Pontifícia Universidade Católica
do Rio de Janeiro

Paulo Henrique Cardoso Alves

Enabling Data Regulation Evaluation through Intelligent and Normative Multiagent Systems Design

Tese de Doutorado

Thesis presented to the Programa de Pós–graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Ciências – Informática.

Advisor : Prof. Hélio Côrtes Vieira Lopes
Co-advisor: Prof^{a}. Clarisse Sieckenius de Souza

Rio de Janeiro
September 2023

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Paulo Henrique Cardoso Alves

Enabling Data Regulation Evaluation through Intelligent and Normative Multiagent Systems Design

Thesis presented to the Programa de Pós–graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Doutor em Ciências – Informática. Approved by the Examination Committee.

Prof. Hélio Côrtes Vieira Lopes
Advisor
Departamento de Informática – PUC-Rio

Profª. Clarisse Sieckenius de Souza
Co-advisor
Departamento de Informática – PUC-Rio

Prof. Bruno Feijó
Departamento de Informática – PUC-Rio

Profª. Simone Diniz Junqueira Barbosa
Departamento de Informática – PUC-Rio

Dr. Guilherme da Franca Couto Fernandes de Almeida
Insper Instituto de Ensino e Pesquisa

Drª. Flavia Maria Santoro
UERJ

Dr. Renato Fontoura de Gusmão Cerqueira
IBM Research

Dr. Fernando Alberto Correia dos Santos Junior
Departamento de Informática – PUC-Rio

Rio de Janeiro, September 29th, 2023

Paulo Henrique Cardoso Alves

Bachelor's in Information Systems (2014) at Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio). Master's in Informatics (2017) at PUC-Rio. Started his doctorate at PUC-Rio in 2018, focusing his research on Ontologies, Negotiation Scenarios, and Intelligent Normative Multiagent Systems applied to Data Protection Regulations.

Bibliographic data

Alves, Paulo Henrique Cardoso

Enabling Data Regulation Evaluation through Intelligent and Normative Multiagent Systems Design / Paulo Henrique Cardoso Alves; advisor: Hélio Côrtes Vieira Lopes; co-advisor: Clarisse Sieckenius de Souza. – Rio de Janeiro: PUC-Rio, Departamento de Informática, 2023.

v., 129 f: il. color. ; 30 cm

Tese (doutorado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática.

Inclui bibliografia

1. Regulação de Dados. 2. Ontologia. 3. Framework. 4. Inteligencia Artificial. 5. Sistemas Multiagentes BDI e Normativos. 6. Assimetria Informacional de Fluxo de Dados. I. Lopes, Hélio Côrtes Vieira. II. de Souza, Clarisse Sieckenius. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. IV. Título.

CDD: 004

Acknowledgments

I would like to express my sincere gratitude to my wife Elen and my children Miguel and Inácio, for their unwavering support and understanding during the most challenging moments of this journey. None of this would have been possible without their love. To my family, I love you.

I am deeply grateful to my parents Raul and Cristina, who have always emphasized that education is the key to success and the foundation for striving toward a better and more just society.

I extend my heartfelt appreciation to my friends Fernando and Isabella, who have been with me throughout my doctoral journey. They are more than just colleagues I met during my master's and doctoral studies; they are individuals who have supported me in various aspects of life, thank you.

I would like to express my gratitude to my advisor Professor Helio and my co-advisor Professor Clarisse for agreeing to guide this research. Countless discussions, guidance, and debates have helped me evolve and progress with the research. With each conversation, we discovered new paths to explore, and your expertise was crucial in guiding us along the best course.

I would also like to thank Professor Marcelo La Rosa for the insightful discussions and connections he facilitated with other professionals during my stay in Australia. In the same vein, I extend my gratitude to my fellow Brazilian colleagues who participated in this exchange program. Lastly, I would like to thank Professor Helio once again for providing contact with Professor Marcelo to arrange my visit to the University of Melbourne. I am also grateful to Nancy, the coordinator of the exchange program at PUC-Rio, for her assistance and encouragement throughout the process, enabling me to participate in the exchange.

I would like to express my appreciation to Professors Gustavo Robichez, Rafael Nasser, and all other members of ECOA Institute for their support and encouragement in both academic and professional development.

I would like to thank Professor Carlos Jose Pereira de Lucena and Marx Leles Vianna for the invaluable knowledge I gained during my master's studies at the Laboratory of Software Engineering, which I have applied and expanded upon in my doctoral research to reach the cutting edge of the field.

Lastly, I would like not only to thank PUC-Rio for funding support but also to express my appreciation to the dedicated staff and coordinators at the PUC-Rio Department of Informatics for their exceptional work in supporting and guiding the students.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Abstract

Alves, Paulo Henrique Cardoso; Lopes, Hélio Côrtes Vieira (Advisor); de Souza, Clarisse Sieckenius (Co-Advisor). **Enabling Data Regulation Evaluation through Intelligent and Normative Multiagent Systems Design** . Rio de Janeiro, 2023. 129p. Tese de doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Sharing and managing personal data are challenging due to the massive amount of data generated, uploaded, and digitalized, informed by data subjects to utilize services, online or not. This challenge disrespects not only the data subjects, but also data controllers and processors, which are responsible for security, privacy, anonymity, and data usage under the legal basis applied and the initial purpose when the data were required. In this scenario, data protection and regulation take place to organize this environment proposing rights and duties to the involved agents. However, each country is free to create and employ its data regulation, e.g., GDPR in European Union and LGPD in Brazil. Therefore, although the goal is to protect the data subjects, the regulations can present different rules based on their jurisdiction. In this scenario, ontologies emerge to identify the entities and relationships to show them at a high abstraction level, facilitating ontology alignment with different regulations. To do so, we developed a metamodel based on GDPR ontologies to enable the LGPD representation focused on the consent legal basis. Moreover, we proposed GoDReP (Generation of Data Regulation Plots) to allow actors to represent their law's interpretation in a specific application scenario. As a result, we set three scenarios to exercise the GoDReP application. Moreover, in this thesis, we also propose an intelligent normative multiagent system architecture (RegulAI) to represent the personal data regulation rights and obligations, as well as the agent's decision-making process. Finally, we developed a use case applying RegulAI in the open banking scenario.

Keywords

Data Regulation; Ontology; Framework; Artificial Intelligence; Normative and BDI Multiagent Systems; Data Flow Information Asymmetry.

Resumo

Alves, Paulo Henrique Cardoso; Lopes, Hélio Côrtes Vieira; de Souza, Clarisse Sieckenius. **Permitindo a Simulação de Cenários na Regulação de Dados através da Aplicação de Sistemas Multiagentes Inteligentes e Normativos**. Rio de Janeiro, 2023. 129p. Tese de Doutorado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

O compartilhamento e o gerenciamento de dados pessoais são atividades desafiadoras devido à grande quantidade de dados gerados, carregados e digitalizados por cidadãos para utilizar serviços, online ou não. Esse desafio afeta não apenas os cidadãos, mas também os controladores e processadores de dados, que são responsáveis pela segurança, privacidade, anonimato e uso de dados fundados em bases legais e no propósito inicial quando os dados foram solicitados. Nesse cenário, a proteção e regulamentação de dados entram em cena para organizar esse ambiente, propondo direitos e deveres aos agentes envolvidos. No entanto, cada país é livre para criar e empregar sua própria regulamentação de dados, como o GDPR na União Europeia e a LGPD no Brasil. Portanto, embora o objetivo seja proteger os cidadãos, as regulamentações podem apresentar regras diferentes com base em sua jurisdição. Nesse cenário, as ontologias surgem para identificar as entidades e relacionamentos e mostrá-los em um nível de abstração elevado, facilitando o alinhamento das ontologias com diferentes regulamentações. Para isso, desenvolvemos um metamodelo baseado em ontologias da GDPR para possibilitar a representação da LGPD com foco na base legal do consentimento. Além disso, propusemos o GoDReP (Geração de Cenários de Regulamentação de Dados) para permitir que os atores representem a interpretação de sua legislação em um cenário de aplicação específico. Apresentamos então três cenários diferentes para exercitar a aplicação do GoDReP. Além disso, nesta tese, também propomos uma arquitetura de sistema multiagente normativo e inteligente (RegulAI) para representar os direitos e obrigações apresentados pela regulamentação de dados pessoais, bem como o processo de tomada de decisão dos agentes. Por fim, desenvolvemos um estudo de caso aplicando o RegulAI no cenário de open banking.

Palavras-chave

Regulação de Dados; Ontologia; Framework; Inteligencia Artificial; Sistemas Multiagentes BDI e Normativos; Assimetria Informacional de Fluxo de Dados.

Table of Contents

List of Figures

List of Tables

1
Introduction

A data protection regulation is crucial to define correct behavior in sharing and managing personal data. Due to widespread goods and services connected to the internet, the massive collection of personal data turns the discussion of regulating personal data into a high-priority item (Mulholland and Frajhof, 2020). It impacts not only the data subjects (DS), *i.e.*, users who might be represented by many synonyms such as user, client, student, patient, and many others, but also the data controllers (DCs) and processors (DPs), *i.e.*, software warehouses, e-commerce platforms, financial institutions, universities, hospitals, among others. Data protection legislation aims to create a structure to regulate the processing of personal data, thus, establishing obligations to DCs and DPs, and rights to DSs. Even though each country has its own jurisdiction to propose its own legislation about this subject, there can be differences and similarities between the norms of different countries.

For instance, the General Data Protection Regulation (GDPR) of the Europe Union was enacted in 2016, but only entered in force in May 2018. This piece of legislation, one of the most important and prominent in the world, comprises ninety-nine articles divided into eleven chapters indicating the DS's rights and the DC and DP duties regarding data processing, management, and deletion. In Brazil, the Brazilian General Data Protection Law (LGPD or *Lei Geral de Proteção de Dados Pessoais*) was enacted in 2018, becoming partially effective in August of 2020[1] (Law 13.709/2018[2]).

The LGPD imposes that whenever personal data are processed, DCs and DPs must observe the law's command, such as its principles, processors and controllers' duties, individual rights, etc. From the controller's perspective, attending to such norms can be defying, as it demands a detailed and holistic knowledge of the data collecting, storing, and processing activities.

From the DS point of view, controlling and following the data flow is also complex, as many entities can be authorized to access and use one's personal data. Thus, the data subject should be able to know if: the informed purpose

[1] The articles referring to penalties that can be applied in case of violation of the LGPD only entered in force in August of 2021.

[2] Lei Geral de Proteção de Dados Pessoais - LGPD - `http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm`

of data use is being observed; the personal data has been shared with other non-informed partners; and once the purpose of the data controller is reached, if the data is still being used, or shared, with a different purpose.

The LGPD is not as detailed as GDPR, and delegated normative competence to the Data Protection National Agency (ANPD, or *Autoridade Nacional de Proteção de Dados*), which will be responsible for defining specificities of relevant legal provision. On the other hand, the GDPR defines rights and duties in more detail. Moreover, LGPD and GDPR established different legal bases in which the DC must justify the processing of personal data, and the consent is one of them. The consent is a legal bases foreseen in both regulations that require the DS interaction. However, this interaction may generate doubts and questions about interpreting information about the data processing activity presented to DS. When consent is used as a legal base, such information is commonly presented in a legal document named Privacy Policy.

There are two important definitions related to privacy policy and consent term. A Privacy Policy is a legal document or statement that outlines how an organization collects, uses, processes, stores, and shares the personal information of its users or customers. On the other hand, a consent term refers to the explicit permission granted by DSs to an organization, allowing the organization to process their personal data for specific purposes. Thus, a privacy policy is a comprehensive document that informs users about a DC's data practices, while a consent term is a specific agreement through which DSs explicitly grant permission for the organization to process their personal data for certain purposes. For example, social networks often provide a Privacy Policy outlining how DSs' data will be utilized and require consent from them to proceed. Another example is in scientific research, where researchers present a specific consent term directed to the DSs. In this thesis, we will use the "Consent Term" in a generic form, encompassing these two definitions.

Thus, when information about the data processing activities are not detailed, explained, direct, and clear, DS may not be able to comprehend what the processing activity encompasses. Moreover, when the processing of personal data takes place, DC must disclose information regarding the purpose, time range of their activity, the DSs, DCs, and DPs identification. Thus, there is a need to present a rich explanation to justify the actions performed in a specific scenario. Last but not least, the DS must actively and expressly give consent; other legal bases do not present such a requirement. We chose consent as the object of this thesis for its interpretability challenges and for the informational asymmetry between DSs, DCs, and DPs that may occurs during negotiation scenarios, where the DS must carefully weigh the advantages and disadvantages of data sharing.

1.1 Research Goals

Ontologies emerge as a semantic proposal for understanding the data protection entities and their relationships regarding the GDPR and the LGPD. To the best of our knowledge, there are two ontologies published regarding the GPDR: PrOnto and GConsent. The former is an ontology proposed by Palmirani et al. (2018) focused on data privacy, and the latter is an ontology proposed by Pandit et al. (2019) focused on the consent term required to allow the data controller to get the data subject's data. Moreover, as GDPR and LGPD have many aspects in common, we made a PrOnto and GConsent extension to create a version considering the particularities of LGPD in regards to the consent term, generating the Ontology for Data Privacy Management (ODPM) Alves et al. (2021). We proposed the Consent Metamodel (CM) based on those three ontologies, gathering the consent definitions and relationships with other entities.

Ontologies are a high abstraction level conceptual model that shows the entities and their relationships; however, there is a gap between the ontology and each application scenario's interpretation. Many studies have presented formal methods to validate ontologies; however, the gap to final users may remain present (Gangemi et al., 2006; Tartir, Arpinar, and Sheth, 2010; Li, Yang, and Ramani, 2009; García-Peñalvo et al., 2012; Dragisic et al., 2016). Therefore, we propose GoDReP (Generation of Data Regulation Plots) to allow the generation of use case scenarios, *i.e.*, pragmatic circumscriptions, to explore the semantic usage.

According to Varici (2013), information asymmetry occurs when one side of the negotiation table has more or better information than the other, which may generate a hazardous environment. In this sense, GoDReP enables CM employment to mitigate the data flow informational asymmetry in an open and live book to record examples of the data regulation behavior. Furthermore, these use case scenarios allow the identification of general attributes and the exploration of the scenario's specificities.

In this context, this thesis proposes three scenarios to explore the developed semantics and the pragmatics following the GoDReP structure. This structure comprises generic scenes and negotiation scenarios to guide the user in building the scenario environment. Hence, constructing the scenarios' repository based on the same semantics could aid DSs, DCs, and DPs in aligning their expectations, rights, and duties regarding the law interpretation, and implementation possibilities. Moreover, community engagement is also crucial to provide "*information delivery, consultation, collaboration in decision-making, empowering action in informal groups or formal partnerships, healthcare delivery and promotion, interaction with various stakeholders*" (Musesengwa, 2017).

Furthermore, this thesis also proposes RegulAI (Artificial Intelligence approach for Data Regulation). This approach aims to apply artificial intelligence techniques to represent the data regulation rights and obligations as well as the agent's decision-making process based on CM and GoDReP specifications. RegulAI employs Normative Multiagent System (NMAS) concepts to represent data regulation constraints and the BDI (Belief-Desire-Intention) reasoning to express the data agent's preferences. Moreover, RegulAI proposes a BDI decision-making process to enable agents to decide whether to comply based on their BDI preferences.

Therefore, based on the challenges mentioned above to mitigate the informational asymmetry between DSs, DCs, and DPs, this thesis presents the following research questions that guided this study:

> Research Question 1
>
> [RQ1] How can data protection regulations be represented?

This research question aims to understand how data protection regulations used to be represented. This RQ presents two assumptions:

[Assumption 1.a] There are representation models that enable data regulation exploration.

[Assumption 1.b] Normative multiagent systems enable modeling scenarios to represent data regulation.

> Research Question 2
>
> [RQ2] What are the general attributes that can be applied across multiple domains?

This research question aims to identify the attributes that are unique and related to the application domain and those that are not, *i.e.*, the attributes that could be reused in most domains.

[Assumption 2] Timing is an attribute that is vital in every domain. Depending on the action's sequence, the interpretation of the validity of such action can be different. For instance, if a DC starts collecting the DS's data before getting the DS's consent, it will be considered a law violation; conversely, if the data collection begins after the informed consent, this action will comply with the law.

Research Question 3

[RQ3] How can data flow information asymmetry be mitigated in a scenario governed by data protection regulations?

This research question investigates how to mitigate the data flow information asymmetry in a data-regulated scenario. This thesis explores mainly the LGPD jurisdiction.

[Assumption 3] A framework development based on an ontology can guide the actions to mitigate this informational asymmetry.

Therefore, in order to drive the presented research questions to an answer, we propose the following contributions: (i) the consent metamodel (CM) based on the literature to aid agents in identifying their major concerns when sharing personal data to satisfy RQ1, (ii) a structure to build use case scenarios in the personal data regulation context to satisfy RQ1 and RQ2, and (iii) an intelligent normative multiagent system architecture to represent the personal data regulation rights and obligations, as well as the agent's decision-making process to satisfy RQ3.

It must be noted that, although the LGPD mentions data minimization and GDPR mentions cryptography algorithms as principles, this thesis does not explore such themes. Moreover, this is not a proposal to automate the law decisions, but an exploration of the multiple interpretations of law based on semantics depending on the application scenario.

1.2 Published Contributions

This thesis gathers the accomplishments conquered during this Ph.D. research. During the research development, we had four academic publications. These publications showcase the contributions in the field of data protection regulation, privacy concerns, and the utilization of innovative technologies like multiagent systems. By addressing the challenges posed by the COVID-19 pandemic and considering the complexities of personal data processing, our work offers valuable insights and solutions to ensure transparency and compliance with data protection laws.

The first publication, titled "Permissioned blockchains: Towards privacy management and data regulation compliance" (Alves et al., 2020), addresses the challenges of data privacy and protection during the COVID-19 pandemic. The paper introduces a data governance model based on the principles of the Governance Analytical Framework. The model focuses on permissioned blockchain technology, providing users with control over their data transparently and securely. It establishes relationships between DSs, DCs, and DPs, ensuring compliance with privacy concerns and the Brazilian General Data Protection Law (*i.e.*, LGPD). The feedback received during the review process led us to redirect our efforts from the governance aspects to a more conceptual model, e.g., to an ontology construction.

In the paper titled “Controlling Personal Data Flow: An Ontology in the COVID-19 Outbreak Using a Permissioned Blockchain” (Alves et al., 2021), we tackle the complexities of complying with data protection regulations. We develop an ontology to identify and establish relationships between entities involved in personal data processing. The ontology aims to foster a common understanding of rights and duties proposed by the Brazilian Data Protection Law within the context of the COVID-19 pandemic. The study also explores permissioned blockchain technology as a solution to manage privacy concerns and enable compliance with the law. Additionally, a conceptual model and data governance approach are presented to enhance the accuracy of data reuse. However, the review process and the feedback during the paper presentation showed that there was a gap between the ontology model and the blockchain discussion. At this time, we focused on providing a technical solution instead of filling this gap.

To bridge this gap, we decided to redirect our efforts from the blockchain perspective to modeling the data regulation environment using NMAS. In the paper titled “A Normative Multiagent Approach to Represent Data Regulation Concern” (Alves et al., 2023a), we recognize the challenges in modeling systems that comply with data protection regulations and propose the use of Multiagent Systems (MAS) combined with Normative MAS. Also, we introduce the DR-NMAS (Data Regulation by NMAS) framework, employing Adaptative Normative Agent - Modeling Language (ANA-ML) and a Normative Agent Java Simulation (JSAN) extension. The paper presents a use case scenario in the Open Banking domain to demonstrate the applicability of the proposed extensions. However, at this point, we still with a gap between the literature ontologies and the NMAS.

The fourth publication, “Designing Intelligent Agents in Normative Systems Toward Data Regulation Representation” (Alves et al., 2023b) delve into the asymmetry of data flow information and its implications for personal data protection. We proposed a consent metamodel, a structure for building use case scenarios, and an intelligent normative multiagent system architecture. These elements enable agents to identify their concerns when sharing personal data, represent rights and obligations outlined by data protection regulations, and make decisions based on their goals and normative rewards and punishments. A use case in the open banking scenario illustrates the capabilities of this system.

Thus, this thesis encompasses the key findings of the research on personal data, data regulation, consent legal bases, and intelligent normative multiagent systems. Furthermore, it considers the valuable feedback received during the review process and the insights gained from participating in conferences, which led to a course adjustment in the research direction.

1.3
Thesis Structure

The remainder of this thesis is organized as follows. The theoretical background is detailed in chapter 2. In chapter 3, we present the related work found in the literature and discuss its contribution to our research. The ontology extension (ODPM) is detailed in chapter 4. In chapter 5, we describe CM and GoDReP concepts. The negotiation scenarios are explained in chapter 6. In chapter 7, we introduce our BDI Normative MAS framework (RegulAI) and presents the open banking use case. In chapter 8, we discuss in regards to outcomes of this research. Finally, we present our conclusion and future work in chapter 9.

2
Background

This chapter aims to clarify aspects regarding the major concepts that will be approached in this thesis. It presents the definition of data protection pegulation, ontology, and BDI Normative Multiagent Systems concepts.

2.1
Data Protection Regulation

The digital transformation movement involves more than digitalized paper-based process. When digital transformation takes place it is an opportunity to rebuild organizational processes, which shall consider the technology state of the art to deliver efficiency and effectiveness (Heavin and Power, 2022; Kraus et al., 2022). Furthermore, digital transformation is strongly related to the current state of an informational and connected society. Many services have migrated to the digital ambience in order to expand their business; others were forced to go digital at the end of 2019 due to the COVID-19 pandemic outbreak (Priyono et al., 2022). Business meetings, medical appointments, and food delivery are examples of activities that had to be changed to the digital sphere. Thus, technology was responsible for mediating and supporting these social relations. Consequently, this interaction means an intense data flow in which data is massively shared, more than ever.

The need to protect data flow is crucial, especially when it involves personal data. In this sense, data protection regulation is essential, since data can be used for several purposes and by different agents (i.e., the State or private entities), which have distinct interests in the use of data. Aware of the importance of guaranteeing the protection of personal data, many countries have issued strong regulations aiming to aid people to protect their personal data and avoid data misuse. For example, the European Union issued, in 2016, the General Data Protection Regulation (GDPR), which entered into force in 2018. The GDPR is one of the most important regulations on the subject, and it inspired other countries to enact data protection regulations, such as Brazil. Furthermore, other countries have enacted their own data protection legislation, such as Australia, in which citizens are supported by the Australian Private Act; Canada, people can rely on PIPEDA, and in Brazil, the population can appeal to the Brazilian General Data Protection Law (LGPD), Federal Law no. 13,709/2018.

However, the mere existence of a valid and effective regulation does not prevent data breaches, or abusive and illegitimate data uses. Citizens must know their rights and understand how to enforce them. Meanwhile, citizens, companies, and governments are performing activities related to personal data, they must comply with the current regulation. It implies understanding their rights and duties, planning their actions, and thinking consequences of their acts in advance. Citizens' knowledge regarding their rights is the first step towards autonomy and a better society. Moreover, this knowledge is essential not only for people that have shared their data but also for organizations that receive such data. Those companies must comply with the data regulation from the jurisdiction where they exercise their activities. Law infringement may imply significant financial losses, administrative and judicial processes, as well as damage to reputation.

Moreover, in personal data regulation, consent is one of the most commonly used legal bases for processing personal data. It allows individuals to control how their data are used and sets the purposes for which their data will be used. Obtaining consent can be critical in situations where the processing of personal data may be considered sensitive, such as health data or data related to a person's sexual orientation or religious beliefs.

Sommers (2020) proposed experiments to understand what is considered valid consent. There are two major lines of thought: (i) regarding fraud on the inducement and (ii) regarding fraud *in the factum*. Briefly, fraud *in the factum* involves misleading someone about the very nature of an action, while fraud in the inducement involves deceiving someone about the motives or reasons behind the action. The experiment conducted by Sommers shows that the interpretation of a valid consent can differ depending on the application domain and the scenario. It is an example that demonstrates the complexity of the discussion around valid consent. Moreover, Demaree-Cotton and Sommers (2022) argue about the validity of consent, considering that, even in cases where a person gives his or her consent, this person might not be able to reason about the terms on the first hand for different reasons, *e.g.*, reduced cognitive ability.

According to LGPD, consent will be valid when freely given, specific, informed, and unambiguous. This means that the individual must clearly understand what they are agreeing to and must not feel pressured or coerced into giving their consent. It is also important that the individual has the option to withdraw their consent at any time. It is the DC's onus (the person or organization collecting and using the personal data) to ensure that they have obtained valid consent from the individual before processing their personal data.

It must be remarked that the Brazilian data protection regulation establishes that individual consent is only one of the legal bases authorizing data processing. In any case, data controllers must abide by the law's principles, rights, and safeguards and act in good faith. Complying with such norms can be a challenge.

Therefore, ontology construction is vital to aid people and organizations in building a more secure, transparent, informed, and fairer environment. Moreover, these agents should have manners to test, explain, and simulate the understanding of data regulation law in certain situations. This will help data processors mitigate the data flow informational asymmetry, and can be used as a compliance tool.

DSs, or individuals whose personal information is collected and processed by organizations, may have many concerns about their data. Some common concerns include:

(i) Privacy: DSs may be concerned about their privacy and an unauthorized disclosure of their personal data. They may worry about who has access to their data and how it is being used;

(ii) Security: DSs may be concerned about the security of their personal data and the potential for it to be stolen or misused. This can include worries about data breaches and cyber-attacks;

(iii) Control: DSs may be concerned about having control over their own data and the ability to access, correct, or delete it if they wish;

(iv) Fairness: DSs may be concerned about whether the collection and use of their personal information are fair and justified and whether they are being treated equitably, and

(v) Transparency: DSs may be concerned about whether they are being informed about how their data is being collected and used, and whether they are being given sufficient information to make informed decisions.

Overall, DSs may have a wide range of concerns related to their personal information and how organizations are handling it.

Thus, providing clear, straightforward, and complete information in a consent term to guarantee the DS's understanding can be challenging for DCs. Moreover, DSs are responsible for authorizing the use of their data, and evaluating all information regarding data processing can be hard for DSs without legal knowledge. It must be noted that legal knowledge must not be required to give consent. Therefore, the consent term must give specific, straight, and unambiguous information to facilitate the DS's comprehension.

2.2
Ontology Theory

This thesis proposes a consent metamodel based on consent ontologies found in the literature to mitigate the informational asymmetry between data agents. Ontologies are representations of a specific domain that aims to create a shareable and reusable model. They are also considered a valuable instrument for reducing conceptual ambiguities and inconsistencies in a specific domain (Staab and Studer, 2010).

Aristotle defined Ontology as the science of "being *qua* being", *i.e.*, the study of attributes that belong to things because of their very nature. Ontology, different from the experimental sciences, which aim at discovering and modeling reality under a particular perspective, focuses on the nature and structure of things (Guarino, Orbele, and Staab, 2009). Moreover, Smith (2012) defines:

> "*Ontology as a branch of philosophy is the science of what is, of the kinds and structures of objects, properties, events, processes, and relations in every area of reality.*"

Still, ontologies are a key factor for elaborating high-quality requirement models for domain exploration (Gharib and Mylopoulos, 2018). A privacy ontology provides developers, users, and service providers an overview of the major entities and their relationships, revealing the actions from a privacy-based perspective. Moreover, an ontology goal is supporting the end-users' decision-making process to evaluate the privacy concerns and requirements for each situation.

Furthermore, as a theory of objects and their ties, an ontology should provide criteria for distinguishing different object types and their connections. According to Corazzon Corazzon (2014), the ontologies can be distinguished into three categories: (i) formal, *i.e.*, the study of the mathematical method of symbolic logic; (ii) descriptive, i.e., the study towards capturing the entities and relationship underlying natural language and human common sense, and (iii) formalized ontologies, *i.e.*, the study that aims to construct a formal codification for the results descriptively. Thus, this thesis aims to address the evaluation of (ii) *descriptive* and (iii) *formalized* ontologies.

2.3 Normative Multiagent Systems

Multiagent Systems (MAS) are distributed computing systems composed of intelligent and autonomous agents able to interact with each other in collaboration to achieve a specific goal in a non-supervised environment without human intervention (Wooldridge, 1999). These agents can take reactive actions, *i.e.*, reactions triggered by other agents' actions or environmental changes.

In a Normative MAS (NMAS), a set of norms defines the environmental boundaries regarding the expected agent's behavior, as well as in the current society, where laws and regulations rule citizens' rights and duties. Regarding citizens' rights, in most jurisdictions, there are sets of laws and regulations to ensure citizens' rights against scenarios of abuses, whether from other people, organizations, or the government itself. In this sense, in NMAS, norms emerge to orchestrate agents' environment without disturbing the agent's autonomous capabilities. The NMAS elements are:

- **Environment**. It is responsible for supplying data to agents to update their beliefs and norms database.
- **Agent**. An agent is composed of its roles and goals.
- **Agent's Role**. It describes the agent's role in the environment.
- **Organization**. It specifies agents into groups and roles.
- **Norm**. It is composed by its *activation, expiration, deontic concept state, rewards* and *punishments* values and specifies to which agent's role this norm is *addressed*.

Moreover, norms can be beneficial or harmful, depending on their alignment with the agent's programmed goals. Therefore, agents must be able to reason about the rewards and punishments defined in an active norm addressed to its role to decide which they should comply with and occasionally violate if it conflicts with other norms or with the agent's private goals (Luck et al., 2013; Alves et al., 2018).

BDI architecture is a model to enable agents to decide how to accomplish their goals and which norms to comply with or violate (Wooldridge, 1999). Figure 2.1 presents the agent's reasoning process. This process starts with the agent's environmental perception, *i.e.*, the environment's sensor updates the environmental attributes and enables the agent to update its beliefs database. Then, based on its beliefs, the agent generates and stores its desires in the desires database. Next, the agent filters its beliefs, desires, and intentions, selecting the actions the agent can perform to achieve its goals.

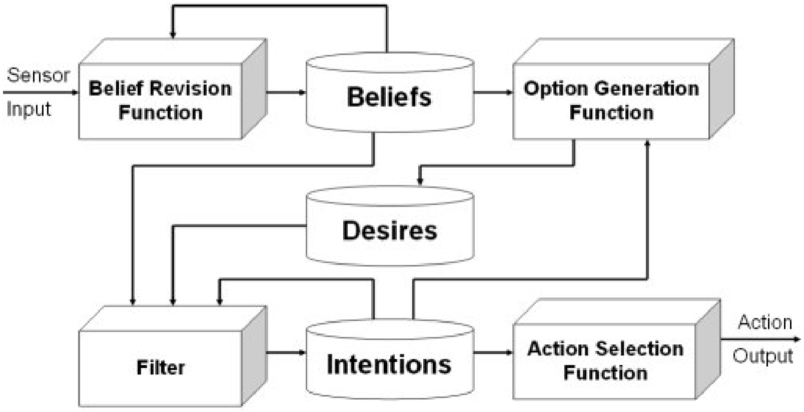


Figure 2.1: BDI architecture (Wooldridge, 1999).

BDI reasoning architecture can complement NMASs since an agent should deliberate whether to comply with norms based on environmental perception and its goals (Neto et al., 2013). The combination of BDI architecture and NMAS allows the representation of the agent's reasoning in a normative data-regulated environment. Thus, NMASs can monitor and automate aspects of data regulation, such as reporting data breaches and sharing data between different organizations. For instance, a DC agent should respect the environmental norms, *e.g.*, GDPR or LGPD, whereas a DS agent reasons regarding its beliefs, desires, intentions, and goals to decide whether share its personal data.

3
Related work

This chapter gathers the existing work found in the literature regarding the topics related directly or indirectly to this thesis. In this sense, this chapter will present studies on data privacy ontologies and data privacy related to our application domains, *i.e.*, studies related to healthcare, education, and open banking.

3.1
Data Privacy Ontology

As mentioned, an ontology is vital to defining high-quality requirement models according to the domain area. In this sense, a privacy ontology can provide to DSs, DCs, and DPs the entities and their ties related to data privacy regulation. Moreover, an ontology may enable the evaluation of differences among the data regulations worldwide, or at least it might mitigate the effort to compare them, providing a structured concept and relationship mapping. Last but not least, a data privacy ontology can be applied in many different domains; we selected three to detail the studies related to this subject in the next section.

Collierf et al. (2010) proposed the descriptive ontology named BioCaster to standardize terms, such as diseases, agents, and symptoms in different languages. Even though such a standard does not support any privacy, data protection, or consent concerns, this work is relevant for its ontology proposal towards the diseases and agents relationships mapping. However, considering data regulation norms, data privacy, protection, and consent are critical in such a domain. Hence, an ontology for privacy, data protection, and consent management is crucial to complement the ontology.

The authors in He et al. (2014) follow a similar thought as presented in Collierf et al. (2010). He et al. presented the biomedical Ontology of Adverse Events (OAE) to propose integration and standards to manage such events. This descriptive ontology defines and classifies adverse events after medical interventions. However, the authors did not present concerns regarding the personal data regulations that might be applied. In this sense, we developed an ontology extension focused on data privacy, data protection, and consent management, pillars of the four data regulations we evaluated.

Based on the GDPR, Fatema et al. (2017) presented an open vocabulary of expressing consent leveraging existing semantic models of provenance, processes, permission, and obligations. Still, they presented a reference architecture for data processing management based on the consent permission in the GDPR context. However, this work highly depends on the application and the adopted use case scenario. In this sense, our work proposes three different use case scenarios to evaluate the ontology particularities and generalizations that could be made.

Still, Mense and Blobel. (2017) proposed standards and components to support the GDPR implementation in the health domain. The authors focused on the companies that provide eHealth services to apply the HL7 (Health Level 7) standards to support security and privacy in handling personal healthcare data. HL7 is a framework that offers a set of standards related to integration, management, exchange, and retrieval of electronic information in healthcare systems. However, the authors did not evaluate the DS's views and how they would be affected. Conversely, GoDReP proposes CM and a framework to allow the stakeholders, *i.e.*, hospitals and patients, to evaluate the scenario possibilities by trying to simulate the expected behaviors from the data controller and the data subject view.

The authors in Kirrane et al. (2018) proposed the SPECIAL system to enable DCs and DPs to comply with consent and transparency obligations in the Europe Union jurisdiction, *i.e.*, under the GDPR. Still, this approach aims to support DSs to control their personal data. To do so, the authors proposed an architecture to evaluate compliance checking based on the log generated by the system application. However, there are no details regarding which ontology was used in the compliance checker module. Even though the use of the application log to verify GDPR compliance is interesting, this work lacks information on the entities evaluated in the log and how it can aid DSs, DCs, and DPs pragmatically. Thus, GoDReP applies the log generation concept to build an initial explanation database, *i.e.*, where the agents' actions are recorded and it can be used to clarify the action motivation.

