The Privacy Policy Permission Model: A Unified View of Privacy Policies

Maryam Majedi[*], **Ken Barker**[**]

[*]Department of Computer Science, University of Toronto, Toronto, ON, M5S 2E4, Canada.

[**]Department of Computer Science, University of Calgary, Calgary, AB, T2N 1N4, Canada.

E-mail: majedi@cs.toronto.edu, kbarker@ucalgary.ca

Received 5 January 2021; received in revised form 22 April 2021; accepted 27 April 2021

Abstract. Organizations use privacy policies to communicate their data collection practices to their clients. A privacy policy is a set of statements that specifies how an organization gathers, uses, discloses, and maintains a client's data. However, most privacy policies lack a clear, complete explanation of how data providers' information is used. We propose a modeling methodology, called the *Privacy Policy Permission Model* (PPPM), that provides a uniform, easy-to-understand representation of privacy policies, which can accurately and clearly show how data is used within an organization's practice. Using this methodology, a privacy policy is captured as a diagram. The diagram is capable of highlighting inconsistencies and inaccuracies in the privacy policy. The methodology supports privacy officers in properly and clearly articulating an organization's privacy policy.

Keywords. Privacy, Privacy policy modeling, Privacy components, Privacy permission, Privacy policy diagram

1 Introduction

We live in the age of data monitoring where our cell phones are personal tracking devices that record every movement, activity, and conversation. Social networks can capture a holistic representation of our lives, even if we are not direct participants. Babies may even have profiles on social networks before they are born. Corporations collect our activity information and permanently store it in their repositories. The colossal amount of gathered information is analyzed for predicting and often influencing our decisions. Modern technology, while promising a richer and easier way of life, provides a means of control of our information by large corporations such as Facebook™, Amazon™, Google™, and Apple™. The reality is that we are being tracked; but this does not always seem to unsettle us until the consequences are revealed through various forms of data breech or misuse. Most of us like our devices tailor-made to provide us with immediate, relevant information. We use wearable gadgets that measure our fitness, sleep, even happiness; or anything else that businesses envision as a value proposition for us. We often willingly trade our information for these services, not realizing how our data is used and analyzed, or considering the consequences. It is time to assess how our privacy has been compromised and what we can do to mitigate the negative consequences.

2 Problem Definition

Privacy policies are difficult to understand because they are typically long, vague, incorporate jargon, and do not clearly define how the data provider's information is used [1],[2], [3]. In addition, organizations might aggregate data to produce new (correct or incorrect) knowledge, about the data providers [4]. Unfortunately, privacy policies are often ignored by users. The process of deciphering privacy policies is difficult, partly because they are written in natural language with all its inherent connotations, so the risk of misunderstanding is high. They may imply access permissions that are illegitimate, which could result in further privacy violations.

We introduce a modeling methodology that provides a standard representation of privacy policies that can accurately and clearly describe how data is used within an organization. Modeling privacy policies, and capturing all the syntactic and semantic aspects within their context, is a critical step to enforcing them.

Organizations use databases to store and manage data for their businesses. Traditional database design begins by developing a conceptual schema using a tool such as Entity Relationship Diagrams (ERD) [5]. Unfortunately, there are no tools that explicitly capture privacy, so a designer wishing to develop a privacy-preserving database must incorporate these in an *ad hoc* way. Our approach is to develop a modeling methodology undertaken as a separate privacy-aware design step. This will explicitly incorporate privacy, independent from the database design, and facilitate the inclusion of privacy policies for all elements in the database.

3 Contributions

Motivated by traditional database design, (i.e. ERD), we develop and design a new privacy modeling methodology. Our approach is independent of the data domain so it can be used within any organization regardless of its activities. It produces a diagram that allows users to understand an organization's privacy policy. We demonstrate how a privacy policy is deconstructed into components and then implemented in a privacy diagram. Since this methodology produces diagrams that directly reflect an organization's privacy policies, privacy officers and administrators can use it to identify and resolve potential violations and contradictions. This modeling methodology is adaptable to policy changes by making the implications of alterations explicit in the resulting diagrams. Our contributions are as follows:

1. We develop a methodology called Privacy Policy Permission Model (PPPM) to depict privacy policies in a structured way including capturing:
   (a) Privacy components:
      - Roles: the categories within and beyond the organization that are granted access to private information.
      - Purposes: the intentions behind the permitted accesses.
      - Data attributes: entities that are collected, accessed, and used.

   (b) Homogeneous connections:
      - Role structure: indicates any hierarchy among roles.
      - Purpose structure: shows how purposes are constructed from ordered sets of tasks.

  - Attribute structure: indicates when attributes are aggregated and generate new information.

  (c) Heterogeneous connections (permissions)

  - Role-purpose connections: indicate permissions for the roles to use purposes.
  - Purpose-attribute connections: capture connections of all purposes' tasks to their required data attributes.

2. We describe how to model each of these components and how they are extracted and presented in a diagram using both a canonical example and a real-world example currently in use.

3. We highlight the method's ability to identify gaps and contradictions by virtue of undertaking a gap analysis of the diagram produced.

We begin by providing a brief review of closely related research.

4 Background

Agrawal *et al.* [6] identify ten principles of privacy when describing their proposal for Hippocratic Databases based on the US Privacy Act of 1974 [7]. These principles are purpose specification, consent, limited collection, limited use, limited disclosure, limited retention, accuracy, safety, openness, and compliance. Similar values are articulated in the European Union Data Protection approach [8] which is followed by Canada [9]. Privacy legislation often requires that data collectors inform data providers about the privacy policies they practice. In some sectors explicit legislation/regulation may be in place such as: the US Health Insurance Portability and Accountability Act (HIPAA) of 1996 [10], Canada's Health Information Act (HIA) [11], Canada's Personal Information Protection Act (PIPA) [12], or Canada's Personal Information Protection and Electronic Documents Act (PIPEDA) [13]. While these acts may be encoded in an organization's privacy policy, which are developed, at least in part, to communicate their practices to their users [14], the policies often do not clearly describe individuals' data usage, but rather focus on protecting the organization from legal consequences, rather than protecting the users themselves [15].

To avoid the risks associated with the misuse of personal data, some individuals provide false information or create several email accounts to protect their privacy [16]. Culnan and Armstrong suggest that organizations must address their clients' privacy concerns to earn their trust [17]. In fact, if clients trust a website, they are more likely to prefer it to purchase merchandise [18]. Therefore, it is important that organizations demonstrate their commitment to protecting their client's privacy [19].

Earp *et al.* [20] compare different methods for presenting privacy policies to online health care customers. In their study, the users' perceptions of privacy policies are examined, and their understanding of categorized policies is measured. Finally, the results are compared to determine which representation is easier to understand. Their study shows that presenting the privacy policies in natural language is the most difficult to understand and is insufficient for conveying information to users. By categorizing the privacy statements, they increase the users' comprehension. Earp *et al.* [20] suggest that additional effort must be made to improve the way privacy policies are presented.

Often people find privacy policies too legalistic [21]. Fabian *et al.* [22] analyze the privacy policies of 50,000 popular websites and determine that the privacy policies are difficult to decipher. The notice-and-consent approach is widely used in the United States, but it is inadequate because it assumes that individuals read the privacy policies and understand the implications of providing data to the data collectors [23].

VenkataSwamy *et al.* [24] define data sets that are subject to the same policies, and maintain permissions using a matrix. Silva *et al.* [25] introduce a multilanguage approach called RSLingo4Privacy to improve privacy policies. In their work, statements are classified to create a logically consistent equivalent, which facilitates a visual representation. The classification is then extracted based on the terms used in the privacy policy. Our methodology models privacy policies independent of the data domain, according to the principle of data independence.

Modeling is an essential step for developing systems. In addition to providing visual presentations, models can be used to define how a system would behave in various situations. For this reason, they can also be used for predicting and understanding potential system gaps. Mai *et al.* [26] propose a modeling method to structure and analyze privacy and security specification requirements in the health care domain, which is useful when developing the software. Context is essential for understanding the privacy-protection being afforded by these models [27]. Our methodology incorporates context to constrain the way that data collectors use and transfer the information collected based on explicit statements in the privacy policy. Chen [5], in his seminal contribution, proposes a diagrammatic technique to model entities and their relationships. This technique is independent of the entities' domains. Our approach is fundamentally inspired by this abstract, generic modeling tool.

5 PPPM: An ERD for Privacy Policies

This section introduces our methodology, Privacy Policy Permission Model, and describes how it can be used to create a unified and natural view of privacy policies. The model achieves a high degree of entity and domain independence. By providing a visual demonstration of privacy policies, our methodology highlights shortfalls to help organizations clearly explain their privacy policy statements. The resulting Privacy Policy Permission Diagram (PPPD) can then be implemented in a privacy catalog (see Figure 1). The privacy catalog, our current project, is a set of database tables that capture a modeled privacy policy. Once the diagram is generated, it is used to populate the privacy catalog in the database. This catalog can then be used for granting or revoking accesses to the data layer, thereby enforcing privacy policies. We will introduce it in future work.

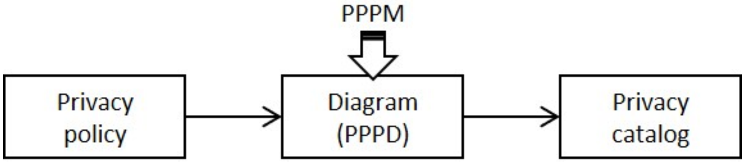


Figure 1: PPPM as a modeling methodology

5.1 Privacy Policy Components and Connections

A privacy policy consists of policy statements. These statements are natural language sentences that define how the data in an organization's custody must be maintained and used. Each privacy statement consists of privacy components. Privacy components are categorized differently by different researchers. Barker *et al.* [28] introduce a data privacy taxonomy to describe key privacy components. According to their taxonomy, the privacy components are purpose, visibility, granularity, and retention. Ni *et al.* [29], [30] consider obligation and condition as privacy policy components as well. We further consider a data attribute a privacy component. In this work, we refer to visibilities as roles. We also develop granularity as functions that apply to data items to prepare them for purposes' usage, and we develop retention as conditions for accessing data. Thus, we do not consider granularity or retention as separate privacy components. Moreover, this methodology does not support obligations at this time. Therefore, the privacy components in this work are identified as roles, purposes, and data attributes.

We now formally describe each component (Section 5.1.1) and explain how connections between homogeneous components are mapped in a *Privacy Policy Permission Diagram* (PPPD) (Section 5.1.2). Finally, we show how the heterogeneous connections, and any conditions associated with them, are added to the diagram (Section 5.1.3).

5.1.1 Components

Roles

Roles define the category of subjects (i.e., individuals, organizations, or agents) who are accessing data. Roles are identified as visibility in the taxonomy developed by Barker *et al.* [28]. An organization's functionality defines the appropriate visibility categories. In this work, a role is an application-specific actor that is identified in the written policy statements, and is a subject accessing data. We denote the set of roles: $R = \{r_1, r_2, \ldots, r_n\}$, where r_i denotes a specific role identified in the policy statement.

Purposes

Purposes are reasons to access data. For example, a bank clerk's purpose for accessing a customer's account information might be 'Issuing a statement'. In other words, purposes are reasons for data collectors to gather and use data from providers. We define the set of all purposes as: $P = \{p_1, p_2, \ldots, p_n\}$, where p_i denotes a specific purpose identified in the policy statement.

