A Preliminary Design of Privacy Ontology for Smart Toys

Otavio de Paula Albuquerque
University of São Paulo
otavioalbuquerque@usp.br

Patrick C. K. Hung
Ontario Tech University
patrick.hung@ontariotechu.ca

Sarajane Marques Peres
University of São Paulo
sarajane@usp.br

Farkhund Iqbal
Zayed University
farkhund.iqbal@zu.ac.ae

Marcelo Fantinato
University of São Paulo
m.fantinato@usp.br

Abstract

Given the technological progress, smart toys have become relevant in the toy market. Toy companies adopt different requirements and web services to create smart toy features in different shapes and purposes. Each company usually has its requirements and implementation process, including semantic information and risk management guidelines. In other words, there is no common knowledge base related to the smart toy domain, in which the organizations could share information and reuse standardized knowledge, mitigating interoperability issues. Our work aims to build a smart toy's privacy context ontology, bringing general concepts and privacy-related, machine-readable, offering organizations and software agents a common knowledge base related to privacy on smart toy's context to reuse for smart toys design and features implementation.

Keywords: smart toys, toy computing, semantic web, ontology, privacy

1. Introduction

A smart toy is a physical toy embedded with devices and sensors, linked to smart mobile services via wireless networks, offering computational power and communication capabilities. This integration boosts interactivity and personalization for children (Rafferty et al., 2017). With smaller processing circuits and reduced costs, smart toys have joined the internet of things (IoT), enabling real-time data collection and processing (Rafferty et al., 2017). The smart toy market is expected to grow substantially, with an estimated 200% increase from 2018 to 2023 (JR, 2018). As smart toys become more popular, they incorporate context-sensitive web services, requiring data collection for full functionality.

Key concerns in this domain pertain to privacy and security. Safeguarding the privacy of collected information constitutes a paramount challenge for smart toys (Hung, Iqbal, et al., 2016), mirroring concerns in the broader IoT landscape. This challenge assumes added complexity due to the primary users of smart toys being children, who typically lack awareness of privacy risks. Moreover, such data may include personal identifiable information (PII), a category encompassing data capable of distinguishing or tracing an individual's identity (Radack, 2010). Instances of privacy breaches associated with smart toys are well-documented in the internet. Notably, Mattel's Hello Barbie cases, an interactive smart doll designed to simulate genuine two-way conversations, was found to collect personal data beyond user interactions and, on occasion, share this data with third parties. This practice contravened the data protection provisions set forth in the Children's Online Privacy Protection Act (COPPA), a United States federal law devised to safeguard the privacy of children under 13 in the context of online data collection. Another real-world example is the case of My Friend Cayla doll, which was banned in Germany due to the perceived risks of privacy breaches and security vulnerabilities.

Companies address privacy risks through various approaches, often relying on internal risk management protocols and guidelines. Regrettably, these measures often fall short of ensuring true privacy protection. Privacy, defined as an individual's right to determine how, when, and to what extent information about themselves could be disclosed to another entity or organization (Liu and Özsu, 2009), is imperiled by unauthorized data collection or disclosures. This risk is amplified when the data in question is of a personal nature. Notable privacy concerns include the potential for eavesdropping, where the toy becomes a data-gathering device, and the risk

URI: https://hdl.handle.net/10125/106596
978-0-9981331-7-1

of remote identification via Wi-Fi data theft (Fantinato et al., 2020). These privacy challenges intersect with interoperability issues stemming from service-oriented architectures used by smart toys, a common paradigm in distributed computing systems that promote maintenance and reusability (Siqueira and Davis, 2022).

Smart toys typically encompass a spectrum of functionalities, with the most prevalent being story-telling and the two-way conversation via speech recognition. Toy companies often integrate third-party service applications from various providers to furnish these features. The integration between smart toys and the features offered by these third parties typically hinges on web and/or mobile services, designed in accordance with each provider's comprehension of the domain. As such, smart toys are susceptible to a range of interoperability challenges, including deficiencies in communication patterns, conflicts in semantic or contextual information, and a lack of transparency in data flow. These issues can culminate in data transport and encryption breaches, misconduct in smart toy requirement design, and misunderstandings regarding privacy policies, among other complications. A concrete example of this challenge is evident in the current landscape of parental control applications. Each company develops its own tool, tailored to the specific requirements of its smart toy offerings, often failing to adequately address the privacy needs of parents (Albuquerque et al., 2022).

We seek to develop an ontology, employing semantic web language, that encapsulates the domain of privacy in the context of smart toys, focusing on data collection, processing, and disclosure. In doing so, we endeavor to establish a common knowledge base for the domain, fostering a shared understanding of distributed systems and enhancing knowledge sharing and use among stakeholders in the smart toy ecosystem. We have two major goals: (1) furnishing stakeholders and software agents with a standardized knowledge base characterized by a semantic information structure pertinent to smart toys and privacy in this context, thereby simplifying the use and analysis of domain knowledge and definitions, and enhancing control over data exchange to address interoperability challenges across applications; and (2) offering information in the form of privacy principles, including associated risks and proposed solutions, to assist organizations in making informed decisions regarding risk management in the development of smart toys.

This paper is organized as follows. Section 2 brings the background of this work. Section 3 reviews related work. Section 4 shows the ontology design and method. Section 5 refers to the smart toy privacy ontology built. Section 6 discusses ontology implementation. Finally, Section 7 concludes the paper.