The authors in Palmirani et al. (2018) presented a GDPR-based formalized ontology focused on data privacy. They defined five main modules: (i) data, (ii) actors and roles, (iii) processing, (iv) legal rules, and (v) legal bases. These modules provide an overview of the major concerns to DSs, DCs, and DPs when faced with the GDPR duties. Even though the authors approached some legal bases present on GDPR, they do not present further details of the applicability and importance of informed consent. However, this ontology delivers entities that could be used in the LGPD, and we will detail this discussion in Chapter 5.

GConsent is a formalized ontology proposed by Pandit et al. (2019) focused on the GDPR consent legal base (GPDR Art. 6). The semantic web ontology proposed aims to represent the consent and compliance requirements. Moreover, such ontology presents new entities not approached in the previous work, such as consent "not given", refused, and withdrawn status. Sill, GConsent introduces the concept of implicit or indirect consent, *i.e.*, the consent is given by a legal person on behalf of another, *e.g.*, when a teenager starts a university course and one of his/her parents has to sign the consent term on his/her behalf. However, this ontology lacks details, other entities related to consent term, and use cases. In regards to entities, PrOnto and GConsent are complementary, and, as well as the PrOnto ontology, GConsent will be detailed in Chapter 4.

Gharib, Giorgini, and Mylopoulos (2021) presented COPri, a Core Ontology for Privacy requirements engineering. The authors argue that privacy concerns should be considered from the early system design phases and propose COPri to elaborate high-quality requirements models to allow system development in compliance with many data regulations, such as GDPR in the Europe Union, the Australian government issued the Privacy Act, PIPEDA (Personal Information Protection and Electronic Documents Act) in Canada, and HIPPA in the United States regarding the healthcare domain. Moreover, the authors exemplified the COPri instantiation in an Ambient-Assisted Living (AAL) system in the healthcare domain. However, COPri aims to assist software engineers only and not users, *i.e.*, data subjects. Still, there is no other case study to validate the ontology application in other domains. In this sense, GoDReP could aid the new use case development giving space to data subjects to contribute with their concerns.

3.2 Normative Multiagent Systems

NBDI is a conceptual framework proposed by Neto et al. (2013) that enables software agents to consider their beliefs, desires, and intentions when evaluating the norm's contribution (positive, negative, or neutral) in an NMAS. In Neto et al. (2013), the authors defined agents as goal-oriented entities to achieve their desires and fulfill the system norms concomitantly. However, respecting the data regulation proposals when managing personal data is also crucial to MAS developed in such context, including normative and BDI agents. In this sense, RegulAI proposes an architecture to address not only the normative BDI agents but also data regulation rules.

BDI4Jade is a framework that aims to enable the use of the BDI reasoning process in MAS (Cunha et al., 2015; Dubey et al., 2020). The authors extended the JADE framework (Bellifemine, Poggi, and Rimassa, 1999) and included BDI capabilities to represent the agent's decision-making process considering their goals and plans. However, they did not explore the BDI capabilities in NMAS. In our work, RegulAI aims to consider the agent's capabilities, i.e., goals and plans, in NMASs.

To support normative agents modeling, Freire et al. (2019) and Viana et al. (2022) proposed the NorMAS-ML (Supporting the Modeling of Normative Multi-agent Systems) and the ANA-ML Adaptative Normative Agent - Modeling Language), respectively, as tools for modeling normative agents. They are extensions of MAS-ML (Gonçalves et al., 2015) that enable modeling normative attributes in MAS. Their metamodel aims to improve the understanding of how agents can change their behaviors to deal with norms and captures interactions between agents' norms and adaptation. However, Freire et al. (2019) neither Viana et al. (2022) considered the reasoning process in their metamodel or data regulation entities. Thus, GoDReP and RegulAI can fit this gap.

To identify environmental norms, Mahmoud, Ahmad, and Mostafa (2019) propose the RNDT (Regulative Norms Detection Technique), which detects norms considering their rewards and penalties. Even though addressing norms challenges is not our focus, the authors proposed a norm taxonomy that classifies norms as follows: (i) regulative, (ii) constructive, and (iii) procedural. Moreover, the authors did not consider the BDI reasoning on the agent's decision-making process, although the regulative term emerged through the deontic concepts. Therefore, RegulAI can fill this gap and represent regulative norms considering the agent's purpose.

In previous work (Alves et al., 2023a,b), we presented an NMAS solution for data regulation. The proposed solution aims to represent data regulation concerns by norms development, employing rewards and punishments for obligations and prohibitions to DC agents who decide to comply or violate them. In such an approach, the deontic concept *permission* represents the DS rights, whereas *obligation* and *prohibition* represent the DC's and DP's duties. However, the agent's goals and cognitive reasoning to define the agents' decision-making process were out of scope, as well as the GoDReP approach to develop use case scenarios. Also, in this previous work, we did not perform the consent legal base evaluation to identify the major entities and their relationships.

3.3 Application Domains

The constant and intense collection of personal data by a myriad of services and goods, and the pan-optical vigilance exercised over our behavior when analyzing this collected data, highlight the importance of ensuring ways to protect personal data. Due to the Brazilian lack of tradition in this subject, it is important to provide society with acculturation and awareness of the importance of protecting personal data.

In Brazil, the LGPD puts forward a set of rules and obligations regulating public and private entities' use of personal data. Thus, controllers and processors must evaluate the legal bases in the law authorizing users' data collection (LGPD Arts. 7 and 11). In this sense, it must be remarked that the Brazilian data protection regulation establishes that individual consent is only one of the legal bases authorizing data processing. DCs must abide by the law's principles, rights, and safeguards and act in good faith.

3.3.1 Healthcare

In regards to the healthcare sector, Phillips (2018) presents an overview regarding the Organisation for Economic Co-operation and Development (OECD) and Council Guidelines Governing the Protection of Privacy and Transborder Flows Of Personal Data (1980) and moves forward to GDPR, HIPAA, and PIPEDA analysis. The author presented broad consent as a limited approach to attending data regulation. He proposed adding additional information regarding the purpose limitation to mitigate the gap to comply with the aforementioned regulations. However, he did not reflect other types of consent, *e.g.*, Dynamic Consent.

Trishan, Mattie, and Celi (2019) presented concerns regarding the amount of health data shared with hospitals, clinics, and companies in the healthcare area. They mentioned not only the importance of the identification related to who owns health data, who is responsible for it, and who can use it, but also the need for specific contracting objects in order to guarantee the data protection required to establish an honest relationship between healthcare organizations and the patients they serve. Last, the authors highlighted the importance of data protection regulations, such as GDPR and California's Consumer Privacy Act. However, well-resourced companies are more prone to bear regulatory compliance costs; hence, it may delay the growth of small healthcare organizations. Thus, the combination of ontology, data regulation rules, and application is vital to empowering companies and patients to offer resources to mitigate the informational asymmetry of data flow. In this sense, we proposed this combination and presented three application scenarios to exemplify our proposal usage.

In previous work, we (Alves et al., 2020) proposed a blockchain data governance to manage health data in compliance with the LGPD. However, the lack of an ontology forbids the employment of such data governance in other jurisdictions. For this reason, we decided to move towards ontology development. In Alves et al. (2021) we proposed the first version of our ontology for controlling personal data flow based on the LGPD. We presented the main concerns regarding DSs, DCs, and DPs' rights and duties. Moreover, we applied such ontology in the pandemic outbreak scenario to exemplify its use and proposed the adoption of blockchain technology to persist the data transparently, distributed, and immutable. However, we decided to improve this ontology in order to detail the consent legal bases entities following PrOnto and GConsent as LGPD presents many aspects in common with GDPR. In this sense, we decided to expand our ontology instead of proposing a new one.

Moreover, Bandara et al. (2021) proposed blockchain-based solutions for contact tracing, and Hardin and Kotz (2021) as well as Miyachi and Mackey (2021) proposed a blockchain solution to perform integration of mHealth systems. Bandara et al. (2021) developed Connect, which is an identity wallet based on blockchain technology to preserve data privacy when sharing health data. Hardin and Kotz (2021) proposed Amanuensis, it is a blockchain system that aims to provide information provenance for mHealth data. Miyachi and Mackey (2021) presented the Hybrid Off-Chain Blockchain System (hOCBS), which is a solution that gathers the on-chain and off-chain benefits into a unified system. However, the authors did not explore the regulation requirements and the legal bases related to such a scenario, *e.g.*, informed consent. Thus, GoDReP could be used to bridge the gap between the DSs and the authors' solution regarding their concerns. Furthermore, the framework could clarify the benefits of using blockchain and how the origin country's data regulation can affect it.

3.3.2 Education

Next, regarding the educational scenario, it presents challenges related to data access. For example, in a university, the students have to choose the disciplines, and the professors would request the list of previous disciplines or more details about the student to propose different activities in class. Moreover, many Brazilian students who subscribe to the university are below eighteen years old, *i.e.*, they are not considered an adult by Brazilian law. Therefore, the educational scenario was selected for presenting these challenges.

Sarabdeen and Ishak (2015) argue that the educational sector has changed the focus on educational tools improvement from the educator-centric to the student-centric model. Still, the authors mention the intersection between the Family Educational Rights and Privacy Act (FERPA), the Children's Online Privacy Protection Act (COPPA) in the United States, and the US Patriot Act that generates a conflict. For foreign citizens, the constitutional right to privacy is not applicable; hence, the Patriot Act allows the US government to monitor foreign citizens' data. In this sense, since performing an ontology alignment, GoDReP can create scenarios to evaluate the intersections and conflicts between the regulations based on the ontology that represents them.

Rosmaini et al. (2018) highlighted the importance of personal data protection in the educational sector in Indonesia. The authors argue that educational institutions require a large amount of personal data to run their business processes to support their activities, such as administration, teaching, learning, and research processes. The authors concluded that educational institutions should create periodical privacy impact assessment, auditory, and proper compliance to enforce personal data protection. Thus, GoDReP could aid the process of creating scenarios to enable DS, DC, and DP to evaluate the privacy terms on their behalf and let these agents try the different possibilities that could emerge in this scenario.

Siibak and Mascheroni (2021) proposed an improvement of the Sarabdeen and Ishak work. Instead of proposing a student-centric approach, Siibak and Mascheroni argue the importance of a child-centered approach to explore social consequences when sharing children's data. Moreover, the authors mention that regulations are controversial. They cited GDPR as an example of a normative framework that addresses children's right to privacy. In the GDPR Art. 8, children are people under the age of 16; however, countries in Europe consider different age limits (13, 14, or 15 years) (Milkaite and Lievens, 2018). In this sense, an ontology alignment between the regulations could identify the conflicting concepts, and the scenario instantiation could show a conflicting situation in practice.

Mishra et al. (2021) proposed a blockchain architecture to reduce security-related issues related to the students' personal data, especially their credentials. However, the authors did not explore the data regulation concerns and how they can be addressed in a blockchain-based solution. In this sense, GoDReP could aid the students and the educational institutions to mitigate the informational asymmetry by exploring the blockchain benefits and concerns towards a data protection regulation.

3.3.3 Open Banking

Third, the open banking scenario was selected for the challenges of sharing personal and transactional data among different financial institutions. Although the Central Banks regulate the processes regarding data sharing, the financial institutions must comply with data protection regulations according to the country's jurisdiction.

Ma et al. (2018) proposed a blockchain-based data privacy management framework in order to address concerns regarding GPDR compliance in the open banking scenario. Even though the authors presented an analysis regarding the attributes that must be informed in the consent term and the data subject data sharing authorization process, they did not follow any ontology to base the framework. Hence, applying this framework in other jurisdictions might not be possible or at least more complex than a framework based on an established ontology. Thus, GoDReP could be used as a pillar for the work presented by Ma et al. to evaluate the emergent technologies to be applied to previously developed scenarios.

Vives (2019) mentioned that digital disruption in the banking scenario could increase the system's efficiency and services, overcoming information asymmetries through big data, artificial intelligence, machine learning techniques, and blockchain technology associated with a straightforward user interface. According to the author, these techniques can improve the DS experience and deliver a less bureaucratic process in favor of the DSs. However, there is a lack between the DSs and the technology employment; the DSs should be able to evaluate their rights according to the local data regulation and consider the possible scenarios they could experience. Thus, GoDReP can bridge such a gap by providing an environment for not only DSs but also for DCs and DPs to evaluate the possible behaviors and law interpretations in a well-defined scenario.

Farrow (2020) mentioned that beyond the data protection regulation, open banking should follow the PSD2 (Payments Services Directive) that regulates the payment-related services to third-party providers. Even though the PSD2 is a group of best practices in APIs, data management, and vendor integration in the European Union, this directive must be translated into law in each specific country to respect the local regulatory jurisdiction.

3.3.4 Other Possible Scenarios

Beyond the use case scenarios above, there are other scenarios in which the agents involved would benefit from using GoDReP. For instance, Fosh-Villaronga et al. (2021) mentioned that the DS should be aware of which personal data is being used in the software decision-making to prevent discrimination. From the GDPR definition, a person's gender is not considered sensitive information, but it can result in discrimination depending on its context. In this sense, GoDReP could be applied to inform DSs about the importance of sharing gender data in a specific use case.

Campanile et al. (2021) proposed a solution for the Internet of Vehicles using blockchain, and they discussed data privacy. However, even providing technical documentation, from use case diagrams to the architecture implementation, the authors recognized that detailed and domain-level documentation could be a promising future work. Thus, GoDReP could enhance the documentation and bring the agents closer to technical aspects combined with regulation concerns.

Still, Eronen et al. (2021) discussed data privacy and data regulation focused on cyberbullying. Even though the authors presented a superficial analysis regarding the regulation concerns, they mentioned the importance of evaluating global policies such as GDPR to avoid abuse in using of personal data. In this sense, GoDReP could be applied to provide agents awareness in regards to the possibilities of personal data use and the expected behavior when the data is misused.

Last but not least, Makhlouf, Zhioua, and Palamidessi (2021) evidence concerns about the use of machine learning on real applications related to fairness. Given the subjectivity of such a term, the authors aimed to explore how fairness is suited to real-world scenarios, such as: college admission, teacher evaluation and promotion, health care, among others. Although the authors did not explore any regulation specifically, they presented concerns regarding data sharing and the automatic decision-making process to show possible unfair judgments. Therefore, GoDReP could aid the agents in evaluating how the decision-making process works, what matters from each agent's perspective, and how the data protection regulation may protect the DS, for instance.

3.4 Related Work Main Takeaways

This section presented an overview of research on data protection regulations, MAS, and the studies regarding use case scenarios in different application domains regarding these two areas. A total of 35 works were evaluated and compared to our approach. Table 3.1 presents the comparison summary classifying the selected works based on eight major concerns.

- [C1] Provides an example in a specific domain (83%).
- [C2] Evaluates the impacts of sharing data (57%).
- [C3] Presents data regulation concerns (54%).
- [C4] Focuses on system development design (46%).
- [C5] Focuses on the consent legal base (20%).
- [C6] Presents concerns dealing with more than one data regulation (17%).
- [C7] Proposes a tool to share personal data (11%).
- [C8] Proposes a framework to bridge the data flow informational gap between agents (only our work).

Papers/ Concerns	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**
Farrow (2020)	x	x	x	x		x		
Ma et al. (2018)	x	x	x	x			x	
Alves et al. (2021)	x	x	x		x	x		
Pandit et al. (2019)	x	x	x		x			
Breuer and Pierson (2021)	x	x	x		x			
Stoilova and Nandagiri (2021)	x	x	x		x			
Sarabdeen and Ishak (2015)	x	x	x			x		
Trishan, Mattie, and Celi (2019)	x	x	x			x		
Mense and Blobel. (2017)	x	x	x					
Alves et al. (2020)	x	x	x					
Fosh-Villaronga et al. (2021)	x	x	x					
Eronen et al. (2021)	x	x	x					
Rosmaini et al. (2018)	x	x	x					
Siibak and Mascheroni (2021)	x	x	x					
Gharib, Giorgini, and Mylopoulos (2021)	x	x		x		x	x	
Campanile et al. (2021)	x	x		x				
Fatema et al. (2017)	x	x			x			
Dougherty (2020)	x	x			x			
Vives (2019)	x	x				x		
Bandara et al. (2021)	x		x	x				
Miyachi and Mackey (2021)	x		x	x				
Hardin and Kotz (2021)	x		x	x				
Mishra et al. (2021)	x		x	x				
Makhlouf, Zhioua, and Palamidessi (2021)	x		x					
Neto et al. (2013)	x			x				
Gonçalves et al. (2015)	x			x				
Freire et al. (2019)	x			x				
Dubey et al. (2020)	x			x				
Viana et al. (2022)	x			x				
Collierf et al. (2010)	x							
He et al. (2014)	x							
Phillips (2018)		x	x		x	x	x	
Kirrane et al. (2018)		x	x		x		x	
Palmirani et al. (2018)		x						
Cunha et al. (2015)				x				
Bellifemine, Poggi, and Rimassa (1999)				x				
Mahmoud, Ahmad, and Mostafa (2019)				x				
Alves et al. (2020)	x	x	x					
Alves et al. (2021)	x	x	x		x	x		
Alves et al. (2023a)	x	x	x	x	x	x		x
Alves et al. (2023b)	x	x	x	x	x	x	x	x

Table 3.1: Comparison summary.

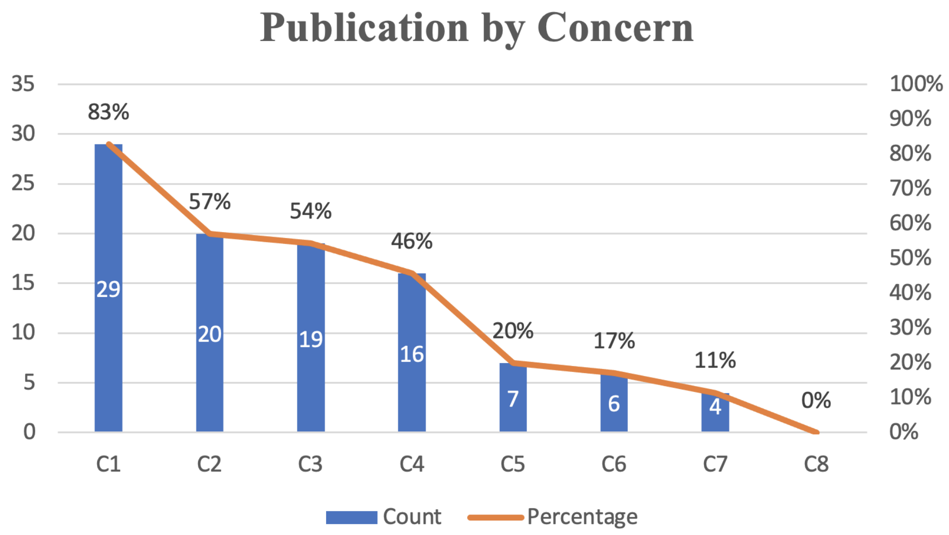


Figure 3.1: Publication by concern.

Although our approach could aid developers in understanding and proposing negotiation scenarios, we decided not to include our published works in the for considering the state of the art before our contributions. Only papers related to technical concerns were marked to this point.

As observed in Figure 3.1 in most of the selected works (83%), providing a use case scenario applied in a specific domain is crucial to enhance the understanding of data regulation operation. Also, according to C2, 57% of the selected works mentioned and detailed the importance of evaluating the impacts of sharing data. Finally, this overview may aid readers in understanding the big picture of data regulation from the agents' perspective.

In this chapter, we presented an overview of research on ontologies in the context of data protection regulations, MAS, and the studies regarding our use case scenarios. Each related work was evaluated and compared to our approaches, presenting the connections and deviations. The selected studies directly impact our work or present an opportunity to be complemented by our approach. Next, we present and detail our ontology extension based on PrOnto and GConsent.

4
Ontology

Ontologies are representations of a specific domain that aims to create a shareable and reusable model. They are also considered a valuable instrument for reducing conceptual ambiguities and inconsistencies in a specific domain (Staab and Studer, 2010).

In this sense, building an ontology is the first step toward defining entities, attributes, and relationships. It should enable the construction of processes and systems in accordance with data protection regulation concerns. Furthermore, an ontology should aid the impact analysis process, *i.e.*, given an internal or external change in one entity instance, the ontology should show which entities and relationships could be affected.

In summary, an ontology would allow people to get at least a brief understanding of the effects of sharing personal data, as well as their rights, under data protection regulations. DSs must know the purpose of data processing, who the controllers are, the responsibilities of the DC, how and if they can revoke access to their information, and limit its use (content and time length of data processing). Disclosing this information is mandatory (LGPD Art. 9). The traceability of the data flow is essential to turn effective the right to informational self-determination, data protection, and privacy (Rodotà, 2008).

4.1
PrOnto - Privacy Ontology

The GDPR introduces the privacy-by-design concept to improve software development, addressing privacy concerns since the beginning. In this sense, the audit and the compliance checking are activities that allow the detection of violations when they occur (Casalicchio et al., 2018). Moreover, GDPR introduces the self-assessment of the digital risks and expresses measures to protect the DS's rights (Palmirani et al., 2018).

In order to support privacy-by-design principles, procedures regarding legal reasoning and semantics can aid companies and even the government's daily activities. To do so, Palmirani et al. (2018) presents the PrOnto (Privacy Ontology), which is an ontology towards data protection regulation. It is important to note that although the authors based the ontology on the GDPR, they aimed to create an ontology that could be extended to other jurisdictions.

In this light, PrOnto defined the data types, privacy agents, processing activities, rights, and duties to model the legal knowledge entities and relationships. Moreover, PrOnto was developed using the MeLOn methodology (Methodology for building Legal Ontology) to decrease the difficulties that law experts use to face when defining a reality model through ontological techniques (Palmirani et al., 2018). This methodology is composed of ten steps that aid ontology development: (i) Describe the goal of the ontology; (ii) Evaluation indicators; (iii) State of the art survey; (iv) List all the relevant terminology; (v) Use usable tools; (vi) Refine and optimize; (vii) Test the output; (viii) Evaluate the ontology; (ix) Publish the document; (x) Collect feedback.

As a result, PrOnto defines five modules: (i) documents and data, (ii) actors and roles, (iii) processing and workflow, (iv) legal rules and deontic formula, (v) purposes and legal bases. Figure 4.1 depicts the modules and their relationships in a high abstraction level. An agent is a physical person with rights regarding his/her data. A processing institution can process these data under a specified time, context, and purpose. Moreover, the data processing must be performed on a legal bases that provides legitimacy. As mentioned before, this thesis aims to explore the legal bases based on the consent term.

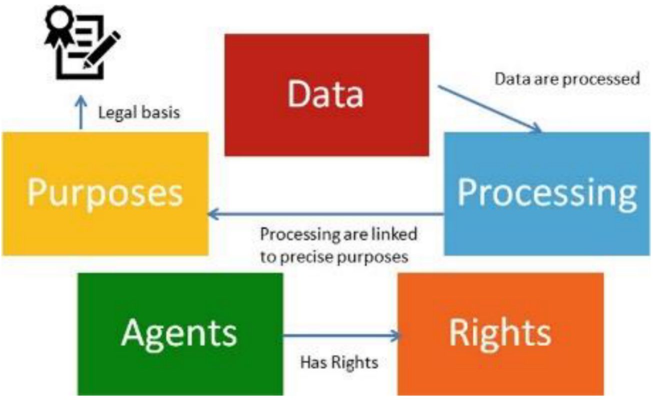


Figure 4.1: PrOnto general modules (Palmirani et al., 2018).

Each module comprises a subset of entities and relationships to explore them in detail. First, it is important to define what personal data are. The term is defined in GDPR Art. 4 (1): "*Personal data are any information which is related to an identified or identifiable natural person.*"[1]. It includes not only name, identification number, location data, but also physical, physiological,

[1] More about in `https://gdpr-info.eu/issues/personal-data/`. Last accessed on April, 2023

genetic, mental, commercial, and cultural or social identity. Furthermore, in practice, GPDR considers any data assigned to a person in any kind of way, such as phone number, credit card, account data, number plate, and so on if combined reveal the person's identity.

PrOnto Data Module. It defines data as personal, non-personal, anonymized, and pseudonymized data. Non-personal data gather anonymized data and data from a legal person, *i.e.*, an individual, company, or other entity which has legal rights and is subject to obligations[2].

Although data encryption and data pseudonymization are boolean attributes, it is crucial to delve deeper into their implications. The exploration should extend beyond their binary nature, as there exists a wide array of encryption and pseudonymization algorithms. However, the DS should be fully aware that this aspect demands thorough attention when making decisions regarding the sharing of personal data. The choice of algorithms implemented can significantly influence the DS's decision to either share or withhold their data.

PrOnto Agents Module. This module defines the difference between a person and an organization, and it establishes the agent's roles. An agent, *i.e.*, a person or an organization, is subjected to an authority and should determine a purpose for the data processing. Still, an agent would play different roles according to the action and event requirements. There are examples of a role: data subject, controller, processor, third party, and DPO (Data Protection Officer) or ANPD (National Data Protection Agency) in Brazil; these roles are fixed by a given time respecting the event duration.

PrOnto Processing Module. It defines a workflow to process data. GDPR Article 35 presents the DPIA (Data Protection Impact Assessment)[3]. The DPIA is a GDPR requirement to address the "protection by design" principle. According to the law: *"Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data."*. In this sense, this module is crucial to perform compliance checking in distinct scenarios. The workflow presented in PrOnto could also be applied in data processing evaluation scenarios under the LGPD context.

[2] Legal person definition: `https://www.lexico.com/definition/legal_person`. Last accessed on Jan, 2023

[3] Impact Assessment Plan is available at: `https://gdpr.eu/data-protection-impact-assessment-template/`

Furthermore, the Processing Module proposes a detailed structure of possible data processing actions. An action can be classified as: (i) deletion, which distinguishes the levels of deletion, such as permanent erase, anonymize, and destroy; (ii) derive; (iii) provide; (iv) observe; (v) store; (vi) communicate, (vii) infer, (vii) transmit, and (viii) consent. However, Palmirani et al. (2018) did not explain how all the actions could be performed; hence, the interpretation of what each action means is unclear.

PrOnto Purposes and legal bases Module. This module defines that the data can be processed according to a specific purpose and time range, as GDPR allows personal data processing in the light of a clear purpose only. Moreover, this module specifies that the data processing must present boolean attributes related to fairness, transparency, and lawfulness to verify the purpose alignment with the legal bases foreseen in the GPDR.

PrOnto Rights. It defines the rights and duties to DSs, DCs, and DPs. The rights and duties are entities that need complements to be used. To do so, the deontic operators can act as complements to produce compliance or violation evidence regarding a right or duty. Moreover, the deontic operators are connected to temporal parameters and jurisdiction to apply the regulation and clauses correctly.

Therefore, PrOnto presented vital concepts to define the entities and their relationships, producing an ontology based on GDPR. Even though this ontology is not extensive and detailed in some aspects, such as transparency, cryptography, fairness, lawfulness, and breachness which are boolean attributes, it presents the crucial concepts that start a regulation analysis regarding the impact of DS, DC, and DP's actions. For this reason, PrOnto and GConsent, which will be detailed in the last section of this chapter, were extended to generate the ODPM ontology based on the LGPD.

4.2 GConsent

As mentioned in the Related Work Chapter, as well as PrOnto, GConsent is an ontology based on the GDPR (Tikkinen-Piri et al., 2018). However, GConsent (Pandit et al., 2019) is focused on the GDPR consent legal bases (Art. 6), which is valid when it is freely given, specific, informed, and unambiguous (Art. 2-11). In order to demonstrate compliance with these obligations, DCs and DPs should record proof of the given consent showing how the consent was collected, used, and changed over time (Mittal and Sharma, 2017).

In this sense, Pandit et al. (2019) proposed the GConsent, a semantic web ontology for representing the consent and compliance needs, considering the flexibility required for expressing entities and their relationship in a standardized, open, and queryable manner. Uniquely the literature (Collierf et al., 2010; He et al., 2014; Bartolini and Muthuri, 2015; Fatema et al., 2017; Kirrane et al., 2018; Palmirani et al., 2018), GConsent supports not only the given consent, but also the “not given”, refused, and withdraw ones. Moreover, there is no reference but GConsent that delivers an approach to consider the proxy consent, *i.e.*, the consent given by a person on behalf of another. In summary, GConsent was developed based on the Consent and Data Management Model (CDMM) (Fatema et al., 2017), and added other relevant entities, for instance, consent status, type, and state. Moreover, except for the aforementioned entities, the remaining entities in CDMM are already presented in the PrOnto ontology.

Figure 4.2 depicts the core entities and their relationships. As consent can be given by a *Person*, which can be the *DataSubject* or a *Minor DataSubject*, *i.e.*, not an adult, if the *DataSubject* is not considered as an adult by the current jurisdiction, the consent term must be given by a delegation relationship, as depicted in Figure 4.3 , *i.e.*, the *Minor DataSubject* must indicate a *DataSubject* to delegate the consent to accept or not the consent term. Still, the consent term must indicate the *Purpose* of using *PersonalData*, who is responsible for *Processing* the data, and what is its *Status*. Complementary, Figure 4.3 shows the entities related to the scenario context in which the consent will be applied, providing aspects of temporality, locality, and medium.

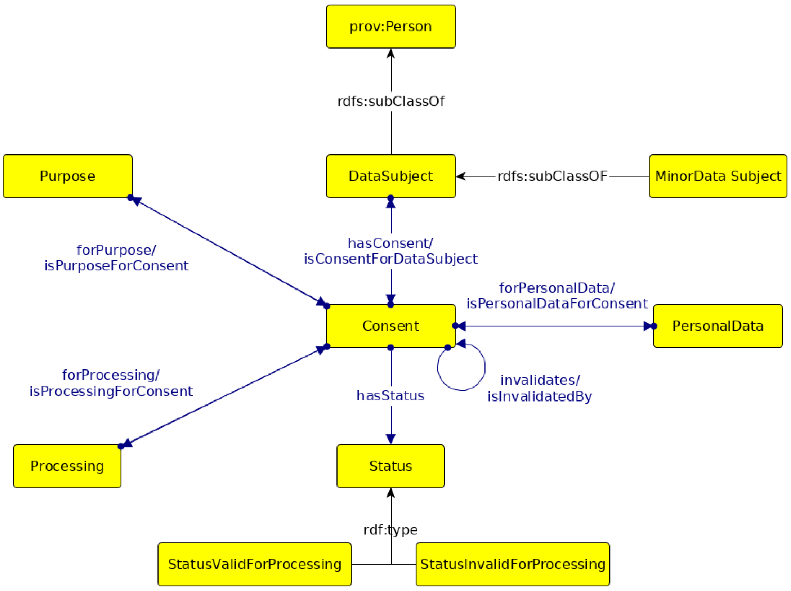


Figure 4.2: GConsent core ontology (Pandit et al., 2019).

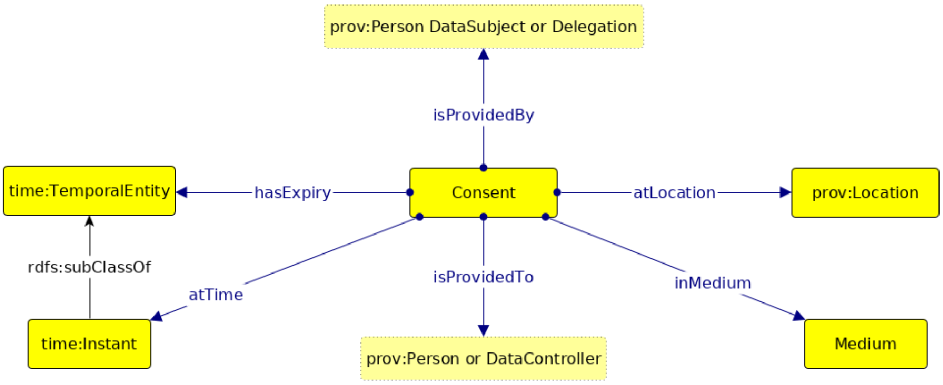


Figure 4.3: GConsent consent context (Pandit et al., 2019).

Furthermore, the authors proposed use case scenarios to represent consent in different contexts, determining the information required towards GDPR compliance. There are fifteen categories of use cases[4], but only four categories were described in the project documentation: (i) Change in Consent State; (ii) Capturing Given Consent; (iii) Capturing Consent Given via Delegation, and (iv) Capturing Consent where Data is shared with a Third Party. Still, the category number (iii), which presents the delegation concept, was also introduced in the paper published by Pandit et al. This use-case scenario describes "*an emergency ward where a nurse provides consent on behalf of the patient*" (Pandit et al., 2019) to exemplify what the authors defined as a consent "implicitly given" using the *Delegation* entity.

However, such a use case does not explain the whole context of the application; for instance, did the hospital try to contact the patient's family? How long was this consent valid? What were the shared data? Who had access to the patient's data? Those are questions that could aid people in understanding how the proposed ontology works. Moreover, each use case can introduce domain particularities; hence, they are vital to validate the ontology.

Overall, PrOnto and GConsent introduce general entities and relationships regarding the GDPR privacy and consent scenarios exploration, but they lack the definitions and their possible applications. Moreover, even if these ontologies were merged, there are particularities in the LGPD that require a new version of such ontologies. These particularities will be explored in the next section.

[4] https://openscience.adaptcentre.ie/ontologies/GConsent/docs/ontology

4.3 ODPM

As well as in PrOnto and GConsent, ODPM (Ontology for Data Privacy Management) seeks to define data regulation entities and their relationships (Alves et al., 2021). It was the first step toward an LGPD ontology. This would allow Brazilian citizen to get a complete understanding of the effects of sharing personal data, as well as their rights, under data protection regulations. DSs must know the purpose of data processing, who are the controllers, what are the responsibilities of the DP and DC, how, and if, they can revoke access to their information and limit its use (content and time length of data processing). Disclosing this information is mandatory (art. 9, LGPD). The traceability of the data flow is essential to turn effective the right to informational self-determination, data protection, and privacy.

In order to control the flow of personal data and decrease informational asymmetry, it must be known (i) the data source and content, (ii) who inserted the data, (iii) when the data were added, (iv) whether the data were changed, and (v) the processing purpose. Hence, an ontology development should consider these concerns to correctly represent this environment's needs by providing proof of the data integrity and provenance. Furthermore, to build a complete ontology, the entities involved should also be considered, as well as the possibility of data auditing. Governments, health organizations, researchers, citizens, and the media should also be able to consult and check the data.