Data Attributes

Data attributes represent specific pieces of sensitive information. A *data item* is an instance of a data attribute. For example, 'Age' is a data attribute, while '26' may be the value of a data item. A data attribute's sensitivity level is relative depending on the data provider's concern/opinion about that data item's value. A data attribute is drawn from the written privacy policies and is an application-specific object used in an environment. For the sake of simplicity, we will use the term 'attribute' rather than the more precise term 'data attribute' throughout this paper.

In an ERD, attributes are an entity's properties, and their values are set when they are instantiated. In a PPPD, there can be an arbitrary number of application-specific attributes (d_i), identified and collected from a policy statement into a set as $D = \{d_1, d_2, \ldots, d_n\}$.

A privacy policy might describe different categories for attributes which we call *attribute groups*. These groups are used to refer to a set of attributes when specifying access permissions in a privacy policy. We denote attribute groups as: $G = \{g_1, g_2, \ldots, g_n\}$, where each g_i denotes an attribute group defined in the privacy policy. Note that each attribute could belong to zero or more groups.

A data item's specificity can be modified by its granularity level. *Granularity* specifies the precision used/needed when data is accessed for a purpose. For example, Barker *et al.* [28] categorize granularity as 'None', 'Existential', 'Partial', and 'Specific'. Granularity can be used to provide enhanced privacy by generalizing or making the data value more abstract. For example, granularity specifies whether 'age' should be provided as the exact age, or as an age range, such as 'Child', 'Teenager', or 'Adult'. Granularity has been addressed in both a deterministic and analytical way in the literature [31] and our contributions will work in either methodology because it is undertaken at the policy level.In our model, granularity is supported by defining conversion functions that alter data precision based on the purpose for which it is accessed (see Purpose structure in the next Section).

5.1.2 Homogeneous Connections

Roles, purposes, and attributes are first modeled independent of their interrelationships so we defined the connections within instances of the same type of components as homogeneous connections. The resulting structures help us clarify how instances of the same component type relate to one another. Each is described next.

Role Structure

The role structure (RS) is based on the connections between roles. RS is a set of ordered pairs, r_i,$r_j \in R$, when r_i is superior[1] to r_j and is depicted as: $r_i \rightarrow r_j$. Thus: $RS = \{(r_i, r_j) \mid r_i \rightarrow r_j\}$ for all r_i and r_j identified in the privacy policy that are connected. This relationship implies that r_i holds at least all the access permitted to r_j. When $RS = \emptyset$ the roles are mutually exclusive with no explicit hierarchical structure.

Purpose Structure

Privacy concerns primarily arise due to access to data. Every such access should be based on its purpose. Purposes can access data in PPPM through *tasks*. Thus, a task describes how a specific attribute value is used for a purpose. Therefore, we designate a task for every data usage, and then compose tasks into sequences that reflect the purpose's data accesses. Tasks can be composed in different ways for different purposes as illustrated in the following.

Thus, a purpose p_i is an ordered set of tasks denoted as $p_i = (t_1, t_2, \ldots, t_n)$ where for each task $t_i \in T$ (where T is the set of all tasks) exactly one attribute d_j is accessed. Note however that a d_j could be accessed by multiple tasks. Furthermore, each task is associated with a granularity function that modifies the attribute's value to match the required granularity level for the purpose's use. This allows us to capture purposes within context, and ultimately clarifies how attributes are used by each purpose.

[1] Superior in this context describes a reports-to relationship such as a manager-employee or a teacher-student model.

Attribute Structure

The attribute instances' structure represents how attributes are combined in an environment. If a privacy policy indicates that specific attributes are combined to generate new information, we establish a connection between them, and add the new information as an attribute. The attributes and their connections form a structure. The attribute structure (DS) is created based on the attributes that are aggregated to generate new information. DS is a set of ordered triples where d_i, d_j, $d_k \in D$ where d_i and d_j are aggregated to generate a new data d_k. This aggregation is depicted as: $(d_i, d_j) \rightarrow d_k$. Thus: $DS = \{(d_i, d_j, d_k) \mid (d_i, d_j) \rightarrow d_k\}$ for all d_i and d_j identified in the privacy policy that aggregate to derive d_k . When $DS = \emptyset$, no data attributes are aggregated in the policy.

5.1.3 Heterogeneous Connections (Permissions)

Heterogeneous connections occur between components of different types. They either provide permissions when roles use purposes, or when purposes access data. Thus, we have: *Role-purpose permissions* and *Purpose-attribute permissions*. Permissions may have conditions, described in the following.

Role-Purpose Permissions

Role-purpose permissions provide permissions for roles to use purposes. If a role's access is conditional, its condition is added to the connecting role-purpose instance permission. Since this permission is effectively applied only to the connection between the role and purpose layers, the condition is independent of any attribute accessed.

We denote role-purpose permission (RP) as a set of triples corresponding to a role $(r_i \in R)$, purpose $(p_j \in P)$, and condition $(c_k \in C)$ where c_k may be null. Thus:

$RP = \{(r_i, p_j, c_k) \mid r_i \textit{ and } p_j \textit{ is permitted under condition } c_k\}$ as stated in the privacy policy.

Purpose-Attribute Permissions

As described in Section 5.1.2, homogeneous connections in the purpose layer create a structure based on purposes' ordered sets of tasks, such that each task requires exactly one attribute. In that stage, tasks were not connected to their attributes because tasks and attributes belong to different, heterogeneous layers. The purpose-attribute permissions in this stage connect the tasks to their required attributes, which capture the heterogeneous connections between the purpose and attribute layers.

Purpose-attribute permissions (PD) can include a condition, which is placed on the corresponding task and attribute connection for that purpose. These conditions are independent of roles. For example, a purpose-attribute condition defining *retention*, could be specified to expire data independent of a role access permission. Conditions must be satisfied in advance of access. This choice effectively eliminates using post-access obligations [30] that are, in general, unenforceable, so this decision provides a stronger privacy guarantee. We leave for future research the incorporation of obligations once techniques are developed to enforce them.

Purpose-attribute permission (PD) is a set of triples corresponding to a purpose $(p_i \in P)$, task $(t_j \in T)$, and condition $(c_k \in C)$ where c_k may be *null*. Thus:

$PD = \{(p_i, t_j, c_k) \mid p_i \textit{ and } t_j \textit{ is permitted under condition } c_k\}$ as stated in the privacy policy.

Conditions

A *condition* is a logical statement that must be satisfied to enact the purpose of a privacy policy. For example, Facebook's™ privacy policy requires that users be 13 years or older for registration. Thus, $'Age \geq 13'$ is a conditional statement for 'Registration'. We denote the set of conditions as: $C = \{c_1, c_2, ..., c_n\}$.

5.2 Extracting Information from the Privacy Policy

We illustrate the process of extracting privacy policy information using an imaginary organization that sells products online. The organization collects client information and ships orders. It analyzes client order lists and uses their age information to send birthday gifts. The complete privacy policy is provided in Appendix A. For the purpose of illustration, we assume this privacy policy is complete, precise, and unambiguous. Some privacy component instances and their connections are directly extracted from the statements. However, one must often infer them from the context. We describe three stages for capturing information from the written privacy policies. In the first stage, we capture components from the statements. The second stage captures homogeneous connections. The final stage defines heterogeneous connections or permissions.

In the following section, we describe each stage using our canonical privacy policy.

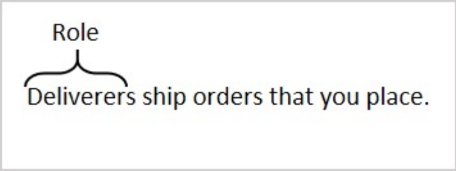
5.2.1 First Stage: Capturing Component Instances

This stage identifies all privacy component instances, and places them into their corresponding layer. In this section, we provide an example instance of each component type, and then identify all other instances occurring in the canonical policy.

Capturing Role Instances

All role instances are captured and placed in the role layer. Example 1 illustrates a statement from Part III of our online shopping privacy policy containing a role. 'Deliverer', in this statement, is the subject that accesses data to complete an action, so it is a role.

Example 1. Role as a subject.



Deliverers ship orders that you place.

All the roles identified in the privacy policy are recorded in Table 1.

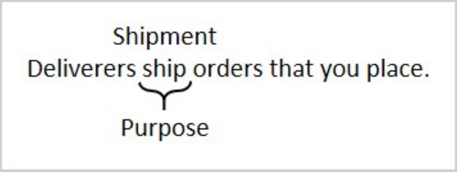
Label	Role
r_1	Manager
r_2	Deliverer
r_3	Analyzer
r_4	Marketer

Table 1: Role instances

Capturing Purpose Instances

Example 2 illustrates the same policy statement in the articulation of the *Shipment* purpose in our sample privacy policy. In this example, 'Shipment' is specified as the reason for accessing customer information by the deliverer.

Example 2. Purpose as an action.



Deliverers ship orders that you place.

Other purposes identified in this privacy policy (found in Part II, Part III, and Part IV) are shown in Table 2.

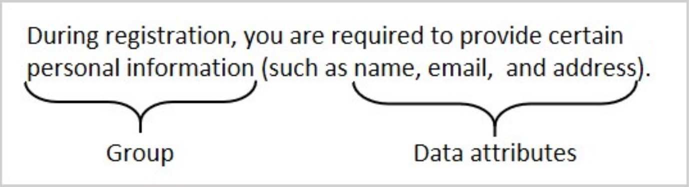
Label	Purpose
p1	Shipment
p2	Marketing
p3	Sending gift
p4	Analyzing

Table 2: Purpose instances

Capturing Attribute Instances

Example 3 shows a privacy policy statement from Part I of the privacy policy that contains three attributes, 'Name', 'Email', and 'Address'. From this statement, we model an attribute group, 'Personal information', which includes all these attributes.

Example 3. Inferring attributes from a policy statement.



During registration, you are required to provide certain personal information (such as name, email, and address).

'DOB' (Date of birth), 'Order list', 'Credit card information', and 'Interest' are other attributes mentioned in Part III and Part IV of the privacy policy. 'Credit card information' also belongs to the 'Personal information' group but other attributes do not belong to any group. This information is recorded in Table 3.

Label	Attribute	Group
d_1	Name	Personal information
d_2	Order list	None
d_3	Credit Card information	Personal information
d_4	Address	Personal information
d_5	Email	Personal information
d_6	DOB	Personal information
d_7	Interest	None

Table 3: Attribute instances

5.2.2 Second stage: Capturing Homogeneous Connections

In this stage the component's layer structure is defined by capturing the connections between homogeneous component instances.

Capturing Role Layer Connections

Part IV in the example privacy policy states that ″The manager supervises analyzers and deliverers″. This statement indicates that the ′Manager′ role is superior to the ′Deliverer′ and ′Analyzer′ roles. Table 4 shows all the relationships/connections between these roles in our sample privacy policy.

Superior	Inferior
Manager	Deliverer
Manager	Analyzer
Analyzer	Marketer

Table 4: Roles connections

Capturing Purpose Layer Connections

The policy statement in Part III ″To ship your orders, deliverers access your name, order list, credit card information, address, and email address to respectively identify you, process your order, charge fees, ship the parcel, and finally, inform you about the shipment″, fully defines the tasks associated with the ′Shipment′ purpose. Table 5 contains all the purpose layer connections to their tasks for the example privacy policy.

Purpose	Tasks
Shipment	Identify client, Process order list, Charge fees, Ship parcel, Inform client
Analyzing	Analyze based on age, Analyze shopping habit, Determine interest
Marketing	Identify client (Age >18), Send advertisement
Sending gift	Check DOB, Identify client, Ship parcel

Table 5: Purpose connections

Task ordering is vital in this process because it describes purposes′ proper data access order during execution.