2. Background

2.1. Smart toy

Smart toys can be categorized as one of the IoT devices, characterized as a physical toy equipped with electronic components, sensors, and software that facilitate wireless communication with other computational systems via a wireless connection. Notable instances of such smart toys include Mattel's Hello Barbie and CogniToys Dino. In the context of a smart toy environment referred to as *Toy Computing*, this ecosystem encompasses three primary components: (1) A conventional *physical toy*; (2) A *mobile device* responsible for offering smart toys with mobile services to enhance their functionalities, which may potentially include parental control features. and (3) a *mobile application* designed to interact with the physical toy, facilitating data processing and storage. The data collected by smart toys are typically stored and processed in cloud services.

2.2. Privacy in the use of smart toys

Toy computing has evolved into a new paradigm that raises specific privacy concerns regarding children (Rafferty and Hung, 2015). Traditional toys are fully self-contained, devoid of processing or networking capabilities to communicate with other devices. When a child engages with traditional toys, parents need not concern themselves with the child's data privacy. However, the advent of electronic toys with embedded systems introduces sensory capabilities and the ability to collect and store data derived from user interactions. Smart toy architectures not only permit but often need information sharing with services and other users. Nevertheless, in terms of data collection capabilities, smart toys introduce numerous risks due to their wireless connectivity.

The environment poses privacy threats and vulnerabilities not previously experienced in the realm of toys (McReynolds et al., 2017). The potential for data leakage or breaches through these connections constitutes a significant contemporary privacy risk. When personal data is involved, these risks become more acute and can even imperil a child's physical safety (Albuquerque et al., 2019). Collecting, storing and sharing personal data, such as child voice recordings, are common practices for smart toys to offer an experience resembling human communication through artificial intelligence (AI) (Fantinato et al., 2020).

Governments worldwide are taking steps to regulate child data privacy protection rules applicable to companies and organizations. The Canadian Personal Information Protection and Electronic Documents Act (PIPEDA) (CA, 2018) pertains to the collection, use,

or disclosure of PII in commercial activities. PIPEDA mandates that organizations must obtain an individual's consent to process their PII. Individuals possess the right to access their PII held by an organization and contest its accuracy if necessary. Moreover, the American Federal Trade Commission (FTC) has established specific regulations regarding PII concerning children's privacy, as governed by COPPA (USA, 2002). COPPA safeguards the online privacy of children under the age of 13 and defines that a child's PII cannot be collected without parental consent. An amendment to COPPA in 2010 extended the definition of PII to encompass geolocation information, photographs, and videos (Hung, Fantinato, and Rafferty, 2016). In 2017, the FTC issued an updated guidance document for COPPA that explicitly covers smart toys.

These acts collectively aim to establish a common understanding by defining the concepts related to privacy, thus addressing any misconceptions regarding compliance with the requirements of these acts.

3. Related work

Numerous research studies in the literature are devoted to the exploration of privacy and security aspects concerning smart toys. These studies endeavor to identify and delineate the essential requirements inherent to this domain (Albuquerque et al., 2022; Carvalho and Eler, 2018; Hung, Fantinato, and Rafferty, 2016; Rafferty et al., 2017); to uncover potential breaches and vulnerabilities present in smart toys available on the market (Albuquerque et al., 2019; Jones and Meurer, 2016; Shasha et al., 2018); and to elucidate the manner in which children's privacy is addressed and adhered to through public policies and smart toy organizations (Chowdhury, 2018; Hung, Iqbal, et al., 2016; McReynolds et al., 2017). However, only a limited number of these studies have attempted to establish a comprehensive framework for defining the most commonly encountered concepts and their associated meanings in this thematic domain.

Rafferty et al. (2017) developed a conceptual model for privacy rules in the realm of smart toys, raising requirements for a parental control tool in this environment. The paper also offers explicit definitions for various concepts pertinent to smart toy privacy, including smart toys themselves, mobile services, devices, and other relevant terms. Additionally, the paper introduces various types of location data, such as absolute location, relative location, and categorical location, and outlines the privacy-related entities embedded in the conceptual model: purpose, recipient, retention, and obligation.

Albuquerque et al. (2019), through a systematic scoping review, established two classifications for risks and proposed solutions in the context of smart toys. This work offers a taxonomy of privacy principles and information security measures that reflect the primary concerns surrounding children's privacy risks in the context of smart toys. The proposed technical solutions used to mitigate these risks are categorized into five privacy-preserving techniques.

Yankson et al. (2019) introduced a privacy-preserving context ontology (PPCO) tailored to smart connected toys (SCTs). PPCO aims to comprehensively grasp the diverse range of context data collection methods employed in the SCT domain. This study defines high-level information objects, including Location, User, Activity, and Computational Object, offering precise definitions for each and delineating their respective class hierarchies. Furthermore, it presents an overview of the privacy-enhancing controls essential for attaining the privacy objectives associated with SCTs.

Gu et al. (2020) worked on the development of a context model ontology using ontology web language (OWL). Their work includes creating a context model capable of supporting and rationalizing various contextual aspects. This endeavor addresses several issues related to semantic context representation, context reasoning, knowledge sharing, context classification, context dependency, and context quality. Recognizing the advantages of ontology development, several studies have embarked on research endeavors focusing on ontologies in privacy contexts across diverse domains, including e-commerce (Hecker et al., 2008) and smart cities (Gheisari et al., 2021). These studies harness OWL to create a collective comprehension of privacy.

4. Ontology design

In pursuit of advancing in the direction of the semantic web using the Protégé framework, this study developed an ontology pertaining to privacy in the smart toy environment. This ontology encompasses concepts related to the smart toy's context itself, as well as privacy principles, including concepts relevant to data flow and manipulation. This section presents the structure of the ontology components, the selected and defined concepts, the vocabularies employed in both ontology creation and text corpus, and case use scenarios to instantiate our proposal for potential real-world issues.