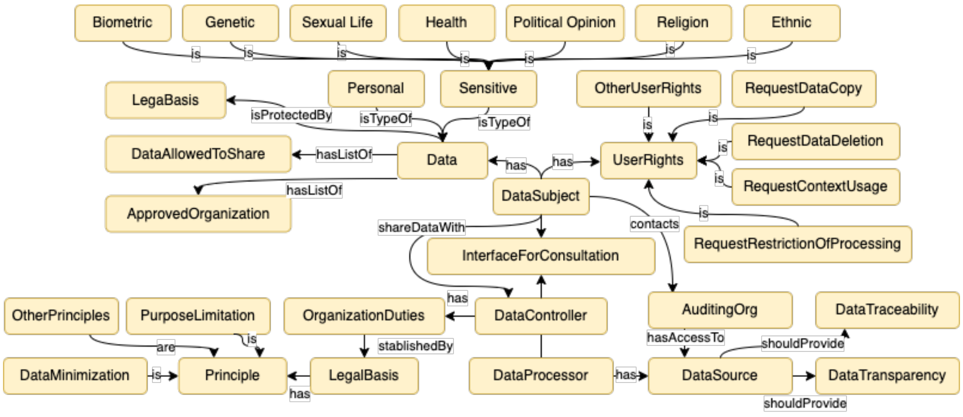


Figure 4.4: Ontology for Data Privacy Management and LGPD Compliance.

In this sense, ODPM aims to identify entities and their relationship for further technological support development and to satisfy the regulatory requirements. Figure 4.4 depicts ODPM, following the description of the ontology concepts in the sequence.

Citizen is the entity responsible for: (i) query information from the data provider, and third parties who received the shared data, and (ii) request validation regarding data and metadata information, such as who and when the data were added to the database. The *Citizen* entity is also safeguarded by its rights and composed of personal data.

UserRights is the entity that represents what citizens can request, such as the copy of stored and processed data, the restriction of processing, the context usage, data deletion, and data correction, for example.

PersonalData is the entity that represents personal and sensitive data collected according to LGPD legal bases. It also includes a list of data that the user agrees to share, a list of organizations that are able to use such data, and the legal bases. We considered personal data as presented by art. 5, I and II, LGPD[5].

DataController entity can process the citizens' data when authorized by one of the legal bases foreseen in arts. 7 and 11. Thus, *DataController* is composed by *OrganizationDuties* and *UserRights.* [6] This entity can process the citizens' data when authorized by one of the following legal bases (art. 7, LGPD): (i) user consent; (ii) to attend a legal or regulatory obligation by the DataController; (iii) by the public administration, for shared purposes and for the execution of a public policy foreseen in law or other legal instrument; (iv) research, implementing data anonymization, when possible; (v) to attend an agreement requirement involving the DS or by his/her request, (vi) to exercise rights foreseen in judicial, administrative or arbitral procedure, (vii) to protect the life or physical state of the DS; (viii) to provide health safeguard in procedures executed by health professionals; (ix) DCs legitimate interests, and (x) credit protection. Moreover, Article 11, which sets the legal bases for processing sensitive personal data, authorizes data processing when based on

[5] Article 5, I: "information relating to an identified or identifiable natural person"; Article 5, II. "Personal data related to racial or ethnic origin, religious conviction, political opinion, membership of a union or organization of a religious, philosophical or political character data, health or sexual data, genetic or biometric data, when associated to a natural person".

[6] Art. 7: (i) user consent; (ii) to attend a legal or regulatory obligation by the DC; (iii) by the public administration, for shared purposes and for the execution of a public policy foreseen in law or other legal instrument; (iv) research, implementing data anonymization, when possible; (v) to attend an agreement requirement involving the DS or by his/her request, (vi) to exercise rights foreseen in judicial, administrative or arbitral procedure, (vii) to protect the life or physical state of the DS; (viii) to provide health safeguard in procedures executed by health professionals; (ix) DCs legitimate interests, and (x) credit protection. Moreover, art. 11 sets the legal bases for processing sensitive personal data, authorizes data processing when based on the following hypothesis: (a) with the user consent; (b) without the user consent in the hypothesis (ii), (iii), (iv), (vi), (vii) foreseen above, and (b.1) to protect one's health, exclusively in procedures performed by healthcare workers, health services or health authority; (b.2) to protect the DS from fraud in identity and authentication registration procedures in electronic systems, preserving DS rights, and except when it is necessary to protect DS's fundamental rights and principles which requires data protection.

the following hypothesis: (a) with the user consent; (b) without the user consent in the hypothesis (ii), (iii), (iv), (vi), (vii) foreseen above, and (b.1) to protect one's health, exclusively in procedures performed by healthcare workers, health services or health authority; (b.2) to protect the DS from fraud in identity and authentication registration procedures in electronic systems, preserving DS rights, and except when it is necessary to protect DS's fundamental rights and principles which requires data protection. Thus, *DataController* is composed by *OrganizationDuties*.

OrganizationDuties is the entity responsible for the legal bases application, defining which one is applicable according to the processing context. For example, as stated by LGPD principles and DS' rights, the *DataController* shall respect *DataMinimization*, *PurposeLimitaion*, *DataDeletion* among others, as well as security and data governance concerns. Once *Citizens* share their data, s/he contributes to populating the database on behalf of society.

InterfaceForConsultation entity is the *DataController* bridge to share data with the *Citizen*. The *DataController* receives the treated information from the *DataProcessor* and discloses data to citizens.

DataProcessor is the entity responsible for processing data strictly in accordance to the *DataController* commands and returning the processed data from the *DataSource* to the *DataController*. The latter can exercise the *DataProcessor* role or delegate to a third party.

AuditingOrg is the entity responsible for auditing the information originated in the *DataSource* and exercising compliance regarding the roles and data addition circumstances, e.g., this entity will evaluate unauthorized data insertion.

DataSource entity represents the database technology. To provide transparency and traceability, it is usually required to check the data provenience. Thus, the database should deliver resources to track data, as well as to provide this information to the *AuditingOrg*.

DataTransparency and ***DataTraceability*** entities represent trackable attributes to provide data transparency and traceability.

Also, according to art. 9, LGPD, the *DataController* must provide some basic information so that the *Citizen* is able to comprehend the data processing and contact the DC. Thus, the *DataController* must provide information in a straightforward manner, structured in a clear, and adequate form, referring to the: (i) specific purpose of the treatment, (ii) form and duration of the processing, (iii) DC identification and contact information, (iv) DC and DP obligations, and (v) DS rights (art. 18). This ontology is designed to empower people to check and claim data privacy and protection, provide knowledge of collective rights (health and social rights), and high-quality information. Also, DCs and DPs are able to efficiently provide accountability.

In this chapter, we presented PrOnto, GConsent, and ODPM ontologies in detail, which answer our *RQ1*, *i.e.*, a data protection regulation can be represented by an ontology, and as LGPD presents many aspects in common with GPDR, an ontology for GDPR could suit the necessities of LGPD. These ontologies are state-of-the-art regarding data protection regulation ontologies. However, we did not find any ontologies related to LGPD specifically, *i.e.*, to the best of our knowledge, no other ontology than ODPM considers the LGPD particularities.

Finally, this ontology aims to model consent's context, state, and provenance. Its scope is limited to consent as defined in the LGPD and the GDPR legal bases. The aim is to assist in modeling information associated with compliance but not determine the compliance itself.

Next, we will present a consent metamodel and a framework to formulate use case scenarios to explore consent-related issues' process, rights, and evaluation. As a result, data subjects, data controllers, and processors can develop their scenarios to perform interpretations possibilities and explainability of each situation. The proposed consent metamodel and the framework will be detailed in the next chapter.

5
Modeling Data Regulation

In this chapter, we introduce CM, which has been developed based on ontologies discussed in chapter 4. Furthermore, we provide a comprehensive overview of GoDReP in this chapter, outlining its procedures and demonstrating the utilization of Prolog and Jupyter Notebooks.

5.1
Consent Metamodel

To describe and produce use case scenarios in a specific domain, first, the data agent should understand the data regulation entities and their relationships. In this sense, we propose the Consent Metamodel (CM) based on the ontologies found in the literature to offer a summarized view of these entities and their relationships to data agents in a data processing context. These entities represent the privacy policy elements that must be included in the consent term to comply with personal data regulations. Complementary, this article proposes GoDReP (Generation of Data Regulation Plots) to allow data agents to describe use cases and their understanding of personal data regulation interpretation enforced by first-order logic sentences based on CM.

As mentioned in Chapter 2, this work focuses on the Consent legal bases and on defining requirements for generating adequate consent. This can be challenging, especially in a globally connected world, *i.e.*, where companies, governments, and citizens can offer and access services worldwide throughout different jurisdictions. In this sense, Kurteva et al. (2021) proposed a survey to explore the consent's state of the art and its best practices based on a table of competency questions related to GDPR. We enhanced this table by addressing the LGPD provisions for each question, generating Table 5.1. This table shows the relevant concepts related to the consent legal bases and addresses where their definitions can be found in GDPR and LGPD.

As presented in Table 5.1, although GDPR and LGPD present different structures, all questions are addressed in both regulations. It means that an ontology built considering the GDPR perspective can be suitable to LGPD, with a few changes, since they present similar concerns. PrOnto and GConsent are ontologies based on the GDPR, just as the ODPM is based on the

Table 5.1: GDPR and LGPD Competency Questions.

Question	**Relevant Concepts**	**GDPR**	**LGPD**
Who collects the data?	DC, DP	Art. 4 (7), Art. 6, Art. 28	Art. 5, (VI, VII, IX)
What is the purpose?	Purpose	Art. 4 (4), Art. 7 (32), Art. 6 (1a, 1f, 4)	Art. 5 (XII), Art. 6 (I), Art. 8 (4)
How to revoke consent?	Status	Art. 17, Rec. 63, Rec. 66	Art. 8 (5,6), Art. 9 (2), Art.15 (III), Art. 18 (IX)
How long does consent last for?	Time Range	Rec. 32, Rec. 42	Art. 15
When was consent given / revoked?	Time Range / Status	Art. 17, Art 19	Art. 7 (I), Art. 8, Art. 9 (2), Art. 15 (III), Art. 18 (IX)
What personal data is collected?	Personal Data Categories	Art. 4 (1), Art. 9	Art. 5 (I, II)
How is the personal data being used?	Processing	Art. 4 (2)	Art. 5 (X)
How is personal data collected?	Data Collection	Art. 12, Art. 13, Art. 14, Rec. 39, Rec. 58, Rec. 62, Rec. 73	Art. 5 (X, XII), Art. 6, Art. 7 (I), Art. 9, Art. 11 (I)
With whom is personal data shared?	Data Processing, Sharing Policy	Art. 4 (7), Art. 6, Art. 28	Art. 5 (XVI), Art 7. (5), Art. 9 (V), Art. 18 (VI), Art. 26
Who is responsible for the personal data?	DC	Art. 24, Rec. 74, Rec. 79	Art. 5 (VI, VII, IX), Art 9. (VI), Art. 37
Where is personal data stored?	Data Storage	Art. 5	Art. 6 (IV, VII, VIII)
Who is the DC?	DC	Art. 4 (7), Art. 28	Art. 5 (VI), Art 6, Art 9. (III, IV)
How to contact the DC?	DC, Contact Channel	Art. 4 (7), Art. 14, Art. 28	Art. 9 (IV)
What are the responsibilities of the DC?	DC, Right	Art. 4 (7), Art 14, Art. 28, Art. 37	From Art. 6 to Art. 50[1]
Who is the DS?	DS	Art. 4 (1)	Art. 5 (VII)
Whom to contact?	Contact Channel	Art. 12, Art. 13, Art. 14	Art. 9 (III, IV)

LGPD, which enables consent knowledge representation. This work proposes the Consent Metamodel (CM) inspired by these three ontologies and the aforementioned competency questions.

CM proposes three modules to mitigate the data flow informational asymmetry, as depicted in Figure 5.1. The *ConsentTerm* module determines the consent legal bases requirements, and it narrows the DS, DC, and DP's actions. The *Action* defines the step execution based on the *ConsentTerm* to accomplish a specific action considering: (i) the jurisdiction, to allow the scenario contextualization; (ii) consent term on which the action is based on; (iii) time frame, (iv) rights based on the jurisdiction applied, and (v) the deontic operator to indicate a normative expression Wright and Henrik (1951). Moreover, the *Action* generates a log of executed actions, and it should explain the action performed. Still, this can be used as evidence to evaluate the consent term compliance.

These modules are essential to provide not only a conceptual view of the consent requirements but also to enable the construction of application scenarios to exercise the data sharing process, the agents' rights and evaluate the impacts in different situations.

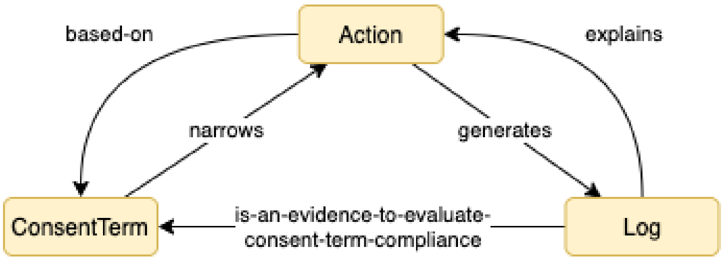


Figure 5.1: Consent Metamodel structure.

As well as GDPR, the LGPD provides rights for DSs and establishes duties and responsibilities for data DCs and DPs. The data protection norms are crucial to define the expected behavior when personal data are shared and processed. However, as a subjective object, the law can be interpreted differently, and the expected behavior can be hazy.

In this sense, in order to develop an ontology that fits GDPR and LGPD, the first step was to evaluate these regulations to identify the main differences. The former is more normative and detailed. As mentioned before, GDPR split ninety-nine articles into eleven chapters indicating the rights and duties strictly for DSs, DCs, and DPs. The latter is generalist and, as law, lets the clauses more open to interpretations case by case.

The LGPD cases can present intersections with other laws in the Brazilian legal scenario. As depicted in Figure 5.2, LGPD has a strong relationship with the Information Access Law - LAI (Law n. 12,527/2011 Lei de Acesso à Informação)[2], which regulates the constitutional right to citizens access public information Teixeira et al. (2017); Oliveira et al. (2020). Thus, the relationship with other laws means that more than one legal provision can be activated beyond the LGPD clauses. However, in the light of this thesis, we restricted the scope to the LGPD context.

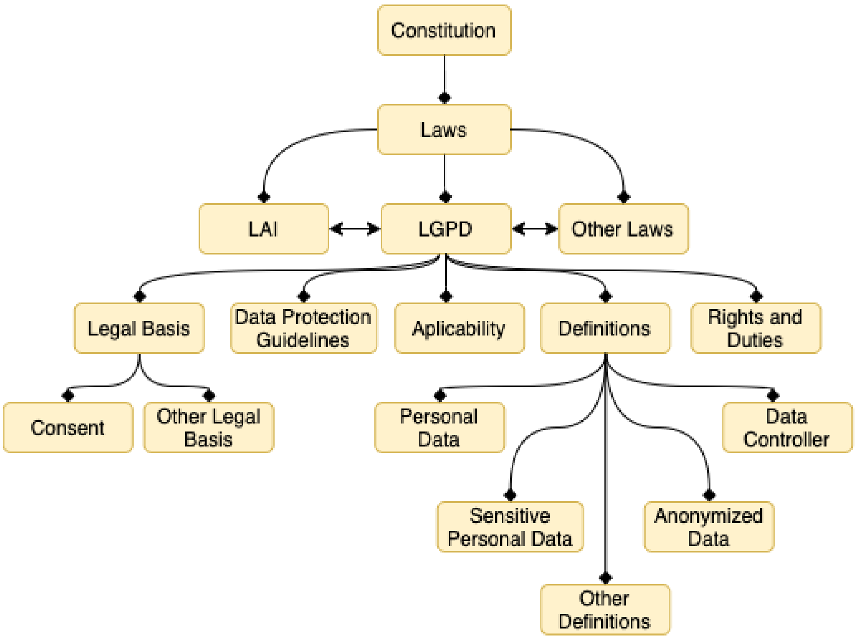


Figure 5.2: LGPD structure.

5.1.1 Consent Module

In this sense, the *Consent* module is depicted in Figure 5.2 and present the following LGPD definitions: (i) legal bases: consent is often used, but there are other legal bases foreseen in the law; (ii) data protection guidelines: general guidelines[3]; (iii) applicability: there are some situations in that LGPD cannot be applied, such as when the data is anonymized; (iv) concepts: LGPD qualifies personal data, sensitive personal data, DC, among others; (v) rights and duties: LGPD sets rights and duties for DSs, DCs, and DPs. These definitions are important to understand how the law is structured.

As mentioned above, even though there are ten legal bases in LGPD, this thesis focused on the consent legal bases. We decided to use consent as a

[2] https://www2.camara.leg.br/transparencia/acesso-a-informacao

[3] Art. 55-J The National Authority has the following duties: III - to elaborate guidelines for the Personal Data Protection and Privacy National Policy. https://iapp.org/resources/article/brazilian-data-protection-law-lgpd-english-translation/

study object because it can be applied in most situations. Furthermore, LGPD defines consent as a free demonstration, informed, and unequivocal in which the DS agrees with data processing under a specific purpose (Art. 5, XII). Thus, consent is the most important entity in our approach.

In a detailed view, Figure 5.3 depicts the relationships between the ontology entities. The yellow entities are those that are present in the PrOnto ontology that fits the LGPD consent legal bases, they are: (i) *ConsentTerm*, which must inform who is the subject and the DC, the purpose limitation, the data that will be collected and processed, and the time range; (ii) *DataSubject*, which has rights to be respected; (iii) *DataController*, which must inform its identification; (v) *DataProcessing*, which is narrowed by the purpose limitation and it can be restricted by a data deletion request; (vi) *DataCollecting*, which is narrowed by the purpose limitation and it can be interrupted by a consent revocation request; (vii) *DataStorage*, that can be restricted by a data deletion request; (viii) *PurposeLimitation*, which must be in the consent term and narrow the data collecting and processing; (ix) *TimeRange*, that will create a time due date to the consent term; (x) *Data*, which can be not only classified as *PersonalData* and *SensitiveData*, but also should have data governance guidelines, sharing policy, and security methods informed, and (xiii) *Right*; which represents the DS's rights based on data privacy regulation.

The blue entities are those inherited from GConsent ontology. These entities enabled the distinction between valid and invalid consent status, and the consent given by the DS or given by proxy, *i.e.*, by delegation to another person. **A *Direct* and *Valid* consent is given when the DS is able to, by the jurisdiction, agree with the consent term and he/she decides to do it.** Conversely, a *Proxy* and *Valid* consent is given when the DS is not considered an adult and requires a legal person in charge to agree with the consent term on his/her behalf. Finally, a consent term is invalid when there is a clause modification, and the DS has not accepted it yet, or when the due date expires.

The green entities are those which were added to fulfill the LGPD needs; however, these entities can also be applied in scenarios ruled by the GDPR without producing inconsistencies or conflicts with the remain entities, they are: (i) *Governance*, (ii) *AccessRestriction*, (iii) *StorageTechnology*, (iv) *SecurityMethod*, (v) *SharingPolicy*, (vi) *DisputeResolution*, (vii) *LossSize*, (viii) *Discrimination*, (ix) *UnauthorizedUse*, (x) *DataBreach*, and (xi) *Anonymization.* For instance, PrOnto and GConsent do not consider the data governance and technology concerns, such as access polices and data storage infrastructure, respectively. Still, these ontologies do not evaluate third-party sharing

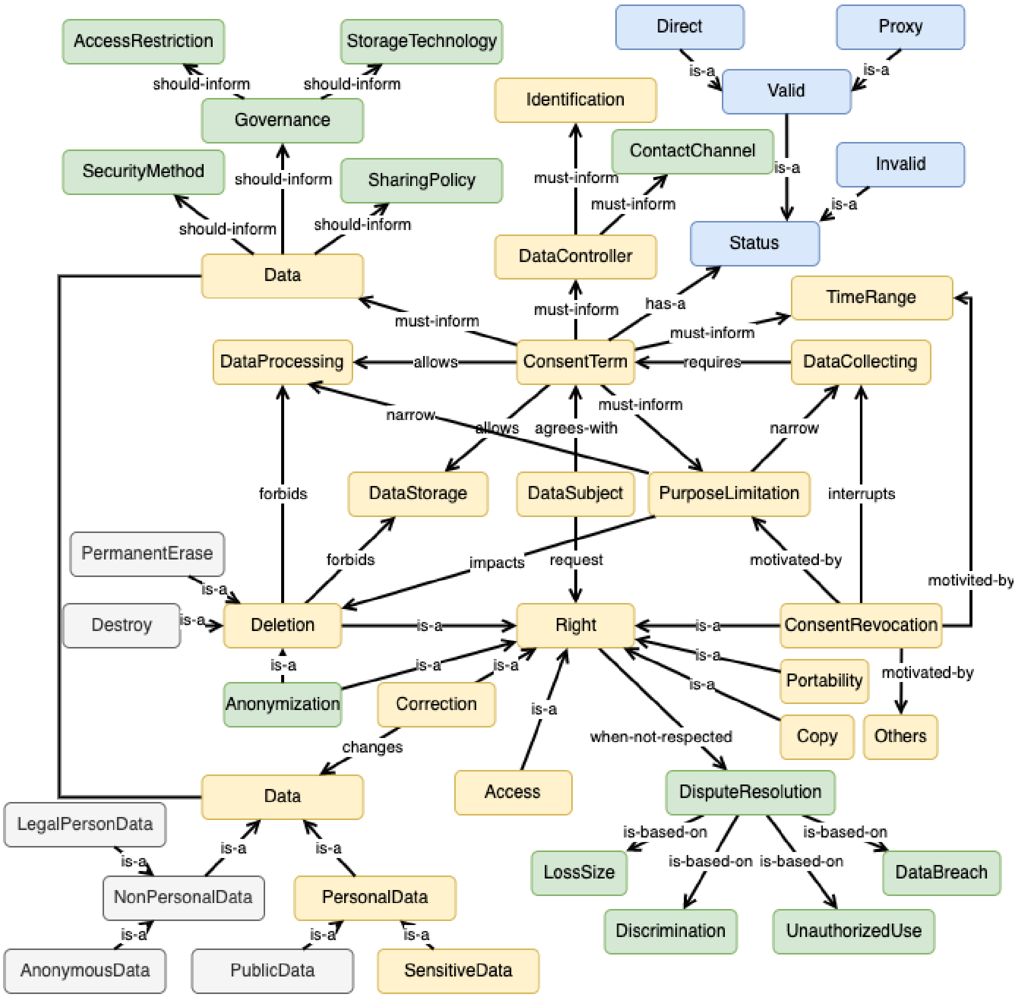


Figure 5.3: Consent Metamodel - Consent module overview.

policies and security methods explicitly in the use cases as well as the dispute resolutions that can emerge from violation of the Law.

However, there are two definitions regarding anonymization. PrOnto considers anonymization a deletion action, and LGPD does not. In this sense, we decide to represent both concepts in our CM. These entities are essential to understand the environmental factors related to the scenario execution and explanation.

The gray entities are those that the LGPD does not consider specifically as a concern. For example, the methods of data deletion and non-personal data are not detailed by LGPD. Furthermore, the Brazilian Citizenship Ministery[4] defines a concept not mentioned on GDPR, i.e., *PublicData*; however, this concept is not defined in the Law n. 13.709/2018, Lei Geral de Proteção de Dados Pessoais. Thus, we decided not to insert this entity into our CM.

[4] https://www.gov.br/cidadania/pt-br/acesso-a-informacao/lgpd/classificacao-dos-dados

In summary, we removed the *NonPersonalData, AnonymousData, LegalPersonData, PublicData, PermanentErase,* and *Destroy* entities as they are not deeply approached in the LGPD.

Overall, as the proposed metamodel is based on the Consent legal bases, the *ConsentTerm* and the *Right* entities are the central points; they have many connections with other concepts. For instance, any change in the consent term will impact many entities around; hence, it will require a new DS's approval. In this sense, depending on the DS will, s/he can disagree, and it will interrupt the data collection. Still, if the data controller does not stop collecting the DS's personal data, it will violate its rights, and fines can be applied to the data controller.

5.1.2
Action Module

As the second module in our CM, the *Action* module was developed based on the PrOnto's Processing module. PrOnto ontology presents many modules to describe the ontology entities; however, it might be hard for common people, aka citizens, to understand the main concerns when they share their personal data and know the reasons behind an action execution. Therefore, the Action module proposed aims to clarify, through examples, the scenario context attributes, and it contributes to the explanation processes. Furthermore, the use of this module enables the structured construction of a broad knowledge database to share the scenarios' particularities. To do so, we gather the PrOnto definitions to create the Action Module, as depicted in Figure 5.4. Our scenario presents five pillars: Agent, Action, Consent Term, Right, and DeonticOperator, which are be detailed below in yellow:

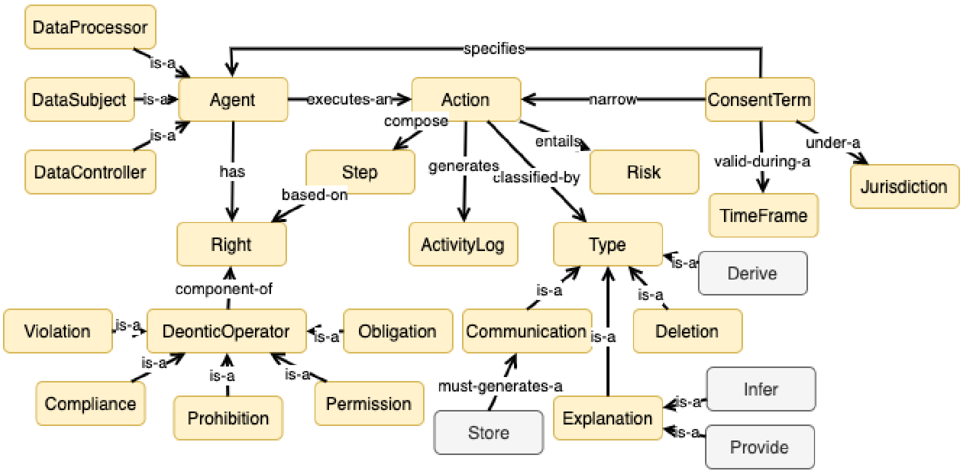


Figure 5.4: Consent Metamodel - Action Module overview.

Agent. The pragmatic circumscription has to define the agents that will be involved, *i.e.*, who are *DataSubject*, *DataController*, and *DataProcessor*. Thus, the *Agent* entity is crucial to define the entities that are performing actions in the proposed scenario.

Action. Actions are narrowed by the consent term, which informs the *TimeFrame* and *Jurisdiction* which the consent term is valid. Moreover, the actions may entail *Risk*, such as the risk of a data leak, and the agent should be aware of the impact of sharing data. Still, actions are composed of *Step*, which are executed based on the current set of *Right* available for the agents and persisted by the *ActivityLog*. Finally, *Action* can be classified by types, which will help the explanation process by filtering the activity log.

Consent Term. As the study object, the consent term has an important role in defining all required information to let the DS be aware of data sharing conditions, narrowing the data controller actions and context of using the DS's information.

Right. The agents may have different rights depending on the classification, time frame, and previous actions. The Deontic Operators complement the rights. For instance, if a DS agent requested for consent revocation and the requisition was accomplished, there is no consent to be revoked. In this case, the agent will not have the right to revoke his/her consent, as one consent term cannot be revoked twice.

Deontic Operator. The deontic concepts define if there *Obligation*, *Prohibition*, and *Permission*. Furthermore, PrOnto includes *Violation* and *Compliance* to express a violation or compliance with the data regulation evaluated. Figure 5.5 depicts the deontic logic construction based on alethic modal notions McNamara (2006); Žarnić and Bašić (2014). In the hexagon of logical relations, the dashed line expresses contradiction, the dotted line indicates contrariety relation, the full line represents subcontrariety, and the arrow represents implication. For instance, this figure shows what *Obrigatory* *p* means, *i.e.*, in our context the action *p* is not-optional and is the opposite of forbidden, *p* is permitted, and *p* is a contradiction of gratuity. For instance, even though a DS has no valid consent with a DC, the Deontic Concepts will be applied indicating that there the DC is prohibited from processing the DS's data. Therefore, the usage of the deontic logic allowed the action classification in the application scenarios.

Even though PrOnto lists the entities *Communicate* and *Observe* in the *Agent* module, these entities were not explained in their documentation nether exemplified their uses. In this sense, we changed the entities' names to *Communication* and *Explanation*, as they are entities related to types of

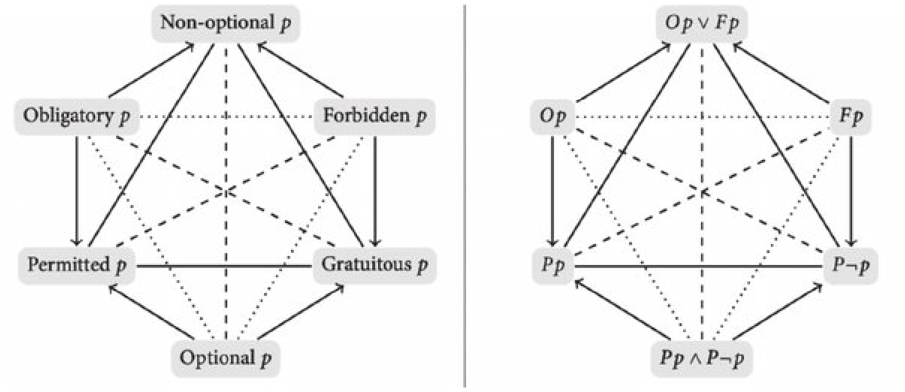


Figure 5.5: Deontic Logic Žarnić and Bašić (2014).

action. The *Communication* entity aims to record the actions and transmits such actions to the environment, *i.e.*, it creates a communication action to persist the operation. The *Explanation* entity aims to record the impacts and consequences of a specific action. After an action is performed, the agents can check the action's motivation and impact. A *Communication* might require an *Explanation*, and the *Explanation* requires a previous *Communication*.

The gray entities were mentioned in the PrOnto ontology, but there are no details regarding their use. Although we could try to infer what they mean, the result could not respond to the original author's perspective; so, we decided to keep those entities in the CM, but not explore them. However, the CM proposed is free for modifications if needed. Its essence is collaborative, and it should be adapted considering the law evolution and the jurisdiction in which the CM is applied.

5.1.3 Log Module

The third module of CM is the *Log* module. This module was proposed to record all actions executed by the agents. It allows query execution to explain the decision-making process until the action execution. Still, this module aids in identifying inconsistencies in application scenarios. For instance, if there is an external agent not mentioned in the consent term and he executed an action, there will be a record in the *Log* indicating such action. Hence, this evidence demonstrates that the *Consent* module or the *Action* module should be reviewed in order to fix such inconsistency.

As the *Action* module, the *Log* module presents the *DeonticOperator* entity in order to build an informative record, as the *ActionDescription* is a not structured text entity. Moreover, the description is in natural language; hence, the entities *ActionTimestamp*, *ActionType*, and *DeonticOperator* deliver struc-

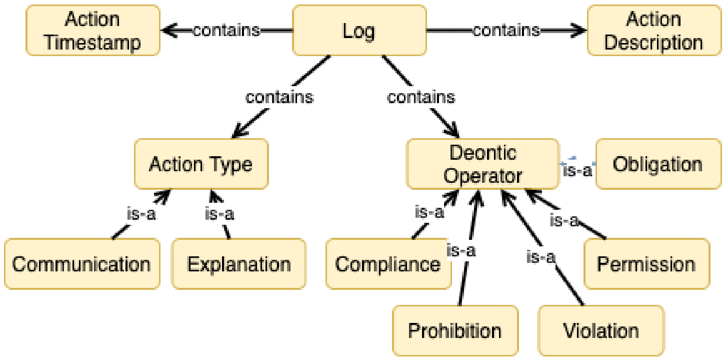


Figure 5.6: Consent Metamodel - Log Module overview.

ture attributes to allow queries' execution. Last but not least, the *ActionType* distinguishes the communication action from the explanation action, *i.e.*, the explanation of impacts and consequences.

5.2
GoDReP - Scenario Generation Structure

In order to create an environment to exercise CM and use case scenarios, we developed Generation of Data Regulation Plots (GoDReP). A framework is a supporting structure around which something can be built, or a system of rules, ideas, or beliefs that are used to plan or decide something[5]. Thus, a framework would run scenarios developed based on the CM. Hence, agents could use it to mitigate the informational data flow asymmetry in scenarios ruled by a data protection regulation exploring the rules, ideas, and beliefs.

In this chapter, we will detail GoDReP. This framework aims to use CM to develop CM in a structured manner to evaluate the computational adherence and to bridge the gap between DSs, DCs, and DPs (*RQ2*). Moreover, GoDReP aims to help the agents to explore a use case scenario at different times in the timeline, *i.e.*, before, during, and after the given consent.

These scenarios will exemplify the understanding of a specific theme to mitigate the informational asymmetry in scenarios ruled by a data protection regulation, such as LGPD and GDPR. Moreover, they enable the evaluation regarding which scenarios' attributes are general and which ones are specific. Still, the use case scenarios can be stored in an open repository to allow agents to contribute by assembling scenarios with different perspectives and concerns. Therefore, a standard to develop such scenarios is vital to follow the semantics proposed in such CM, and GoDReP conducts the agents towards this principle.

[5] Dictionary definition of framework `https://dictionary.cambridge.org/pt/dicionario/ingles/framework`

5.2.1 Scenario Instantiation

To create a scenario, as depicted in Figure 5.7, GoDReP proposes five macro processes: (i) *Scenario Description*, which aims to identify the agents, purpose, time range, personal data, storage technology, security methods, access restrictions, third-party sharing policies jurisdiction, consent compliance requirements based on the jurisdiction; (ii) *Macro Process Definition*, *i.e.*, the step-by-step design to be executed by the agents; (iii) *Process Execution*, *i.e.*, the record of the scenario's facts seeking regulation compliance; (iv) *Impact Exploration*, *i.e.*, the evaluation of the impacts after the *Process Execution*, and (v) *Advanced Exploration*, which aims to explore other scenarios to offer evaluation regarding different possible situations.

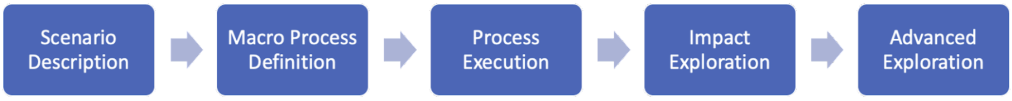


Figure 5.7: GoDReP macro process.

Moreover, Figure 5.8 depicts the *Advanced Exploration* process, which proposes the insertion of a new fact and environment impact analysis. At the end of each advanced sub-scenario, the new facts are removed, and the environment turns back to the basic scenario state. Hence, the advanced scenarios are independent of each other, based on the same basic environment.

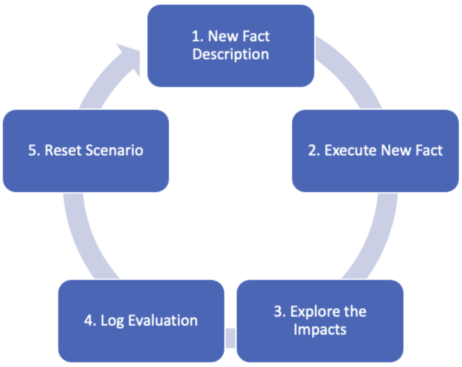


Figure 5.8: GoDReP advanced exploration.

