IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 9, 2020, accepted March 11, 2020, date of publication April 17, 2020, date of current version May 5, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2988566

TR-Model. A Metadata Profile Application for Personal Data Transparency

THIAGO ADRIANO COLETI[1], PEDRO LUIZ PIZZIGATTI CORRÊA[1], (Member, IEEE), LUCIA VILELA LEITE FILGUEIRAS[1], AND MARCELO MORANDINI[2]

[1]Escola Politécnica/School of Engineering, University of São Paulo, São Paulo 05508-010, Brazil
[2]School of Arts, Science and Humanities, University of São Paulo, São Paulo 03828-000, Brazil

Corresponding author: Thiago Adriano Coleti (thiagocoleti@usp.br)

This work was supported by the Fundação de Amparo à Pesquisa do Estado de S. Paulo (FAPESP) under Project 2018/06017-6.

ABSTRACT People's usage of social networks, mobile applications, websites, sensor networks and other computer systems leads to a massive production of personal data about their behaviors and preferences. Personal data are used by organizations in business and marketing tasks. However, details about personal data usage are often not accessible or clear to data subject, raising concerns about privacy and security. Presentation of information about personal data usage needs improvement towards Personal Data Transparency. Thus, this paper aims to present the TR-Model, a Metadata Application Profile guideline that intends to propose a standardization on information to be considered minimally necessary to Personal Data Transparency as well as a set of specifications to guide developers on how to present this data. TR-Model elements are focused providing Personal Data Transparency in a user-friendly and high quality format. TR-Model presents a set of specification based on entities, metadata, metaevents and descriptions. The model evaluation was based on user testing in several scenarios of usage of personal data in a gym application tool. The information presented was created based on the TR-Model metadata, metaevents and descriptions. Participants evaluated transparency considering dimensions of Human-Computer Interaction and Information Quality. Participants' opinions were recorded in surveys and analyzed with descriptive statistics; the results indicate that the TR-Model was effective in supporting the production of friendly, understandable and relevant Transparency for data subjects, in compliance with regulations like GDPR.

INDEX TERMS Human-data interaction, Metadata Application Profile, Personal Data Transparency, personal infovis, user-friendly transparency.

I. INTRODUCTION

The use of Personal Data produced as a result of interaction between people and hardware/software resources has become common practice [1]. According to General Data Protection Regulation (GDPR) [2], Personal Data are *any information relating to an identified or identifiable natural person ('data subject'). An identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as name, identification number, location data, online identifier or one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*.

The associate editor coordinating the review of this manuscript and approving it for publication was You Yang.

The interest in Personal Data had increased significantly as the use of social media and mobile applications based on the internet had grown significantly the last few years [3], [4]. So, the use of technological resources is responsible for data generation that, some how, reflect the behavior, preferences or data subject' features. Examples include: being part of a community on a social network can tell about a person's interest in a particular subject; the repeated e-commerce portals access may indicate the desire for buying specific products; or even the use of credit card in a region or type of commerce may define buying interests and profiles [5], [6].

Many companies monetize Personal Data as they can provide important insights about existing ad potential consumers [1]. These insights can be used by companies to obtain advantages for increasing financial gains such as: (1) understand the data subjects' profiles; (2) know / understand

routine behaviors; and (3) process text messages to obtain a strategic information for a purpose.

These actions (among many others) are carried out even if, sometimes, they are done without data subject's knowledge and / or prior authorization; this may cause concerns about privacy, freedom and security of the data subject's personal life [5], [7], [8]. According to the GDPR, Art. 4, Alinea 1, Data Subject is *an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person*.

The abusive and unauthorized use of Personal Data is already a reality and is discussed by [9]. In this context, the concept of Personal Data Transparency arises as a resource to allow Data Subjects visualizing and understanding events and agents involved in the use of their Personal Data and thus, acting to ensure their privacy, freedom and security.

Personal Data Transparency (Transparency) is the *ability to enable data subject to access and understand information about why, how and by whom their Personal Data are used* [3], [5]. Bellamy and Allonso [10] explain Transparency as a *fundamental concept and requirement in many data privacy laws around the globe. It requires organization handling Personal Data to be open with and inform data subjects of their data uses and practices*. And Murmann and Fischer-Hubner [11] highlight that Transparency is an important factor to ensure confidence in the application and its controllers.

Some initiatives to work with Transparency are being conducted by companies and researchers in order to provide data owners with the ability to know what happens to their data. As an example, Transparency Enhancing Tools (TETs) are software or browsers plugins that provide PD information such as recipient Internet IP, or allow data subject to manage who is able to access his/her data in social network [12] and [13].