To build the ontology, we followed the method steps specified by Noy and McGuinness (2001). Additionally, we used the framework recommended by the method authors, Protégé, which is an open-source ontology editor developed by Stanford University. Protégé employs OWL 2 with RDF to create the entities or components

involved in the process, including classes, individuals, object properties, and data properties. The method employed in this project involves the reuse of concepts from other ontologies and the use of concept definitions from common vocabulary sources such as encyclopedias and glossaries. Reuse is considered a best practice in ontology development, involving the incorporation of relevant portions of related ontologies rather than an entire ontology (Fernández et al., 2009).

A *class* is the fundamental component representing, in an abstract manner, entities in the specific domain existing in the real world. Classes are structured in a superclass-subclass hierarchy, where a superclass represents a parent class (the principal class), and a subclass is a child class (belonging to a parent class). In the context of ontology development, each class, whether super or sub, represents a concept that must be defined in the context to create a specific vocabulary. This, in turn, aids distributed computer systems in comprehending the context in which they operate, facilitates the sharing of common knowledge, and enhances interoperability, which are the primary objectives of the semantic web (Taye, 2010). The first class, titled "own:Thing" is the default root class automatically generated by the framework, representing the overarching domain to which all other classes will be subclassed. OWL class naming conventions adhere to CamelCase notation, wherein class names starts with a capital letter and should not contain spaces if the class comprises multiple words.

4.1. Vocabulary and definition

Ontology enables interoperation among web applications from different domains, technologies and perspectives into a single application. Consequently, it becomes imperative to establish mappings between concepts in different ontologies to capture their semantic correspondence (Taye, 2010). A good practice in the semantic web is the reuse of concepts from existing ontologies to build new ones. Nevertheless, other methods were also employed to select terms and their meanings, including text corpus selection and empirical knowledge. Given the novel domain of this ontology, there are no gold standard ontologies to serve as references. Nonetheless, several ontologies exist pertaining to the IoT, with only a few associating IoT with privacy or security. However, these ontologies serve different purposes, do not comprehensively cover the subject matter, and encompass various types of IoT devices. Consequently, our Smart Toy Ontology has selectively adopted relevant portions concerning data collection, processing, and disclosure.

To define a wide range of privacy-related concepts, vocabularies related to smart toy privacy and data flow (including data collection, processing, and disclosure) were meticulously researched and chosen. A variety of ontologies and text corpora from diverse domains were employed, including privacy, information security, data processing, technology, and computing. Among the ontologies used, GDPRtex (Pandit et al., 2018) and GConsent (Pandit et al., 2019), both linked to the General Data Protection Regulation (GDPR), were instrumental in offering a taxonomy of data-related concepts and their meanings, owing to their extensive coverage of data type concepts.

Furthermore, ontologies pertaining to privacy and data processing were incorporated. The data privacy vocabulary (DPV) (Pandit, 2021) encompassed general terms such as *request*, *response*, *user*, among others. The platform for privacy preferences specification (P3P) (W3C, 2002) included general terms but introduced privacy entities that describe data practices in a granular manner. It encompasses four additional privacy entities, namely *purpose*, *recipient*, *retention*, and *obligation* in addition to the commonly used entities *subject*, *object*, and *operation*. The use of P3P is highly recommended in the smart toy environment (Albuquerque et al., 2022; Rafferty et al., 2017).

Other types of vocabularies were also consulted, including comprehensive glossaries covering privacy, security, technology, and computer-related terms, such as the Computer Security Resource Center (CSRC) (NIST, 2021) and the Open Digital Rights Language (ODRL) (Iannella et al., 2018). The WordNet resource (Princeton, 2010) was employed for additional reference. Technical reports related to security, like ISO/IEC 27000:2009 on information security management systems (ISO/IEC, 2009), were considered. Encyclopedias from National Geographic, which offer insights into physical locations and boundaries concepts (NG, 2011a, 2011b), were also referenced. Scientific papers concerning smart toy privacy, primarily used to build concepts related to privacy principles and privacy-preserving techniques (Albuquerque et al., 2019; Rafferty et al., 2017), were incorporated into the vocabulary selection process.

To accurately define domain terms and their meanings, several vocabularies were employed as text corpora to aid concept selection. These auxiliary vocabularies were used to assess the semantic similarity of concepts already included in the ontology by empirically comparing their labels, meanings, and characteristics as synonyms. The support text corpus encompassed various vocabularies aligned with the proposed ontology's scope, such as legislation (EU, 2016), technical reports (Hu et al., 2013), ontologies (Francart et

al., 2019; McRoberts and Doncel, 2017; Pandit, 2019), glossaries (Iannella et al., 2018), and encyclopedias (Jajodia et al., 2011; Liu and Özsu, 2009; Schintler and McNeely, 2021).

In total, our work encompassed twenty-one distinct vocabularies, in both the ontology and text corpus scope. These vocabularies were categorized into six types: papers, legislation, technical reports, ontologies, glossaries, and encyclopedias.

4.2. Use case scenario

Smart toys, driven by technological advancements, are a promising market poised to incorporate more devices and sensors, including AI, leading to increased web service usage. Building a shared knowledge base could improve third-party web service communication, address semantic conflicts, and streamline implementation. This aids in crafting concise privacy policies, lessening disparities in distributed systems and privacy misconceptions. Moreover, it can guide mitigating known risks and suggesting technical solutions.