Therefore, in order to answer our RQ2, GoDReP was developed using the combination of Prolog language and the Jupyter Notebook tool. Prolog is a descriptive and prescriptive programming language based on first-order logic and formal logic to express relations and represent facts and rules (Clocksin and

Mellish, 2003). The compliance between the ties, facts, and rules is achieved by running queries over these relations, evaluating which relationships are "true", and which formal relationships and objects occur in the proposed environment.

5.2.2 Computational Representation

In this sense, Prolog is convenient for exploring rule-based logical queries, and it can support data regulation interpretation. For instance, when we say "John agrees with the consent term", we communicate that a relationship, or an agreement, exists between one object "John" and another individual object "consent term". Moreover, Prolog allows asking questions, such as "Did John agree with the consent term?", to determine this relationship value.

However, some relationships do not always mention all the involved objects. For instance, in "John is an adult" we specify a relationship, called "being considered an adult", which involves John. However, no one mentioned who considered John an adult or why, and Prolog allows it; what the computer will accomplish depends on the amount of detail provided.

Another example developed in Prolog is depicted in Figure 5.9. The query presented in the such figure represents the question "What is the specific purpose in the consent term that has Bank B as DC, John as the DS, and offer_products_and_services as purpose", then the program will return the answer "create_specific_offers".

```
?- specificPurpose('Bank B',                              % Data Controller
                   'John',                                % Data Subject
                   'offer_products_and_services',         % Purpose
                   SPECIFICPURPOSE).                      % What we are looking for
```

```
>> Output: SPECIFICPURPOSE = create_specific_offers .
```

Figure 5.9: Prolog query example.

Prolog is a valuable programming language that allows fast rules and facts instantiation development. Thus, we built the relationships in Prolog based on the CM. Additionally, the agents could add other relationships, and even new entities could be used in the program as well as other data protection regulations since the ontology alignment happens.

Even though Prolog allows the creation of rule-based logical queries, there could be a gap between the program and the final user, who is not obligated to know the specificities of software development. Moreover, even good code documentation may not be clear enough to the final user. Therefore, to deliver more context and documentation tools to improve user awareness, we developed a Prolog program using the Jupyter Notebook, which is a tool often applied in a data science context and in the learning processes to support the workflow of scientific computing (ClocKluyver et al., 2016; Randles et al., 2017; Perkel, 2018; Cardoso, Leitão, and Teixeira, 2018).

Notebooks enable the interactive exploration to publish a detailed record of the computational execution. The code in a notebook is organized into a markdown structure and cells, *i.e.*, chunks that can be individually modified and run. The cell outputs are located directly below each cell, and they are stored as part of the document ClocKluyver et al. (2016).

Additionally, Jupyter is an open-source project, which allows the users to access the program by browsers and execute different programming language codes based on a kernel. Many kernels have already been developed, such as C++, Python, Bash, and Prolog. Still, the browser enables the use of the same interface locally or on a remote server, allowing access from people who do not have a server to run the code.

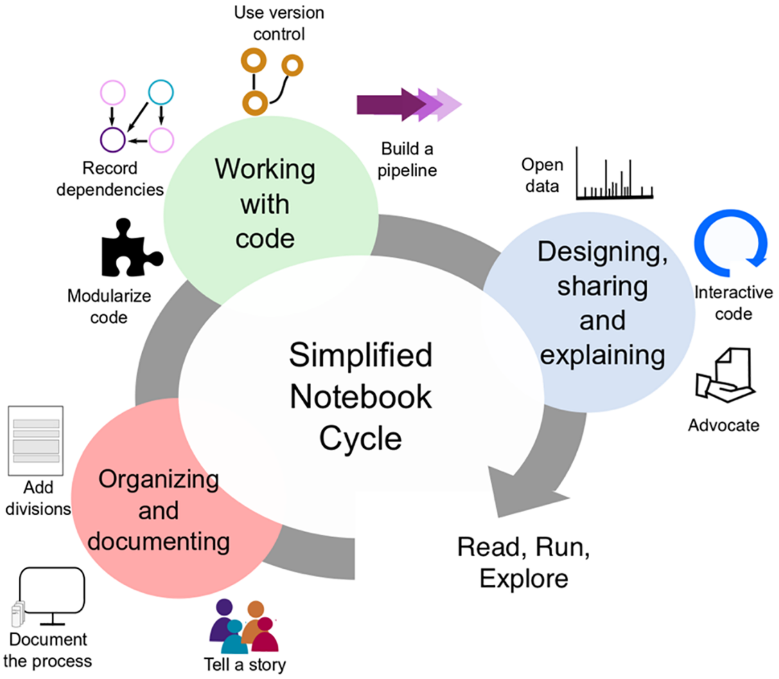


Figure 5.10: Jupyter Cycle (Rule et al., 2019).

One of Jupyter Notebook's goals is to turn scientific findings reproducible, *i.e.*, users could access and rebuild the code from raw source and get the same result (Wang et al., 2020). In this sense, as depicted in Figure 5.10, Rule et al. (2019) proposed ten rules for writing and sharing computational analyses in Jupyter Notebook: (i) tell a story for an audience, *e.g.*, the GoDReP *Scenario Description* process; (ii) document the process, not just the results, *e.g.*, the GoDReP *Macro Process Definition* process; (iii) use cell divisions to make steps clear; this rule was addressed internally in the notebook using the markdown language; (iv) modularize code, *e.g.*, we developed functions to avoid duplicating code, as the authors recommended in this rule; (v) record dependencies, *e.g.*, we manage our dependencies using Pip and Virtual Env[6]; (vi) use version control, *e.g.*, we used Git in a open source repository; (vii) build a pipeline, *e.g.*, as demonstrate in GoDReP macro process (Figure 5.7); (viii) share and explain your data; (ix) design your notebooks to be read, run, and explored, and (x) advocate for open research. These rules ensure that the construction of the notebook is reproducible and aligns with the GoDReP macro process.

Therefore, Prolog and Jupyter Notebook offer pragmatic and semantic computing resources to enable the instantiation of the GoDReP framework respecting the CM and delivering a possibility to generate high-level documentation and reproducible code. However, although the framework may indicate a monotonic process execution, the application scenarios might require changes in the internal processes, *i.e.*, in these cases, the internal process construction will be different from the previously created. Moreover, these changes require a user able to change the Prolog code, *i.e.*, a user with programming logic skills.

5.2.3 Notebook Scenes

The notebook scenes respect the GoDReP macro process as well as the advanced exploration. They have a basic module and an advanced module. The former is divided into scenes and the latter explores the insertion of a new set of information to evaluate their impact. The basic module was designed following the structure below.

[**Scene 0**]. Describe the pragmatic circumscription in natural language. This description aims to contextualize the reader regarding the environment, and it includes the macro process definition to show an overview of the other scenes. For instance, Figure 5.11 depicts the scene macro process.

[**Scene 1**] Set consent term. This scene describes the consent term highlighting the consent requirements and it requires that whoever is building the consent term knows the data protection regulation in which the scenario is being created. Moreover, this scene initiates the Prolog code construction, representing the consent term in this programming language code.

[**Scene 2**]. Simulate a DS agreement. In this scene, the DS can verify programmatically if the consent term has all items foreseen in the data

[6] `https://packaging.python.org/en/latest/guides/installing-using-pip-and\-virtual-environments/`

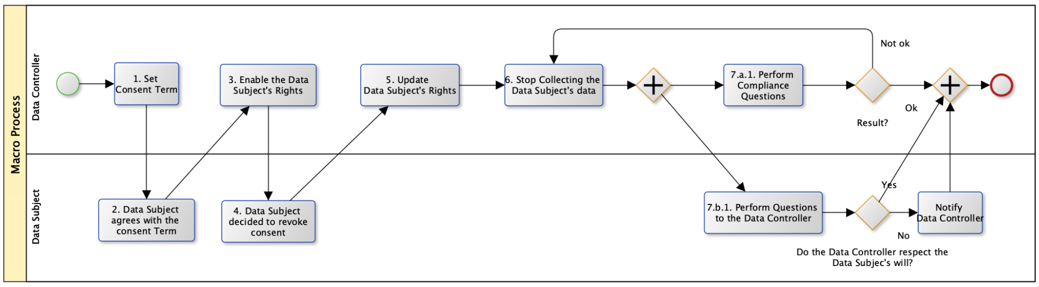

Macro Process
Data Controller
Data Subject
1. Set Consent Term
3. Enable the Data Subject's Rights
5. Update Data Subject's Rights
6. Stop Collecting the Data Subject's data
7.a.1. Perform Compliance Questions
Not ok
Ok
Result?
2. Data Subject agrees with the consent Term
4. Data Subject decided to revoke consent
7.b.1. Perform Questions to the Data Controller
Yes
No
Notify Data Controller
Do the Data Controller respect the Data Subjec's will?

Figure 5.11: Scene macro process.

protection regulation and simulates an agreement. Moreover, questions about whether the DS is, indeed, able to agree with the consent term can emerge in this scene. Depending on the jurisdiction, the DS can not be considered an adult and may request an acceptance of a legal person in charge.

[**Scene 3**]. Defining the DS's rights. This scene defines the DS's rights based on the data protection regulation. The agent can change any right depending on the data protection regulation he is evaluating.

[**Scene 4**] Revoke consent. In this scene, the DS request to revoke his/her consent. Hence, the DC has to abide by the DS's request and stop collecting and processing the DS's data.

[**Scene 5**] Impact evaluation. This scene explores the impacts of the previous scenes performing questions in Prolog to evaluate the facts. As depicted in Figure 5.12, the red entities suffered impact directly or indirectly when the consent is revoked. First, the data controller must stop collecting personal data immediately. Next, the data controller must update the sharing policies and access restrictions to prevent unauthorized access or new data processing. Still, the consent status will change to "invalid", as the controller cannot use this consent anymore.

The advanced module explores the negotiation scenarios with parameters other than the basic module. Hence, this module is composed of cause-effect scenes to evaluate access and processing confirmation, compliance, and information about consent term. Those questions would aid DSs and DCs to exercise their understanding regarding possible scenarios and the evaluation of their actions. Moreover, for each "true" or "false" returned in the Prolog query, we added an explanation pointing to the activity that motivated such a result.

To do so, we proposed nine cause-effect scenes to perform the advanced explorations:

Consent revocation not respected. In this scene, the agents can explore how they could identify if the consent revocation request was not

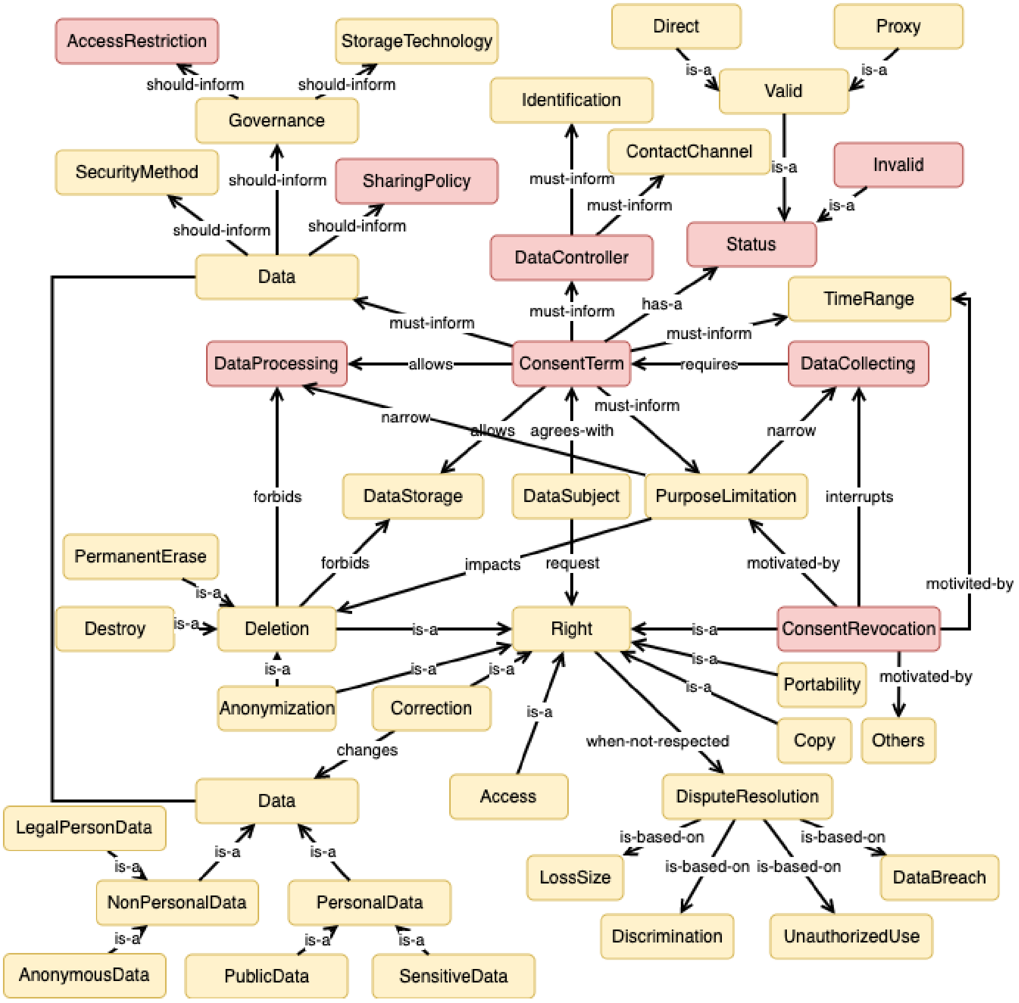


Figure 5.12: Consent revocation impact.

respected and the consequences in this case. Figure 5.13 depicts the revocation process. After the consent revocation request, fines should be applied to the data controller if it was not accomplished. To do so, some questions could be performed in Prolog. Figure 5.14 shows a question regarding data processing, *i.e.*, if the data controller is processing data. This query returned false, and the explanation can be found in the log, *i.e.*, the log will report that the DS requested to revoke his consent. Moreover, the agents can perform other questions to the environment in order to produce more explanation possibilities.

Data breach, what to do? In this scene, we proposed a process to enable the DS to verify which companies have his/her data and check the consent term to confirm if there is a sharing clause that allows the DC to share such data. Moreover, the DC must inform the national authority and the DS when a data breach occurs that may cause risks or damage to him/her. Such communication has to be done as soon as possible and should inform: (i)

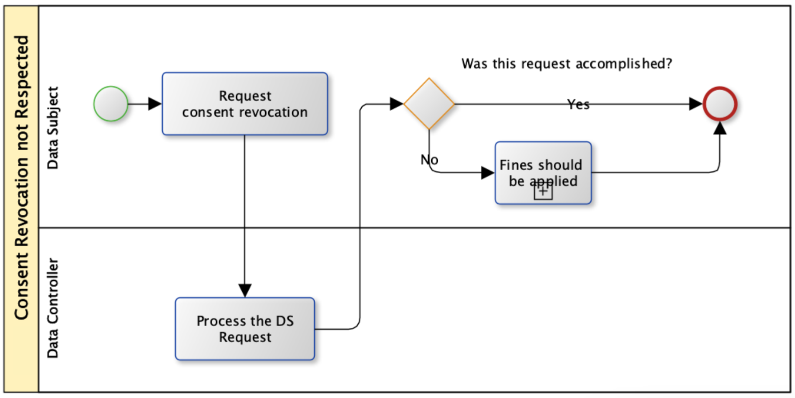
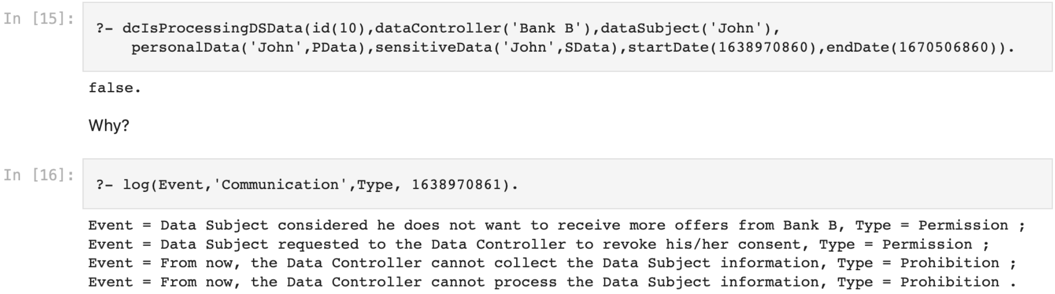

Consent Revocation not Respected
Data Subject
Data Controller
Request consent revocation
Process the DS Request
Was this request accomplished?
Yes
No
Fines should be applied

Figure 5.13: Consent revocation process.



```
In [15]:  ?- dcIsProcessingDSData(id(10),dataController('Bank B'),dataSubject('John'),
              personalData('John',PData),sensitiveData('John',SData),startDate(1638970860),endDate(1670506860)).

          false.

          Why?

In [16]:  ?- log(Event,'Communication',Type, 1638970861).

          Event = Data Subject considered he does not want to receive more offers from Bank B, Type = Permission ;
          Event = Data Subject requested to the Data Controller to revoke his/her consent, Type = Permission ;
          Event = From now, the Data Controller cannot collect the Data Subject information, Type = Prohibition ;
          Event = From now, the Data Controller cannot process the Data Subject information, Type = Prohibition .
```

Figure 5.14: Prolog consent revocation question example.

personal data category; (ii) what data were leaked; (iii) what were the technical and security; (iv) measures used to protect data; (v) the risks related to the incident; and (vi) what the data controller will do to revert or mitigate the damage.

Furthermore, the omission of any fact related to informing the DSs about unauthorized access or neglecting the system security could result in fines applied to the data controller. Last but not least, if the data controller notices a data breach, once informed, the data controller has to act immediately (LGPD Art. 48). Depending on the incident severity, the Data Controller will have to disclose such an event in high-impact communication media. In this sense, as depicted in Figure 5.15, we proposed a situation that the data controller suffered from a hacker attack and DS's personal data were leaked on social media, and he/she is receiving a few calls from different numbers. So, DC is obligated to inform the incident to ANPD (Brazilian Data Protection National Agency) and inform the DS that his/her phone number was leaked. Furthermore, even if the DS has revoked his/her consent, s/he has to be informed regarding the data breach as his/her data is still on the data controller's database. Thus, this scene will impact the entities as depicted in

Figure 5.16. First, as mentioned before, a data breach event must be informed to all agents impacted. This message must contain the security methods and the storage technologies applied to avoid a data breach as depicted by the red entities. However, this event could trigger other impacts, which are represented in blue. For instance, after a data breach, the DS could enter into a dispute resolution claiming discrimination, loss, and unauthorized use of his/her data. Furthermore, the DS might request changes in the consent term, impacting the sharing policies and access restrictions. Also, the DS might request consent revocation and data deletion, which affects data collecting, processing, and storing.

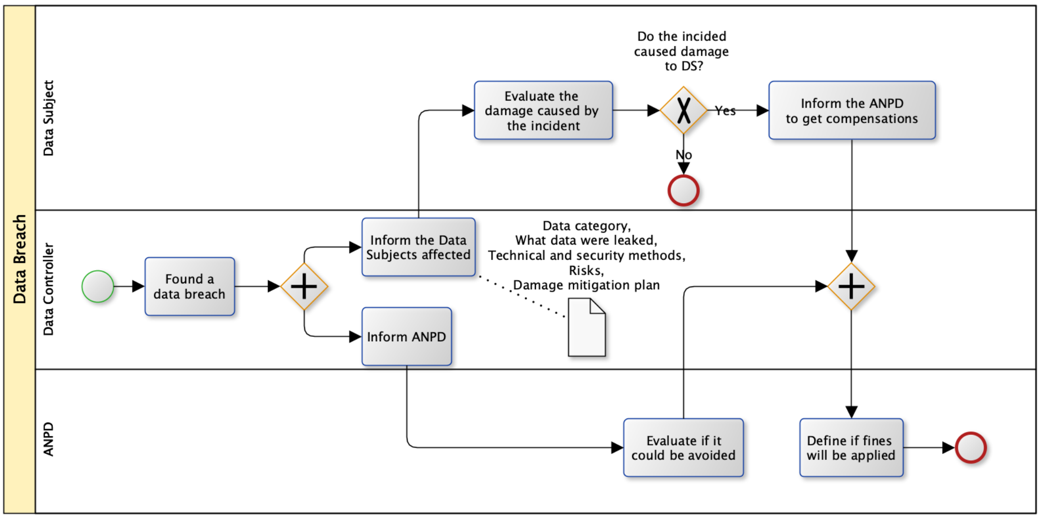

Data Breach
Data Subject
Data Controller
ANPD
Do the incided caused damage to DS?
Evaluate the damage caused by the incident
Yes
No
Inform the ANPD to get compensations
Found a data breach
Inform the Data Subjects affected
Inform ANPD
Data category,
What data were leaked,
Technical and security methods,
Risks,
Damage mitigation plan
Evaluate if it could be avoided
Define if fines will be applied

Figure 5.15: Data breach mitigation process.

Evidencing data leak. In this scene, the agents can verify how the agents could behave if the data was leaked. We considered the same process presented in the data breach, where Figure 5.17 shows the process of evidencing data leaks. To create concrete evidence that a Data Controller leaked a DS's data, first, it is important to verify who has such data. If there is just one data controller legally storing such data; hence, the chances that such a DC has leaked personal data are elevated. Moreover, the data controller is obligated to inform if personal or sensitive data is stored in the database. The DS can request such information for each data controller. This scene will impact the same entities as depicted in Figure 5.16.

Requesting data correction. In this scene, the DS request to change the shared data and verifies if it was accomplished, as depicted in Figure 5.18. Moreover, even if the DS revokes his/her consent, the data will not be deleted – an express data deletion request is required. So, to check if the data correction request was accomplished, the DS should call the *Data Access*

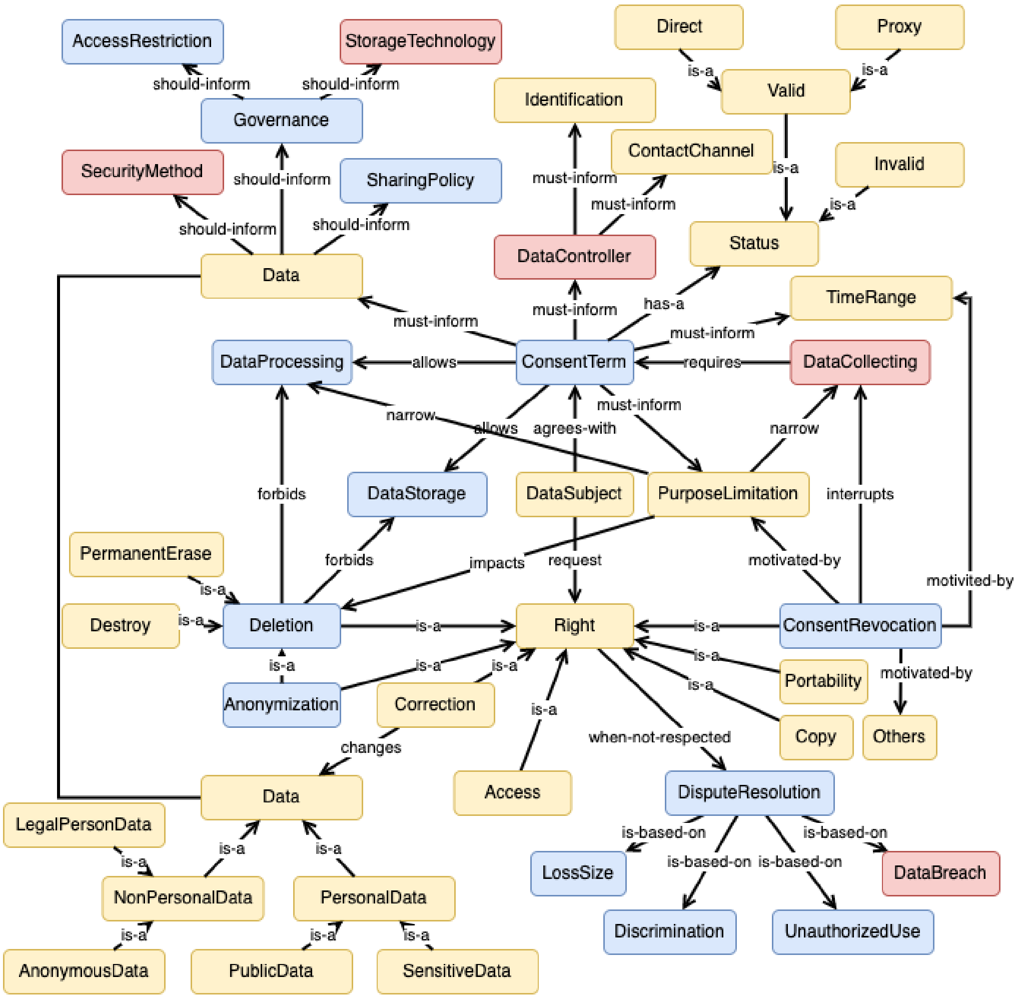


Figure 5.16: Data breach impact.

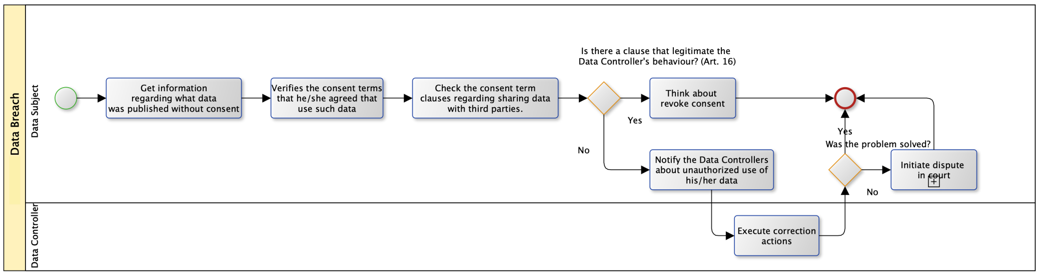

Data Breach
Data Subject
Data Controller
Get information regarding what data was published without consent
Verifies the consent terms that he/she agreed that use such data
Check the consent term clauses regarding sharing data with third parties.
Is there a clause that legitimate the Data Controller's behaviour? (Art. 16)
Yes
Think about revoke consent
No
Notify the Data Controllers about unauthorized use of his/her data
Yes
Was the problem solved?
No
Initiate dispute in court
Execute correction actions

Figure 5.17: Data leak identification process.

right. The DC is obligated to abide by the DSs' requests as correction as data access. Also, the controller is obligated to inform all processors regarding the correction. Thus, this scene will impact the entities as depicted in Figure 5.19. The data correction scene impacts data processing and storage, as the personal or sensitive data were changed. Therefore, DCs and DPs should also verify if the DS's copy is updated.

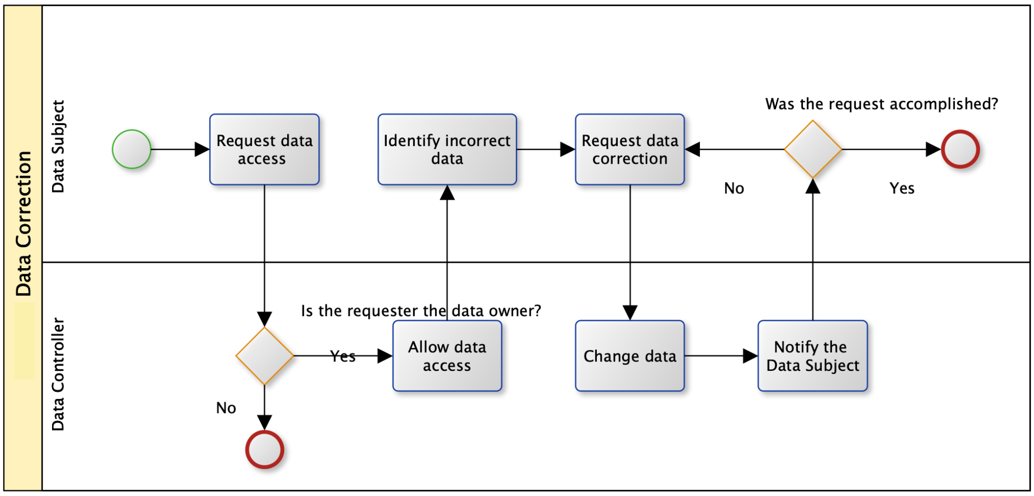


Figure 5.18: Data correction process.

Requesting data anonymization. In this scene, Figure 5.20 depicts the DS request to anonymize the shared data. Hence, once the data is anonymized, the data controller will not have the resources to provide details about such data, including correction. After this request, the data controller is not obligated to comply with requests that should involve reidentification actions. Moreover, the controller is obligated to inform all processors regarding the anonymization. Finally, questions regarding the anonymization algorithms could emerge, but this is not the focus of this work. This work focuses on the causes and consequences, understanding possible scenarios. Finally, this scene will impact the entities as depicted in Figure 5.21. Data anonymization impacts almost all DS's Rights. The anonymization process may turn the personal data not identifiable anymore if it is made properly, i.e., avoiding reidentification. Hence, the anonymized data is out of LGPD's scope. In this sense, requests related to data access, deletion, correction, portability, or copy, may not be answered by the DC, as the Controller might not identify the DS anymore.

Data deletion. In this scene, Figure 5.22 depicts the DS request to data deletion; however, the LGPD Art. 16 legitimizes the data controller to keep the personal data stored in the database in the following situation: (1) compliance with a legal or regulatory obligation by the controller; (2) study by a research institution, ensuring, whenever possible, the anonymization of personal data;

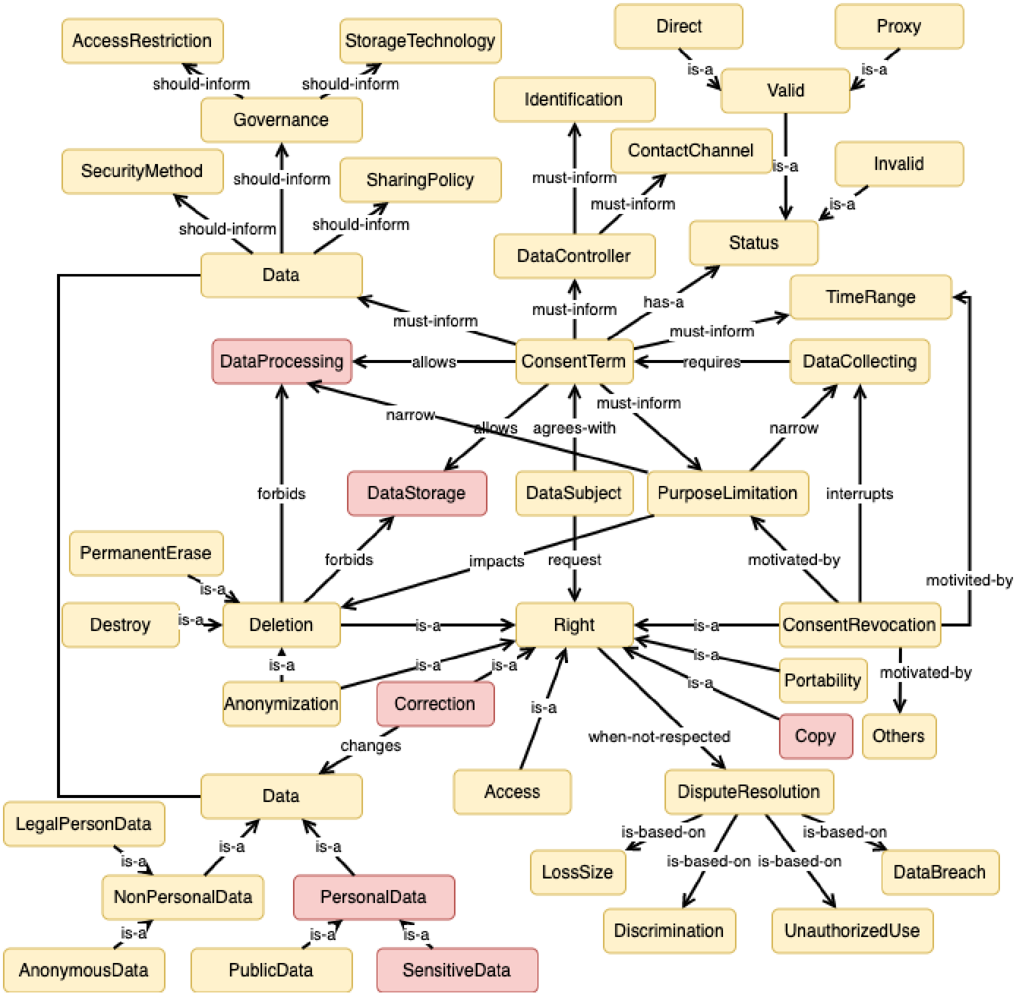


Figure 5.19: Data correction impact.

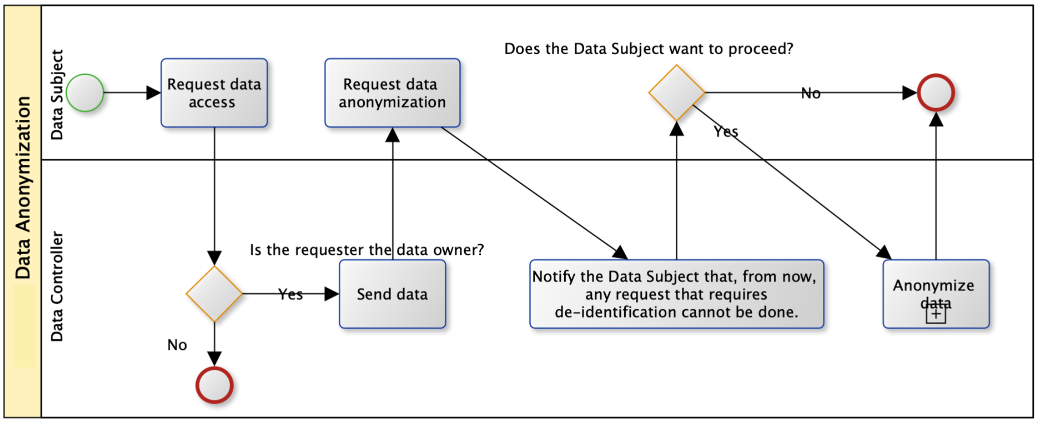

Data Anonymization
Data Subject
Data Controller
Request data access
Request data anonymization
Does the Data Subject want to proceed?
No
Yes
Is the requester the data owner?
Yes
Send data
No
Notify the Data Subject that, from now, any request that requires de-identification cannot be done.
Anonymize data

Figure 5.20: Data anonymization process.