Capturing Attribute Layer Connections

Attribute connections are established if the policy specifies that they are combined. The resulting connected attributes create a structure. Attribute combinations often generate new information that an organization might use for its own purposes. For example, in Part IV, the policy states: ″Our analyzers combine your date of birth, and shopping history to better understand your shopping habits, and predict your interests″[2], so ′DOB′ and ′Order list′ are combined and analyzed to predict customer ′Interest′. Table 6 illustrates combined attributes from the example policy.

Attribute 1	Attribute 2	New Attribute
DOB	Order list	Interest

Table 6: Attribute connections

5.2.3 Third stage: Capturing Heterogeneous Connections (Permissions)

Connections between roles and purposes, and the permissions required for purposes to access attributes are called heterogeneous connections. We now describe how to capture these permissions and their conditions from the sample privacy policy.

Capturing Role-purpose Permissions

Role-purpose permissions must use statements indicating the purpose that a role has for access. These statements define permission connections between roles and purposes. In our example privacy policy, the statement ″Deliverers ship orders that you place″, specifies that there is a connection between the role ′Deliverer′ and the purpose ′Shipment′. These statements may also include conditions. For example, the statement ″Marketing staff members will send you advertisements within business hours,″ places a condition on the permission for the ′Marketer′ role when using the purpose ′Marketing′. The condition specifies that the role ′Marketer′ can perform ′Marketing′ only between 8 am and 5 pm. This condition is independent of the attributes utilized.

Table 7 illustrates the completed role-purpose permissions list from our example. The privacy policy also indicates that an ′Analyzer′ is allowed to use the purpose ′Analyzing′, and that the ′Marketer′ can perform the ′Sending gift′ purpose.

Role	Purpose	Condition
Deliverer	Shipment	
Analyzer	Analyzing	
Marketer	Marketing	8 am <now() <5pm
Marketer	Sending gift	

Table 7: Role-purpose permissions

[2] A customer′s ′order history′ is defined as the combination of previous ′order lists′.

Capturing Purpose-attribute Permissions

Statements that specify a purpose using an attribute define purpose-attribute permissions. These permissions are captured by adding connections between the purpose's tasks and corresponding attributes, which may include conditions. Consider the statement in Part IV that "[a] customer's name can be used for marketing if the customer is over 18 years old." This statement indicates permission for the purpose 'Marketing' to allow access to the 'Name' attribute. The 'Marketing' purpose includes tasks 'Identify client' and 'Send advertisement' according to the policy. This infers that the tasks 'Identify client' must access the 'Name' attribute. In the next step, we place the 'Age >18' condition on the purpose-attribute permission where the task 'Identify client' accesses 'Name'.

Other permissions are listed in Table 8, where the first column contains the tasks, the second contains the data attributes that the tasks require, the third contains any conditions, and the last column contains the granularity function used to modify the data attribute for the tasks' usage. For example, the task 'Analyze based on Age' requires the age calculated from the attribute 'DOB'.

Label	**Task**	**Attribute**	**Condition**	**Granularity**
t_1	Identify client	Name	Age >18	
t_2	Process order list	Order list		
t_3	Charge fees	CC Info		
t_4	Ship parcel	Address		
t_5	Inform client	Email		
t_6	Send advertisements	Email		
t_7	Check DOB	DOB		
t_8	Analyze based on Age	DOB		Date2Age
t_9	Analyze shopping habit	Order list		
t_{10}	Determine interest	Interest		

Table 8: Purpose-attributes permissions

5.3 Developing the Diagram

We now develop the diagram for the example privacy policy using the extracted information from Section 5.2. We first create the three component layers and their homogeneous connections, and then complete the diagram by adding the permissions across heterogeneous component layers.

5.3.1 Role Layer Diagram

Recall that Table 1 contains the four roles required to create the role layer diagram, whose connections, defined in Table 4 , are captured in Figure 2. Directed edges from superior roles to inferior ones must be defined to complete the role layer diagram. To simplify this diagram, a node label is defined for each role and a legend is provided, listing the corresponding roles' instances. For example, in Figure 2 label r_1 represents the role of 'Manager'. This practice is used in all the figures presented in the balance of the paper.

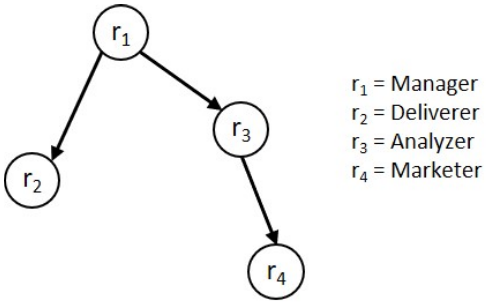


Figure 2: Role layer diagram

5.3.2 Purpose Layer Diagram

The purpose layer contains purposes, represented as nodes (see Figure 3), which are connected to a sequence of tasks defined by the attributes they access. When a purpose's tasks (represented by solid dots) are specified in the privacy policy, they connect to their purpose through directed edges. If no tasks are specified for a particular purpose, only the purpose itself is included. The directed edges between task nodes capture the task order for a particular purpose. Recall that Table 5 identifies the purposes, along with their tasks and their orders. Figure 3 uses 'p's to represent purposes, which are composed of 't's, which represent tasks. For clarity, we use colour to illustrate different task sequences.

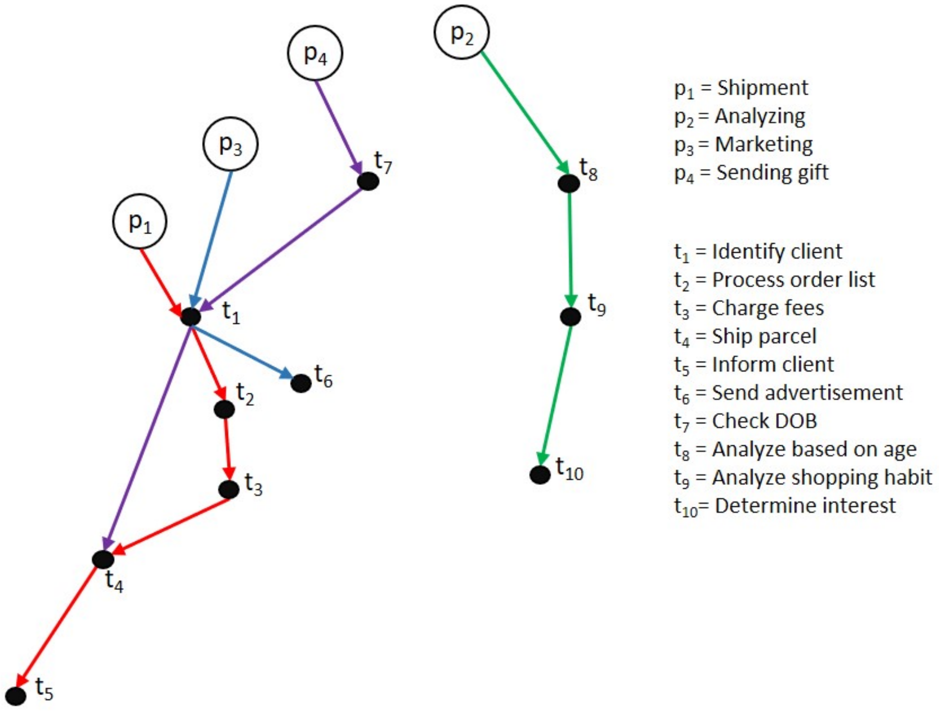


Figure 3: Purpose layer diagram

5.3.3 Attribute Layer Diagram

In the attribute layer, attribute instances are nodes and their connections are edges. When attribute groups are specified, for the purpose of illustration, they are represented by surrounded areas containing their members. Recall that Table 3 identifies the attributes in the policy. The 'Personal information' group, containing its attributes, is illustrated as a surrounded area in Figure 4.

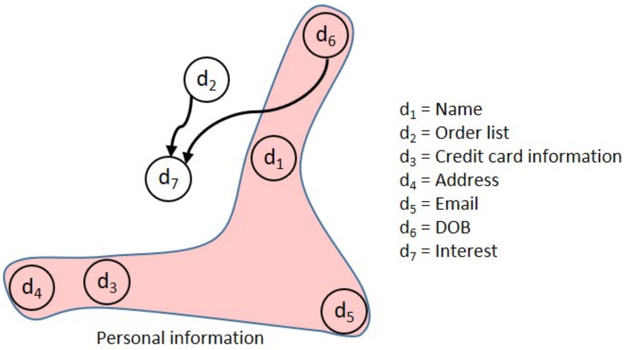


Figure 4: Attribute layer diagram

Attribute layer connections capture statements in the privacy policy that indicate that the organization combines attributes in some way, which may give rise to implied attributes. From Table 6, the 'Order list' and 'DOB' are combined, which gives rise to another attribute 'Interest'. Thus, the 'Interest' attribute is added to the diagram as a node and the arrows depict the two attributes that their combination produced. Figure 4 illustrates the attribute layer which contains this combination.

5.3.4 Role-purpose Permissions Diagram

Role-purpose heterogeneous connections and any corresponding conditions must be added to the diagram. Table 7 identifies these connections, which are illustrated in Figure 5 with dashed-lines. For example, the 'Deliverer' role and the 'Shipment' purpose are connected, thereby illustrating that a deliver has legitimate permission to undertake shipment. Similarly, the 'Analyzer' role has permission to undertake the 'Analyzing' purpose; and the 'Marketer' role is connected to both the 'Marketing' and the 'Sending gift' purposes. To illustrate the use of conditions, the permission for the 'Marketer' role is conditional in that it must occur during the workday. These conditions are attached to the edges between roles and purposes as illustrated between r_4 and p_3 in Figure 5.