4.2.1. Hello Barbie's Scenario In 2016, Mattel faced a lawsuit due to the lack of information in Hello Barbie's privacy policy. A section of the ToyTalk privacy policy described how the recorded data could be used: "use, store, process, convert, transcribe, analyze or review recordings in order to offer, maintain, analyze and improve the functioning of the services, to develop, test or improve speech recognition technology and AI algorithms, or for other research and development and data analysis purposes." In another section, it mentioned with whom the data would be shared: "vendors, consultants, and other service providers." All the offered information was too vague, lacking explanations, suggesting that all collected data would be used for all operations and purposes and shared with all third parties. This lack of specificity raises privacy concerns. Furthermore, the information, as presented, is illegal, as data such as PII cannot be shared for advertising purposes according to COPPA.

Using the knowledge base to build the privacy policy documentation can mitigate misunderstandings, as the information will be explicitly detailed, identifying and describing the concepts and instantiating them. For example, using the aforementioned section of the Hello Barbie privacy policy and the concepts related to data flow entities: the **Object** (a passive entity that contains or receives information) such as **voice recording** will perform **Operations** (the set of access modes or action types performed by a subject on objects of the system) like **use, store, process, and sharing** by the **Subject** (a human or non-person entity representing the entity requesting to perform an operation upon the object) **ToyTalk**. ToyTalk will share it with the **Recipient** (the legal entity or domain, beyond the subject and its agents where data may be distributed) like the **speech recognition service provider** for **Purposes** (the purposes for data processing) related to **development, maintenance, and testing of the speech recognition functionality**, and so on. This approach can be applied to all data types collected by the smart toy.

4.2.2. My Friend Cayla's Scenario In 2014, My Friend Cayla's smart doll was released as "the first internet-connected doll," evoking mixed feelings of disruptive innovation for children, but also security and privacy concerns. In early 2015, security researchers revealed vulnerabilities in its software design that allowed the doll to be hacked. Unauthorized devices could connect to Cayla's toy via Bluetooth and gain remote control access to the microphone and speaker. Moreover, Cayla had weaknesses in data encryption (Munro, 2015). A few months later, the toy company released a software update to address reported issues. However, this update proved insufficient, and in the subsequent months, various methods for taking control of the doll were discovered. Also in 2015, Mattel released the Hello Barbie doll, and Cognitoys introduced Dino, two of the smart toy industry's references. These newer toys incorporated improvements related to security and privacy requirements compared to My Friend Cayla, but they still exhibited significant vulnerabilities such as data encryption flaws (Valente and Cardenas, 2017).

Adopting a common knowledge base among toy companies and sharing and reusing information could improve privacy issues in smart toys. This approach could prevent the recurrence of issues identified and reported in older smart toys. Additionally, it could serve as a resource for accessing proposed technical solutions for each issue, thereby improving decision-making regarding smart toy requirement design by toy companies and enhancing public perceptions of their security and privacy measures.

5. Smart toy privacy ontology

In our smart toy ontology, we created eleven superclasses presented in alphabetical order: Concept, Data, Entity, Location, Preserving Technique, Privacy Principle, and Rule. Most superclasses have subclasses defined by the concept's name as a label and its corresponding meaning as the description. This approach ensures the ontology contains meaningful information. To enhance comprehension, we describe the classes in

a different order than presented previously. However, to improve the understanding of the classes present in this section, first, it is necessary to understand the relationships that involve them.

The ontology relationships include the *"is a"* relationship, typically encompassing relationships like *"part of"* or *"subclass of"*. These relationships denote the semantic link between a hypernym (superclasses) and a hyponym (subclasses) (as shown in Table 1 through axioms (1) and (2)). Another relationship defined herein is *"equivalent to"*, which establishes a similarity property between equivalent concepts sharing the same meaning. To define this relationship, a semantic similarity measure was employed, wherein two concepts are considered similar if they share common meanings and characteristics, such as synonymy, hyponymy, and hypernymy concepts (Raad et al., 2015) (as shown in Table 1 through axiom (3)).

To render data machine-readable, one of the foremost goals of semantic web, metadata must be assigned to each concept, as mentioned previously. In our ontology, we used the following metadata fields in the framework: "Label" to define the concept's name, "Comment" to elucidate the concept's meaning, "Is Defined By" to specify the information source, and "See Also" to link the information to a uniform resource locator. All of these metadata elements are defined as RDF's metadata, linked to their respective concepts as defined in the OWL code via a unique uniform resource identifier.

The **Concept** class (Fig.1) encompasses the most relevant concepts in the smart toy environment. These range from generic concepts describing the physical components of smart toys (e.g., **Smart toy**, **Device**, and **Sensors**) to those related to data flow (e.g., **Request**, **Response**, and **Web Service**). This class also includes privacy-related concepts, which can have varying definitions depending on the domain. Examples of these concepts include **Consent**, **User**, **Data Owner**, **Data Controller**, **Data Subject**, **Data Requester**, **Service Provider**, and **Service**.

For instance, **User** and **Data Owner** are equivalent in meaning in most application domains. However, in the context of smart toys, a *user* refers to a child interacting with the smart toy, while a *data owner* is the child's parent who must agree to the end-user license agreement and becomes the owner of the data generated by the toy. Identifying equivalent concepts is crucial for building an efficient ontology and ensuring proper understanding. This equivalence relationship is denoted by double arrows in opposite directions between sibling subclasses. In Figure 1, you can observe this relationship, such as **Data Controller** having the same meaning as **Service Provider**, and likewise for **Data Subject** and **Data Owner**.

The class **Entity** (Fig. 2) defines the entities present in data exchange in smart toys, based on the privacy rule conceptual model for smart toys (Rafferty et al., 2017). Seven subclasses have been defined: **Subject**, represents the permission requester who gains access to data; **Object**, denotes the resource that is desired to be accessed, it encompasses various types of objects, including images, locations, voices, and text; **Operation**, signifies the action to be performed by the Subject on the Object.