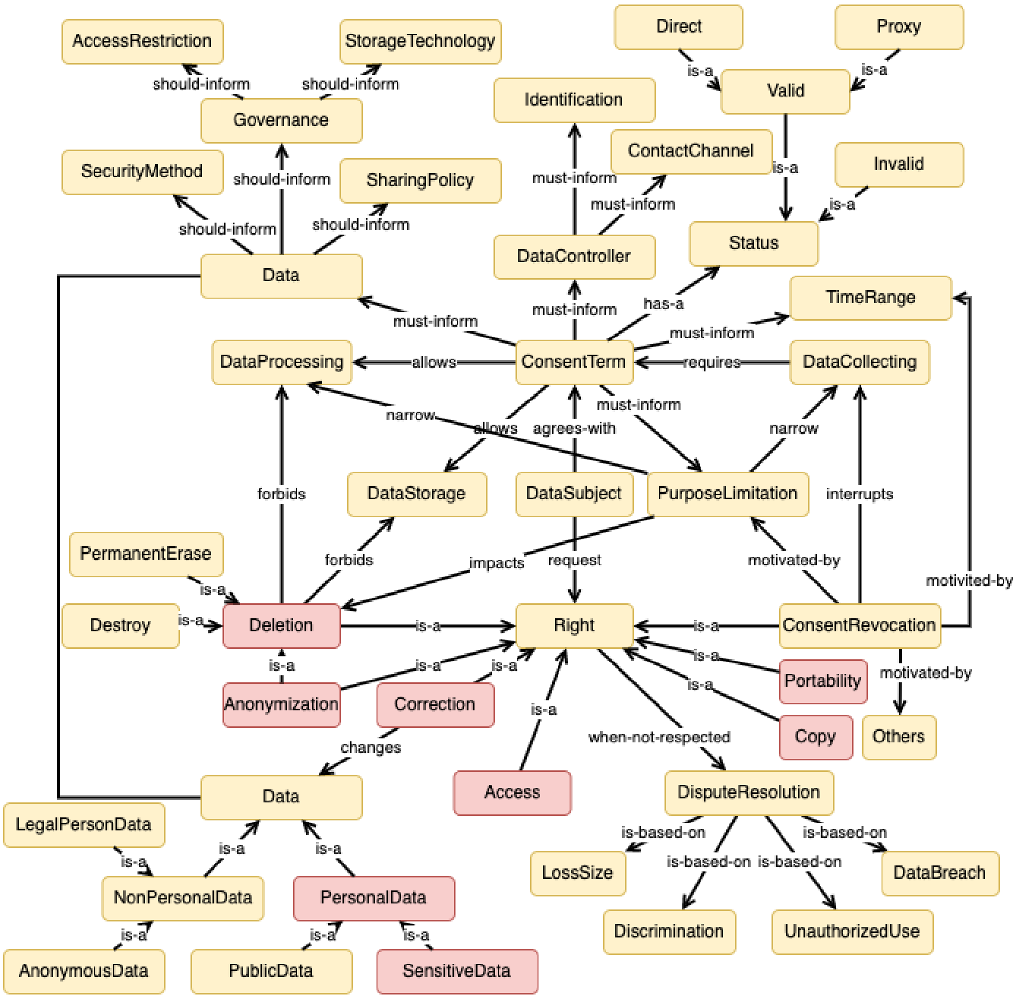


Figure 5.21: Data anonymization impact.

(3) transfer to a third party, provided that the data processing requirements set out in this Law is respected; or (4) exclusive use of the controller, its access by a third party is prohibited, and anonymization is also required. As depicted in Figure 5.23, the request for data deletion may impact differently depending on the purpose limitation. For example, the data storage may have to anonymize the data. Moreover, if the data is deleted or anonymized, the Data Controller cannot achieve requests related to data correction, portability, and copy anymore.

Technology unavailability. This scene evaluates the consequences when a system is offline. Companies are vulnerable to technical faults, unavailability, or security breach. In this sense, DSs might be impacted by technology troubles. In some cases, the technology unavailability may not impact DSs but only internal companies' processes. As depicted in Figure 5.24, in this scenario, we propose a simulation of a technology unavailability event, *i.e.*, that Company B's cloud server, which has personal data storage, is offline. Inter-

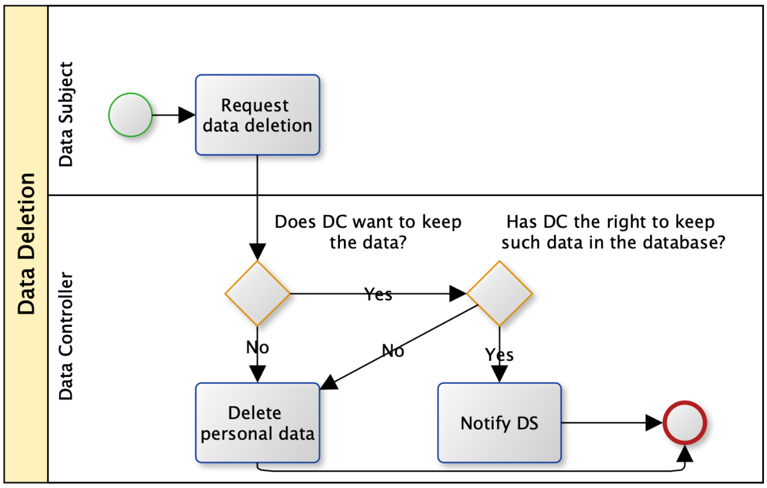


Figure 5.22: Data deletion process.

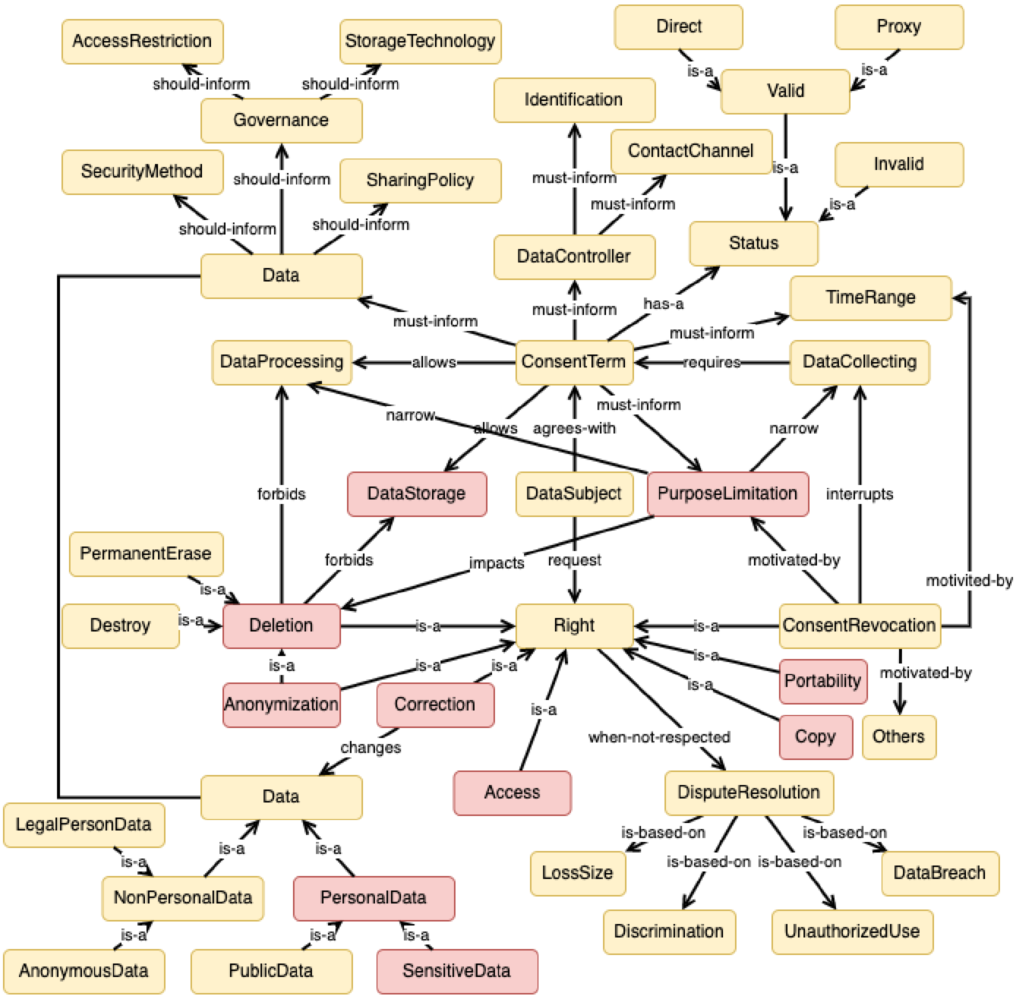


Figure 5.23: Data deletion impact.

nally, Company B suffered a high impact of this unavailability; all systems that depend on this database are offline, *i.e.*, the internal data governance is jeopardized/ compromised. Hence, employees and clients cannot access any internal system. As depicted in Figure 5.25, besides the governance, data unavailability may directly impact the users' rights. For example, the DC and DP cannot delete or execute data corrections if the system is unavailable. Moreover, if a data controller requests for portability, anonymization, or portability, the data controller will not be able to attend to such requests as fast as expected; a considerable delay is expected, instead. Furthermore, fines can be applied depending on the delay, but they should be evaluated case by case. Finally, no previous work proposed this scene; however, given the network and energy instability that may emerge, we proposed this new scene topic.

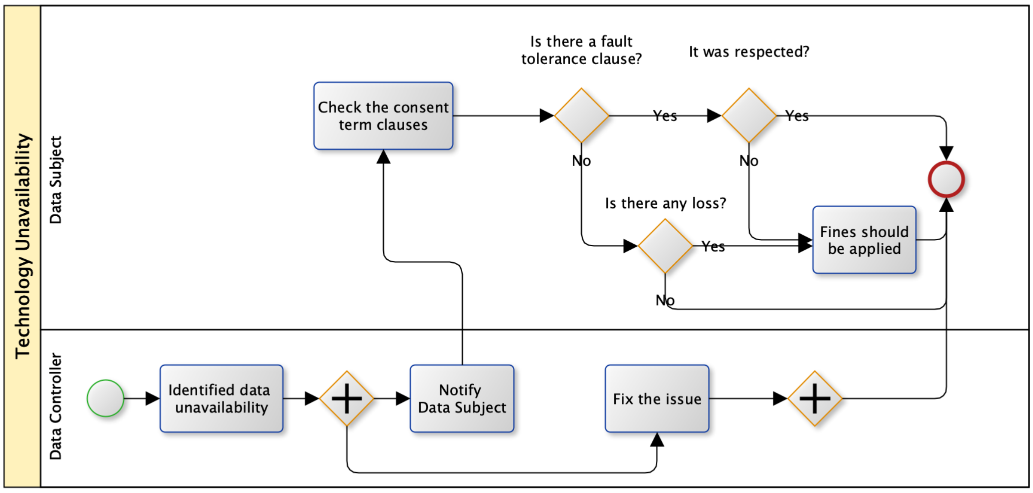


Figure 5.24: Technology unavailability process.

In this chapter, we will detail GoDReP. This framework aims to use the CM to develop use case scenarios in a structured manner to evaluate the computational adherence and to bridge the gap between DSs, DCs, and DPs (*RQ2*). Moreover, GoDReP aims to help the agents to explore the circumscription at different times in the timeline.

Inconsistent behavior. As well as the previous scene, this is a new scene. As depicted in Figure 5.26, this scene proposes the evaluation when the DS presents inconsistent behavior, *i.e.*, s/he agrees and revokes the consent term frequently in a short time. These actions may indicate that the DS is performing some illegal or immoral action. This scene will impact the entities as depicted in Figure 5.27. The inconsistent behavior may cause multiple requests in the data controller's servers and may turn the servers offline depending on the volume of requisitions. The latter, combined with multiple users, bots, or zombie machines, would characterize a distributed denial of service attack

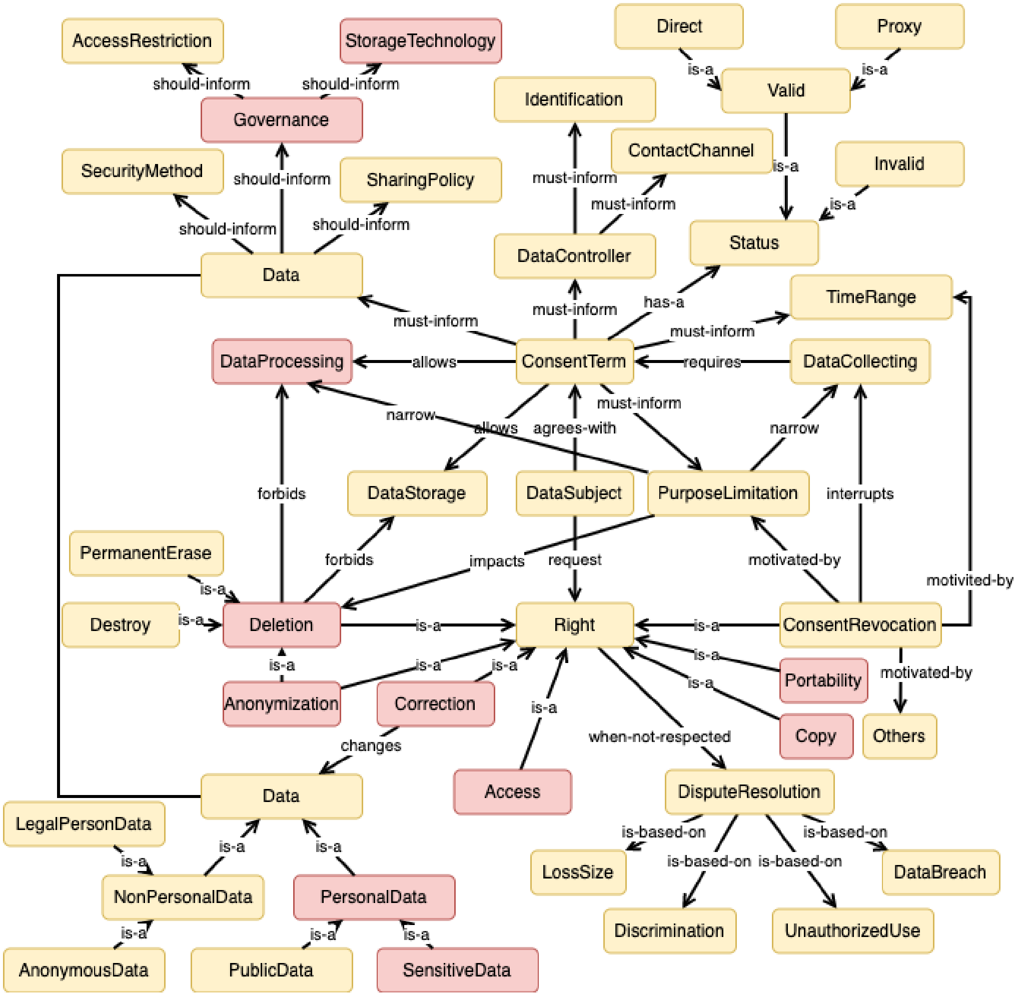


Figure 5.25: Technology unavailability impact.

(DDoS) Yu et al. (2013). Thus, this behavior would impact data processing, security methods, and storage technology. Hence, the DS's rights will also be impacted if the system is offline.

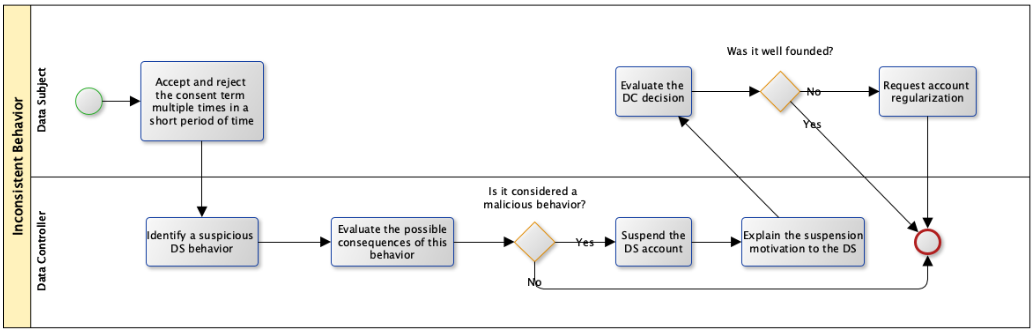

Inconsistent Behavior
Data Subject
Accept and reject the consent term multiple times in a short period of time
Was it well founded?
Evaluate the DC decision
No
Yes
Request account regularization
Data Controller
Identify a suspicious DS behavior
Evaluate the possible consequences of this behavior
Is it considered a malicious behavior?
Yes
No
Suspend the DS account
Explain the suspension motivation to the DS

Figure 5.26: Inconsistent behavior process.

Data portability. This scene can be explored in at least two ways. First, as cellphone companies, data portability means migrating the DS phone

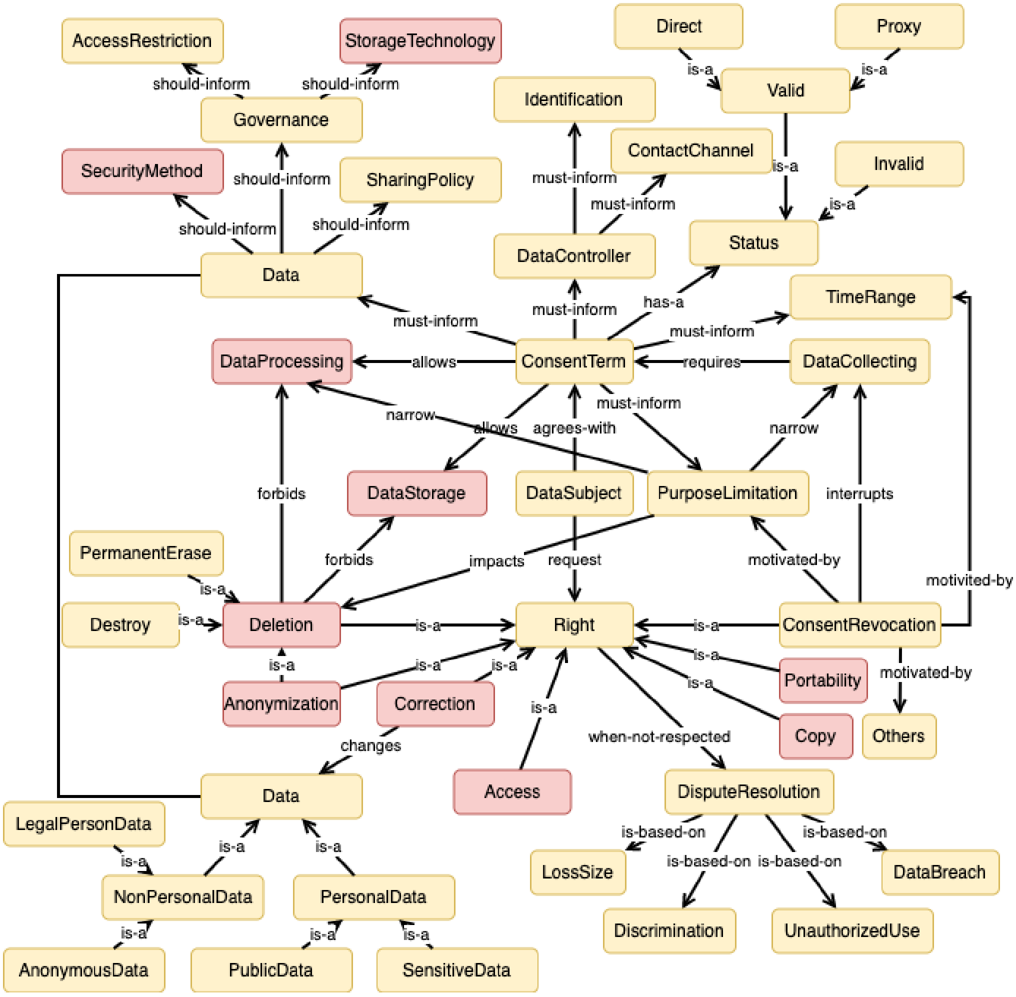


Figure 5.27: Inconsistent behavior impact.

number to another company. The client information should be migrated from one company to another. Second, like streaming video companies, data portability may mean just the act of copying the data to another company. Both companies would have the same client data at the moment of data portability request.

Figure 5.28 depicts the portability process between two data controllers. The process starts with the DS will to share the data with a new data controller. After the DS accepts the data controller consent term, s/he will request the data portability to the company that already has his/her data. Then, this company sends the data to the new DC, and the DS should verify if the data are correct. However, depending on the context, the DS may decide not to share all data from the old data controller. In this case, the DS should be able to define which data s/he wants to share.

Furthermore, all entities will be impacted as there is a new consent term with the new data controller. The DS should evaluate the new consent term to be sure regarding the data portability. Figure 5.29 shows almost all entities in red to indicate that all entities should be reviewed in this scenario. The yellow entities are not considered personal data and will not be affected.

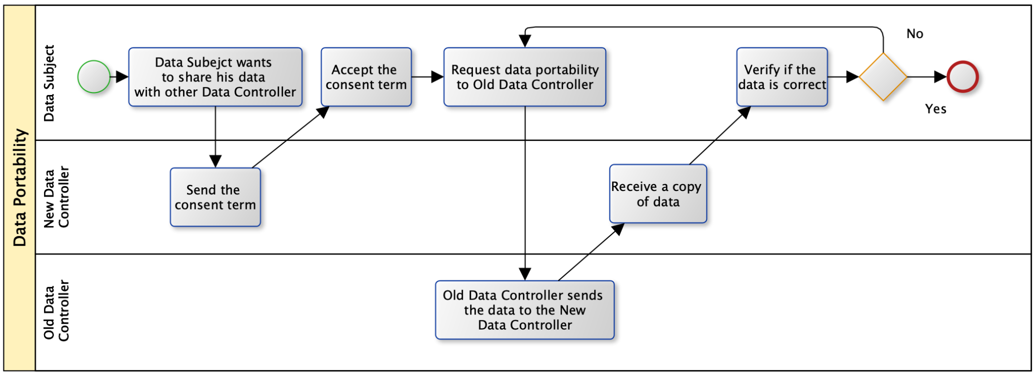

Data Portability
Data Subject
New Data Controller
Old Data Controller
Data Subejct wants to share his data with other Data Controller
Accept the consent term
Request data portability to Old Data Controller
Verify if the data is correct
No
Yes
Send the consent term
Receive a copy of data
Old Data Controller sends the data to the New Data Controller

Figure 5.28: Data portability process.

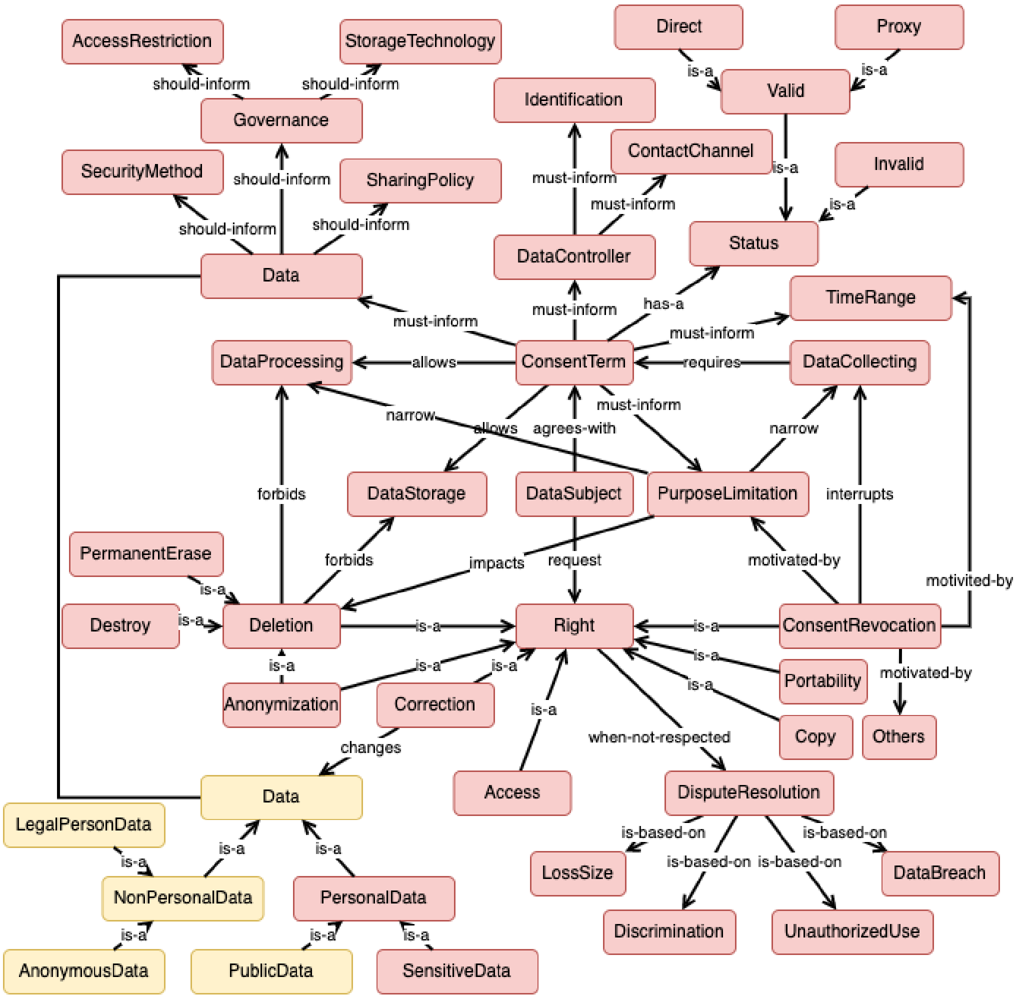


Figure 5.29: Data portability impact.

Pluggable consent. The pluggable consent can be applied when there are minor actions to be performed under a major consent term as depicted in Figure 5.30. For instance, a University can request a consent term from a student, and a professor can request a pluggable consent from a student to participate in his class. The pluggable consent is more specific and has to be executed in a period inside the time range of the major consent. As depicted in Figure 5.31, the access restriction and sharing policy will be affected. Moreover, as described before, the time range and data access will also be changed. The data controller identification, data processing, and data collecting could also present new attributes. Last but not least, the purpose of the pluggable consent must be different from the major consent. In a dispute resolution case, the DS should verify if there is discrimination or unauthorized use related to this pluggable consent.

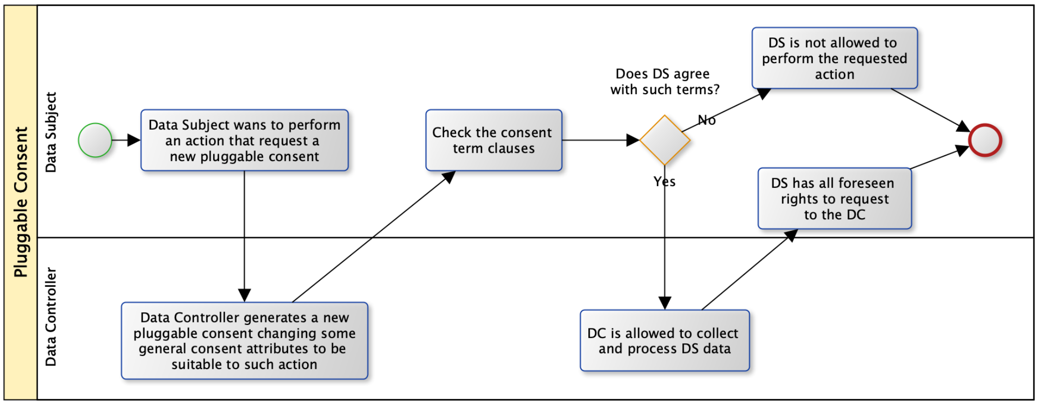

Pluggable Consent
Data Subject
Data Controller
Data Subject wans to perform an action that request a new pluggable consent
Check the consent term clauses
Does DS agree with such terms?
No
Yes
DS is not allowed to perform the requested action
DS has all foreseen rights to request to the DC
Data Controller generates a new pluggable consent changing some general consent attributes to be suitable to such action
DC is allowed to collect and process DS data

Figure 5.30: Pluggable consent process.

In general, these scenes, except *Technology unavailability* and *Inconsistent behavior* scenes, were based on the fifteen cases presented in the GConsent use cases and in the LGPD rights[7]. Moreover, the Brazilian government produced a customer guide that summarizes the main concepts regarding LGPD[8].

In summary, GoDReP proposes a structure to be reused, and eventually changed, to construct negotiation scenarios to mitigate the informational asymmetry related to data privacy, rights, and duties according to jurisdiction. These negotiation scenarios seek to clarify doubts between agents simulating the expected behaviors in specific cases. Moreover, GoDReP allows the insertion of new clauses related to the domain particularities, and agents can use GoDReP to contribute to constructing an open repository. Instead of building

[7] `https://www.gov.br/cidadania/pt-br/acesso-a-informacao/lgpd/direitos-do-titular`

[8] `http://www.mds.gov.br/webarquivos/acesso_informacao/LGPD/guia_do_consumidor_v5_5.pdf`

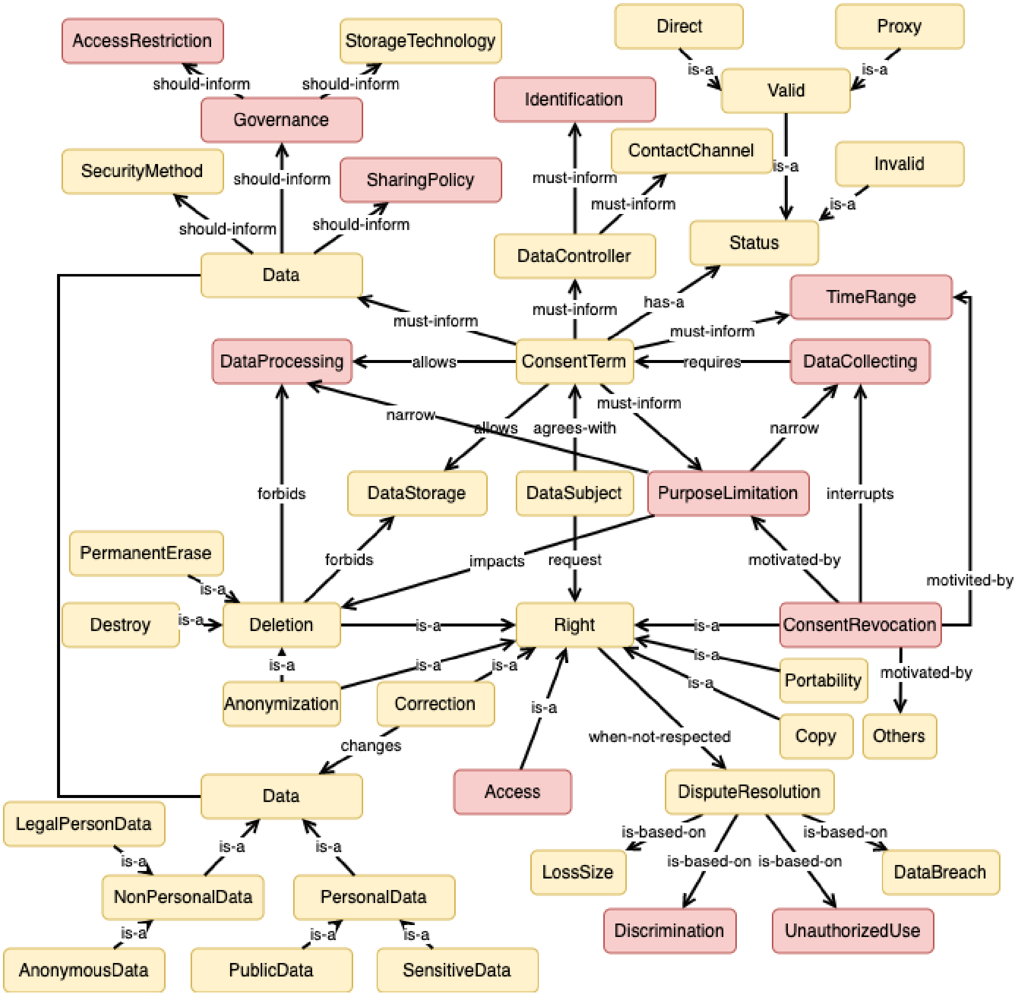


Figure 5.31: Pluggable consent impact.

it from scratch, this repository will allow other agents to create the pragmatic circumscription based on a previous instantiation.

Moreover, the Jupyter Notebook scenes are data that contribute to society regarding expectations alignment between agents and as an object of discussion related to the interpretation in other jurisdictions. As mentioned before, the latter requires an ontology calibration considering the regulation differences and particularities. Furthermore, these notebooks can represent case laws, reflecting the decision's interpretation regarding a case.

In this chapter, we presented CM based on those three ontologies. CM is focused on Consent management and it is composed of three modules *Consent*, *Action*, and *Log*. Combined, these three modules offer resources to represent the application scenarios and evaluate the agents' behaviors considering the LGPD particularities. Furthermore, as this metamodel inherited entities from PrOnto, GConsent, and ODPM, CM supports GDPR as well. Even though we did not use *LegalPersonData*, *NonPersonalData*, *AnonymousData*, *PublicData*, *PermanentErase, Destroy, Derive, Store, Infer,* and *Provide* entities, they do not conflict with any concept presented by the CM. Hence, those entities can be used if needed.

Also, we presented the GoDReP framework and detailed the processes and introduced the computational representation using the Prolog programming language and the Jupyter notebooks. The GoDReP instantiation in three different use case scenarios will be detailed in the next chapter.

6
Negotiation Scenarios with GoDReP

In this chapter, we aim to use GoDReP in three different application domains to employ the proposed framework. Moreover, these circumscriptions will aid in identifying which are general attributes and which are domain particularities (RQ3). Still, the outcome of generating the scenarios using GoDReP can be used to mitigate the data flow information asymmetry (RQ4). The whole code developed is available in an open repository[1].

The detailing of scenarios and the focus are at the discretion of the scenario creators, depending on what they consider to be the most relevant topics. In our scenarios, we emphasize points related to consent and personal data rather than the technical details of computational implementation. For instance, we provide a thorough examination of how personal data is handled in our data systems, but we do not delve into the specifics of the algorithms used for data encryption.

6.1
Health Scenario

The constant and intense collection of personal data by a myriad of services and goods, and the pan-optical vigilance exercised over our behavior when analyzing this collected data, highlight the importance of ensuring ways to protect our personal data. Due to the Brazilian lack of tradition in this subject, it is important to provide society with acculturation and awareness of the importance of protecting personal data in general.

In Brazil, the LGPD puts forward a set of rules and obligations that regulate personal data by public and private entities. In the pandemic scenario, controllers and processors must evaluate which legal bases are foreseen in the law authorizing user data collection (articles 7 and 11 of the LGPD). In this sense, it must be remarked that the Brazilian data protection regulation establishes that individual consent is only one of the legal bases authorizing data processing. In any case, data controllers must abide to the law's principles, rights, safeguards, and act in good faith.