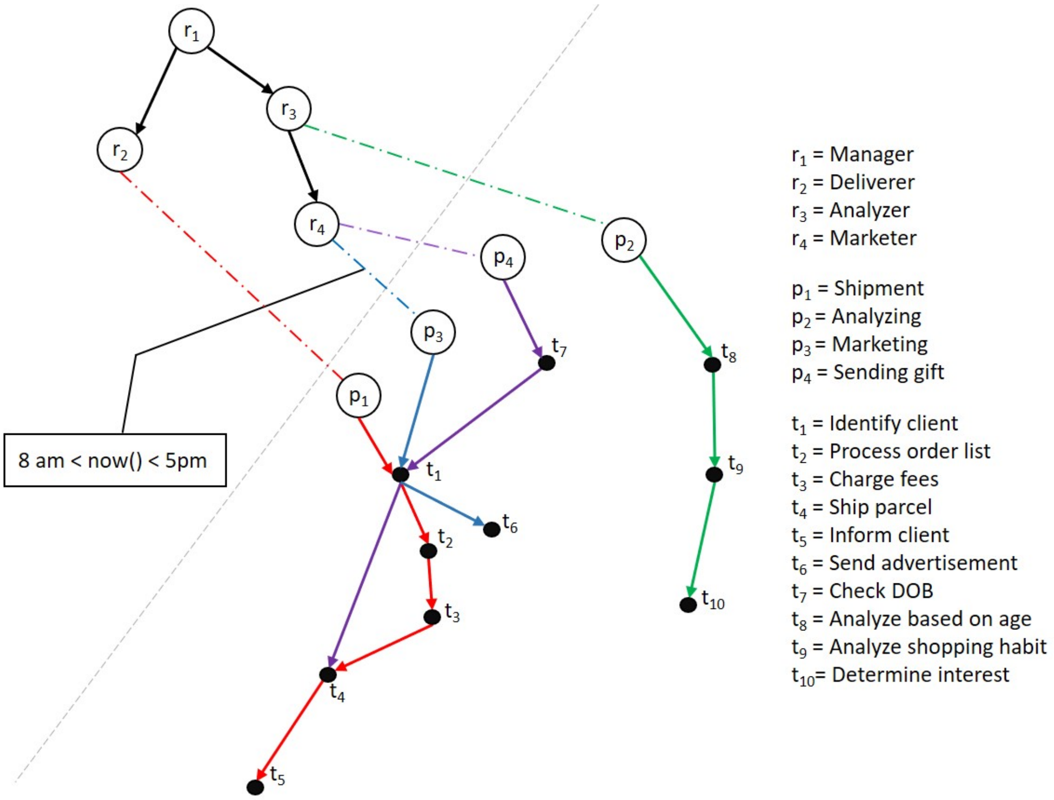


Figure 5: Role-purpose diagram

5.3.5 Purpose-attribute Permissions Diagram

Purpose-attribute heterogeneous permissions provide access to attributes for identified purposes. Since attributes are accessed by explicitly identified tasks, which are only accessed for defined purposes, the connection of purposes to attributes is via purposes' corresponding tasks. Figure 6 illustrates these connections, listed in Table 8, for our example privacy policy. For example, the task 'Identify client' (t_1) is connected to the 'Name' attribute (d_1), with the condition that, within the 'Marketing' purpose (p_3), the corresponding age must be over 18. An example of how an attribute can be connected to multiple purposes is illustrated with the 'Order list' attribute (d_2). The tasks 'Process order list' (t_2) and 'Analyze shopping habit' (t_9) connect to the attribute 'Order list' (d_2), so connections to this attribute exist for purposes 'Shipment' (p_1) and 'Analyzing' (p_2). For the 'Analyze based on Age', the 'DOB' attribute's value must be converted to age which is added to the corresponding connection in the diagram.

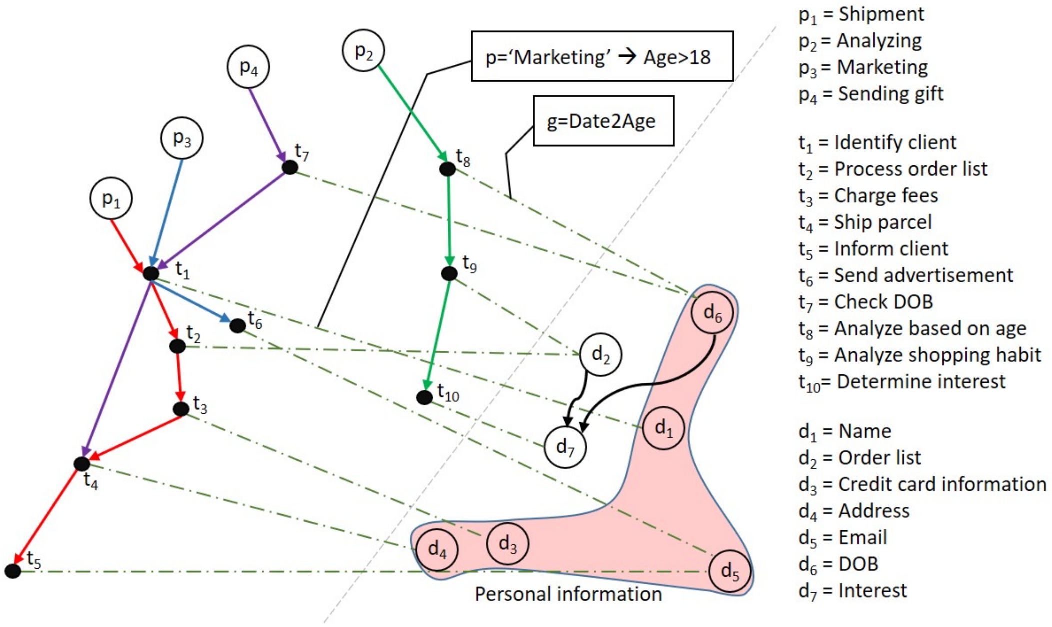


Figure 6: Purpose-attribute diagram

5.3.6 Final Diagram for the Example Privacy Policy

Figure 7 shows the complete PPPD, including components, and both homogeneous and heterogeneous connections identified in the example privacy policy in Appendix A.

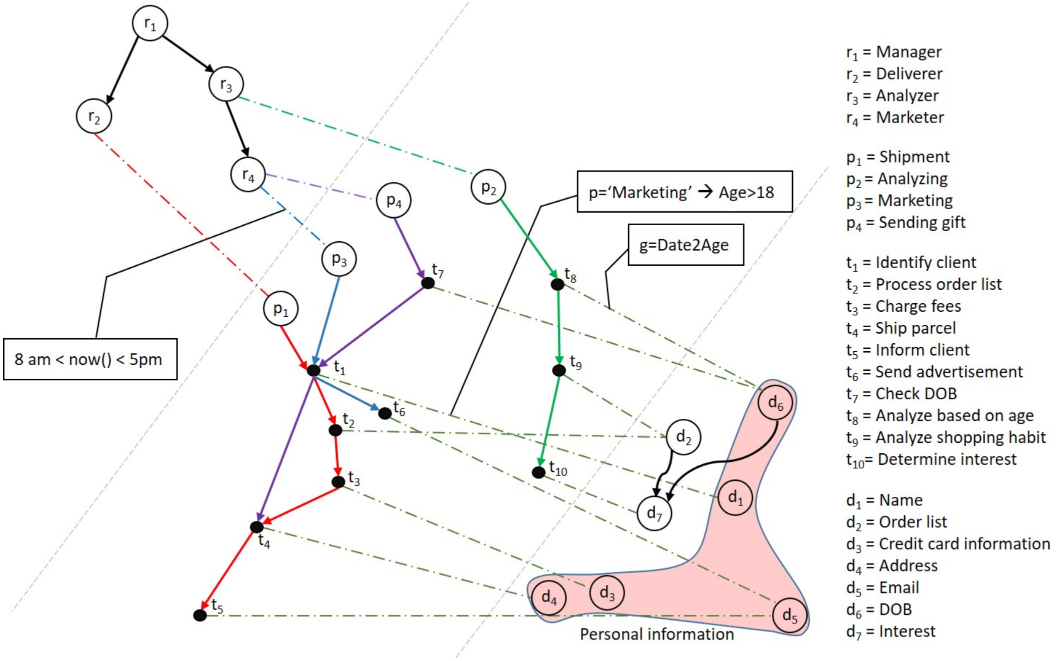


Figure 7: Final diagram for the example privacy policy

Thus, to develop a PPPD we must first identify all of the components described in the privacy policy including roles, purposes, and data attributes (and attribute groups when applicable), and each is placed in their corresponding layer.

Connections between the homogeneous component instances are identified to capture the structure in each layer. Operational features are defined by tasks that structure how purposes access data in the purpose layer. Finally, attribute aggregation and implicit attributes are captured in the attribute layer. These fully define all homogeneous connections for a particular policy.

The final step is to capture the heterogeneous connections, including those between roles and purposes and between purposes and attributes. Ultimately, these connections represent permissions for roles to access purposes, and permissions for purposes to use attributes.

6 Applying the PPPM

By applying the PPPM, we develop a PPPD for the ChatterBaby™ application's privacy policy [32]. This application, developed at Cambridge University [3], collects recordings of infants crying to interpret their needs. The collected audio files are also used to assess autism risk factors. We are not concerned about the utility, ethics, or capabilities of their application, but rather evaluate their privacy policy using our methodology. We generate the diagram and identify the privacy policy shortfalls by evaluating the result.

Figure 8 illustrates the role-purpose layer of the privacy policy. (All component instances and their connections identified from the privacy policy are listed in Appendix B).

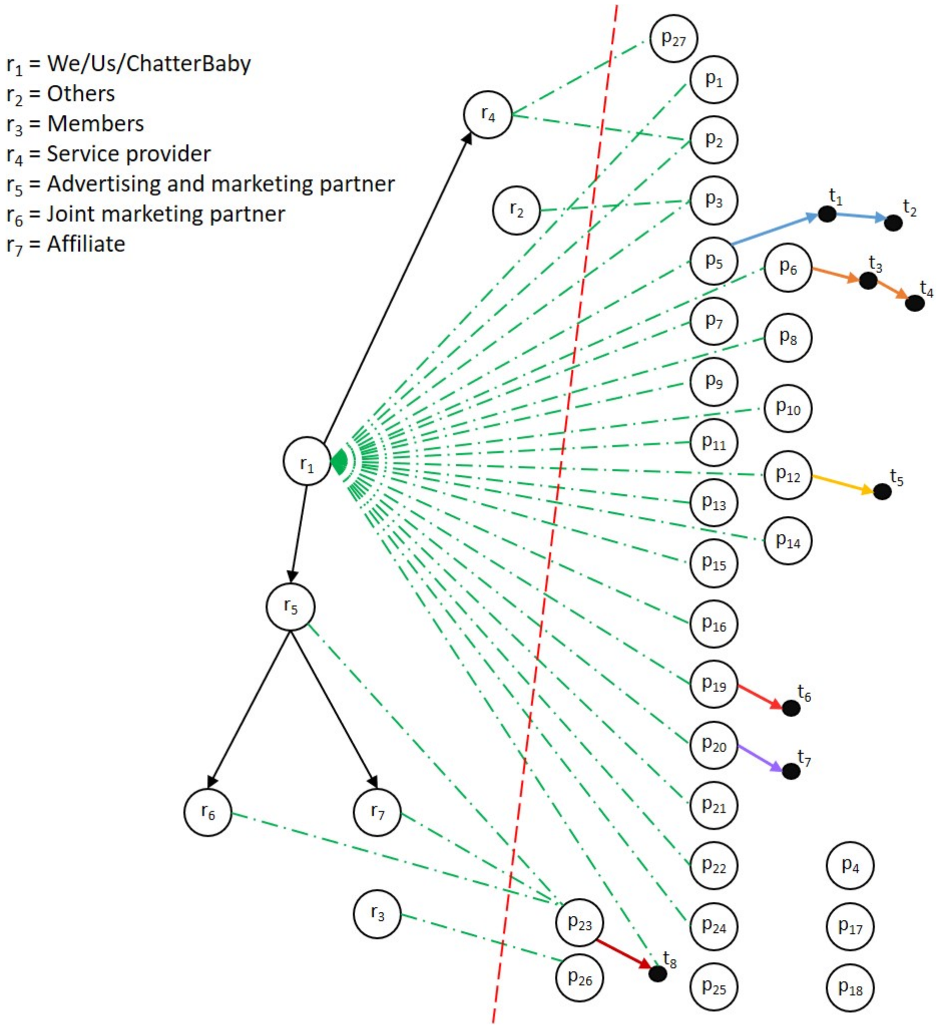


Figure 8: ChatterBaby™ role-purpose permission diagram

[3] ChatterBaby™'s privacy policy exists as a pdf document so it is clearly intended to be a living document, which is completely appropriate for a such an application. Our assessment was undertaken based on its December 26^{th}, 2020 version and it is available at the following URL: `https://chatterbaby.org/files/view/download_files/Privacy_Policy_IRB.pdf`

It is noteworthy that the role 'We/Us/ChatterBaby' has permission for many purposes, including purpose 'Any' (p_{24}). Although this approach of identifying the organization as a whole is often found in corporate privacy policies, it is problematic because no distinction can be made about which roles can actually access sensitive data because the policy explicitly allows anyone access within the organization. This single all-encompassing role, with universal permission, is an ideal illustration of how the PPPM approach highlights risky privacy policy features. As a counterpoint to this challenge, ChatterBaby™ also defines a role called 'Others' (r_2), which has permission to access the 'Identify' purpose (p_3). Unfortunately, this vague purpose is not defined, so the statement and its implications should be reviewed.