The other four entities are privacy-based entities developed by P3P. These entities are designed to offer fine-grained access control for complex environments, such as toy computing involving context-sensitive services: **Purpose**, represents one or more purposes defined by the Subject in the access request; **Recipients**, refers to internal or external entities that may have access to the Objects permitted for access; **Obligations**, signifies the legal commitments that the Subject must adhere to in order to maintain access to the data (e.g., if a subject is required to comply with the regulation "do not share data with third parties," they must adhere to this legal requirement); and **Retention**, indicates the duration for which the Subject will have access to the Object. Mitigating misunderstandings of these terms is of utmost importance because they play a crucial role in governing data exchange in smart toy features.

In Fig. 3, we present the **Data** class, a generic class encompassing various types of data. Offering categories for data is essential to assist smart toy organizations in adhering to regulations and building comprehensive privacy policies. The significance of this class is heightened in the smart toy environment due to the limited awareness of privacy among its users and the extensive use of contextual data.

This class adheres to the taxonomy of data outlined in the GDPR due to its well-defined categorization. The GDPR classifies data into three groups: **Anonymous Data**, data is considered anonymous when it cannot be linked or associated with any individual or individuals who have offered or are associated with it; **Pseudo-anonymous Data**, this type of data can no longer be attributed to a specific data subject without the use of additional information; **Personal Data**, refers to any information pertaining to an identified or identifiable natural person. In this category, there is a subclass called **Sensitive Data**, which necessitates careful protection. Sensitive Data is further classified into subcategories, including **Racial**, **Religious**, **Medical**, **Criminal**, and **Biometric** data. Such data could potentially be used for discriminatory purposes against particular social groups.

The PII category has been included in the Sensitive Data class. PII is a data type demanding meticulous han-

Table 1. Axioms and description logic expression representation

Axiom	Logical expression
(1) Domain of Entity is owl:Thing	Entity.$\top \sqsubseteq$ owl:Thing
(2) Subject, Object, Obligation, Purpose, Recipient, and Retention are subclasses of Entity	Subject $\sqcap$ Object $\sqcap$ Obligation $\sqcap$ Purpose $\sqcap$ Recipient $\sqcap$ Retention $\sqsubseteq$ Entity
(3) Data Subject is equivalent to User	DataSubject $\Leftrightarrow$ User

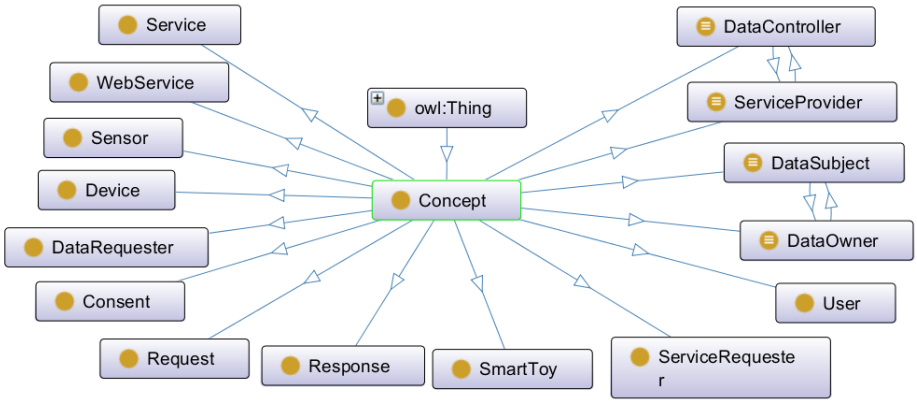

Figure 1. Smart toy concept class

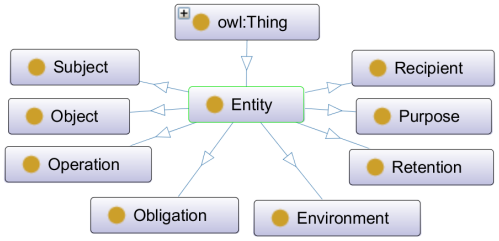

Figure 2. Entity class

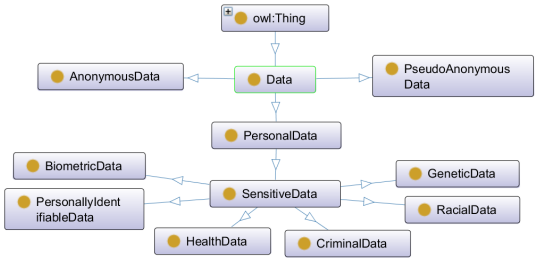

Figure 3. Data class

dling in the realm of smart toys due to its potential risks to privacy and security, especially concerning data such as location. However, this information might be essential for the functionality of smart toys, such as identifying physical characteristics and user preferences to enhance interaction with the user. Hence, it is imperative to identify and clearly define the meanings of these data categories so that all parties involved with the toy, including organizations and the parents of toy users, are informed about the nature of the data being processed. Additionally, each of these data groups can take various formats, including images, voice recordings, and text.

The **Location** class (Fig. 4) offers a comprehensive definition of geographical regions and geopolitical entities, including **County**, **City**, **Province**, **State**, **Country**, **Continent**, and **Trade Bloc**. It draws upon information from the National Geographic encyclopedia concerning physical locations and boundaries (NG, 2011a, 2011b). This class's primary purpose is to facilitate data collection decisions for smart toy organizations, considering specific regional rules and regulations. Additionally, it enhances the reasoning abilities of software agents by establishing connections between location and rule classes. The inclusion of the **Trade Bloc** concept addresses situations where regulations cover territories governed by intergovernmental bodies, potentially involving multiple countries and not necessarily corresponding to an entire continent.