Companies that use Personal Data seek to review their Transparency strategies as discussed in [10]. Specifically, scientific innovations such as [14], which the authors discussed about how developers usually access and work with personal data and how they have difficulties to offer transparency to make their apps more privacy friendly. Aiming to provide a more user-friendly Transparency, the authors propose the PrivacyStream, a function programming frameworks that transform raw data from source as sensors and database in a standard-format stream. The output information can be used to be analyze with single processing methods (avoiding external states and complex croos reference) and using flow graphs to describe how personal data process is conducted in a app.

Also, Patrick [6] presented a set of privacy principles with it respective HCI requirements. With this parameter, the author proposed possible solutions, for example, for transparency privacy principle, the HCI requirement describe as: *data subject must be aware of the transparency options, and feel empowered to comprehend and control how their data are handled*. As possible solution, the author proposed the transparency information explanation through examples and tutorials. The author prototype and modeled with UML a software as a concept proof to exemplify the solutions and to be used for future works in meeting data subject's privacy needs.

Also, Privacy and Security Policies (PSP) are usually used by several companies to inform their costumers about the use of data. However, PSP are complex, written in legal language as a long, verbose contract, making it difficult for data subjects to read and understand [10].

Different strategies encourage and/or enforce provision of clear and understandable information for Data Subjects about the use of his/her Personal Data, in different levels of demand. Examples are

- (1) Personal Data Regulations as the General Data Protection Regulation (GDPR) created by the European Union [2]; the Brazilian General Law for Data Protection (LGPD) [15]; the Personal Information Protection and Electronic Documents Act (PIPEDA) from Canada [16];
- and (2) the Privacy and Security Policies (PSP) is a document that describes how a company handle the Personal Data regarding the privacy, security and usage [10], [17].

However, all the presented strategies should be focused on providing legible and relevant information, thus allowing the analysis and understanding by data subjects in order to support their personal data monitoring. Thus, they have issues calling for improvements:

1) The standardization of Transparency, by the definition of metadata that can be used by data subjects to understand the use of their data and that allows analysis whether the process is in compliance with the national regulation;
2) Transparency Information that can present meaningful information for data subject, once presented strategies can provide information that are difficult to use or useless for data subject; and
3) User-friendly presentation of Transparency resources, to allow the data subject performing the correct analysis of the processing and usage of their Personal Data as well as actions in case he/she identifies any unwanted action.

The first improvement, towards standardization of transparency metadata was discussed by [18], who sees advantages like providing a relationship between the data subject and applications and also in reducing costs to provide transparency. Transparency metadata can be considered as a set of information about agents and processes involved in the use of personal data for a purpose.

Standardization is a challenge because Transparency metadata are significantly heterogeneous and can be application-context driven. Work by different researchers point to this

heterogeneity. For example, Transparency metadata can include:

- programming level information: technical information about low-level systems' processes, such as services, methods, data storage data encryption and the usage of hardware resources like microphones or cameras as presented by Li *et al.* [14] and Rieder *et al.* [19]. These data are usually available in computer-readable formats and the abstraction for human-readable formats can be complex;
- information about the collection and the use of behavior data that are usually collected from smartphones devices as GPS location, contact agenda, telephone call records, sensors working in houses and cars and can provide [20], [21]. These information could provide valuable inputs to support controllers to know about people common tasks.
- knowledge obtained by data subject with personal data processing: describes how personal data are processed and which information are created and how the information is applied in data subject daily life [22];
- information about the commercial use of personal data: the purpose of use, sharing and disclose details, data recipients, controllers and processors that access data [23];
- data subject rights information: information about how they can exercise their rights in order to request data copy, cancel or update permission of use, report misuse of their data, access data protection office.

Thus, it is assumed that differences in information between applications may cause Data Subject to have difficulty in using and understanding Transparency, since it can allow to be considered that the information is being concealed intentionally. In this way, the need for a standard is justified once all systems can display a minimum set of Transparency satisfies that the data owners' needs, reduces implementation costs, is relevant to the understanding of the data and is in accordance with some use of Personal Data regulation.

The second improvement, which to some extent is a consequence of the diversity of metadata, refers to the relevance of information. An issue may arise when a data subject requests an unavailable Transparency Information but also, when he/she is required to analyze Transparency Information irrelevant to their needs. The effort to understand irrelevant information is detrimental to Transparency for it may lead the data owner to an attitude of lack of interest and disengagement.

The third improvement aims to provide information that is easily comprehensible by data subjects. Li's cited proposal of presenting data processing details as a graph [14], for example, may result in serious difficulties for people with low educational level. Alternatives such as using symbols and plain language can facilitate the comprehension of data events, even if they are technical actions.