Figure 8 highlights another key value of the PPPM. There is no role connected to 'Collect, measure, process autism risk' (p_4), 'Fighting spam/malware' (p_{17}), or 'Facilitate data collection' (p_{18}). This implies that any data attributes accessed for these purposes are not explicitly connected to a responsible role. This arises in the ChatterBaby™ policy because the relevant statements are written using a passive voice, so it is unclear who is responsible. While passive statements may not pose legal risks, and may even have advantages if challenged legally, they lack clarity about who accesses the data and for what purpose.

The complete list of all data attributes identified in the ChatterBaby™ privacy policy are found in Appendix B, and are organized into the corresponding attribute groups, as provided in Table B3. Note that the 'Individual' attribute group, which contains explicit personally identifying information, can be accessed for several purposes. In fact, the entire group of 'Personal' attributes contains what most would consider sensitive or personal data.

Figure 9 depicts another part of ChatterBaby™'s PPPD, which highlights additional concerns. The policy statement: "From time to time, we may use your Personal Information to send important notices, such as communications about purchases and changes to our terms, conditions, and policies" provides purpose 'Send notice' (p_{12}) with access to all attributes in the 'Personal' group. On the other hand, the statement "if we believe that the changes are material, we'll let you know by [...] sending you an email or message about the changes" indicates that p_{12} only has task 'Send email' (t_5), which accesses the 'Email' (d_5) attribute. Therefore, the permission of p_{12} to access the personal group has no justification. In general, if a purpose is not connected to the attributes through tasks, then it is unclear how the data is used. This contradiction could be corrected by limiting the access available to purpose p_{12}.

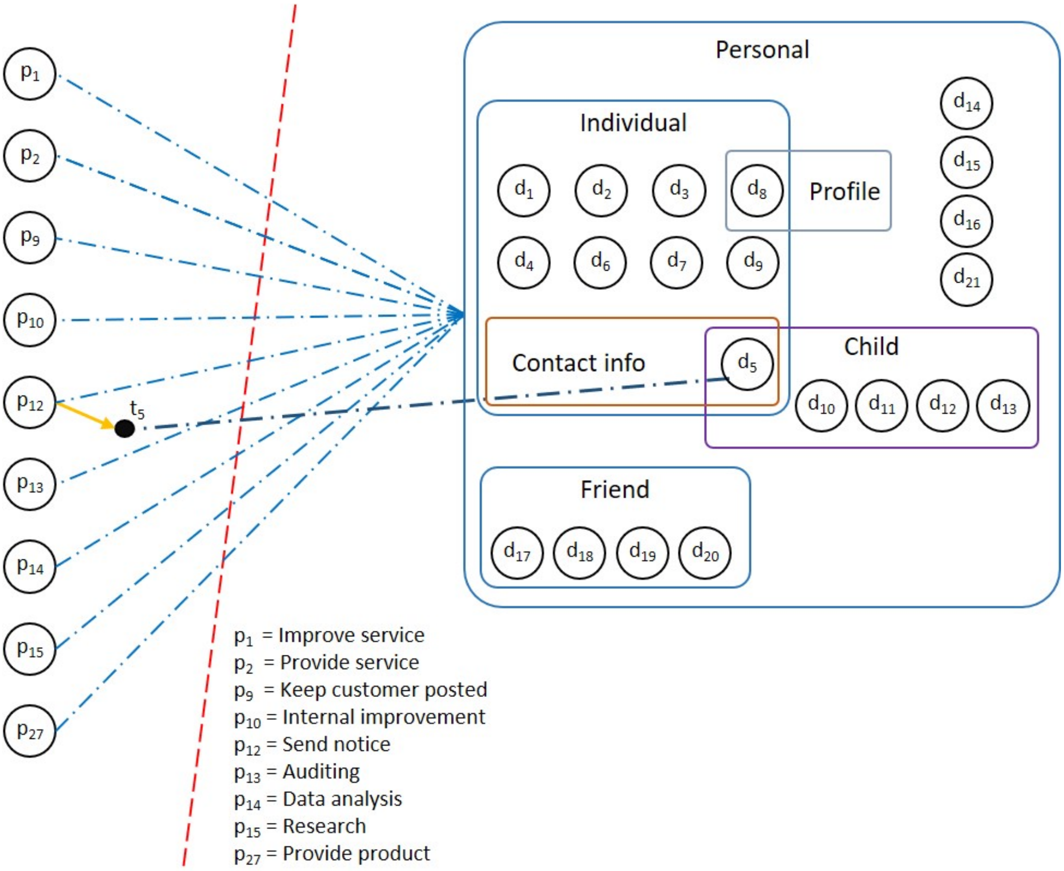


Figure 9: ChatterBaby™ heterogeneous permissions diagram to access the 'Personal' group

Figure 10 illustrates the data included in the ’non-personal’ group. Although it is unclear how medical or location information can be considered ’non-personal’ information, we will use this label, but note that the policy is likely implying that this data is anonymized sufficiently to be considered non-identifying. We wish to provide a PPPD reflective of the policy, but this terminology is misleading, and should be flagged through the PPPD definition process to highlight such narrative inconsistencies.

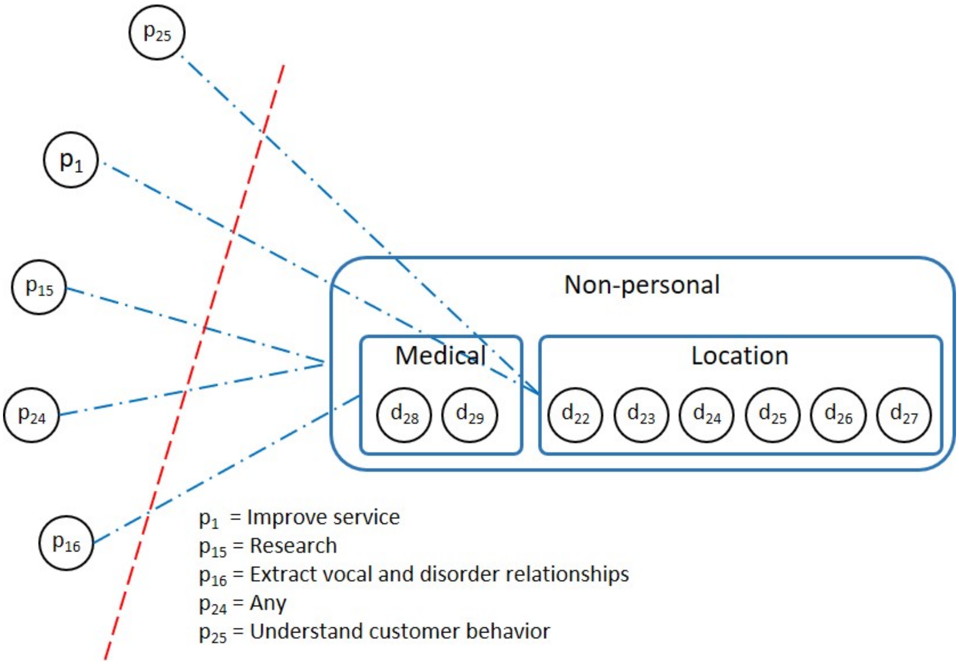


Figure 10: ChatterBaby™ heterogeneous permissions diagram to access the ’Non-personal’ group

Setting the nomenclature aside for the moment, Figure 10 highlights another concern in the policy. Purpose p_{24}, which represents ’Any’ purpose has permission to use all the attributes in the ’Non-personal’ group. This permission is explicit in the policy statement: ”We may collect, use, transfer, and disclose non-personal information for any purpose.” Given that medical and location information are subgroups of the ’Non-personal’ attribute group, this would effectively allow for unrestrained access to this sensitive information. Our assumption that this data is anonymized is critical in such a situation, but it is based on a more tenuous assumption that the anonymization process is fully protective. It is also noteworthy that the use of universal access for ’Any’ purpose is a clear violation of best practices in privacy protection and such a purpose should be flagged as a severe privacy risk.

7 Discussion and Directions

7.1 PPPM Advantages

Most of privacy policies today, do not specifically explain how the data is used or combined to generate additional information, and that is why they are vague. In our sample privacy policy, we clearly describe how data is combined to create new knowledge. For example, in the statement, "Our analyzers combine your date of birth, and shopping history to better understand your shopping habits, and predict your interests", we specifically specified that DOB and shopping history are combined to find customer's interests. This type of statement is not common in privacy policies, and we argue that this is a problem; therefore, we use PPPM to show this gap throughout the diagram. The power of the PPPM is its ability to highlight ambiguities and shortfalls in privacy policies, which can then be actioned by privacy officers, policy developers, and lawful organizations to correct policy shortfalls efficiently. We have provided a few examples of the shortfalls in Section 6. The PPPD development process is structured to systematically identify and define privacy components. The component connections identify only explicit permissions associated with those components. This allows an organization to specify the components and their connections clearly and flag any inconsistent accesses once the system becomes operational.

PPPM also provides an easy way to demonstrate an organization's practices and capture how client data is used. The resulting diagram could also help clients to understand the organization's privacy policy in a visual way. PPPM is an abstract tool and does not require conforming to a rigorous definition of privacy or an abstraction of access. For example, earlier work required that the privacy context be defined, as illustrated by privacy conformance efforts in social networks, which first had to define classes of visibility such as 'Data provider', 'Friends', 'Friends of friends', 'Third parties', and 'All/world' [28]. Unfortunately, the work on social networks using this categorization could not be ported to a different environment such as a hospital. The PPPM approach does not require the predefinition of roles, purposes, or data attributes and types.

PPPM also specifies the connections between different components. Using some instances of components together in a privacy policy may result in impracticable policies or privacy violations. For example, any use of a universal access such as 'Any' purpose or 'All' data attributes is highlighted in the formation of the PPPD and can be carefully reviewed for potential (possibly unintentional) privacy risks.

We believe that this modeling methodology would also allow for changes to privacy policies. As new statements are added or existing statements are adjusted, they could be evaluated using the existing PPPD, and more quickly assessed for nascent privacy risks. Developing a tool to allow for this evaluation and assessment to occur automatically is left as a future direction.

7.2 Limitations

A clear limitation of the PPPM approach is that it only reviews how data is handled with respect to the privacy policy. If the policy does not disclose how data might be transferred to third-parties or used outside of the scope of the policy, there is no way to model potential privacy issues. This is not a fault of PPPM itself, but rather a pragmatic limitation of the policy. Understanding how to model the external flow of data beyond an organization's boundaries would be an interesting extension to the PPPM, so it is left as an opportunity for future research.

PPPM itself is highly valuable but it relies on the organization to enforce the principles involved in the privacy policy. Another valuable direction would be to develop mechanisms that allow the PPPD to be used as input for an enforcement system to ensure that operational data only flows in accordance with the model.