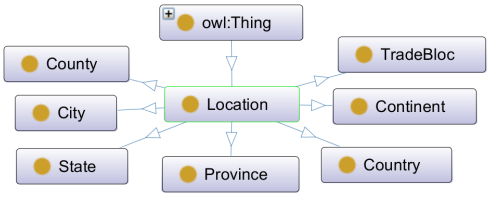

Figure 4. Location class

In Fig. 5, the **Rule** class categorizes various types of regulatory actions related to data manipulation, either entirely or partially concerning children. The concepts in this class, including the overarching **Rule** category, share semantic similarity and essentially function as synonyms. The use of **Rule** as a superclass concept was chosen due to its broad and generic nature. Subclass concepts such as **Regulation**, **Law**, and **Act** were introduced to account for the lack of consensus in terminology used in contemporary laws, including the United States' Act (COPPA), the European Union's Regulation (GDPR), Brazil's Law Data Protection (BR, 2018), or South Africa's Protection of Personal Information Act (POPIA) (SA, 2013).

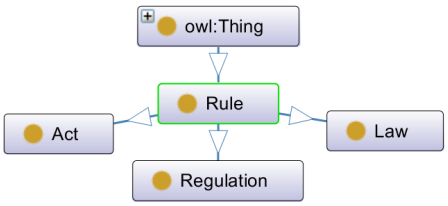

Figure 5. Rule class

The last two classes rely on the privacy principle taxonomies from the study "Privacy in Smart Toys: Risks and Proposed Solutions" (Albuquerque et al., 2019). These taxonomies classify privacy principles relevant to privacy risks and privacy-preserving techniques in the smart toy domain. They were derived from a systematic scoping review that compiled data from 26 primary studies in the literature.

In the **Privacy Principles** class (Fig. 6), there are ten principles serving as subclasses, which address researchers' concerns regarding privacy risks involving children in the smart toy domain. These principles include **Consent and Choice**, **Purpose Legitimacy and Specification**, **Collection Limitation**, **Data Minimization**, **Use Retention and Disclosure Limitation**, **Accuracy and Quality**, **Openness, Transparency and Notice**, **Individual Participation and Access**, **Privacy Compliance**, and **Information Security**. The **Information Security** class further subdivides into six subclasses related to security principles: **Confidentiality**, **Integrity**, **Availability**, **Authentication**, **Accountability**, and **Authorization**.

The **Preserving Technique** class encompasses concepts related to technical solutions aimed at mitigating privacy risks identified in the literature related to smart toys. In this class, there are five crucial proposed solutions often used in the smart toy context: **Noise Addition**, **Anonymization**, **Access Control**, **Cryptography**, and **Blockchain** (Fig. 7). The primary purpose of these two classes and their respective concept definitions is to offer smart toy organizations with a knowledge base pertaining to privacy in the smart toy domain. This knowledge base is intended to serve as a guideline for the development process of smart toys.

6. Discussion

Over time, the perspective is that smart toys will gain more functionality and sensors, increasing their power in terms of connectivity. They will be able to connect with other smart toys for multiplayer functionality and with other devices such as smartphones and smart homes. They will process and use larger amounts of data and integrate more AI and natural language processing (NLP) models into their functionalities. Thinking about it, creating ontologies related to the domain is an important mechanism to harness the benefits that the semantic web offers for the construction of knowledge bases and to aid in interoperability, both semantically and technically, between systems and functionalities. This will improve decision-making at all stages of the smart toy implementation process, from design to enhancing smart toy processing autonomy.

As we do not have an ontology that could serve as a reference for this domain, the concepts were identified and added to the ontology, as the domain coverage needs to be introduced. The first concepts selected for the ontology were concepts already established in the field of smart toys, taken from articles in the related literature, and which offered, in addition to the concepts, their respective meanings, as was the case with the Privacy Principle classes, Preservation techniques of privacy, and Entity, in addition to the most basic concepts of the domain or generics such as smart toy, device, sensor, web services, request, response, user, etc. From this, ontologies were selected, all related to privacy or data processing, in which other terms of the more general class (Concept class) were reused, in addition to establishing the structure of concepts of the Data class. Finally, texts and common types of vocabulary were selected, where other concepts were introduced, enabling them to be machine-readable.

The construction of the ontology considered different approaches to the terms of privacy in smart toys, covering different goals. Concepts related to privacy in data collection processing, and disclosure such as the Concepts, Entities, and Data classes, were defined aiming at making the data flow more transparent and functional, offering a standard semantic dictionary for communication between entities. The concepts of privacy risks and proposed solutions to mitigate such risks were covered as guidelines for decision-making by organizations in