[1] GoDReP repository. Available at: `https://github.com/phalves/GoDReP`. Last accessed on August 31, 2023.

Biomedical research, including epidemiological research as the novel coronavirus disease (COVID-19), and research in public health in general, often require the participants' informed consent to collect and process his/her personal data (Godard et al., 2003; Tolley et al., 2016; Budin-Ljøsne et al., 2017; Ma et al., 2020; Velmovitsky et al., 2020). The consent aims to ensure that the participants and patients, *i.e.*, the data subjects, are informed about the research goals and risks. Moreover, consent has evolved in the last few years. In the recent past, some studies showed that informed consent did not increase the patient's comprehension in biomedical research as presented by Paris et al. (2010).

In this sense, the consent discussion also includes ethical responsibility to maintain the participants updated in regards to further uses of the collected data and privacy management (Alves et al., 2020). Since medical research involves the processing of personal and sensitive data (such as health data), data regulations around provisions must be respected, such as: (i) the European Union General Data Protection Regulation (GDPR) (Fatema et al., 2017; Marelli, Lievevrouw, and Van Hoyweghen, 2020), (ii) the Brazilian Data Regulation Law (LGPD) (Mulholland and Frajhof, 2020), (iii) the Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada (Office of the Privacy Commissioner of Canada, 2019) and (iv) the Health Insurance Portability and Accountability Act (HIPAA) also in Canada (Banerjee, Hemphill, and Longstreet, 2018; McGhin et al., 2019). Therefore, depending on the country where public health research is being conducted, researchers must be aware of the local personal data regulation.

Furthermore, according to Carrol (2003), scenarios highlight the goals suggested by the system's appearance and behavior; what people try to do with the system; what procedures are embraced, or left out, successfully or incorrectly carried out; and what explanations people make of what occurred to them. Building and manipulating scenarios drive agents beyond the static answers. Scenarios must be concrete and flexible at once, *i.e.*, tangible enough to not be considered shallow and avoid indeterminacies, and flexible enough to allow agents to think about other possible ramifications.

In order to mitigate the informational asymmetry between DSs, DCs, and DPs, we developed a use case scenario using GoDReP in the healthcare domain considering the CM detailed in chapter 5.

6.1.1 Scenario Description

Moreover, as described in chapter 5 section 5.2, the first step (scene 0) mentioned in GoDReP is to describe the use case scenario in natural language. For this scenario, we proposed the following description:

"*Data Subject agrees with the Data Controller consent term, but then decides to revoke his/her consent.*

The Data Controller RioHealth wants to use John's (Data Subject) personal data and health data to research genetic factors related to COVID-19 from Wednesday, May 26, 2021 1:21:00 PM to Thursday, November 26, 2021 1:21:00 PM (180 days). Also, RioHealth will apply cryptographic algorithms and access policies to avoid data breaches and unauthorized access. The personal and sensitive data will be stored in a private cloud where RioHealth has complete control of applied technologies. Furthermore, RioHealth is committed to sharing the data with third parties if the purpose is vaccination prioritization information.

To do so, the Data Controller must send the consent term to the Data Subject. The consent term must present all the information defined in the LGPD art. 9.

However, after accepting the consent term, the Data Subject decides to revoke his consent on Saturday, June 26, 2021 1:07:55 PM. "

6.1.2 Scene 1 - Consent Description

Next, scene 1 is responsible for describing the consent term, and then the Prolog is applied to generate such consent term in a programming language code. Scene 1 presents the following description: " *The **Data Subject John** allows the **Data Controller RioHealth** to access, store, and process his **phone number** and **blood factor/type** to perform **research** regarding **genetic factors related to COVID-19** using **statistical analysis** for **180 days**. However, the phone number **will not be public available** and will be used only in emergency situations.*

*RioHealth will apply **cryptographic algorithms** and **access policies** to avoid data breaches and unauthorized access. The personal and sensitive data will be **stored in a private cloud** where RioHealth has complete control of applied technologies.*

*The Data Controller is allowed to **share** the Data Subject data **only** with the **vaccination prioritization purpose**.*

To make any request, please use the Data Controller communication channel by ***email lgpd@riohealth.br.***"

Figure 6.1 depicts the developed Prolog function to create a consent term with the essential parameters according to LGPD. Still, Figure 6.2 shows the consent instantiation.

```
% Description: This function defines a consent term including all required
%  information described in the LGPD Art. 9

createConsentTerm(ID,DC,             % Consent ID, Data Controller
                DS,                  % Data Subject
                PData,SData,         % Personal Data, Sensitive Data
                Purpose,             % Purpose
                SpecificPurpose,     % Specific Purpose
                Form,                % Form
                TimeLength,          % Time length of processing
                ThirdPartyPurpose,   % The purpose when sharing data with
                                     %  others
                Channel,             % Communication channel to the DS
                                     %  request
                DCContact,           % Data Controller contract
                CA,                  % Cryptography Algorithm
                AP,                  % Access Policies
                SP) :-               % Storage Platform

                assertz(id(ID)),
                assertz(dataSubject(DS)),
                assertz(dataController(DC)),
                assertz(personalData(DS,PData)),
                assertz(sensitiveData(DS,SData)),
                assertz(purpose(DC,DS,Purpose)),
                assertz(specificPurpose(DC,DS,Purpose,SpecificPurpose)),
                assertz(form(DC,DS,Purpose,SpecificPurpose,Form)),
                assertz(timeLength(DC,DS,Purpose,SpecificPurpose,TimeLength)),
                assertz(thirdyPartySharingPurpose(DC,DS,Purpose,
                    SpecificPurpose,TimeLength,ThirdPartyPurpose)),
                assertz(channelToProvideInformation(DC,DS,Channel,DCContact)),
                assertz(criptographyAlgoritm(CA)),
                assertz(accessPolitics(AP)),
                assertz(storagePlatform(SP)).
```

Figure 6.1: Healthcare Scene 1 - Consent creation function.

```
% This is a function call that defines a consent term with the informed params

?- createConsentTerm(10,'RioHealth','John','9999-9999','A+','research',
                'genetic_factors_related_to_COVID-19',
                'statistic_analysis',
                15811200,
                'vaccination_priorization',
                'e-mail',
                'lgpd@riohealth.br',
                'SHA256',
                'Authorized researchers can access the data only',
                'RioHealth private cloud').
```

```
>> Output: true .
```

Figure 6.2: Healthcare Scene 1 - Consent call.

6.1.3 Scene 2 - Consent Acceptance

Next, the DS should verify if all the crucial elements are described in the consent term presented by the DC. If so, the program will set that the consent term is ok, *i.e.*, it has all the required information. Figure 6.3 depicts the function related to the acceptance act. This function aims to verify the consent term clauses and persists the data related to the data controller's rights to collect, access, store, and process the data. Still, the program will generate logs to be evaluated when the agents perform questions to the environment. Figure 6.4 shows the function call with the scenario parameters.

```
% Description: This function sets that the Data Subject agreed with
%  the consent term.

setThatDSAgreeWithConsentTerms(id(ID),        % Consent ID
                    dataSubject(DS),          % Data Subject
                    dataController(DC),       % Data Controller
                    requestFormat(RF,DS,LPC), % Request format
                                              % LPC (Direct/Implicit)
                    personalData(DS,PData),   % Personal Data
                    sensitiveData(DS,SData),  % Sensitive Data
                    startDate(StartTS),       % Start Date - Timestamp
                    endDate(EndTS))           % End Date - Timestamp
                    :-

    consentTermStatus(id(ID),dataController(DC),dataSubject(DS),status('Valid')),

    assertz(origin(id(ID),dataSubject(DS),dataController(DC),requestFormat(RF,DS,LPC))),
    assertz(requestFormat(RF,DS,LPC)),
    assertz(dsAgreeWithConsentTerms(dataSubject(DS),dataController(DC),startDate(TS),endDate(TS))),
    assertz(log('Data Subject agrees with consent term','Communication','Compliance',StartTS)),

    assertz(dcIsCollectingDSData(id(ID),dataController(DC),dataSubject(DS),personalData(DS,PData),sensitiveData(DS,SData),startDate(StartTS),endDate(EndTS))),
    assertz(log('Data Controller can collect the Data Subject information','Explanation','Permission',StartTS)),

    assertz(dcIsStoringDSData(id(ID),dataController(DC),dataSubject(DS),personalData(DS,PData),sensitiveData(DS,SData),startDate(StartTS),endDate(EndTS))),
    assertz(log('Data Controller can store the Data Subject information','Explanation','Permission',StartTS)),

    assertz(dcIsProcessingDSData(id(ID),dataController(DC),dataSubject(DS),personalData(DS,PData),sensitiveData(DS,SData),startDate(StartTS),endDate(EndTS))),
    assertz(log('Data Controller can process the Data Subject information','Explanation','Permission',StartTS)).
```

Figure 6.3: Healthcare Scene 2 - Acceptance function.

```
% This is a function call returns true in case of success.

?- setThatDSAgreeWithConsentTerms(id(10),
                                 dataSubject('John'),
                                 dataController('RioHealth'),
                                 requestFormat('Direct','John','null'),
                                 personalData('John',9999-9999),
                                 sensitiveData('John','A+'),
                                 startDate(1622035260),
                                 endDate(EndDate)), EndDate is 1622035260+15811200.
```

```
>> Output: EndDate = 1637846460 .
```

Figure 6.4: Healthcare Scene 2 - Acceptance call.

6.1.4 Scene 3 - Data Subject Rights

According to the LGPD Art. 18, when the data subject is sharing data with a DC, s/he has the following rights, among others: (i) Data Access, (ii) Data Copy, (iii) Data Correction, (iv) Data Anonymization, (v) Data Portability, (vi) Data Deletion, (vii) Information regarding the data sharing with a third party, and (viii) Request consent revocation. To do so, Figure 6.5 shows the function which sets the data subject rights and an example of a call.

```
% Description: This function sets all Data Subject right's foreseed in the LGPD.
% This function receives the params:
%  i. Data Subject
% ii. Data Controller

setDSRights(dataSubject(DS),dataController(DC),startDate(StartTS)) :-
    assertz(dsRight(dataAccess,dataSubject(DS),dataController(DC))),
    assertz(dsRight(dataCopy,dataSubject(DS),dataController(DC))),
    assertz(dsRight(dataCorrection,dataSubject(DS),dataController(DC))),
    assertz(dsRight(dataAnonymization,dataSubject(DS),dataController(DC))),
    assertz(dsRight(dataPortability,dataSubject(DS),dataController(DC))),
    assertz(dsRight(dataDeletion,dataSubject(DS),dataController(DC))),
    assertz(dsRight(dataSharingInformation,dataSubject(DS),dataController(DC))),
    assertz(dsRight(requestConsentRevocation,dataSubject(DS),dataController(DC))),
    assertz(log('Data Subject can now have all foressen rights','Explanation',
    'Permission',StartTS)).
```

```
% This is a function call that returns true if all Data Subject's right was associated to him/her.

?- setDSRights(dataSubject('John'),
                dataController('RioHealth'),
                startDate(1622035260)).
```

```
>> Output: true .
```

Figure 6.5: Healthcare Scene 3 - Data subject rights function and call.

6.1.5 Scene 4 - Consent Revocation

As mentioned in the scenario's description, the DS decided to revoke his consent. The DS considered that the purpose limitation was not adequate. Once performed, the action of requesting the consent revocation cannot be executed again, and the DC is forbidden to continue to collect the DS's data. Figure 6.6 depicts the consent revocation function.

```
% Description: This function revoke the Data Controller's action of collecting
     the Data Subject's data.

setDSRevokeConsent(id(ID),                    % Consent ID
                   dataSubject(DS),           % Data Subject
                   dataController(DC),        % Daa Controller
                   personalData(DS,PData),    % Data Subject's Personal Data
                   sensitiveData(DS,SData),   % Data Subject's Sensitive Data
                   now(Date),                 % Current Date
                   startDate(StartTS),        % Start Date
                   endDate(EndTS)             % End Date
                   ) :-

    requestFormat('Direct',DS,'null'),
    not(dsRight(requestConsentRevocation,dataSubject(DS),dataController(DC))),
    assertz(log('Data Subject tried to revoke his/her consent, but fail','
        Explanation','Prohibition',Date));

    retract(dsRight(requestConsentRevocation,dataSubject(DS),dataController(DC
        ))),
    assertz(log('Data Subject requested to the Data Controller to revoke his/
        her consent','Communication','Permission',Date)),

    retract(dcIsCollectingDSData(id(ID),dataController(DC),dataSubject(DS),
        personalData(DS,PData),sensitiveData(DS,SData),startDate(StartTS),
        endDate(EndTS))),
    assertz(log('From now, the Data Controller cannot collect the Data Subject
         information','Communication','Prohibition',Date)),

    retract(dcIsProcessingDSData(id(ID),dataController(DC),dataSubject(DS),
        personalData(DS,PData),sensitiveData(DS,SData),startDate(StartTS),
        endDate(EndTS))),
    assertz(log('From now, the Data Controller cannot process the Data Subject
         information','Communication','Prohibition',Date)),

    retract(consentTermStatus(id(ID),dataController(DC),dataSubject(DS),status
        ('Valid'))),
    assertz(consentTermStatus(id(ID),dataController(DC),dataSubject(DS),status
        ('Invalid'))),
    assertz(log('From now, consent is not valid to be used by the data
        controller','Explanation','Prohibition',Date)).
```

Figure 6.6: Healthcare Scene 4 - Consent revocation function.

6.1.6 Scene 5 - Exploring Cause-Effect Scenes

Once the scenario is configured, DSs as DCs and DPs will be able to perform queries to the environment to speculate the motivation of the query result. For instance, as the DS requested the DC to revoke his consent, a query related to the data processing should return *false*, as depicted in Figure 6.7.

```
?- dcIsProcessingDSData(id(10),dataController('RioHealth'),dataSubject('John')
    ,personalData('John',PData),sensitiveData('John',SData),startDate
    (1622035260),endDate(1637846460)).
```

```
>> Output: false .
```

```
% Why?
?- log(Event,'Communication',Type, 1624712875).
```

```
>> Output:
    Event = Data Subject considered that the purpose limitation is not
        adequate, Type = Permission ;
    Event = Data Subject requested to the Data Controller to revoke his/her
        consent, Type = Permission ;
    Event = From now, the Data Controller cannot collect the Data Subject
        information, Type = Prohibition ;
    Event = From now, the Data Controller cannot process the Data Subject
        information, Type = Prohibition .
```

Figure 6.7: Healthcare Scene 5 - Exploring general scenario aspects.

Consent revocation not respected. Considering that the DS requested the consent revocation, the DC can no longer collect or process data. To check if there is any regulation violation regarding such request, the agent should: (i) verify if the data controller is still collecting data, (ii) verify if there is a valid consent term, and (iii) verify if there is a request to revoke the consent term. Furthermore, the generated log could be used to recover the actions performed in such a scenario in order to deliver more information to the agent that is testing the information asymmetry possibilities.

Data breach, what to do? The DC must inform to national authority and the DS when a data breach that may cause risks or damage to the DS occurs. Such communication has to be done as soon as possible and should inform: (i) personal data category, (ii) what data were leaked, (iii) what were the technical and security measures used to protect data, (iv) what were the reasons for the communication delay, when applied, (v) the risks related to the incident, and (vi) what the data controller will do to revert or mitigate the damage. The DC will have to disclose such an event in high-impact communication media, according to the severity of the incident.

Moreover, if the DC suffered from a hacker attack and the DS's personal data were leaked on social media, then, the DC is obligated to inform the incident to **ANPD** and inform the DS that his phone number was leaked. Even if the DS has revoked his consent, he must be informed of the data breach as his data is still on the DC's database.

Evidencing data leak. The first step before suspecting data leaked is to verify which DCs have the data, if possible. In our scenario, we can execute the query depicted in Figure 6.8, and it will return that RioHealth is the only controller that has the data that were leaked and accessed by other institutions.

```
?- dcIsStoringDSData(id(ID),dataController(DC),dataSubject('John'),
    personalData('John',9999-9999),sensitiveData('John','A+'),startDate
    (1622035260),endDate(1637846460)).
```

```
>> Output: ID = 10, DataControlle = RioHealth .
```

Figure 6.8: Healthcare scenario - Data controller query.

As RioHealth is the only institution that DS John shared his data, if other companies have the same data as RioHealth, it means that there is a possibility that RioHealth has violated the consent term and shared the data with another institution.

Requesting data correction. The DC is obligated to abide by the DSs' requests as correction as well as data access. Also, the controller is obligated to inform all processors regarding the correction. Thus, after such a request, the DS should verify if the data were updated. If the request was not accomplished, the DS can request the ANPD intervention.

Requesting data anonymization. Once the data is anonymized, the DC will not have the resources to provide details about such data, including correction. Hence, after this request, the DC is not obligated to comply with requests that should involve reidentification actions. Still, the controller is obligated to inform all processors regarding the anonymization. Figure 6.9 depicts an example of anonymization return in Prolog.

```
?- dcIsStoringDSData(id(_),dataController(riohealth),dataSubject(DataSubject),
    personalData(_,_),sensitiveData(_,'A+'),startDate(1622035260),endDate
    (1637846460)).
```

```
>> Output: DataSubject = Variable(70) .
```

Figure 6.9: Healthcare scenario - Anonymization return.

Thus, the DC should anonymize the data and inform the processors. However, as mentioned before, the anonymization methods are out of this thesis's scope.

Data deletion. As mentioned before, there are four cases in which the DC can keep the data even with a request from the DS to delete the data (LGPD Art. 16). As RioHealth informed that the purpose is to perform research, the data deletion request will not be fulfilled.

Technology unavailability. In an urgent situation, technology unavailability might cause consequences for the DS who needs the data in the short term. Hence, it will directly impact the rights related to request data copy and data portability.

Inconsistent behavior. To the best of our knowledge, the inconsistent behavior in the healthcare scenario could be related to an attempt of the **DDoS attack**, which would impact the entities as described in chapter 5.

Data portability Unlikely the open banking scenario, there are no specific rules related to healthcare data portability[2] . However, we assumed that the DS desired to migrate all his data to another institution. To do so, first, the DS must accept the consent term from the other institution to request the data portability right. If there is no valid consent term with the new institution, the old DC cannot share the data.

In this section, we presented the healthcare scenario instantiated using GoDReP. This scenario was divided into scenes and coded in Prolog in a Jupyter Notebook. This and the other scenarios are available in an open repository.

6.2 Educational Scenario

In this scenario, we aim to explore the particularities of the educational environment faced with LGPD concerns. The educational scenario will approach situations related to an adult starting a university course. Moreover, we exemplified the pluggable consent concept, which allows the professors to request a new consent term for students to participate in their classes.

6.2.1 Scenario Description

As described in chapter 5, GoDReP scene 0 describes the use case scenario in natural language. Thus, we proposed the following description:

"*The Data Subject John is a 17 years old person, and he is going to start taking classes at XYZ University. However, as John is below 18 years old, i.e., he is considered a teenager under Brazilian law, and he is not emancipated, he needs his parents to accept the consent terms to start the academic activities.*

Therefore, the University has to get John's personal data and request the consent term acceptance to his parents. John has to inform the following data:

- *Full Name - used to identify the person.*

[2] It is crucial to observe that the Open Health movement, *e.g.*, Open Health Brazil (`https://www.openhealthbr.com/`), necessitates caution and diligence to ensure that health workers and patients are fully informed and aware of the advantages and significance of data integration. It is important to understand their responsibilities and obligations, however, this case is different from ours and would be developed with the assistance of GoDReP.

- *Address - used to keep communication by mail.*
- *Email - used to keep communication by the internet.*
- *Gender - used to create University's reports <Choose as many as you like: Male, Female, Non-binary, Transgender, Intersex, I prefer not to say>.*
- *Birth - used to decide if the student is a person that can respond by their acts or if his/her parents have to sign on behalf of the student, i.e., if the student is below 18 years old (in Brazil).*
- *Personal Identification Number - used to check the person's identity*
- *Educational Transcripts - used to prove that the student has the minimum requirements to become the University's student.*
- *Legal Person in Charge - used to request legal action while the student is below 18 years old.*

Therefore, from Wednesday, May 26, 2021, 1:21:00 PM to Thursday, November 25, 2021, 1:21:00 PM, when the student becomes above 18 years old, the student's parents will be the legal persons who will respond on his behalf. The University will share the student's data following the government guidelines, but it will not share any information with unauthorized third parties. The University's professors will be able to get all the discipline scores when the student subscribe to their new disciplines. If needed and justified, the professors can request such information when creating the discipline. Also, the University will apply cryptographic algorithms and access policies to avoid data breaches and unauthorized access. The personal and sensitive data will be stored in a private cloud where the University has complete control of applied technologies.

To do so, the University, i.e., the Data Controller, must send the consent term to John and his parents, i.e., the Data Subject. The consent term must present all the information defined in the LGPD art. 9.

General Best practices:

- *The University should require a new consent when he/she turns 18, i.e., when the student becomes an adult legally.*
- *When the student becomes an adult legally, the University should communicate to his/her parents, notifying them that they are no longer the student's legal representative.*
- *XYZ University should not request any data without a justification.*

Semestral Consent Term Best Practices:

Each semester, the University should require the student's consent to remember that the data might be shared with the university professors to which the student has subscribed. Moreover, the University's professor should be able to:

- *Require the student course information in the consent term.*
- *Set the student's transcripts as required information to subscribe to their classes. If a professor requests the transcripts, he/she has to inform why this information will be collected.*"

6.2.2
Scene 1 - Consent Description

Next, scene 1 is responsible for describing the consent term, and then the Prolog is applied to generate such consent term in a programming language code. Scene 1 presents the following description:

The ***Data Subject John*** *allows the* ***Data Controller University XYZ*** *to access, store, and process his* ***transcripts*** *and* ***personal information*** *in order to* ***improve the class dynamics***, ***allowing professors to design the class activities better***.

Such information will be shared with the university's employees under strict governance policies that guarantee that only the necessary information to execute their functions will be shared. The employees will respond to any unauthorized data access, leak, or other activities that may expose or cause any loss to the data subject. ***The transcripts will be available to professors to whom the data subject had subscribed***. *No information will be publicly available without previous consent acceptance.*

The personal and sensitive data will be available, stored, and processed while the data subject has an active registration number in the university. ***A new consent term will be required in two situations***:

- *in a new term, .i.e., when the data subject has to subscribe to a new discipline, and*
- *when the data subject finished its course*

Therefore, this consent term is ***valid for one term***, *and must be renewed by term. In Brazil, each term is represented by a semester, i.e., six months.*

Last but not least, if the data subject is not an adult, i.e., if the data subject is under eighteen years old in Brazil, the data subject must be represented *by one of his/her parents or a person legally in charge. This representation will be automatically changed when the data subject is considered an adult.*

*University XYZ will apply **cryptographic algorithms** and **access policies** to avoid data breaches and unauthorized access. The personal and sensitive data will be **stored in a private cloud** where University XYZ has complete control of applied technologies.*

*The Data Controller is **not allowed to share** the Data Subject data, except for cases that the government or courts requires such data.*

*To make any request, please use the Data Controller communication channel by **emailing to lgpd@univerisyxyz.br**.*

Figure 6.10 shows the code that creates such consent term with the parameters described above.

```
% This is a function call that defines a general consent term with the
    informed params

?- createConsentTerm(10,'universityXYZ','John','john@mail.com','transcripts',
                'improve_class_dynamics',
                'design_class_activities',
                'statistic_analysis',
                15811200,
                'none',
                'e-mail',
                'lgpd@universityxyz.br',
                'SHA256',
                'Authorized employees can access the data only',
                'University XYZ private cloud').
```

```
>> Output: true .
```

Figure 6.10: Educational Scene 1 - Consent call.

6.2.3 Scene 2 - Consent Acceptance

The consent acceptance in this educational scenario requires more information than in the healthcare scenario described previously. As the scenario description presents, DS John is not considered as an adult from the LGPD perspective. Hence, the consent cannot be given directly, *i.e.*, John needs a legal representantive to accept the consent term on his behalf. Thus, this characterizes proxy consent, as depicted in Figure 6.11.

6.2.4 Scene 3 - Data Subject Rights

To the best of our knowledge, the educational scenario does not differ regarding the rights foreseen in LGPD compared with the healthcare scenario.

```
% This is a function call returns true in case of success.

?- setThatdsAgreeWithConsentTerms(id(10),
                                dataSubject('John'),
                                dataController('universityXYZ'),
                                requestFormat('Proxy','John','Mary'),
                                personalData('John','john@mail.com'),
                                sensitiveData('John','transcripts'),
                                startDate(1622035260),
                                endDate(EndDate)),
                                EndDate is 1622035260+15811200.
```

```
>> Output: EndDate = 1637846460 .
```

Figure 6.11: Educational Scene 2 - Proxy consent call.

6.2.5 Scene 4 - Consent Revocation

The consent revocation scene required a change in the code to consider that the DS may not be responsible for his acts in some cases. For example, in our educational scenario, the DS must request his legal representative to sign this request, as depicted in Figure 6.12.

```
% This call store the Data Subject's motivation to request the cosent
    revocation.
?- assertz(log('Data Subject considered that the purpose limitation is not
    adequate','Communication','Permission',1624712875)).

% This is a function call returns true if all Data Subject's request was
    successfully performed.
?- setDSRevokeProxyConsent(id(10),
                        dataSubject('John'),
                        dataController('universityXYZ'),
                        legalPersonInCharge('Mary'),
                        personalData('John','john@mail.com'),
                        sensitiveData('John','transcripts'),
                        now(1624712875),
                        startDate(1622035260),
                        endDate(EndDate)
                        ),
                        EndDate is 1622035260+15811200.
```

```
>> Output: true .
>>         EndDate = 1637846460 .
```

Figure 6.12: Educational Scene 4 - Consent revocation call.

6.2.6 Scene 5 - Exploring Cause-Effect Scenes

In general, the cause-effect scenes are similar even when the pragmatic circumscription presents different natures. Most of the cause-effect scenes in the educational scenario are similar; the agent should change the parameters only to adapt to the context. However, we will detail the main differences below. Last but not least, in this pragmatic circumscription, we experienced the need for a new concept, the pluggable consent.

Requesting anonymization. In this scenario, the data anonymization request may impact the student discipline subscriptions. Considering a professor who needs the student data to design the class activities, if a student requests data anonymization, such a professor will not be able to recover the student's data. A possible behavior is a professor denying the student participation, as there will be no data regarding such a student. However, there are other possibilities that an agent could evaluate in this scenario.

Technology unavailability. Depending on the moment of the technology unavailability, the impact would be more, or less, severe. For instance, if the students are in the subscription moment, they may lose the timing to do their class subscriptions. Hence, it could generate many issues, for instance, related to: (i) classes size measurement, *i.e.*, students per class; (ii) available physical space, which depends on the class size, and (iii) the university should provide another moment to students do their class subscriptions if they were affected. Moreover, if the students are at the end of the term, they may request to delay the final exam, and the professors may delay the final grade. Conversely, if the unavailability occurs in the middle of the term, students may not be impacted.

Inconsistent behavior. In our educational scenario, as some professors need the students' transcripts to diversify the working groups, the student might perform such inconsistent behavior to be allocated into a better working group. In this case, if identified by the class professor, the professor could not accept the student to his/her class. Even though a student may have performed such inconsistent behavior accidentally, the DC has mechanisms to identify such a situation. Therefore, the DC should look for inconsistency motivation, for instance, if it is just a user testing his/her possibilities in the system, or if there is a bug in the system, or if a malicious person is trying to damage the DC.

Pluggable consent. In the educational scenario, we developed the concept of pluggable consent, *i.e.*, a new consent term under a major consent term. Thus, the pluggable consent time range must be equal to or lower than the value in the major consent term. Moreover, we had to develop a consent relationship to link the pluggable and the major consent term. Thus, the agents can explore what happens with the pluggable consent if the DS revokes the major consent term and vice versa.

In this section, we presented the educational scenario instantiated using GoDReP. So, such a scenario was divided into scenes and coded in Prolog in a Jupyter Notebook.

6.3 Open Banking Scenario

In this scenario, we aim to explore the particularities of the open banking scenario faced with the LGPD concerns. The open banking scenario will approach situations regarding sharing financial data between financial institutions.

Open banking is a practice of enabling data sharing between financial institutions. This practice came to allow banking interoperability by APIs (Application Programming Interfaces). For example, the DS can request a credit card from bank A, a financial loan from bank B, and buy assets from bank C. Moreover, open banking allows, for instance, the DS to open an account just by requesting his/her data from an institution with that s/he has an account previously. Therefore, open banking turns data sharing more agile, transparent, and secure, by providing resources to the data subject chose: (i) which data s/he wants to share; (ii) when; (iii) how long, and (iv) with whom s/he wants to share.

In this sense, open banking acts as a kind of data portability foreseen in the LGPD. However, there are strict rules set by the Central Bank to enable data exchange between financial institutions. For instance, the consent term related to open banking must not be provided by paper or an adhesion contract, by forms with agree option filled by default, or without an explicit will of acceptance from the data subject.

A DS that wants to participate in the open banking ecosystem, has to agree with a consent term that allows the institution to share his/her data. Following the LGPD and the GDPR rules, the institution must offer to the DS an option to revoke his/her consent at anytime. Moreover, particularly in the Open Banking, such consent term must be valid by 12 months at most, *i.e.*, the institution must request a new consent term every year to confirm the DS's wills. There is an extensive list of: (i) personal data, (ii) enterprise data, and (iii) transactional data:

- Personal Data: Full Name, Document ID, Residential Address, Phone Number, E-mail Addresses, Social Name, Parents' Names, Marriage Status, Born date, Gender, Nationality, Income, Patrimony, Occupation, Relationship Start Date, Products and Services Hired, Agency and Account Number, Legal Person in Charge Name and Id Number (if exists).

- Enterprise Data: Company's Name, Identification Number, Address, Latitude and Longitude, Phone Number, E-mail Addresses, Owners' Names and their Identification Numbers, Administrators, Society Rates, Start Date, Activity Field, Income, Patrimony, Relationship Start Date, Products and Services Hired, Agency and Account Number.
- Transactional Data: Account Balance, Credit Card Type and Identification, Limit, Transactions, Credit Card Bill, Credit Services (Ex: loan and investments).

DS must agree with the proposed consent term to share the above data. The process of using the open finance service is free of charge. Moreover, the open banking process is composed of six steps as follows:

(i) DS should start the process showing Bank B his/her intention to get his/her data from Bank A.

(ii) DS should verify the purpose of the data usage from Bank B and go to the next step if he/she has no objection to the informed purpose.

(iii) DS should choose: (i) the origin institution, *i.e.*, Bank A, to request the data, (ii) the data that he/she wants to share, and (iii) the time frame that must be twelve months at most[3].

(iv) Bank B will redirect the DS to Bank A where the DS will be able to verify his/her identity as well as confirm his/her intent to share the selected data.

(v) Bank A will redirect the DS to Bank B. Bank B will notify the DS if the authorization process is concluded.

(vi) Finally, Bank B will be able to request the authorized data from Bank A.

It is important to note that the DS can revoke his/her consent at any time. To do so, the DS should access Bank A communication channels and request the consent revocation. Another case is when the consent expire automatically, *i.e.*, in twelve months. In this case, the DS will be able to choose to renew his/her consent to continue sharing his/her data. Moreover, the consent revocation may imply stopping the receival of services and product offers from Bank B.

In joint account cases, the authorization can be done individually. The transactional data will be available to anyone who can manage the Bank B account without the others. If there is a requirement to request authorization from more than one account, both banks must provide information regarding how to do that.

[3] JOINT RESOLUTION Nº 1, 4th, May 2020 art. 10, par. 1, item III

6.3.1 Scenario Description

As described in chapter 5, GoDReP scene 0 describes the use case scenario in natural language. Thus, we proposed the following description:

John wants to use the open banking feature to share his data from Bank A to Bank B. The shared data will be used to offer products and services that fit with the data subject's profile or comply with other legal bases and obligations, such as money laundry, fraud, and risk evaluation, including credit risk. Moreover, the data will be used to create and improve the Bank B services, products, and processes.

From the LGPD perspective, the data will be shared for twelve months at most; therefore, from Wednesday, December 08, 2021, 10:41 a.m. to Thursday, December 08, 2022, 10:41 a.m. Bank B will be allowed to get John's data from Bank A.

Bank B will share only the data allowed by John, and will use them to comply with government laws and propose new products and services according to John's profile. Bank B will not share any information with third parties without contacting John. If Bank B updates its consent term, Bank B must inform John regarding this update and request a new approval.

Best practices based on the user experience with two financial institutions using open banking:

- *Bank A should allow John to select which data he wants to share with bank B.*
- *Bank A should allow John to set the time range that he wants to share his data, considering a maximum of twelve months.*

6.3.2 Scene 1 - Consent Description

Next, scene 1 is responsible for describing the consent term, and then the Prolog is applied to generate such consent term in a programming language code. Scene 1 presents the following description:

*The **Data Subject John** allows the **Data Controller Bank B** to access, store, and process his personal and transactional data from Bank A in order to **offer products and services, allowing Bank B to send offers appropriately based on John's data**.*

John's personal and transactional data will be shared with Bank B under strict governance policies that guarantee that only the information required to execute their functions will be shared. The employees will respond to any *unauthorized data access, leak, or other activities that may expose or cause any loss to the data subject. None information will be publicly available without a previous consent acceptance.*

The personal and transactional data will be available, stored, and processed while the data subject has an active consent term with Bank A and Bank B. ***A new consent term will be required in two situations:***

- *when there is an update on the consent term;*
- *when the data subject changes the data that he wants to share or change the time range;*
- *when the due date is accomplished.*

Therefore, this ***consent term is valid for twelve months at most considering the open banking rules****. The data subject may renew the consent or revoke it at any time. Moreover, Bank B will apply* ***cryptographic algorithms*** *and* ***access policies*** *to avoid data breaches and unauthorized access. The personal and transactional data will be* ***stored in a private cloud*** *where Bank B has complete control of applied technologies.*

The Data Controller Bank B is not allowed to ***share*** *the Data Subject data, except if such data is requested by a Court. To make any request, please use the Data Controller communication channel by e***mail lgpd@bankb.br.**

6.3.3 Scene 2 - Consent Acceptance

Next, the DS can read the consent term and accept following the same process presented in the previous scenarios.

6.3.4 Scene 3 - Data Subject Rights

To the best of our knowledge, the open banking scenario does not differ regarding the rights foreseen in LGPD compared with the healthcare scenario.

6.3.5 Scene 4 - Consent Revocation

As mentioned in the scenario's description, the DS decided to revoke his consent. The DS considered that he does not want to receive offers from Bank B. Once performed, the action of requesting the consent revocation cannot be executed again, and the DC is forbidden to still collect the DS's data, as well as described in the healthcare scenario.