Using PPPM, the resulted diagram becomes more complex if the privacy policy is long. To address this issue, we are working on a user-friendly tool to give the user the ability to define "Zones" for different parts of the privacy policies. Nevertheless, the underneath concept of the tool will stay based on PPPM.

7.3 Summary

Often privacy policies lack a complete and clear explanation of how data is used. In many privacy policies, an organization's internal processes are explained in vague or ambiguous language and contain gaps and/or contradictions. PPPM can model privacy policies that allow for the identification of these shortfalls. A formally modeled privacy policy also allows for managing access permissions and information usage in an organization. This methodology is not domain-dependent, so when using PPPM, an organization's database design does not need to be modified to accommodate the capturing of privacy policies. The database design maintains and expands on an ERD in that the PPPM creates a separate diagram representing the privacy policy. PPPM provides the privacy officers and policy designers with a diagram of their privacy policy. This diagram could enable the organization's clients to better understand the privacy policy that they may sign.

References

[1] L. Wu, M. Majedi, K. Ghazinour, and K. Barker, "Analysis of Social Networking Privacy Policies," in *Proceedings of the 2010 EDBT/ICDT Workshops*, 2010, pp. 1–5.

[2] S. Wilson, F. Schaub, A. A. Dara, F. Liu, S. Cherivirala, P. G. Leon, M. S. Andersen, S. Zimmeck, K. M. Sathyendra, N. C. Russell *et al.*, "The Creation and Analysis of a Website Privacy Policy Corpus," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016, pp. 1330–1340.

[3] A. Oltramari, D. Piraviperumal, F. Schaub, S. Wilson, S. Cherivirala, T. B. Norton, N. C. Russell, P. Story, J. Reidenberg, and N. Sadeh, "PrivOnto: A Semantic Framework for the Analysis of Privacy Policies," *Semantic Web*, vol. 9, no. 2, pp. 185–203, 2018.

[4] A. Kim, L. J. Hoffman, and C. D. Martin, "Building Privacy into the Semantic Web: An Ontology Needed Now," in *Proc. of semantic web workshop, hawaii, usa*, 2002.

[5] P. P. Chen, "The Entity-Relationship Model-Toward a Unified View of Data," *ACM transactions on database systems (TODS)*, vol. 1, no. 1, pp. 9–36, 1976.

[6] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Hippocratic Databases," in *VLDB'02: Proceedings of the 28th International Conference on Very Large Databases*. Elsevier, 2002, pp. 143–154.

[7] A. A. Bushkin and S. I. Schaen, *The Privacy Act of 1974: A Reference Manual for Compliance*. System Development Corporation McLean, Va., 1976.

[8] "General Data Protection Regulation (GDPR)," European Commission, 2018. [Online]. Available: https://gdpr-info.eu/

[9] J. DeCew, "Privacy," 2018. [Online]. Available: https://plato.stanford.edu/archives/spr2018/entries/privacy/

[10] "Health Insurance Portability and Accountability Act of 1996 (HIPAA)," p. 191, 1996. [Online]. Available: https://ca.practicallaw.thomsonreuters.com/1-501-6222?transitionType=Default&contextData=(sc.Default)&firstPage=true&bhcp=1

[11] "Health Information Act (HIA)," 2018. [Online]. Available: http://www.qp.alberta.ca/documents/Acts/H05.pdf

[12] "Personal Information Protection Act," 2019. [Online]. Available: https://www.qp.alberta.ca/documents/Acts/P06P5.pdf

[13] "The Personal Information Protection and Electronic Documents Act (PIPEDA)," 2019. [Online]. Available: https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/

[14] C. Jensen, C. Potts, and C. Jensen, "Privacy practices of Internet users: Self-reports versus observed behavior," *International Journal of Human-Computer Studies*, vol. 63, no. 1-2, pp. 203–227, 2005.

[15] A. Anton, J. B. Earp, D. Bolchini, Q. He, C. Jensen, and W. Stufflebeam, "The Lack of Clarity in Financial Privacy Policies and the Need for Standardization," *North Carolina State University Technical Report# TR-2.*, 2003.

[16] I. Pollach, "What's wrong with online privacy policies?" *Communications of the ACM*, vol. 50, no. 9, pp. 103–108, 2007.

[17] M. J. Culnan and P. K. Armstrong, "Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation," *Organization science*, vol. 10, no. 1, pp. 104–115, 1999.

[18] S. Y. Hong and H. Rim, "The influence of customer use of corporate websites: Corporate social responsibility, trust, and word-of-mouth communication," *Public Relations Review*, vol. 36, no. 4, pp. 389–391, 2010.

[19] N. Olivero and P. Lunt, "Privacy versus willingness to disclose in e-commerce exchanges: The

effect of risk awareness on the relative role of trust and control," *Journal of economic psychology*, vol. 25, no. 2, pp. 243–262, 2004.

[20] J. B. Earp, M. Vail, and A. I. Anton, "Privacy Policy Representation in Web-based Healthcare," in *2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07)*. IEEE, 2007, pp. 138–138.

[21] G. R. Milne and M. J. Culnan, "Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices," *Journal of interactive marketing*, vol. 18, no. 3, pp. 15–29, 2004.

[22] B. Fabian, T. Ermakova, and T. Lentz, "Large-Scale Readability Analysis of Privacy Policies," in *Proceedings of the International Conference on Web Intelligence*, 2017, pp. 18–25.

[23] H. Nissenbaum, "A Contextual Approach to Privacy Online," *Daedalus*, vol. 140, no. 4, pp. 32–48, 2011.

[24] M. VenkataSwamy, S. Ramaswamy, and N. Agarwal, "CBPM: Context Based Privacy Model," in *2010 IEEE Second International Conference on Social Computing*. IEEE, 2010, pp. 1050–1055.

[25] A. R. da Silva, J. Caramujo, S. Monfared, P. Calado, and T. Breaux, "Improving the Specification and Analysis of Privacy Policies: The RSLingo4Privacy Approach ," *ICEIS 2016*, p. 336, 2016.

[26] P. X. Mai, A. Goknil, L. K. Shar, F. Pastore, L. C. Briand, and S. Shaame, "Modeling Security and Privacy Requirements: a Use Case-Driven Approach," *Information and Software Technology*, vol. 100, pp. 165–182, 2018.

[27] C. Lachner, T. Rausch, and S. Dustdar, "Context-Aware Enforcement of Privacy Policies in Edge Computing," in *2019 IEEE International Congress on Big Data (BigDataCongress)*. IEEE, 2019, pp. 1–6.

[28] K. Barker, M. Askari, M. Banerjee, K. Ghazinour, B. Mackas, M. Majedi, S. Pun, and A. Williams, "A Data Privacy Taxonomy," in *British National Conference on Databases*. Springer, 2009, pp. 42–54.

[29] Q. Ni, D. Lin, E. Bertino, and J. Lobo, "Conditional Privacy-Aware Role Based Access Control," in *European Symposium on Research in Computer Security*. Springer, 2007, pp. 72–89.

[30] Q. Ni, E. Bertino, and J. Lobo, "An Obligation Model Bridging Access Control Policies and Privacy Policies," in *Proceedings of the 13th ACM symposium on Access control models and technologies*, 2008, pp. 133–142.

[31] N. Venkataramanan and A. Shriram, *Data privacy: principles and practice*. CRC Press, 2016.

[32] "The ChatterBaby privacy policy," 2017. [Online]. Available: https://chatterbaby.org/files/view/download_files/Privacy_Policy_IRB.pdf

[33] "Eider House's privacy policy." [Online]. Available: http://www.eiderhouse.com/privacypolicy.pdf

Appendices

A ImaginaryOnlineShopping Privacy Policy

This privacy notice discloses the privacy practices for (ImaginaryOnlineShopping.com)[4]. This privacy notice applies solely to information collected by this website. It notifies you of the following:

1. What personally identifiable information is collected from you through the website, how it is used and with whom it may be shared.
2. Who uses your data in our company.
3. What choices are available to you regarding the use of your data.
4. The security procedures in place to protect the misuse of your information.
5. How you can correct any inaccuracies in the information.

Part I Information Collection, Use, and Sharing

We are the sole owners of the information collected on this site. We only have access to/collect information that you voluntarily provide via email or other direct contact from you. We will not sell or rent this information to anyone. We will use your information to respond to you, regarding your purchases. We will not share your information with any third party outside of our organization, unless necessary to fulfill your request, e.g. to ship an order.

In order to use our website, you must first complete the registration form. During registration, you are required to provide certain personal information (such as name, email, and address). This information is used to contact you about the products/services on our site in which you have expressed interest. You may also provide personal demographic information (such as date of birth) about yourself, but it is not required.

Part II Access to your data

In our organization, only employees who need information to perform a specific job (for example, shipping, sending gift, or analyzing) are granted access to personal information.

Part III Orders and Shipment

We request additional information from you on our order form. To buy from us, you must provide contact information (like name and email address) and financial information (like credit card number). Deliverers ship orders that you place. To ship your orders, deliverers access your name, order list, credit card information, address, and email address to respectively identify you, process your order, charge fees, ship the parcel, and finally, inform you about the shipment.

We use an outside shipping company to ship orders, and a credit card processing company to bill users for goods and services. These companies do not retain, share, store, or use personally identifiable information for any secondary purposes beyond filling your order.

[4] This privacy policy is a modified version of Eider House's privacy policy [33].

Part IV Marketing

From time-to-time, our analyzers perform analyses on your shopping history and date of birth to enhance our services. Our analyzers combine your date-of-birth and shopping history to better understand your shopping habits, and predict your interests; our marketers will then suggest products that might interest you. A marketer is an employee with a valid contract term, and they work under Analyzers' supervision; the manager supervises analyzers and deliverers. Marketing staff members will send you advertisements within business hours. To send advertisements, first marketers identify you by name, then use your email to send you suggestions and ads that might interest you. Our marketers will also send you a gift on your birthday. To send birthday gifts, marketers check your date of birth, identify you, and send a gift to your address. A customer's name can be used for marketing, if the customer is over 18 years old.

Part V Your Access to and Control Over Information:

Unless you ask us not to, we may contact you via email in the future to tell you about specials, new products or services, or changes to this privacy policy. You may opt out of any future contacts from us at any time. You can do the following at any time by contacting us via the email address or phone number given on our website:

- See what data we have about you, if any.
- Change/correct any data we have about you.
- Have us delete any data we have about you.
- Express any concern you have about our use of your data.

Part VI Security

We take precautions to protect your information. When you submit sensitive information via the website, your information is protected both online and offline. Wherever we collect sensitive information (such as credit card data), that information is encrypted and securely transmitted to us. You can verify this by looking for a lock icon in the address bar and looking for "https" at the beginning of the address of the Web page. While we use encryption to protect sensitive information transmitted online, we also protect your information offline. The computers/servers in which we store personally identifiable information are kept in a secure environment.

B ChatterBaby™'s Privacy Policy Components and Connection Lists

Table B1 lists all of the roles mentioned in this privacy policy. Several statements in ChatterBaby™'s privacy policy introduce 'We', 'Us', and 'ChatterBaby™' as a single role, for the first time, in the privacy policy. Although 'We/Us/ChatterBaby™' seem like proper names, which refer to the same specific, distinct, legal entity, and choice-of-words is designed to be non-legalistic and friendly for the consumer, they do not precisely introduce the role in the company; rather, they introduce the role as the whole company.