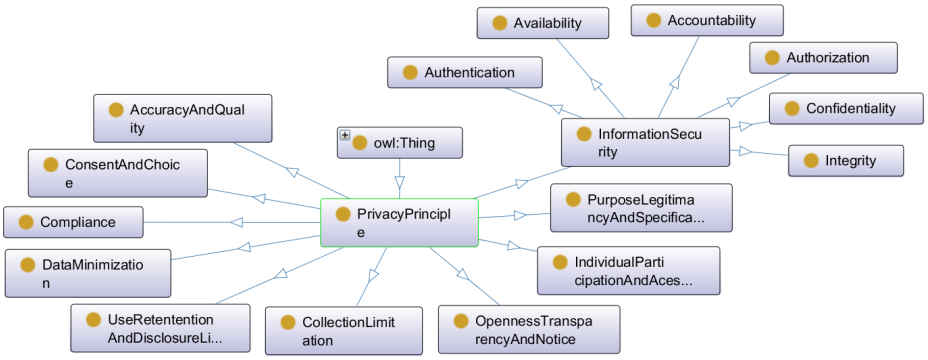

Figure 6. Privacy principles class

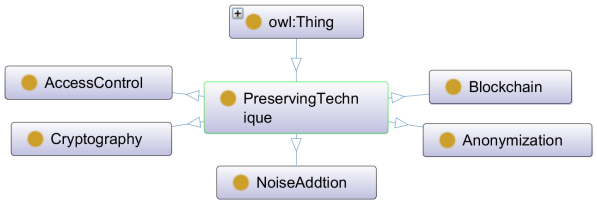

Figure 7. Privacy preserving techniques class

planning requirements and implementing a smart toy. In addition, all concepts are machine-readable, which helps the reasoning of intelligent software agents, improving the accuracy of inferred information, and consequently the automation of decision-making in data processing, including what to collect or not.

7. Conclusion and Future work

There are just a few studies in the literature related to smart toys and the semantic web yet, but the creation of a common knowledge base related to this domain, which organizations and intelligent agents can share and reuse, seems more and more necessary. The regulations related to children's privacy or data protection, such as the COPPA, GDPR, or POPI, present a knowledge base of the general concepts and their meanings, not covering or covering shallow the children's data protection necessities of the smart toys environment.

This work aimed to develop a high-level concept ontology related to privacy in smart toys, with the main goal of sharing a knowledge base. This offers a standardized machine-readable semantic dictionary for organizations and software agents to use. Moreover, it offers guidelines on privacy principles tied to the risks in this domain and potential technical solutions, thereby simplifying risk management in toy implementations.

As a limitation, the ontology built in a high-level concept structure used just two types of relationships, the "is-a" for superclass-subclass relation and "Equivalent to" to a few siblings classes. The concepts defined here were evaluated in an empirical way, comparing the concepts selected for the ontology, to the text corpus selected as vocabulary support.

In future works, we will conduct an ontology evaluation, using the Corpus-based evaluation method to compare the Ontology's concepts with a large text corpus related to the IoT and privacy, using algorithms of NLP to measure the similarity between concepts, measuring, mainly, criteria as Accuracy to see if the concepts comply with the domain knowledge; Completeness to measure the appropriately covered of the domain; Conciseness to identify possible irrelevant or redundant concepts in the ontology presented here. The relationship between concepts will be implemented aim to create autonomous software agents for smart toys, trying to improve their reasoning over data flow.

8. Aknowledgments

This study is partially supported through Research Incentive Fund (RIF) (activity code: R20090) research office, Zayed University, Dubai. Also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

Albuquerque, O. d. P., Fantinato, M., Hung, P. C., Peres, S. M., Iqbal, F., Rehman, U., & Shah, M. U. (2022). Recommendations for a smart toy parental control tool. *The J. of Supercomputing*, 1–39.

Albuquerque, O. d. P., Fantinato, M., Kelner, J., & Albuquerque, A. P. (2019). Privacy in smart toys: Risks and proposed solutions. *Electronic Commerce Research and Applications*, *39*, 1–15.

BR. (2018). Lei geral de proteção de dados (LGPD) [Brazilian National Congress]. https://www.serpro.gov.br/lgpd

CA. (2018). Personal information protection and electronic documents act (PIPEDA) [Min. of Justice, CA]. https://laws-lois.justice.gc.ca/PDF/P-8.6.pdf

Carvalho, L. G., & Eler, M. M. (2018). Security requirements and tests for smart toys. *19th Int'l Conf. on Ent. Inf. Sys.*, 291–312.

Chowdhury, W. (2018). Toys that talk to strangers: A look at the privacy policies of connected toys. *Conf. of Fut. Tech. Conf.*, 152–158.

EU. (2016). General data protection regulation [European Union]. https://gdpr.eu/tag/gdpr

Fantinato, M., Albuquerque, O. D. P., De Albuquerque, A. P., Kelner, J., & Yankson, B. (2020). A literature survey on smart toy-related children's privacy risks. *53rd Hawaii Int'l Conf. on Sys. Sci.*

Fernández, M., Overbeeke, C., Sabou, M., & Motta, E. (2009). What makes a good ontology? A case-study in fine-grained knowledge reuse. *Asian Semantic Web Conf.*, 61–75.

Francart, T., Dann, J., Pappalardo, R., Malagon, C., & Pellegrino, M. (2019). The european legislation identifier. *Knowledge of the Law in the Big Data Age*, *317*, 137–148.

Gheisari, M., Najafabadi, H. E., Alzubi, J. A., Gao, J., Wang, G., Abbasi, A. A., & Castiglione, A. (2021). OBPP: An ontology-based framework for privacy-preserving in IoT-based smart city. *Future Generation Computer Systems*, *123*, 1–13.

Gu, T., Wang, X. H., Pung, H. K., & Zhang, D. Q. (2020). An ontology-based context model in intelligent environments. *arXiv preprint arXiv:2003.05055*.

Hecker, M., Dillon, T. S., & Chang, E. (2008). Privacy ontology support for e-commerce. *IEEE Internet Computing*, *12*(2), 54–61.

Hu, V. C., Ferraiolo, D., Kuhn, R., Friedman, A. R., Lang, A. J., Cogdell, M. M., Schnitzer, A., Sandlin, K., Miller, R., Scarfone, K., et al. (2013). Guide to attribute based access control (ABAC) definition and considerations (draft). *NIST Special Publication*, *800*(162), 1–54.

