleml: Design principles and foundations," in *Reasoning Web International Summer School*, Springer, 2015, pp. 151–188.

[31] O. Sacco and A. Passant, "A privacy preference ontology (ppo) for linked data," in *LDOW*, 2011.

[32] O. Sacco, J. G. Breslin, and S. Decker, "Fine-grained trust assertions for privacy management in the social semantic web," in *2013 12th IEEE international conference on trust, security and privacy in computing and communications*, IEEE, 2013, pp. 218–225.

[33] L. T. Van Binsbergen, L.-C. Liu, R. van Doesburg, and T. van Engers, "Eflint: A domain-specific language for executable norm specifications," in *Proceedings of the 19th ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences*, 2020, pp. 124–136.

[34] L. T. van Binsbergen, "Reflections on the design and application of eflint," 2022.

[35] D. Ferraiolo, R. Chandramouli, R. Kuhn, and V. Hu, "Extensible access control markup language (xacml) and next generation access control

(ngac)," in *Proceedings of the 2016 ACM International Workshop on Attribute Based Access Control*, 2016, pp. 13–24.

[36] A. Anderson, A. Nadalin, B. Parducci, *et al.*, "Extensible access control markup language (xacml) version 1.0," *Oasis*, 2003.

[37] Z. K. Confidential, "Privacy rights markup language specification (prml), technical report.," 2001.

[38] S. Hada and M. Kudo, "Xml access control language: Provisional authorization for xml documents," 2000.

[39] P. Ashley, S. Hada, G. Karjoth, C. Powers, and M. Schunter, "Enterprise privacy authorization language (epal)," *IBM Research*, vol. 30, p. 31, 2003.

[40] S. Kasem-Madani and M. Meier, "Security and privacy policy languages: A survey, categorization and gap identification," *arXiv preprint arXiv:1512.00201*, 2015.

[41] J. Leicht and M. Heisel, "A survey on privacy policy languages: Expressiveness concerning data protection regulations," in *2019 12th CMI Conference on Cybersecurity and Privacy (CMI)*, IEEE, 2019, pp. 1–6.

[42] B. Esteves, V. Rodrıguez-Doncel, H. J. Pandit, N. Mondada, and P. McBennett, "Using the odrl profile for access control for solid pod resource governance," in *European Semantic Web Conference*, Springer, 2022, pp. 16–20.

[43] R. Ianella, "Open digital rights language (odrl)," *Open Content Licensing: Cultivating the Creative Commons*, 2007.

[44] B. Esteves, H. J. Pandit, and V. Rodrıguez-Doncel, "Odrl profile for expressing consent through granular access control policies in solid," in *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, IEEE, 2021, pp. 298–306.

[45] M. G. Kebede, G. Sileno, and T. V. Engers, "A critical reflection on odrl," in *AI Approaches to the Complexity of Legal Systems XI-XII*, Springer, 2020, pp. 48–61.

[46] H.-P. Lam and M. Hashmi, "Enabling reasoning with legalruleml," *Theory and Practice of Logic Programming*, vol. 19, no. 1, pp. 1–26, 2019.

[47] M. D. Vos, S. Kirrane, J. Padget, and K. Satoh, "Odrl policy modelling and compliance checking," in *International Joint Conference on Rules and Reasoning*, Springer, 2019, pp. 36–51.

[48] N. Fornara and M. Colombetti, "Operational semantics of an extension of odrl able to express obligations," in *Multi-Agent Systems and Agreement Technologies*, Springer, 2017, pp. 172–186.

[49] N. Fornara and M. Colombetti, "Using semantic web technologies and production rules for reasoning on obligations, permissions, and prohibitions," *AI Communications*, vol. 32, no. 4, pp. 319–334, 2019.

[50] G. A. Teixeira, M. M. da Silva, and R. Pereira, "The critical success factors of gdpr implementation: A systematic literature review," *Digital Policy, Regulation and Governance*, vol. 21, no. 4, pp. 402–418, 2019.

[51] I. G. P. Team, *EU general data protection regulation (gdpr)–an implementation and compliance guide*. IT Governance Ltd, 2020.

[52] F. Bieker, M. Friedewald, M. Hansen, H. Obersteller, and M. Rost, "A process for data protection impact assessment under the european general data protection regulation," in *Annual Privacy Forum*, Springer, 2016, pp. 21–37.

[53] P. Lambert, *The Data Protection Officer: Profession, Rules, and Role*. CRC Press, 2016.