Label	Role
r_1	We/Us/ChatterBaby
r_2	Others
r_3	Member
r_4	Service provider
r_5	Advertising and marketing partner
r_6	Joint marketing partner
r_7	Affiliate

Table B1: ChatterBaby™'s roles

Table B2 shows the list of all the purposes and their labels.

Label	Purpose	Label	Purpose
p_1	Improve service	p_{15}	Research
p_2	Provide service	p_{16}	Extract vocal and disorder relationships
p_3	Identify	p_{17}	Fighting spam/malware
p_4	Collect, measure, & process autism risk	p_{18}	Facilitate data collection
p_5	Extracting acoustic features	p_{19}	Identify web browser
p_6	Send alert	p_{20}	Find site visit statistics
p_7	Fulfill request	p_{21}	Marketing
p_8	Anti-fraud	p_{22}	Send research participation request
p_9	Keep customer posted	p_{23}	Promote service
p_{10}	Internal improvement	p_{24}	Any
p_{11}	Send service information	p_{25}	Understand customer behavior
p_{12}	Send notice	p_{26}	Facilitate interaction
p_{13}	Auditing	p_{27}	Provide product
p_{14}	Data analysis		

Table B2: ChatterBaby™'s purposes

Table B3 shows the attribute list. The last column contains the attribute groups according to ChatterBaby™'s privacy policy. In this privacy policy, attributes are directly categorized into three main groups: 'Personal information', 'Non-personal information', and 'Other information'. Each group also has sub-groups. According to the privacy policy, 'Personal information' includes 'Individual', 'Friend', 'Child', 'Profile', and 'Contact information' groups. The 'Non-personal information' group includes 'Medical', and 'location' groups. Finally, the 'Other information' group includes 'Device-specific', and 'Browser' groups. In the statement "Where we use your data for direct marketing purposes...", the policy refers to accessing customers' data. Therefore, we consider 'Data' as a group that includes all the collected data about customers.

Label	**Attribute**	**Group**
d_1	Name	Data, Personal, Individual
d_2	Age	Data, Personal, Individual
d_3	Mailing address	Data, Personal, Individual
d_4	Phone number	Data, Personal, Individual
d_5	Email address	Data, Personal, Individual, Child, Contact information
d_6	Contact preferences	Data, Personal, Individual
d_7	Credit card information*	Data, Personal, Individual
d_8	Username	Data, Personal, Individual
d_9	Password	Data, Personal, Individual
d_{10}	Child's name	Data, Personal, Child
d_{11}	Child's date of birth	Data, Personal, Child
d_{12}	Week of delivery	Data, Personal, Child
d_{13}	Child's gender	Data, Personal, Child
d_{14}	Audio recording	Data, Personal
d_{15}	Video data	Data, Personal
d_{16}	Information from services	Data, Personal
d_{17}	Friend's name	Data, Personal, Friend
d_{18}	Friend's mailing address	Data, Personal, Friend
d_{19}	Friend's email	Data, Personal, Friend
d_{20}	Friend's phone number	Data, Personal, Friend
d_{21}	Written contents	Data, Personal
d_{22}	Language	Data, Non-Personal, Location
d_{23}	Zip code	Data, Non-Personal, Location
d_{24}	Area code	Data, Non-Personal, Location
d_{25}	Referrer URL	Data, Non-Personal, Location
d_{26}	Location	Data, Non-Personal, Location
d_{27}	Time zone	Data, Non-Personal, Location
d_{28}	Medical history	Data, Non-Personal, Medical
d_{29}	Autism risk factors	Data, Non-Personal, Medical
d_{30}	IP address	Data, Other, Browser
d_{31}	Cookies	Data, Other, Browser
d_{32}	Device identifier	Data, Other
d_{33}	Network information	Data, Other
d_{34}	Hardware model	Data, Other
d_{35}	Device interaction	Data, Other

Table B3: ChatterBaby™'s attributes

* ChatterBaby™'s privacy policy includes two statements regarding collecting billing information. The statement "we may collect a variety of information, including your name, age, mailing address, phone number, email address, contact preferences, credit card information, username and password" specifies that 'Credit card information' might be col-

lected. Alternatively, the statement ”We will not collect billing information, as our service is free” specifies that no billing information is collected. Nevertheless, we include the ’Credit card information’ in our diagram. This contradiction is highlighted though the process of extracting the information.

Table B4 shows all of the roles’ connections in the privacy policy. These connections are used to form a structure for the role layer.

Superior	**Inferior**
We	Service Provider
We	Advertising and marketing partner
Advertising and marketing partners	Joint marketing partner
Advertising and marketing partner	Affiliate

Table B4: ChatterBaby™’s role connections

Purposes and their tasks are listed in Table B5 if they are specified in the privacy policy.

Purpose	**Tasks**
Extracting acoustic features	Process audio recordings, process video data
Send alert	Process service information, Email alert
Send notice	Email notice
Identify web browser	Identify web browser
Find site visit statistic	Find site visit statistic
Promote service	Disclose information

Table B5: ChatterBaby™’s purpose connections

In ChatterBaby™’s privacy policy, no statements provide accurate information about whether any attributes are combined, or what new attributes are created. Therefore, we are not able to establish any connections between the attributes.

The privacy policy contains statements that specify permissions for roles to use purposes. Table B6 shows these permissions.

Role	Purpose	Condition
We	Provide service	
We	Improve service	
Others	Identify	
We	Identify	
We	Extracting acoustic features	
We	Send alert	
ChatterBaby(We)	Fulfill request	
ChatterBaby(We)	Anti-fraud	
We	Keep customer posted	
We	Internal improvement	
We	Send service information	
We	Send notice	
We	Auditing	
We	Data analysis	
We	Research	
We	Any	
We	Understand customer behavior	
We	Extract vocal and disorder relationships	
We	Identify web browser	
We	Find site visit statistics	
We	Marketing	
We	Send research participation request	
Member	Facilitate interaction	
Service provider	Provide service	
Service provider	Provide product	
Advertising and marketing partner	Promote service	
Joint marketing partner	Promote service	
Affiliate	Promote service	

Table B6: ChatterBaby™'s role-purpose permissions

Table B7 lists purposes' permissions for accessing attribute groups, and B8 lists the tasks and their required attributes.

Purpose	**Attribute group**	**Condition**
Improve service	Personal, Location	
Provide service	Personal	
Identify	Individual	
Collect, measure, and process autism risk	Child	
Extracting acoustic features	Acoustic	
Provide service	Friend	
Fulfill request	Friend	
Anti-fraud	Friend	
Keep customer posted	Personal	
Internal improvement	Personal	
Send service information	Contact information	Consent = True
Send notice	Personal	
Auditing	Personal	
Data analysis	Personal	
Research	Personal	
Research	Non-personal	
Understand customer behavior	Location	
Extract vocal and disorder relationships	Medical	
Fighting spam/malware	Browser	
Facilitate data collection	Browser	
Identify web browser	Cookies	
Find site visit statistics	Cookies	
Marketing	Data	Subscription = True
Send research participation request	Contact information	
Facilitate interaction	Profile	
Provide product	Personal	
Promote services	Service information	

Table B7: ChatterBaby™'s purpose-attribute permissions

Label	**Task**	**Attribute**	**Condition**	**Granularity**
t_1	Process audio recording	Audio recording		
t_2	Process video data	Video recording		
t_3	Process info from services	Information from services		
t_4	Send alert	Email		
t_5	Send notice	Email		
t_6	Identify web browser	Cookies		
t_7	Find site visit statistics	Cookies		
t_8	Disclose information	Service information		

Table B8: ChatterBaby™'s Purpose-attribute (tasks) permissions

C ChatterBaby™'s Privacy Policy Permission Diagram

By using the information in Table B1, we create a node for each role in the privacy policy. We then use Table B4 to add the connections between the roles. Diagram C1 depicts the complete role layer.

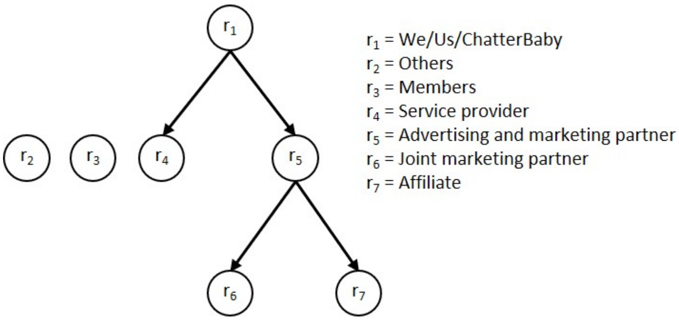


Figure C1: ChatterBaby™'s role structure diagram

The ChatterBaby™ privacy policy diagram's purpose layer is created by adding a node for each purpose listed in Table B2. If a purpose's tasks are provided, we illustrate them in the diagram, using the information in Table B5. Figure C2 depicts the complete purpose layer.

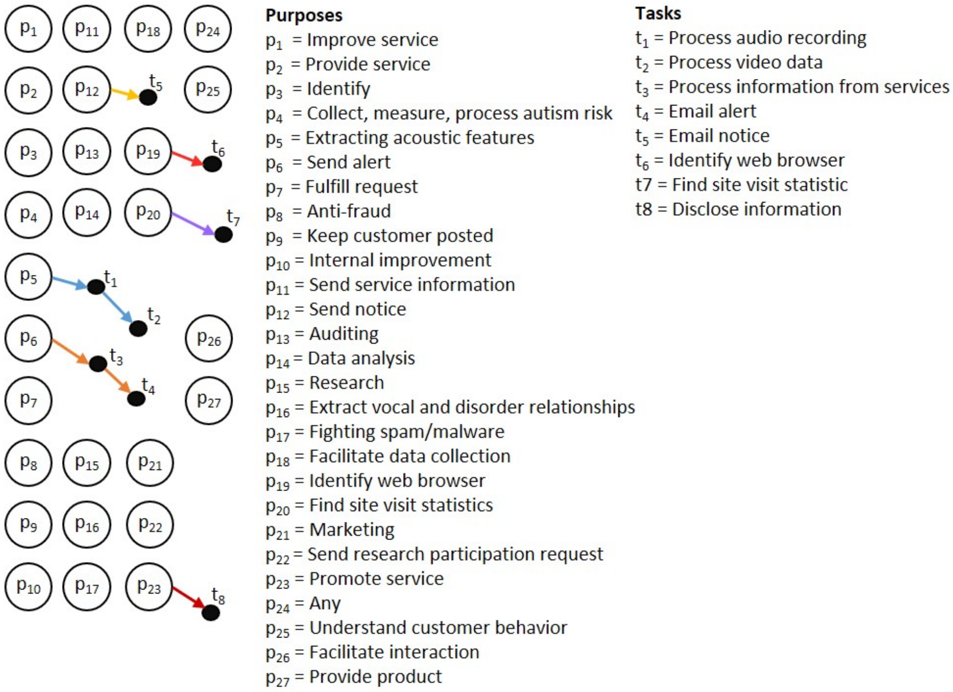
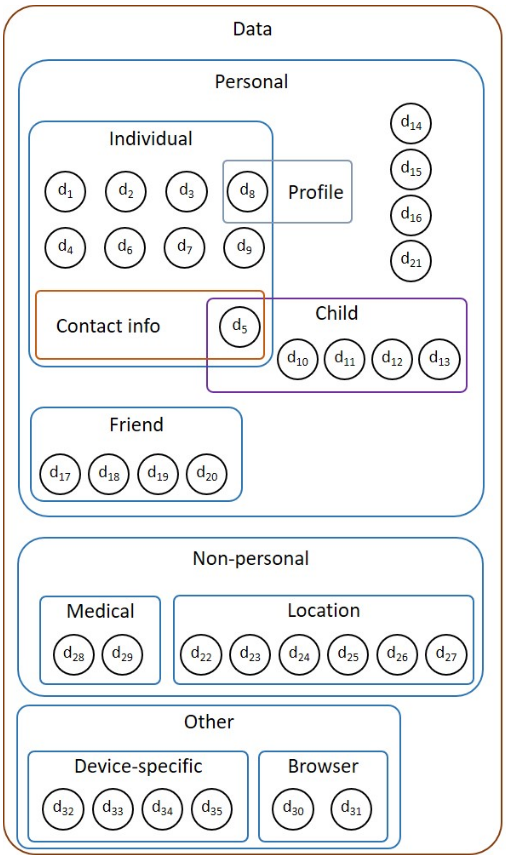