6.3.6 Scene 5 - Exploring Cause-Effect Scenes

As presented in the previous scenarios, the cause-effect scenes are similar in general. Thus, we will highlight scenes that show the differences only.

Evidencing data leak. In the open banking scenario, the presented decision-making process depicted to aid the DS in figuring out who leaked the data is ineffective. As two DCs have access to John's data, the action of verifying who leaked the data is even more difficult and may turn the process inconclusive.

Requesting anonymization. In this scenario, once the data are anonymized, the DC will not have the resources to provide details about such data, including correction. Hence, after this request, the DC is not obligated to comply with requests that should involve reidentification actions. The data anonymization request may imply stop receiving offers and products from Bank B. Moreover, as the anonymization right turns not possible for Bank B to execute its purpose, *i.e.*, create and send specific products and services based on John's data, Bank B should ask if he would like to revoke his consent. Last but not least, if the DS requested anonymization and did not request consent revocation, the new data would not be anonymized.

Inconsistent behavior. In the open banking scenario, this behavior might indicate that the client is confused about sharing his/her information or trying to manipulate the data to get advantages. The unusual behavior can be caught by analyzing the event log. Depending on the magnitude, this kind of event may cause damage to the bank system. Moreover, as Bank B needs John's data to offer products and services, John might perform such inconsistent behavior to try getting advantages hiding specific data, which may compromise his reputation. In this case, if Bank B identifies such behavior, the bank may request additional information before starting offering new products and services.

Data portability. As the open banking scenario is a kind of data portability, *i.e.*, the DS request from Bank A to send his data to Bank B, this scene was not evaluated in this scenario.

In this section, we presented the open banking scenario instantiated using GoDReP. So, such a scenario was divided into scenes and coded in Prolog in a Jupyter notebook.

6.4 Scenarios Main Takeaways

The exploration through a defined pragmatic contextualizes the semantic, *i.e.*, the personal data regulation. Once it is frozen, the pragmatic will be reduced to a semantic. Hence, time is a crucial attribute in delivering semantic contextualization (RQ3). Moreover, the order of the performed actions is essential to understand if an action is legal or not. For instance, a DC can collect data from a DS whether a consent term is filled by the DS informing that this action was approved. However, if we invert the order, *i.e.*, the DC starts collecting data without an agreed consent term, the DC will be violating the data protection regulation.

Moreover, the deontic concepts are essential in the use case scenarios to set the compliance directives (RQ3), as explained in chapter 4. Last but not least, even though the cause-effect scenes can be considered as attributes that must be present independently of the domain, the domain may require changes to adapt the context to some scenes. For instance, in the educational scenario, we added the pluggable consent scene to represent a particularity of this domain.

However, the specific aspect of each domain were not considered in its totality, as it is an exponential problem, *i.e.*, defining all particular characteristics of all domains is not possible due to its magnitude (RQ3). In this case, resources mean identifying people able to define and detail all possible scenes in all application domains already known. Thus, the domain questions can be explored as far as the agent is able to express them in natural language following the ontology guidelines and in a programming language to build the logical algorithm.

Finally, the agents can mitigate data flow information asymmetry by developing use case scenarios using GoDReP, creating a dialog between them (RQ4), *i.e.*, a DS, DC, and DP dialog to align their understanding and concerns regarding each possible scenario. Moreover, Table 6.1 shows a summary of significant changes in the proposed scenarios. The healthcare scenario was the first scenario developed; hence, we had to create all scenes from scratch. Next, we developed the educational scenario, and as mentioned before, this scenario required significant changes in the three scenes. Still, this scenario required the development of a new scene, the Pluggable consent scene. Last but not least, the open banking scenario was the last scenario to be developed, and it required major changes on four of the ten proposed scenes.

Table 6.1: Summary of scenarios changes.

Exploring Cause-Effect Scenes	Healthcare	Educational	Open Banking
Consent revocation not respected	x		
Data breach, what to do?	x		
Evidencing data leak	x		x
Requesting data correction	x		
Requesting data anonymization	x	x	x
Data deletion	x		
Technology unavailability	x	x	
Inconsistent behavior	x	x	x
Data portability	x		x
Pluggable consent		x	

In this chapter we defined three use case scenarios to exercise the GoDReP framework and evaluate each scenario's particularities. Therefore, we concluded that using GoDReP, the application scenarios present many aspects in common related to the consent term construction, and many exploration scenes are very similar. Moreover, the scenes evaluation process is crucial to explore the scenario possibilities. GoDReP could be used to aid agents in building the scenarios based on an existing scenario and not from scratch.

Next, we will present the Artificial Intelligence approach for Data Regulation (RegulAI) framework. Moreover, we will also discuss the RegulAI applicability considering the DS and DC's perspective.

7
Designing Intelligent Agents in Normative Systems Toward Data Regulation Representation

RegulAI, the Artificial Intelligence (AI) approach for Data Regulation, seeks to utilize AI techniques to depict the rights and obligations associated with data regulation, along with the decision-making process of the agent as explained in GoDReP. This framework proposes applying NMAS to regulate agents' behavior considering data regulation constraints.

As mentioned in chapter 2, NMAS is responsible for defining the *Environment*, *Agents* and their *Roles*, *Norms*, and *Organizations* parameters to ensure that the data regulation will be respected when it emerges in the collection, storage, and use of data, otherwise, agents will suffer punishments. A new norm can be added into the environment at run-time, and the software agents analyze if such legal command is activated and addressed to them. Next, they will evaluate if they shall comply or not based on the rewards and punishments.

In the data regulation context, *Norm*'s deontic concept defines if a norm is an obligation, permission, or prohibition (Žarnić and Bašić, 2014). From the DCs and DPs' perspective, norms set their obligations foreseen by a certain data regulation. On the other hand, from the DSs' perspective, the norms set which are their rights, and allow them to exercise them. The addressed agents can decide whether to comply with a norm; they must evaluate the rewards, punishments, and goals to make a decision. Rewards and punishments can be from distinguish nature depending on the use case and the simulation goal. For instance, rewards can be related to increasing reputation and accessing DSs data to agents who comply with a norm. From the punishment's perspective, they can be related to decreasing reputation and issuing fines to agents that decide to violate a norm. Moreover, a *Norm* is activated, or deactivated, if a condition is triggered, turning the norm state to active or inactive.

Agent and *Agent Role* represent DSs, DCs, and DPs entities. *Environment* represents the application domain where the agents reside and provide data to contribute to agents' decision-making process, i.e., agents read the *Environment's* available data and then, based on their goals, decide which action they will perform. *Organization* groups agents that present common goals, e.g., DC agents from a company can be grouped in the same organization.

From the DS's perspective, the BDI decision-making process represents the DS's reasoning. Figure 7.1 depicts the normative BDI architecture for designing data regulation representation. This approach aims to provide an explanation for data agents regarding data regulation concerns and the decision-making process when they are involved in a data-sharing plot. Moreover, the proposed architecture is based on two major layers: (i) BDI decision-making process, and (ii) legal bases representation. The former provides cognitive intelligence to data agents following the BDI architecture. The latter represents data regulation rights and obligations by norm generation.

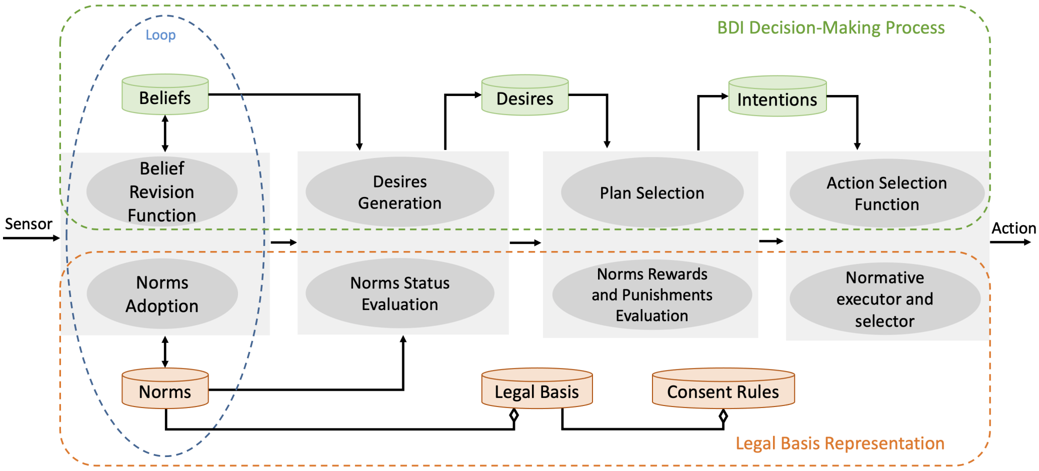


Figure 7.1: Normative BDI architecture for data regulation representation.

In the next sections, we will detail the DS and DC's perspectives when using the RegulAI architecture considering the data regulation norms and the agents' preferences.

7.1 DS's Perspective

From the DS's perspective, the RegulAI process starts when agents are active and observe the environment for events. The *sensors* are responsible for reading the environment's changes and sending them to agents. Next, based on the *sensor*'s returns, DS updates its *Beliefs* and *Norm*'s database, evaluating if any norm is addressed to its role — the system's architecture defines the repetition frequency. Then, DS defines its desires based on its beliefs considering the norms' status addressed to him. The generated desires are stored in the *Desire*'s database.

As we focus on the Consent legal bases, we created a representation of Consent as

$$C =< P, E, S, DC, DS >, \tag{7-1}$$

where P means the purpose limitation, E means the expiration date, and S represents the data sharing policies to provide clear, straightforward, and complete information.

Then, DS can select a plan based on the Consent Evaluation (CE) and on the Consent Compatibility Index (CCI). CE is defined by

$$CE_{DS} =< D, P, E, S, R_{DC} >, \tag{7-2}$$

where D is the DS's desire; P, E, S are the DS's preferences, and $0 \leq R_{DC} \leq 9$ is the minimum reputation value acceptable by DS. The DC's reputation is built according to the respected norms, i.e., according to the rewards received.

This representation considers that DS is responsible for providing its preferences (DSP) related to *D, P, E, S,* and R_{DC}, setting weights for each one. DSP is defined as,

$$DSP =< W, X, Y, Z >, \tag{7-3}$$

where:

- w, if $D_{DS} = P_{DC}$,
- x, if $E_{DS} \subseteq E_{DC}$,
- y, if $S_{DS} \supseteq S_{DC}$,
- z, if $R_{DC_{DS}} \geq R_{DC_{DC}}$,
- $\{w, x, y, z\} \in [0, 9]$,
- $0 \leq sum(w, x, y, z) \leq 10$.

The Consent Compatibility Index (CCI) is a number between 0 and 9 generated from Eq 7-4. DS can set a minimum score to define an acceptable CCI value according to its preferences and consider this value when deciding whether to share its data. Next, DS should evaluate the norms' rewards and punishments.

$$CCI = \sum_{i=0}^{n(DSP)} DSP_i \tag{7-4}$$

Norms define rights (permission) and duties (obligation or prohibition) for agents to execute their goal in a particular context and during a given time. In this sense, deontic concepts can represent data regulation constraints in a normative system. Let $(op \in O, P, F)$, it defines a norm as an obligation (O), a permission (P), or a prohibition (F). Obligation and prohibition are concepts the agent must comply with when such a norm is activated and addressed to him. Otherwise, sanctions or future litigation claiming damages can be happen. Conversely, the permission concept allows agents to comply with such a norm facultatively. Thus, a norm follows Eq. 7-5 construction, let

$$N =< Ad, Ac, Ex, Re, Pu, Op, St >, \tag{7-5}$$

where *Ad* represents addressees, *Ac* represents the activation trigger, *Ex* represents the expiration trigger, *Re* represents the norm's reward, *Pu* represents the norm's punishments, *Op* represents the deontic concept, and *St* represents the norm's state.

Thus, the normative contribution (NC) considers the active norms addressed to the Software Agent (SA) to measure the agent desires (D), and norms rewards (Re) and punishments (Pu) as

$$\forall n \in N \begin{cases} NC_n, & \text{if } St = \mathit{Active} \,\&\, \mathit{Ad} = \mathit{AgRole} \\ 0, & \text{otherwise} \end{cases} \tag{7-6}$$

where $NC_n = D + Re_n - |Pu_n|$.

Finally, the agent's intentions are represented by

$$I =< B, D, CCI, NC, SA_{Plans} > \tag{7-7}$$

It is important to note that *CCI* is an element addressed to DSs only, where SA_{Plans} are the available plans considering the software agent's *Beliefs* and *Desires*. Next, the agent decides the compatible action based on its intentions to achieve the selected desire.

7.2 DC's Perspective

From the DC's perspective, software agents must evaluate the environmental norms to decide whether to comply with the current regulation. This agent role will follow the legal bases Representation layer depicted in Figure 7.1 and defined by its elements in Eq. 7-1.

First, the agent will read the environmental norms and check which ones are addressed to him. Second, the agent will verify which are the active norms (Eq. 7-6). Third, the agent will evaluate which norms comply based on the rewards and punishments. Fourth, and finally, the agent will execute his action considering the decision related to the norms that he will comply with or not.

Thus, a DC can define the norms and the consent attributes under a specific data regulation to model the rules that DSs, DCs, and DPs should follow, evaluating the pros and cons of sharing and managing personal data. Moreover, NMAS enables the development of a simulation environment for DSs to experience the defined rules and impacts when sharing data.

Last but not least, NMAS can be used to monitor the use of data by organizations and to flag any instances where the data is being used in ways that are not in accordance with the preferences and expectations of the DS. This can aid in protecting the DSs' privacy and ensure that their data is only used in scenarios that they are comfortable with.

7.3 OpenBanking Scenario With RegulAI

Consent is one of the legal bases that authorizes data treatment in many regulations, *e.g.,* in the General Data Protection Regulation (GDPR) and in the LGPD. Typically, consent requires that DS are informed of how the data treatment will occur and interact to accept or not the DC terms. This information is presented in a privacy policy format to DS decide whether to accept it or not. The non-acceptance of these terms often implies the DS's non-allowance to access the requested service or goods. This interaction may raise doubts and questions related to the interpretation of the consent term if not detailed, explained, and experienced (Dougherty, 2020), generating information asymmetry.

Varici (2013) defines that information asymmetry occurs when one side of the negotiation table has more or better information than the other, which may generate a hazardous environment. For instance, a company may have more information regarding how that data is being used than individuals. This asymmetry can undermine individuals' ability to protect their rights and interests and lead to a lack of trust in organizations and government bodies responsible for protecting their data. Thus, modeling consent entities and their relationship is a crucial step toward improving data agents' knowledge about how their personal data will be treated.

The challenge goes beyond jurisdictions, as an illustration, in 2018, the global bank HSBC failed to implement effective controls to prevent misuse of its services, which led to a $1.9 billion settlement with the U.S Department of Justice, and regulatory fines and penalties in other jurisdictions (Naheem, 2016). One year after, the same bank was fined £33.6 million (thirty-three million) Pounds by the UK's data protection regulator, the Information Commissioner's Office (ICO), for failing to protect customers' personal data. The ICO found that the bank had failed to implement appropriate security measures to protect personal data[1]. These cases demonstrate the challenges that multinational financial companies may face in complying with data

[1] `https://www.bbc.com/news/technology-46117963`

protection regulations, and the severe consequences of non-compliance, which can include significant fines and penalties, as well as reputational damage.

In this sense, user agents can shape and manage personal data available to be collected at the point at which that data is inserted, by whom, how, and with which constraints in the system (Cranor, Guduru, and Arjula, 2006; Berjon, 2021). Following this affirmative, Multiagent System (MAS) is an Artificial Intelligence (AI) paradigm (Wooldridge, 2009) that enables representing data agents as autonomous agents in a shared environment (Shang, 2021; Alves et al., 2023a). As agents cohabit in a shared environment, Normative MAS (NMAS) can orchestrate their behaviors by proposing rewards, punishments, obligations, prohibitions, and permissions to make agents contribute and coexist in society. Moreover, BDI (Belief-Desire-Intention) is a reasoning architecture (Wooldridge, 1999) that enables agents to decide how to accomplish their goals based on their preferences. Combining NMAS and BDI architecture can be an instrument to represent data agents' preferences and data regulation norms in order to clarify and aid agents in their decision-making process, particularly when they are faced with questions related to data sharing and what are the limits of data treatment, which was informed and consented by the DS.

In order to materialize the employment of CM, GoDReP, and RegulAI, this section presents a use case scenario in the open banking application domain. Open banking is a financial system that allows DSs to migrate their data between institutions to receive more credit, better interest rates, and fewer fees. This system provides third-party data access through application programming interfaces (APIs). Once allowed by DSs, the financial institution will be able to access the DSs' data for a specific time range. This authorization is given under the acceptance of a consent term, which defines which data will be shared, with whom, and for how long it will take. This consent term must follow the current data protection regulation according to the DC and DS location.

As mentioned in chapter 2 section 2.1, the consent is any freely given, specific, informed, and unambiguous demonstration of the DS's desire by a statement or by a clear affirmative action that signifies agreement to the processing of personal data relating to him or her. DS can revoke its consent at anytime by requesting such action for DC, following the communication channel provided in the consent term. In the open banking scenario, the consent may present different expiration dates depending on the country from DS and DC. For instance, EU sets the expiration should be ninety days at most, while Brazil determines twelve months.

This section presents a use case scenario in the open banking application domain, this use case proposes the employment of GoDReP to design an open baking use case scenario where there are two agents, John as a DS agent and Bank-B as a DC agent. John aims to share his data from Bank-A, located in an EU country under GDPR jurisdiction, to Bank-B, located in Brazil (LGPD jurisdiction), to receive offers for better interest rates, as mentioned in chapter 6 section 6.3. In this case, RegulAI will reproduce the DS's decision-making process as well as the DC norms. Next, this use case will experience the creation of a new Bank-A branch in the EU; hence, the new branch will have to follow the Brazilian open banking rules and, thus, new norms will be developed.

So, first, the scenario description was developed to contextualize the readers, providing the open banking goals and particularities. Second, the macro process was defined considering that the consent term presents all attributes foreseen in the data regulation, such as the purpose of collecting data, expiration date, sharing policies, and communication channels. Moreover, for this scenario, we consider that John (i) has given his consent to Bank-A, (ii) will give his consent to Bank-B, and (iii) then will decide to revoke his consent. Third, the process execution starts with developing the Prolog sentences to: (i) check the consent term attributes provided by Bank-B. For instance, Bank-B's purpose is to offer the best interest rates, the consent is valid for twelve months, and the data will be shared with Bank-B partners, for the same purpose, (ii) simulates the acceptance by John, (iii) settle John's rights, such as data copy and data portability requests, and (iv) finish with John revoking his consent. Fourth, and finally, the impact evaluation describes, for instance, the data breach scenario.

After the scenario development using GoDReP, the next step is to start building the RegulAI environment, defining two agents, John and Bank-B as Normative BDI agents to represent John's and Bank-B's decision-making process. Table 7.1 presents the DS agent (John agent) attributes:

However, besides the BDI attributes definition, RegulAI requires the legal bases definition and its attributes to be considered in John's decision-making process. As this scenario requires the Consent legal bases, John has to inform his consent preferences to generate the *CCI* defined by Eq. 7-4, and the *CCI*'s minimal score, i.e., the minimum acceptable value so that John can give his consent. For this use case scenario, we will consider minimum $CCI = 6$. Besides, to reach the highest score, the DC must present a consent term with:

- Purpose equals John's desire, then DSP(w) = 4,
- Expiration date equals 365, then DSP(x) = 2,

Table 7.1: John as a Normative BDI agent.

Beliefs (B_J)	*Bank-A allows DSs to share their data through the Open Bank process*
	Bank-B offers receiving data from DSs through the Open Bank process
	Bank-B offers the best interest rates on the market
Desires (D_J)	*Investing saving's balance for a year in with the best returns*
	Share financial data with Bank-B and Bank-C only
Intentions (I_J)	*it will be generated after John evaluates the consent provided by Bank-B*
Plans ($Plans_J$)	*Open an account on the new bank*
	Transfer money to this new account
	Invest this money

- Sharing policy equals to "*Share financial data with Bank-B and Bank-C only*", then DSP(y) = 1,
- DC's reputation bigger than 8, then DSP(z) = 3.

Once defined John's preferences, the next step is representing Bank-B's consent terms and the environmental norms. As defined in the scenario described using GoDReP, Bank-B offers receive data from other banks to allow DSs to create new accounts and migrate their investments. To do so, Bank-B requests the DS's consent. This consent term presents the following attributes and values:

- Purpose: *Offer the best interest rates*,
- Expiration: *12 months*,
- Sharing Policy: *Organization with the same purpose only.*

Moreover, as Bank-B is a new bank, its reputation will be considered zero. Thus, these attributes' definitions enable John to calculate the *CCI*. As $D_{DS} = P_{DC}$, $E_{DS} = E_{DC}$, $S_{DS} \neq S_{DC}$, and $R_{DS_{DC}} \neq R_{DC_{DC}}$, then $CCI = 6$. Thus, as CCI is equal to the cut score informed by John, and at this point, there is no norm addressed to John, then John has all elements to evaluate his intention defined by Eq. 7-7. First, the current beliefs enable John to follow his desire. Second, John's desire is compatible with Bank-B's terms, i.e., John will give his consent and, hence, John will be able to execute his plans as intended.

Since there is a valid consent term, Bank-B must follow what was proposed and respect the obligations foreseen in the data regulation. In order to represent the data regulation obligations, permissions, and prohibitions, Table 7.2 shows a group of norms proposed to this use case scenario following the format defined by Eq. 7-5. As well as the DS agent, from time to time, the DC agent will verify if a new norm is addressed to him, as depicted in Fig. 7.1. Then, the DC agent will verify if there is an active norm.

Table 7.2: Brazilian Open Banking Norms.

Norm Att	**Consent Request**	**Consent Revocation**	**Consent Renew**	**Data Breach**	**Data Copy**
Addressees	Bank-B	Bank-B	Bank-B	Bank-B	John
Deontic Concept	Permission	Obligation	Permission	Prohibition	Permission
Rewards	Access to DS's data	Reputation +1	Continue accessing	None DS's data	Reputation +1
Punishments	None	Reputation -3. Fine 10.000	Reputation -4. Fine 10.000	Reputation -9. Fine 20.000	Reputation -2. Fine 5.000
Activation	When requested by a DS	When requested by a DS	After 90 days, or there is a purpose update	When Bank-B access DS's data without consent	When requested by a DS
Deactivation	When DS revokes or 365 days	When data collection stops	When DS decides to renew or not	When Bank-B fix the open breach	When John receive the requested data
Purpose Limitation	Account creation	Access revocation	Access to DS's data	Safeguards DS's data	Access financial data only
Application Domain	Open Banking	Open Banking	Open Banking	Open Banking	Open Banking

As described in the GoDReP scenario, after John gives his consent, he decides to revoke it. This action activates the *Consent Revocation* norm. Then, we will begin modeling the DC's BDI attributes and the environmental norms. Table 7.3 the DC agent (Bank-B agent) attributes.

Table 7.3: Bank-B as a Normative BDI agent.

Beliefs (B_B)	*Bank-B is open to receiving new accounts request*
	John gave his consent
	John request consent revocation
	Bank-B reputation is 0
Desires (D_B)	*Avoid receiving sanctions and fines*
	Improve the reputation score
Intentions (I_B)	*it will be generated after the norms evaluation*
Plans ($Plans_B$)	*Revoke John's consent immediately if requested*
	Stop collecting and processing John's data if John withdraws his consent
	Stop sharing John's data with third parties if John withdraws his consent

The RegulAI architecture proposes constant beliefs and norms revision to verify if the sensor identifies any environment's change. This step will identify the Bank-B beliefs and the norms addressed to it. Next, the *Desires Generation* will identify which are the desires enabled considering the available beliefs.

In *Norms Status Evaluation*, Bank-B will identify that the *Consent Revocation* (CR) norm is active. In this sense, as Bank-B's desires are (i) *Avoid receiving sanctions and fines* and (ii) *Improve the reputation score*, then *NC* can be calculated as defined in Eq. 7-6. Moreover, Eq. 7-8 demonstrates the NC evaluation, i.e., $NC = 3$ if Bank-B decides to fulfill the *Consent Revocation* norm, or $NC = -1$ if Bank-B decides to violate this norm; hence, Bank-B will decide to comply with this norm.

$$NC_{CR} = \begin{cases} 2 + 1 - 0, & \text{if Bank-B decides to fulfill it} \\ 2 + 0 - 3, & \text{otherwise} \end{cases} \tag{7-8}$$

Following the RegulAI architecture, the next step is selecting the plans considering *NC*. As Bank-B decides to obey *CR* norm, Bank-B will be able to execute all plans foreseen before. Thus, Bank-B has all elements to evaluate its intention defined by Eq. 7-7. First, the current beliefs enable Bank-B to follow its desires. Second, Bank-B's *NC* allows Bank-B to execute its plans. Then, Bank-B will perform the actions needed based on his plans.

Furthermore, we designed other norms for the Open Banking scenario. For instance, *Consent Renew* is an obligation norm that requires Bank-B to request new consent from DSs. The Brazilian Open Banking regulation sets that after twelve months DC must request DS to renew his consent; otherwise, the DC must revoke the DS's consent automatically. Moreover, if there is an update in any consent attributes, DC must also request a consent renewal.

Another designed norm is the *Data Breach* norms, which defines that Bank-B is prohibited from contributing actively or passively to a data breach incident. It means that Bank-B must provide security actions to avoid a data breach; otherwise, its reputation will decrease, and it will be a target for fines and sanctions.

Finally, the *Data Copy* norm was designed to mirror the data copyright foreseen in many data regulations, such as GDPR and LGPD. This norm sets John's right to request a copy of his data from Bank-B. As a right, this norm is optional to John, i.e., John is permitted to request his data.

In another scenario explored in this context, we considered that Bank-A states in Brazil and aims to open a new bank branch in EU. Hence, Bank-A must comply with EU and Brazilian financial regulations. Following GoDReP and the NMAS modeling, this new branch can be represented as an *Organization* entity. This environment requires Bank-A to: (i) change the norm's punishment to update the fines' values, and (ii) change the norm's deactivation related to the consent expiration date, i.e., the Brazilian Open Banking foresees that consent is valid for one year, whereas the EU Open

Banking sets the limit of ninety days.

All fines' values must be updated to address the EU regulation. Moreover, the *Consent Renew* norm allows Bank-A to renew John's consent to continue accessing his data. As mentioned previously, the *Consent Renew* norm defines that Bank-A is obligated to send a renewal request if the previous consent is expired or there is an update in any consent term attributes. However, the EU Open Banking regulation sets that the consent is valid for ninety days, instead of 365 foreseen by Brazilian regulation.

Next, we will present a discussion in regard to GoDReP and RegulAI employment, the DSs, DCs, and DPs benefits, and the challenges to using this proposal. Moreover, we will also discuss other concerns that GoDReP and RegulAI could be applied, such as dynamic consent and privacy calculus. Finally, we will argue about the study's limitations.

8
Discussion and Limitations

In this chapter, we present a discussion regarding CM, GoDReP, and RegulAI employment to mitigate the data flow informational asymmetry and address the research questions. Moreover, we present the thesis limitations considering the domain and technical challenges.

8.1
Discussion

GoDReP, as a framework, allows the agents to explore the negotiation actions on many occasions based on the CM. As mentioned in the background section and in the related works, ontologies can represent data protection regulations (RQ1/ Assumption 1.a). However, even though we found PrOnto and GConsent, which are related to the GDPR, we figured out that there are no ontologies related specifically to the LGPD. In relation to *Assumption 1.b*, RegulAI demonstrated its ability to simulate scenarios that embody the principles of data regulation using NMAS concepts as observed in (Alves et al., 2023a) article.

Evaluating the weight of the pros and cons of sharing personal data is a well-known problem. For example, Lin, Yeh and Yu (2016), Kim et al. (2019), Jozani et al. (2020), and Meier and Krämer (2022) evaluated the willingness to provide personal information for personalized services based on privacy calculus, *i.e.*, based on the trade-off between the risk of data breach and the benefits of using such a service. Furthermore, they argue that personal information has economic value as an exchange between information and benefits. Thus, GoDReP and RegulAI would be used to improve the privacy calculus; hence, helping the data subject decision-making process.

Furthermore, GoDReP would enable the data subjects to find out their rights and how they can be requested in many situations. They can also explore the limits of their rights, stressing the possible data controller's behaviors. Still, Dynamic Consent enables personalized online consent by using communication platforms Budin-Ljøsne et al. (2017). Teare, Prictor, and Kaye Teare, Prictor, and Kaye (2021) defines dynamic consent as:

> "*an approach to informed consent that allows communication and*

> *engagement through a secure digital portal in ways that have not been possible before, with individuals being able to revisit and review consent decisions and preferences over time, as and when they choose.*"

In this sense, our framework would be applied to guide a digital portal development based on a data protection regulation. To do so, an ontology alignment must be done to check if any concern was not identified in the previously evaluated regulations.

In this sense, GoDReP was developed to create a computational adherence to CM and enable application scenarios representation (RQ1). According to the related works, computational adherence can be reached by an ontology instantiation tool. Such a tool can follow the ontology formalization processes and/or the application in different domain scenarios (*Assumption 1*). Moreover, the GoDReP instantiation using Prolog and Jupyter Notebook is not the unique implementation possible to materialize the metamodel concepts. However, we selected Prolog for its simplicity of environment construction and Jupyter Notebooks to improve the documentation process to approximate the agents to the metamodel instantiation.

We presented three different scenarios where we could explore their particularities. These use case scenarios are vital to employ the developed semantic, *i.e.*, the metamodel. Moreover, GoDReP would be used to align the agents' expectations before signing a contract (RQ1). For instance, if a company from European Union decides to move to Brazil, based on the company's experience, this company could use GoDReP to express its understanding and align this comprehension with Brazilian agents to evaluate if some LGPD concept was misunderstood. Also, the metamodel and GoDReP would address the questions mentioned by Sommers (2020) and Demaree-Cotton and Sommers (2022) related to the understanding of what is valid consent. Thus, such a company could perform cause-effect studies and evaluate impacts to generate a piece of evidence regarding the rights and duties interpretations. Last but not least, this company could reuse the existing scenario to base a new one.

Still, these examples present benefits to the data subject, as he/she would evaluate the scenes to decide if he/she feels confident and comfortable accepting the consent term (RQ1). The simulation of these negotiation processes could be used for the sake of mitigating the data flow informational asymmetry between the agents involved. Moreover, this simulation allows agents to evaluate the impact of their actions considering different periods.

In a preliminary study, we identified blockchain technology as a possible solution to mitigate the data flow informational asymmetry Alves et al. (2020, 2021). Such a technology emerges as a possible solution to build a unified, distributed, and trusted database. Such technology can be an entirely private ecosystem or a hybrid environment. The former allows invited entities only to read and write data. The latter enables flexible rules. Thus, permissioned blockchain platforms can be applied when the environment requires privacy concerns related to business secrecy and sensitive data. The data immutability provided by the consensus mechanisms ensures unified historical information. Also, the data distribution among the worldwide network participants guarantees high data availability. Moreover, the cryptography used in most blockchain platforms has performed satisfying results regarding data storage and transaction security. Last but not least, permissioned blockchain applications allow personalized data sharing; the data subjects can set access rules and set which data should be public, private, or accessed under case-by-case authorization.

However, even though blockchain or other technologies could be employed to mitigate the data flow informational asymmetry (RQ3), they could present a high learning curve and concerns that would extrapolate the goal of presenting a computational adherence. For instance, technical issues could emerge and it is not the focus of this work. For this reason, we evaluated blockchain technology as it is a novel approach that is supposed to deliver data immutability, process transparency, and a high availability environment.

Therefore, we proposed GoDReP to guide the generation of domain scenarios to explore some possible interpretation of the law on different occasions (*Assumption 4*). Moreover, GoDReP highlights the importance of log generation to improve the actions' explainability in the scenarios. Each "true" or "false" returned should contain information justifying such a result. In order to verify the transparency impact, the scenario's writer can clean the log to disable transparency and then verify if the explainability was compromised. Also, GoDReP allows agents to create new scenarios in order to improve the scenarios already developed. They can set conditionals to explore, for instance, the data deletion right, which depends on the purpose in the consent term. However, such creation would require a software developer and a lawyer to code the further instructions in the simulation scenario.

It is important to consider that this work does not execute decision automation or automatic law judgment. Instead, we focused on the scenarios explanation in a defined jurisdiction to deliver more resources to evaluate the environment and decisions.

Moreover, as mentioned before, we developed three application scenarios to employ GoDReP and, hence, the ontology extension, in order to identify the general attributes that could be applied in any domain (RQ2). We identified that not only *time* is a vital attribute, as supposed in *Assumption 2*, but also the compliance attributes are essential to identify circumstances where the law is not respected. Hence, the deontic operators are attributes that must be part of the scenario's elaboration. Moreover, considering the scenarios developed, we identified that they must present: (i) the purpose; (ii) time length of the processing; (iii) the data subject, controllers, and processors identification, and if the data subject is not an adult, a legal person in charge must be indicated to perform actions on the data subject's behalf; (iv) communication channel; (v) access policies, and (vi) the technology employed to store the data. However, defining all particular characteristics of all domains is not possible due to its magnitude, and it is a limitation of this thesis.

Furthermore, we experienced using the open banking feature in the real world between two well-known financial institutions after studying the Brazilian Central Bank rules and the GoDReP instantiation. This experience shows how far the ideal world could be from reality. In practice, the open banking feature did not allow us to choose which data we would share. Almost all data were mandatory to share. Moreover, regarding the consent due date, one company did not allow us to choose how many days we would like to use the feature, and the other company offered three options for the due date. Although the data subject is allowed to revoke his/her consent at anytime, enabling the data subject to preset a due date could be a good practice to be adopted.

In summary, GoDReP presents the potential to deliver (RQ3): (i) a tool to explore negotiation scenarios ruled by LGPD and templates to be applied in other domains; (ii) cause-effect exploration scenarios; (iii) impact analysis around the consent term entities and relationships, (iv) a piece of evidence regarding the interpretation in a particular moment. Moreover, particularly to the data subjects, GoDReP can aid them to: (i) exploring their rights; (ii) simulating when the rights can be triggered in different situations, and (iii) exploring the limits of their rights to establish the boundaries according to the domain and the jurisdiction. Last but not least, especially regarding the data controllers and processors, GoDReP allows: (i) an extensive exploration of their duties and rights; and (ii) the consent vulnerabilities exploration and pieces of evidence to aid them in mitigating such weaknesses. Last but not least, the pragmatic circumscription development using GoDReP would generate an epistemic effect related to the agents' learning, *i.e.*, the agents would learn while deliberating in regards to the possible law's interpretation. These are all benefits of our approach listed until now, which validates the *Assumption 4*. However, as GoDReP is an open-source framework, it would be changed to address other concerns related to data privacy regulation.