Hung, P. C. K., Fantinato, M., & Rafferty, L. (2016). A study of privacy requirements for smart toys. *20th Pacific Asia Conf. on Information Systems*, 1–7.

Hung, P. C. K., Iqbal, F., Huang, S.-C., Melaisi, M., & Pang, K. (2016). A glance of child's play privacy in smart toys. *Int'l Conf. on Cloud Comp. and Sec.*, 217–231.

Iannella, R., Steidl, M., Myles, S., & Rodríguez-Doncel, V. (2018). Open digital rights language vocabulary. https://www.w3.org/TR/odrl-vocab

ISO/IEC. (2009). ISO/IEC 27000:2009 - Information technology - Security techniques - Information security management systems - Overview and vocabulary. https://standards.iteh.ai/catalog/standards/iso/f09ef98e-d070-4c69-942a-e2fe1b1eb0ce/iso-iec-27000-2009

Jajodia, S., Samarati, P., & Yung, M. (2011). *Encyclopedia of cryptography, security and privacy*. Springer.

Jones, M. L., & Meurer, K. (2016). Can (and should) Hello Barbie keep a secret? *IEEE Int'l Symp. on Ethics in Engin., Sci. and Tech.*, 1–7.

JR. (2018). Smart toy revenues to grow by almost 200% from 2018 to $18 billion by 2023 [Juniper Research]. https://www.juniperresearch.com/press/smart-toy-revenues-grow-almost-200pc-by-2023

Liu, L., & Özsu, M. T. (2009). *Encyclopedia of database systems* (Vol. 6). Springer.

McReynolds, E., Hubbard, S., Lau, T., Saraf, A., Cakmak, M., & Roesner, F. (2017). Toys that listen: A study of parents, children, and internet-connected toys. *CHI Conf. on Human Factors in Comp. Sys.*, 5197–5207.

McRoberts, M., & Doncel, V. (2017). Open digital rights language 2.2 ontology. https://www.w3.org/ns/odrl/2

Munro, K. (2015). Making children's toys swear. https://www.pentestpartners.com/security-blog/making-childrens-toys-swear

NG. (2011a). Encyclopedia Entry: Boundary [National Geographic]. https://www.nationalgeographic.org/encyclopedia/boundary/12th-grade

NG. (2011b). Encyclopedia Entry: Province [National Geographic]. https://www.nationalgeographic.org/encyclopedia/province

NIST. (2021). Computer security resource center glossary [National Institute of Standards and Technology]. https://csrc.nist.gov/glossary

Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology. http://protege.stanford.edu/publications

Pandit, H. J. (2019). GDPRov – the GDPR provenance ontology. https://openscience.adaptcentre.ie/ontologies/GDPRov/docs/ontology

Pandit, H. J. (2021). Data privacy vocabulary (DPV). https://dpvcg.github.io/dpv

Pandit, H. J., Debruyne, C., O'Sullivan, D., & Lewis, D. (2019). Gconsent-a consent ontology based on the GDPR. *European Semantic Web Conf.*, 270–282.

Pandit, H. J., Fatema, K., O'Sullivan, D., & Lewis, D. (2018). GDPRtEXT-GDPR as a linked data resource. *European Semantic Web Conf.*, 481–495.

Princeton. (2010). Wordnet [Princeton University]. http://wordnetweb.princeton.edu/perl/webwn

Raad, J., Bertaux, A., & Cruz, C. (2015). A survey on how to cross-reference web information sources. *Science and Inform. Conf.*, 609–618.

Radack, S. (2010). Guide to protecting personally identifiable information [ITL Bulletin, National Institute of Standards and Technology, Gaithersburg, MD]. https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=905656

Rafferty, L., Hung, P., Fantinato, M., Peres, S. M., Iqbal, F., Kuo, S., & Huang, S. (2017). Towards a privacy rule conceptual model for smart toys. *50th Hawaii Int'l Conf. on Sys. Sci.*, 1–10.

Rafferty, L., & Hung, P. C. K. (2015). Introduction to toy computing. In P. C. K. Hung (Ed.), *Mobile services for toy computing* (pp. 1–7).

SA. (2013). Protection of personal information act (POPIA) [Information Regulator, South Africa]. https://popia.co.za

Schintler, L. A., & McNeely, C. L. (2021). *Encyclopedia of big data*. Springer.

Shasha, S., Mahmoud, M., Mannan, M., & Youssef, A. (2018). Playing with danger: A taxonomy and evaluation of threats to smart toys. *IEEE IoT J.*, *6*(2), 2986–3002.

Siqueira, F., & Davis, J. G. (2022). Service computing for industry 4.0: State of the art, challenges, and research opportunities. *ACM Comp. Surveys*, *54*(9).

Taye, M. M. (2010). Understanding semantic web and ontologies: Theory and applications. *J. of Comp.*

USA. (2002). Children's online privacy protection act (COPPA) [Federal Trade Commission, US Congress, USA]. https://www.ftc.gov/sites/default/files/documents/rules/children%5C%E2%5C%80%5C%99s-online-privacy-protection-rule-coppa/coppasurvey.pdf

Valente, J., & Cardenas, A. A. (2017). Security & privacy in smart toys. *Wks. on IoT Sec. and Priv.*, 19–24.

W3C. (2002). The platform for privacy preferences 1.0 specification [W3C Consortium]. https://www.w3.org/TR/P3P

Yankson, B., Iqbal, F., Aleem, S., Shah, B., Hung, P. C., & De Albuquerque, A. P. (2019). A privacy-preserving context ontology (ppco) for smart connected toys. *12th CMI Conf. on Cybersecurity and Privacy*, 1–6.