p_1 p_{11} p_{18} p_{24}
p_2 p_{12} t_5 p_{25}
p_3 p_{13} p_{19} t_6
p_4 p_{14} p_{20} t_7
p_5 t_1 t_2
p_6 t_3 t_4 p_{26}
p_7 p_{27}
p_8 p_{15} p_{21}
p_9 p_{16} p_{22}
p_{10} p_{17} p_{23} t_8
Purposes
p_1 = Improve service
p_2 = Provide service
p_3 = Identify
p_4 = Collect, measure, process autism risk
p_5 = Extracting acoustic features
p_6 = Send alert
p_7 = Fulfill request
p_8 = Anti-fraud
p_9 = Keep customer posted
p_{10} = Internal improvement
p_{11} = Send service information
p_{12} = Send notice
p_{13} = Auditing
p_{14} = Data analysis
p_{15} = Research
p_{16} = Extract vocal and disorder relationships
p_{17} = Fighting spam/malware
p_{18} = Facilitate data collection
p_{19} = Identify web browser
p_{20} = Find site visit statistics
p_{21} = Marketing
p_{22} = Send research participation request
p_{23} = Promote service
p_{24} = Any
p_{25} = Understand customer behavior
p_{26} = Facilitate interaction
p_{27} = Provide product
Tasks
t_1 = Process audio recording
t_2 = Process video data
t_3 = Process information from services
t_4 = Email alert
t_5 = Email notice
t_6 = Identify web browser
t7 = Find site visit statistic
t8 = Disclose information

Figure C2: ChatterBaby™'s purpose structure diagram

Figure C3 shows the attribute layer in which attributes are categorized into their groups using information in Table B3. Since the privacy policy does not explicitly specify the attributes' connections, there are no edges in this layer.



Personal information
d_1 = Name
d_2 = Age
d_3 = Mailing address
d_4 = Phone number
d_5 = Email address
d_6 = Contact preferences
d_7 = Credit card information
d_8 = Use name
d_9 = Password
d_{10} = Child's name
d_{11} = Child's date of birth
d_{12} = Week of delivery
d_{13} = Child's gender
d_{14} = Audio recording
d_{15} = Video data
d_{16} = Information from services
d_{17} = Friend's name
d_{18} = Friend's mailing address
d_{19} = Friend's email
d_{20} = Friend's phone number
d_{21} = Written contents

Non-personal information
d_{22} = Language
d_{23} = Zip code
d_{24} = Area code
d_{25} = Referrer URL
d_{26} = Location
d_{27} = Time zone
d_{28} = Medical history
d_{29} = Autism risk factors

Other information
d_{30} = IP address
d_{31} = Cookies
d_{32} = Device identifier
d_{33} = Network information
d_{34} = Hardware model
d_{35} = Device interaction

Figure C3: ChatterBaby™'s attribute structure diagram

The diagram in Figure C4 shows that if the customer subscribes, all their data can be used for the ’Marketing’ purpose. This permission is a result of the statement ”Where we use your data for direct marketing purposes, you can always object using the unsubscribe link in such communications or changing your account settings.” This statement has ’We’ as a role, ’Marketing’ as a purpose, and ’Data’ as a group that includes all attributes.

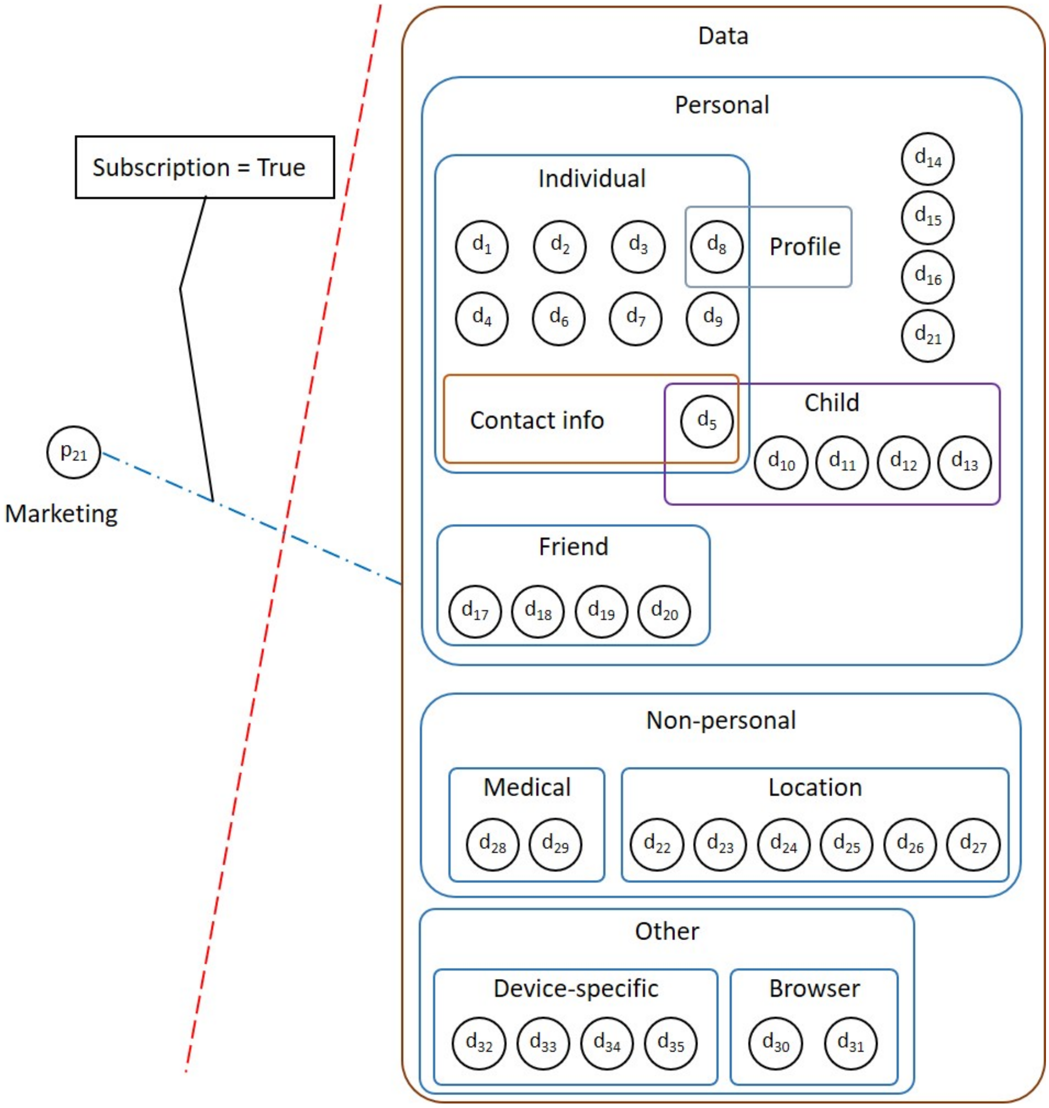
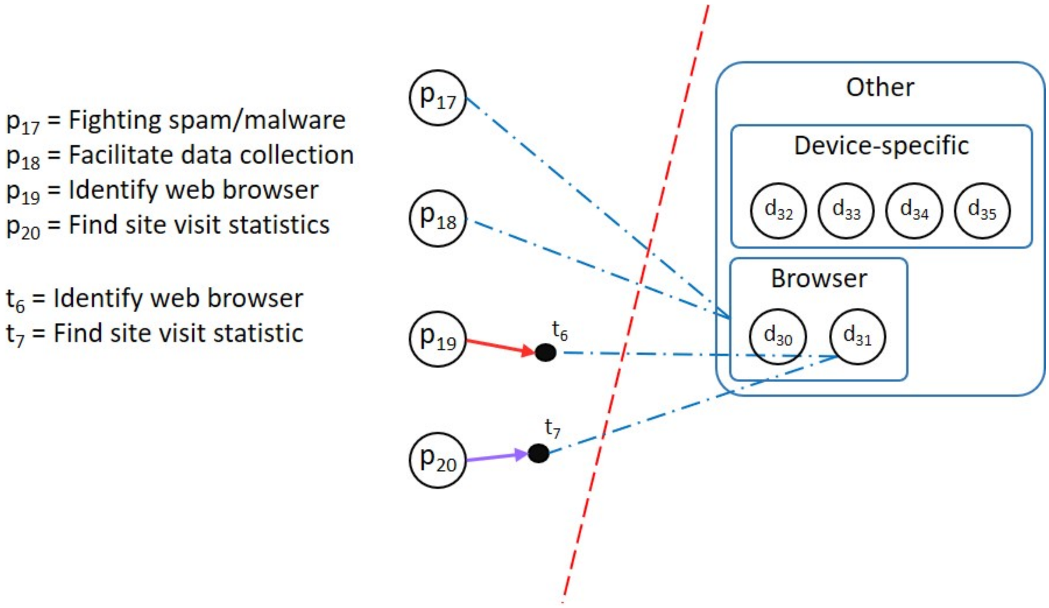

Data
Personal
Individual
Profile
Contact info
Child
Friend
Non-personal
Medical
Location
Other
Device-specific
Browser
Subscription = True
p_{21}
Marketing
d_1 d_2 d_3 d_4 d_5 d_6 d_7 d_8 d_9 d_{10} d_{11} d_{12} d_{13} d_{14} d_{15} d_{16} d_{17} d_{18} d_{19} d_{20} d_{21}
d_{22} d_{23} d_{24} d_{25} d_{26} d_{27} d_{28} d_{29} d_{30} d_{31} d_{32} d_{33} d_{34} d_{35}

Figure C4: ChatterBaby™’s heterogeneous permissions to access the ’Data’ group

Figures C5 shows permissions for the ’Other’ group and its subgroups, capturing the statement ”We may collect some device-specific information if you access the Services using a mobile device. Device information may include but is not limited to unique device identifiers, network information, and hardware model, as well as information about how the device interacts with our Services.” While the data items of the attributes in the ’Device-specific’ group may be collected, no statement in the privacy policy specifies a purpose for collecting them. This shortfall is illustrated in Figure C5, where the attributes in the ’Device-specific’ group are not connected to any purpose.



p_{17} = Fighting spam/malware
p_{18} = Facilitate data collection
p_{19} = Identify web browser
p_{20} = Find site visit statistics
t_6 = Identify web browser
t_7 = Find site visit statistic
p_{17}
p_{18}
p_{19}
p_{20}
t_6
t_7
Other
Device-specific
d_{32}
d_{33}
d_{34}
d_{35}
Browser
d_{30}
d_{31}

Figure C5: ChatterBaby™’s heterogeneous permissions to access the ’Other’ group