8.2 Limitations

8.2.1 CM and GoDReP

As mentioned before, LGPD recommends data anonymization, data minimization, and cryptography employment to safeguard personal data. First, many anonymization techniques could be applied, such as data masking, pseudonymization, generalization, data swapping, data perturbation, and synthetic data (Murthy et al., 2019; Majeed and Lee, 2020). However, this thesis focused on informing which anonymization technique is applied to preserve the data subject's privacy and not evaluating or recommending a specific technique.

Second, the data minimization mentioned in GDPR and LGPD requires that the collected data must be adequate, relevant, limited to the informed purpose, and limited to what is necessary concerning the purposes that they are processed (Biega et al., 2020; Goldsteen et al., 2021; Podda and Vigna, 2021; S. Bargh et al., 2021). Moreover, identifying the minimum data set to allow the data controller and processor to collect and process data is not trivial. Thus, as well as the evaluation of the data anonymization technique, the discussion of which is the most suitable data minimization method is out of our scope; however, other studies would explore this subject deeply.

Third, the cryptography techniques are also objects to be discussed on behalf of the data subject's privacy, and many studies have presented different approaches to explore this area, *e.g.*, using an asymmetric cryptography key-pair in a blockchain environment (Truong et al., 2019), post-quantum cryptography techniques (Malina et al., 2021), and zero-knowledge proofs or homomorphic encryption (Limniotis, 2021), among others. Thus, this thesis focused on letting the agents know the need to inform the applied cryptography technique, but we did not profoundly explore this question.

Even though GoDReP is designed to use CM, GoDReP could be utilized in use-case scenarios based on GPDR as well, because CM is aligned with GDPR ontologies and LGPD concerns. Thus, other data privacy regulations could use GoDReP, but an ontology alignment is required to tune CM. Moreover, when constructing consent terms, different jurisdictions may require different vocabularies to meet comprehensibility and explainability requirements. This can lead to potential issues when moving a scenario from one jurisdiction to another, even if the scenarios have many aspects in common. For example, migrating from GDPR to LGPD requires agents to translate and adapt different terms, which can increase their cognitive load. Kelley, Bresee, Cranor, and Reeder (2009) propose a labeling technique that may fill this gap of comprehensibility and explainability.

Furthermore, as GoDReP proposes the evaluation of data privacy regulation and the development of the use-case scenarios, to use GoDReP correctly is recommended to proceed with at least one person from the Law sector and one from the IT sector, or someone with programming skills. The pair programming would improve the scenario development from the law and the IT perspective, perhaps producing a complete description based on the law and a high-quality programming code. Still, even though this thesis considered use-case scenarios of different natures, other scenarios would require the development of new functions or even changes in the framework structure.

Finally, the application scenarios were not explored exhaustively, given the many possible forks that a scenario would have. However, to the best of our knowledge, we tried to produce the scenarios as complete as possible, considering our expertise in the respective domain areas. Still, as we focused on the CM and the computational representation of metamodel, we did not conduct qualitative studies to evaluate the agents' adherence to the developed framework.

8.2.2 RegulAI

RegulAI aims to represent the scenarios elaborated using GoDReP by addressing data regulation concerns in NMAS. However, to work correctly, the agent's desires and goals must be compared with the consent's purpose. However, since both desires and purposes are expressed in natural language, the automatic comparison may be challenging. A possible solution would be the usage of communication templates with a limited vocabulary to represent these sentences as program commands. Otherwise, the comparison would really focus solely on Natural Language Processing techniques.

Furthermore, eventually, norms can conflict, and this thesis does not propose a normative conflict resolution in this case. However, there are numerous normative resolution techniques, and they require an in-depth study focused on this point (Vasconcelos, Kollingbaum, and Norman, 2009; Santos et al., 2017; Alves et al., 2018; Silvestre et al., 2019).

Last but not least, as well as mentioned as a GoDReP's limitation, this thesis is focused on the Consent Legal Bases. However, we did not evaluate other Legal Bases that could benefit from GoDReP, and RegulAI proposals, if applicable. Another similar limitation is changing the data regulation, which was not considered in this article. One could argue that there are other data regulation, or NMAS, relevant aspects that were not emerged and addressed in our research.

In this chapter, we presented the discussion regarding the use of the proposed framework as well as the possible impacts in the dynamic consent and privacy calculus studies. We also presented the limitations considering the technical and the domain challenges. Next, we present the conclusions and future work.

9
Conclusions and Future Work

Data flow informational asymmetry is a global challenge faced by many data protection and regulation worldwide. GDPR and LGPD are examples of data regulation in the European Union and Brazil, respectively. However, the DSs will exercise their rights superficially for its complexity, lack of studies, and lack of empirical studies. Also, DCs and DPs can be penalized if they do not comply with the regulation employed in their jurisdiction.

Thus, the first step to mitigate the data flow informational asymmetry was to analyze studies regarding GDPR ontologies to propose changes if needed. As a result, we found PrOnto and GConsent, ontologies based on GDPR. We proposed ODPM and CM considering these two ontologies, inserting the LGPD specificities.

Next, in order to offer a structured manner to create and evaluate use case scenarios, we proposed a framework based on the extended ontology. GoDReP is a framework that aims to provide a method to build scenarios in different domains for DSs, DCs, and DPs.

The use-case scenarios are vital to increasing agents' knowledge of data protection requirements, rights, and duties. Therefore, we presented three different domains to exercise GoDReP's use. The generated material can be used by any agents that wish to build new scenarios, even in other domains. Also, this material would be used as evidence of the deal between the agents involved. This may represent the clauses' understanding and the expected behavior for each pragmatic circumscription.

Based on the scenario description developed in GoDReP, RegulAI enables data agents to represent their goals, plans, and environmental norms by employing BDI reasoning architecture in a NMAS to express data regulation concerns and expectations regarding the collection, storage, and use of their data. The BDI architecture represents the agent's preferences, and the NMAS defines the data regulation norms that agents must evaluate whether they comply with or not, considering the norm's rewards and punishments.

RegulAI defines a CCI to aid DS agents in evaluating their preferences versus the consent term purpose. Once the preferences are aligned with the consent term purpose, the CCI will return a value, and the DS agent will choose whether to share personal data based on the minimum score defined previously.

For future work, we believe that developing scenarios in other domains will be a valuable contribution. People will have more examples of GoDReP employment, which may facilitate the new constructions considering that there will be more material to reuse. For instance, regarding data portability, considering a DS that aims to move his/her data from Facebook to Twitter, the agent could ask Twitter how the data will be processed considering the limitation of 280 characters.

Still, whether a DS decides to move his/her data from Facebook to a life insurance company, the DS should be aware of the possibility of selecting the posts and photos to be shared. If so, the life insurance company should be aware that the DS could hide data that could influence the company's offer. Conversely, if the DS cannot choose the data that he/she wants to share, sensitive information could be shared unnecessarily. Those scenarios could be explored by agents using GoDReP to clarify the possibilities available in order to mitigate the data flow information asymmetry.

Another future work is related to the use of GoDReP by Judges to document case laws. For example, judges could feed the scenario repository to generate a live acknowledgment. Also, this repository would be used to evaluate and understand the impacts of law changes. Also, Natural Language Processing (NLP) and Machine Learning (ML) techniques would be applied to evaluate the consent's purpose and the agent's goals and plans to improve the compatibility between them. For instance, NLP and ML can be used in healthcare to analyze the language used in consent forms and patient communication, identify gaps in information, and personalize consent based on patient preferences or limitations. This improves the compatibility between patient consent purposes and the agent's goals.

Furthermore, as mentioned in the limitation section, norms may conflict, and deciding which norm to comply with is not trivial. Thus, other future work is on the normative resolution direction. In this context, another future work is developing an in-depth study on the reputation systems to improve the agent's reputation capabilities. Finally, RegulAI would be used, for instance, to monitor systems and notify DSs, DCs, and DPs when a data breach occurs or the DS's personal data is used inappropriately.

Finally, RegulAI provide simulation scenarios where there are different DS's rewards and goals combined with reinforcement learning to improve the simulation environment and test the agents' behaviour, including the impact of these behaviours on the DC and DPs perspective.

Bibliography

ALVES, P. H. C.; VIANA, M. L. ; DE LUCENA, C. J. P.. **An architecture for autonomous normative bdi agents based on personality traits to solve normative conflicts.** In: INTERNATIONAL CONFERENCE ON AGENTS AND ARTIFICIAL INTELLIGENCE (ICAART), p. 80–90, 2018.

ALVES, P. H.; FRAJHOF, I. Z.; CORREIA, F. A.; DE SOUZA, C. ; LOPES, H.. **Permissioned blockchains: Towards privacy management and data regulation compliance.** In: JURIX, p. 211–214. IOS Press, 2020.

ALVES, P. H.; FRAJHOF, I. Z.; CORREIA, F. A.; DE SOUZA, C. ; LOPES, H.. **Controlling personal data flow: An ontology in the covid-19 outbreak using a permissioned blockchain.** In: PROCEEDINGS OF THE 23RD INTERNATIONAL CONFERENCE ON ENTERPRISE INFORMATION SYSTEMS - VOLUME 2: ICEIS, 2021.

ALVES, P. H.; CORREIA, F.; FRAJHOF, I.; SIECKENIUS DE SOUZA., C. ; LOPES, H.. **A normative multiagent approach to represent data regulation concerns.** In: PROCEEDINGS OF THE 15TH INTERNATIONAL CONFERENCE ON AGENTS AND ARTIFICIAL INTELLIGENCE - VOLUME 1: ICAART, p. 330–337. INSTICC, SciTePress, 2023.

ALVES, P. H.; CORREIA, F.; FRAJHOF, I.; DE SOUZA, C. ; LOPES, H.. **Designing intelligent agents in normative systems toward data regulation representation.** IEEE Access, 0:1–16, 2023.

BANDARA, E.; LIANG, X.; FOYTIK, P.; SHETTY, S.; HALL, C.; BOWDEN, D.; RANASINGHE, N. ; DE ZOYSA, K.. **A blockchain empowered and privacy preserving digital contact tracing platform.** Information Processing & Management, 58(4):102572, 2021.

BANERJEE, S. S.; HEMPHILL, T. ; LONGSTREET, P.. **Wearable devices and healthcare: Data sharing and privacy.** Information Society, 34(1):49–57, jan 2018.

BARTOLINI, C.; MUTHURI, R.. **Reconciling data protection rights and obligations: An ontology of the forthcoming eu regulation.** In:

WORKSHOP ON LANGUAGE AND SEMANTIC TECHNOLOGY FOR LEGAL DOMAIN (LST4LD), 09 2015.

BELLIFEMINE, F.; POGGI, A. ; RIMASSA, G.. **Jade–a fipa-compliant agent framework.** In: PROCEEDINGS OF PAAM, volumen 99, p. 97–108. London, 1999.

BERJON, R.. **The fiduciary duties of user agents.** Available at SSRN 3827421, p. 1–15, April 2021.

BIEGA, A. J.; POTASH, P.; DAUMÉ, H.; DIAZ, F. ; FINCK, M.. **Operationalizing the legal principle of data minimization for personalization.** In: PROCEEDINGS OF THE 43RD INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, p. 399–408, 2020.

BREUER, J.; PIERSON, J.. **The right to the city and data protection for developing citizen-centric digital cities.** Information, Communication & Society, 24(6):797–812, 2021.

BUDIN-LJØSNE, I.; TEARE, H. J.; KAYE, J.; BECK, S.; BENTZEN, H. B.; CAENAZZO, L.; COLLETT, C.; D'ABRAMO, F.; FELZMANN, H.; FINLAY, T. ; OTHERS. **Dynamic consent: a potential solution to some of the challenges of modern biomedical research.** BMC medical ethics, 18(1):1–10, 2017.

CAMPANILE, L.; IACONO, M.; MARULLI, F. ; MASTROIANNI, M.. **Designing a gdpr compliant blockchain-based iov distributed information tracking system.** Information Processing & Management, 58(3):102511, 2021.

CARDOSO, A.; LEITÃO, J. ; TEIXEIRA, C.. **Using the jupyter notebook as a tool to support the teaching and learning processes in engineering courses.** In: INTERNATIONAL CONFERENCE ON INTERACTIVE COLLABORATIVE LEARNING, p. 227–236. Springer, 2018.

CARROLL, J. M.. **Making use: scenario-based design of human-computer interactions.** MIT press, 2003.

CASALICCHIO, E.; CARDELLINI, V.; INTERINO, G. ; PALMIRANI, M.. **Research challenges in legal-rule and qos-aware cloud service brokerage.** Future Generation Computer Systems, 78:211–223, 2018.

KLUYVER, T.; RAGAN-KELLEY, B.; PÉREZ, F.; GRANGER, B. E.; BUSSONNIER, M.; FREDERIC, J.; KELLEY, K.; HAMRICK, J. B.; GROUT, J.; CORLAY, S. ; OTHERS. **Jupyter notebooks-a publishing format for reproducible computational workflows.** In: INTERNATIONAL CONFERENCE ON ELECTRONIC PUBLISHING, 2016.

CLOCKSIN, W. F.; MELLISH, C. S.. **Programming in PROLOG**. Springer Science & Business Media, 2003.

COLLIERF, N.; GOODWIN, R. M.; MCCRAE, J. P.; DOAN, S.; KAWAZOE, A.; CONWAY, M.; KAWTRAKUL, A.; TAKEUCHI, K. ; DIEN, D.. **An ontology-driven system for detecting global health events**. In: PROCEEDINGS OF THE 23RD INTERNATIONAL CONFERENCE ON COMPUTATIONAL LINGUISTICS (COLING 2010), p. 215–222, 2010.

CORAZZON, R.. **Theory and history of ontology**. 2014.

CRANOR, L. F.; GUDURU, P. ; ARJULA, M.. **User interfaces for privacy agents**. ACM Transactions on Computer-Human Interaction (TOCHI), 13(2):135–178, 2006.

CUNHA, F.; MARX, L.; ROSEMBERG, M. ; LUCENA, C.. **Verifying the behavior of agents in bdi4jade with aspectj**. WESAAC, 2015.

DEMAREE-COTTON, J.; SOMMERS, R.. **Autonomy and the folk concept of valid consent**. Cognition, 224:105065, 2022.

DOUGHERTY, T.. **Informed consent, disclosure, and understanding**. Philosophy & Public Affairs, 48(2):119–150, 2020.

DRAGISIC, Z.; IVANOVA, V.; LAMBRIX, P.; FARIA, D.; JIMÉNEZ-RUIZ, E. ; PESQUITA, C.. **User validation in ontology alignment**. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, p. 200–217. Springer, 2016.

DUBEY, A.; ABHINAV, K.; JAIN, S.; ARORA, V. ; PUTTAVEERANA, A.. **Haco: a framework for developing human-ai teaming**. In: PROCEEDINGS OF THE 13TH INNOVATIONS IN SOFTWARE ENGINEERING CONFERENCE ON FORMERLY KNOWN AS INDIA SOFTWARE ENGINEERING CONFERENCE, p. 1–9, 2020.

ERONEN, J.; PTASZYNSKI, M.; MASUI, F.; SMYWIŃSKI-POHL, A.; LELIWA, G. ; WROCZYNSKI, M.. **Improving classifier training efficiency for automatic cyberbullying detection with feature density**. Information Processing & Management, 58(5):102616, 2021.

FARROW, G. S.. **Open banking: The rise of the cloud platform**. Journal of Payments Strategy & Systems, 14(2):128–146, 2020.

FATEMA, K.; HADZISELIMOVIC, E.; PANDIT, H. J.; DEBRUYNE, C.; LEWIS, D. ; O'SULLIVAN, D.. **Compliance through informed consent: Semantic based consent permission and data management model.** In: PRIVON@ ISWC, 2017.

FOSCH-VILLARONGA, E.; POULSEN, A.; SØRAA, R. A. ; CUSTERS, B.. **A little bird told me your gender: Gender inferences in social media**. Information Processing & Management, 58(3):102541, 2021.

FREIRE, E. S. S.; CORTÉS, M. I.; JÚNIOR, R. M. D. R.; GONÇALVES, Ê. J. T. ; DE LIMA, G. A. C.. **Normas-ml: Supporting the modeling of normative multi-agent systems**. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 8(4):49–81, 2019.

GANGEMI, A.; CATENACCI, C.; CIARAMITA, M. ; LEHMANN, J.. **Modelling ontology evaluation and validation**. In: EUROPEAN SEMANTIC WEB CONFERENCE, p. 140–154. Springer, 2006.

GARCÍA-PEÑALVO, F. J.; COLOMO-PALACIOS, R.; GARCÍA, J. ; THERÓN, R.. **Towards an ontology modeling tool. a validation in software engineering scenarios**. Expert Systems with Applications, 39(13):11468–11478, 2012.

GHARIB, M.; GIORGINI, P. ; MYLOPOULOS, J.. **Copri v. 2—a core ontology for privacy requirements**. Data & Knowledge Engineering, 133:101888, 2021.

GHARIB, M.; MYLOPOULOS, J.. **A core ontology for privacy requirements engineering**. arXiv preprint arXiv:1811.12621, 2018.

GODARD, B.; SCHMIDTKE, J.; CASSIMAN, J.-J. ; AYMÉ, S.. **Data storage and dna banking for biomedical research: informed consent, confidentiality, quality issues, ownership, return of benefits. a professional perspective**. European Journal of Human Genetics, 11(2):S88–S122, 2003.

GOLDSTEEN, A.; EZOV, G.; SHMELKIN, R.; MOFFIE, M. ; FARKASH, A.. **Data minimization for gdpr compliance in machine learning models**. AI and Ethics, p. 1–15, 2021.

GONÇALVES, E. J. T.; CORTÉS, M. I.; CAMPOS, G. A. L.; LOPES, Y. S.; FREIRE, E. S.; DA SILVA, V. T.; DE OLIVEIRA, K. S. F. ; DE OLIVEIRA, M. A.. **Mas-ml 2.0: Supporting the modelling of multi-agent systems with different agent architectures**. Journal of Systems and Software, 108:77–109, 2015.

GUARINO, N.; OBERLE, D. ; STAAB, S.. **What is an ontology?** In: HANDBOOK ON ONTOLOGIES, p. 1–17. Springer, 2009.

HARDIN, T.; KOTZ, D.. **Amanuensis: Information provenance for health-data systems**. Information Processing & Management, 58(2):102460, 2021.

HE, Y.; SARNTIVIJAI, S.; LIN, Y.; XIANG, Z.; GUO, A.; ZHANG, S.; JAGANNATHAN, D.; TOLDO, L.; TAO, C. ; SMITH, B.. **Oae: the ontology of adverse events**. Journal of biomedical semantics, 5(1):1–13, 2014.

HEAVIN, C.; POWER, D. J.. **Challenges for digital transformation–towards a conceptual decision support guide for managers**. Journal of Decision Systems, 27(sup1):38–45, 2018.

JOZANI, M.; AYABURI, E.; KO, M. ; CHOO, K.-K. R.. **Privacy concerns and benefits of engagement with social media-enabled apps: A privacy calculus perspective**. Computers in Human Behavior, 107:106260, 2020.

KELLEY, P. G.; BRESEE, J.; CRANOR, L. F. ; REEDER, R. W.. **A "nutrition label" for privacy**. In: PROCEEDINGS OF THE 5TH SYMPOSIUM ON USABLE PRIVACY AND SECURITY, SOUPS '09, New York, NY, USA, 2009. Association for Computing Machinery.

KIM, D.; PARK, K.; PARK, Y. ; AHN, J.-H.. **Willingness to provide personal information: Perspective of privacy calculus in iot services**. Computers in Human Behavior, 92:273–281, 2019.

KIRRANE, S.; FERNÁNDEZ, J. D.; DULLAERT, W.; MILOSEVIC, U.; POLLERES, A.; BONATTI, P. A.; WENNING, R.; DROZD, O. ; RASCHKE, P.. **A scalable consent, transparency and compliance architecture**. In: EUROPEAN SEMANTIC WEB CONFERENCE, p. 131–136. Springer, 2018.

KRAUS, S.; DURST, S.; FERREIRA, J. J.; VEIGA, P.; KAILER, N. ; WEINMANN, A.. **Digital transformation in business and management research: An overview of the current status quo**. International Journal of Information Management, 63:102466, 2022.

KURTEVA, A.; CHHETRI, T. R.; PANDIT, H. J. ; FENSEL, A.. **Consent through the lens of semantics: State of the art survey and best practices**. Semantic Web, (Preprint):1–27, 2021.

LI, Z.; YANG, M. C. ; RAMANI, K.. **A methodology for engineering ontology acquisition and validation**. AI EDAM, 23(1):37–51, 2009.

LIMNIOTIS, K.. **Cryptography as the means to protect fundamental human rights**. Cryptography, 5(4):34, 2021.

LIN, C.-Y.; YEH, J.-Y. ; YU, Y.-T.. **The influence of privacy calculus, user interface quality and perceived value on mobile shopping**. Journal of Economics, Business and Management, 4(10):567–572, 2016.

LUCK, M.; MAHMOUD, S.; MENEGUZZI, F.; KOLLINGBAUM, M.; NORMAN, T. J.; CRIADO, N. ; FAGUNDES, M. S.. **Normative agents**. In: AGREEMENT TECHNOLOGIES, p. 209–220. Springer, 2013.

MA, S.; GUO, C.; WANG, H.; XIAO, H.; XU, B.; DAI, H.-N.; CHENG, S.; YI, R. ; WANG, T.. **Nudging data privacy management of open banking based on blockchain**. In: 2018 15TH INTERNATIONAL SYMPOSIUM ON PERVASIVE SYSTEMS, ALGORITHMS AND NETWORKS (I-SPAN), p. 72–79. IEEE, 2018.

MA, X.; WANG, Y.; GAO, T.; HE, Q.; HE, Y.; YUE, R.; YOU, F. ; TANG, J.. **Challenges and strategies to research ethics in conducting covid-19 research**. Journal of Evidence-Based Medicine, 13(2):173–177, 2020.

MAHMOUD, M. A.; AHMAD, M. S. ; MOSTAFA, S. A.. **Norm-based behavior regulating technique for multi-agent in complex adaptive systems**. IEEE Access, 7:126662–126678, 2019.

MAJEED, A.; LEE, S.. **Anonymization techniques for privacy preserving data publishing: A comprehensive survey**. IEEE Access, 2020.

MAKHLOUF, K.; ZHIOUA, S. ; PALAMIDESSI, C.. **Machine learning fairness notions: Bridging the gap with real-world applications**. Information Processing & Management, 58(5):102642, 2021.

MALINA, L.; DZURENDA, P.; RICCI, S.; HAJNY, J.; SRIVASTAVA, G.; MATULEVIČIUS, R.; AFFIA, A.-A. O.; LAURENT, M.; SULTAN, N. H. ; TANG, Q.. **Post-quantum era privacy protection for intelligent infrastructures**. IEEE Access, 9:36038–36077, 2021.

MARELLI, L.; LIEVEVROUW, E. ; VAN HOYWEGHEN, I.. **Fit for purpose? The GDPR and the governance of European digital health.** Policy Studies, 41(5):447–467, sep 2020.

MCGHIN, T.; CHOO, K. K. R.; LIU, C. Z. ; HE, D.. **Blockchain in healthcare applications: Research challenges and opportunities**, 2019.

MCNAMARA, P.. **Deontic logic.** In: HANDBOOK OF THE HISTORY OF LOGIC, volumen 7, p. 197–288. Elsevier, 2006.

MEIER, Y.; KRÄMER, N. C.. **The privacy calculus revisited: an empirical investigation of online privacy decisions on between-and within-person levels.** Communication Research, p. 00936502221102101, 2022.

MENSE, A.; BLOBEL, B.. **Hl7 standards and components to support implementation of the european general data protection regulation.** European Journal for Biomedical Informatics, 13(1):27–33, 2017.

MILKAITE, I.; LIEVENS, E.. **Counting down to 25 may 2018: Mapping the gdpr age of consent across the eu.** Better Internet for Kids, 2018.

MISHRA, R. A.; KALLA, A.; BRAEKEN, A. ; LIYANAGE, M.. **Privacy protected blockchain based architecture and implementation for sharing of students' credentials.** Information Processing & Management, 58(3):102512, 2021.

MITTAL, S.; SHARMA, P.. **The role of consent in legitimising the processing of personal data under the current eu data protection framework.** Asian Journal of Computer Science And Information Technology, 7(4):76–78, 2017.

MIYACHI, K.; MACKEY, T. K.. **hocbs: A privacy-preserving blockchain framework for healthcare data leveraging an on-chain and off-chain system design.** Information Processing & Management, 58(3):102535, 2021.

MULHOLLAND, C.; FRAJHOF, I. Z.. **A LGPD e o novo marco normativo no Brasil.** Arquipelago, 1st edition, 2020.

MURTHY, S.; BAKAR, A. A.; RAHIM, F. A. ; RAMLI, R.. **A comparative study of data anonymization techniques.** In: 2019 IEEE 5TH INTL CONFERENCE ON BIG DATA SECURITY ON CLOUD (BIGDATASECURITY), IEEE INTL CONFERENCE ON HIGH PERFORMANCE AND SMART COMPUTING,(HPSC) AND IEEE INTL CONFERENCE ON INTELLIGENT DATA AND SECURITY (IDS), p. 306–309. IEEE, 2019.

MUSESENGWA, R.; CHIMBARI, M. J. ; MUKARATIRWA, S.. **Initiating community engagement in an ecohealth research project in southern africa**. Infectious diseases of poverty, 6(1):22, 2017.

NAHEEM, M. A.. **Risk of money laundering in the us: Hsbc case study**. Journal of Money Laundering Control, 2016.

DOS SANTOS NETO, B. F.; DA SILVA, V. T. ; DE LUCENA, C. J. P.. **Developing goal-oriented normative agents: The nbdi architecture**. In: Filipe, J.; Fred, A., editors, AGENTS AND ARTIFICIAL INTELLIGENCE, p. 176–191, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

OFFICE OF THE PRIVACY COMMISSIONER OF CANADA. **PIPEDA fair information principles**, 2019.

OLIVEIRA, G. C.; DIAS, C. A.; SANTOS, A. ; SANTOS, C.. **The law of access to information and its applicability in brazilian municipalities**. International Journal Of Development Research, 10(3):34478–34483, 2020.

PALMIRANI, M.; MARTONI, M.; ROSSI, A.; BARTOLINI, C. ; ROBALDO, L.. **Pronto: Privacy ontology for legal reasoning**. In: INTERNATIONAL CONFERENCE ON ELECTRONIC GOVERNMENT AND THE INFORMATION SYSTEMS PERSPECTIVE, p. 139–152. Springer, 2018.

PANDIT, H. J.; DEBRUYNE, C.; O'SULLIVAN, D. ; LEWIS, D.. **Gconsent-a consent ontology based on the gdpr**. In: EUROPEAN SEMANTIC WEB CONFERENCE, p. 270–282. Springer, 2019.

PARIS, A.; BRANDT, C.; CORNU, C.; MAISON, P.; THALAMAS, C. ; CRACOWSKI, J.-L.. **Informed consent document improvement does not increase patients' comprehension in biomedical research**. British journal of clinical pharmacology, 69(3):231–237, 2010.

PERKEL, J. M.. **Why jupyter is data scientists' computational notebook of choice**. Nature, 563(7732):145–147, 2018.

PHILLIPS, M.. **International data-sharing norms: from the oecd to the general data protection regulation (gdpr)**. Human genetics, 137(8):575–582, 2018.

PODDA, E.; VIGNA, F.. **Anonymization between minimization and erasure: The perspectives of french and italian data protection authorities**. In: INTERNATIONAL CONFERENCE ON ELECTRONIC GOVERNMENT AND THE INFORMATION SYSTEMS PERSPECTIVE, p. 103–114. Springer, 2021.

PRIYONO, A.; MOIN, A. ; PUTRI, V. N. A. O.. **Identifying digital transformation paths in the business model of smes during the covid-19 pandemic**. Journal of Open Innovation: Technology, Market, and Complexity, 6(4):104, 2020.

RANDLES, B. M.; PASQUETTO, I. V.; GOLSHAN, M. S. ; BORGMAN, C. L.. **Using the jupyter notebook as a tool for open science: An empirical study**. In: 2017 ACM/IEEE JOINT CONFERENCE ON DIGITAL LIBRARIES (JCDL), p. 1–2. IEEE, 2017.

RODOTÀ, S.. **A vida an sociedade de vigilância: a privacidade hoje**. Renovar, 2008.

ROSMAINI, E.; KUSUMASARI, T. F.; LUBIS, M. ; LUBIS, A. R.. **Study to the current protection of personal data in the educational sector in indonesia**. In: JOURNAL OF PHYSICS: CONFERENCE SERIES, volumen 978, p. 012037. IOP Publishing, 2018.

RULE, A.; BIRMINGHAM, A.; ZUNIGA, C.; ALTINTAS, I.; HUANG, S.-C.; KNIGHT, R.; MOSHIRI, N.; NGUYEN, M. H.; ROSENTHAL, S. B.; PÉREZ, F. ; OTHERS. **Ten simple rules for writing and sharing computational analyses in jupyter notebooks**, 2019.

S. BARGH, M.; MEIJER, R.; VAN DEN BRAAK, S.; LATENKO, A.; VINK, M. ; CHOENNI, S.. **Embedding personal data minimization technologies in organizations: needs, vision and artifacts**. In: 14TH INTERNATIONAL CONFERENCE ON THEORY AND PRACTICE OF ELECTRONIC GOVERNANCE, p. 71–79, 2021.

SANTOS, J. S.; ZAHN, J. O.; SILVESTRE, E. A.; SILVA, V. T. ; VASCONCELOS, W. W.. **Detection and resolution of normative conflicts in multi-agent systems: a literature survey**. Autonomous agents and multi-agent systems, 31:1236–1282, 2017.

SARABDEEN, J.; ISHAK, M. M. M.. **Impediment of privacy in the use of clouds by educational institutions**. 2015.

SHANG, Y.. **Consensus formation in networks with neighbor-dependent synergy and observer effect**. Communications in Nonlinear Science and Numerical Simulation, 95:105632, 2021.

SIIBAK, A.; MASCHERONI, G.. **Children's data and privacy in the digital age**. 2021.

SILVESTRE, E. A.; DA SILVA, V. T.; DA SILVA, O. D. S.; OLIVEIRA, L. F. ; RAMOS, W. S.. **A platform for detection and resolution of conflicts among multiple norms in mas (multi-agent systems)**. Revista Inova Ciência & Tecnologia/Innovative Science & Technology Journal, p. 74–83, 2019.

SMITH, B.. **Ontology**. In: THE FURNITURE OF THE WORLD, p. 47–68. Brill, 2012.

SOMMERS, R.. **Commonsense consent**. Yale Law Journal, 2020.

STAAB, S.; STUDER, R.. **Handbook on ontologies**. Springer Science & Business Media, 2010.

STOILOVA, M.; NANDAGIRI, R. ; LIVINGSTONE, S.. **Children's understanding of personal data and privacy online – a systematic evidence mapping**. Information, Communication & Society, 24(4):557–575, 2021.

TARTIR, S.; ARPINAR, I. B. ; SHETH, A. P.. **Ontological evaluation and validation**. In: THEORY AND APPLICATIONS OF ONTOLOGY: COMPUTER APPLICATIONS, p. 115–130. Springer, 2010.

TEARE, H. J.; PRICTOR, M. ; KAYE, J.. **Reflections on dynamic consent in biomedical research: the story so far**. European Journal of Human Genetics, 29(4):649–656, 2021.

TEIXEIRA, D. A.; PROCOPIUCK, M.; REZENDE, D. ; ANDRADE, P.. **Public administration: A critical analysis of the brazilian law on access to public information**. Revista Juridica, 2:493–506, 2017.

TIKKINEN-PIRI, C.; ROHUNEN, A. ; MARKKULA, J.. **Eu general data protection regulation: Changes and implications for personal data collecting companies**. Computer Law & Security Review, 34(1):134–153, 2018.

TOLLEY, E. E.; ULIN, P. R.; MACK, N.; ROBINSON, E. T. ; SUCCOP, S. M.. **Qualitative methods in public health: a field guide for applied research**. John Wiley & Sons, 2016.

PANCH, T.; MATTIE, H. ; CELI, L. A.. **The "inconvenient truth" about ai in healthcare**. NPJ digital medicine, 2(1):1–3, 2019.

TRUONG, N. B.; SUN, K.; LEE, G. M. ; GUO, Y.. **Gdpr-compliant personal data management: A blockchain-based solution**. IEEE Transactions on Information Forensics and Security, 15:1746–1761, 2019.

VARICI, I.. **The relationship between information asymmetry and the quality of audit: An empirical study in istanbul stock exchange.** International Business Research, 6(10):132, 2013.

VASCONCELOS, W. W.; KOLLINGBAUM, M. J. ; NORMAN, T. J.. **Normative conflict resolution in multi-agent systems.** Autonomous agents and multi-agent systems, 19:124–152, 2009.

VELMOVITSKY, P. E.; SOUZA, P. A. D. S. E.; VAILLANCOURT, H.; DONOVSKA, T.; TEAGUE, J.; MORITA, P. P. ; OTHERS. **A blockchain-based consent platform for active assisted living: Modeling study and conceptual framework.** Journal of Medical Internet Research, 22(12):e20832, 2020.

VIANA, M.; ALENCAR, P.; GUIMARÃES, E.; CIRILO, E. ; LUCENA, C.. **Creating a modeling language based on a new metamodel for adaptive normative software agents.** IEEE Access, 10:13974–13996, 2022.

VIVES, X.. **Digital disruption in banking.** Annual Review of Financial Economics, 11:243–272, 2019.

WANG, J.; TZU-YANG, K.; LI, L. ; ZELLER, A.. **Assessing and restoring reproducibility of jupyter notebooks.** In: 2020 35TH IEEE/ACM INTERNATIONAL CONFERENCE ON AUTOMATED SOFTWARE ENGINEERING (ASE), p. 138–149. IEEE, 2020.

WOOLDRIDGE, M.. **Intelligent agents.** Multiagent systems: A modern approach to distributed artificial intelligence, 1:27–73, 1999.

WOOLDRIDGE, M.. **An introduction to multiagent systems.** John wiley & sons, 2009.

VON WRIGHT, G. H.. **Deontic logic.** Mind, 60(237):1–15, 1951.

YU, S.; TIAN, Y.; GUO, S. ; WU, D. O.. **Can we beat ddos attacks in clouds?** IEEE Transactions on Parallel and Distributed Systems, 25(9):2245–2254, 2013.

ŽARNIĆ, B.; BAŠIĆ, G.. **Metanormative principles and norm governed social interaction.** Revus. Journal for Constitutional Theory and Philosophy of Law/Revija za ustavno teorijo in filozofijo prava, (22):105–120, 2014.