Addressing these issues, this paper presents the TR-Model. TR-Model was aimed to be used to support the development of tools to provide Transparency in a understandable and practical strategy for data owner. We claim that TR-Model promotes Transparency by solving issues inherent to the lack of usable information about personal data usage by final data subjects and providing resources to address the three presented lines of improvements:

- Standardized model of Transparency, with entities, metadata and relations that can be used to provide Transparency in applications and websites. The Transparency domain model and its features can support different purposes of Personal Data usage, in compliance with GDPR and according to data subjects expectations. The metadata and relations may be considered relevant to support Data Subject to understand and act about his/her personal data;
- Provide a metadata description based on data subject experience design and information quality assets, to provide usable Transparency, by providing Controllers with guidance on developing usable Transparency tools and understandable Transparency information for Data Subjects.

TR-Model was created based on Metadata Application Profile (MAP), that is an Information Science approach to provide a set of elements, strategies, guidelines and vocabulary that are defined for a specific domain to guarantee that the application achieves its functional requirements [24], [25].

The activities for developing TR-Model were performed in five phases.

1) The analysis of artifacts (papers, regulations, technical reports etc) regarding to Transparency, which were used to identify Transparency issues focused on delivery information for Data Subject is presented in Section II;
2) Definition of a Domain Model for Transparency;
3) Definition of a set of relevant Transparency attributes to be analyzed by data owners;
4) Proposition of a set of specifications to guide the development of Transparency Visualization Tools; and
5) Validation of TR-Model that aimed to verify whether TR-Model was able to provide relevant and understandable Transparency information for Personal Data owner.

We believe that the TR-Model model can be used as Metadata Profile Application (MAP) to support the development and/or the evaluation of Transparency software. We believe that the use of TR-Model to support software development can improve the interaction capacity between Data Subjects and Transparency Data regarding the quality and reliability of the information, as the Data Subject can access a set of standardized information and describe the ones that may make easier to understand the Transparency and also ensure that the minimum Transparency information is displayed. Also, for evaluation processes, we consider TR-Model useful to support Transparency inspection guidelines and parameters for evaluating user experience in users testing.

In order to present TR-Model as well as the basis for its development, we structure this paper in five sections. Section II is dedicated to presenting the fundamental concepts related to Personal Data Transparency and its regulation. Also, the foundations of the Metadata Application Profile are also presented. Section III presents the literature review on Personal Data Transparency which provided information on requirements for TR-Model. Section IV is the core of this paper, in which we detail the TR-Model specification, which includes its requirements, its domain model and main concepts definitions. Section V is dedicated to presenting our work on assessing the effectiveness of TR-Model, which we performed using two methods - a Transparency coverage inspection and a user experience evaluation. Section VI presents our conclusions and future work.

II. CONCEPTS AND DEFINITIONS

This section present concepts and definitions that supported the development of TR-Model. The subjects is focused on laws/regulations for Personal Data usage, Metadata and Metadata Profile Application.

A. PERSONAL DATA

Mortier *et al.* [5] explains Personal Data as *electronic records of the transactions or activities of a particular person that became him/her identifiable and/or make able to be analyzed by system in order to learn about his/her individual (data subject) behavior*. Mortier cites as examples of Personal Data are: person's financial transactions; telephone and Internet usage; Social Networks events; GPS Location etc.

Personal data are created by the several ways of interaction among people and computational resources and it changed the way of human lifestyle significantly [26]. Mostafa *et al.* highlights that technology companies identified that, processing personal data, it can obtain valuable insights about people and use it to guide commercial tasks such as creating customer profiles or developing intelligent software tools.

The massive personal data exploration occurs frequently after years 2000, but there is no deadline for it. Schneier [27] reports that Personal Data usage limits are pointed as computational or internet accesses options and there are no well defined variables that can make this usage difficult.

As mentioned, personal data are created for devices as smartphones and sensors. The personal data can have different data type and formats according to controller's purpose of use. One of the main features that influence Personal Data usage is the level of detail for the collecting and processing phases. The level of detail can determinate which information the controller are able to produce and what kind of knowledge about the data subject can be obtained [14]. In this paper, the detail level are named as granularity.

For example, it can be considered the scenario in which a company wants any specifically/individual information related to credit card purchases. So, it can be said that the Personal Data is the **card payment record**, whereas the level of detail (graininess) may vary according to the need for the information. Example of granularity is shown in Figure 1 based on [14].

TABLE 1. Personal data granularity example. Adapted by [14].

Personal Data: Purchase registration with credit card	
Fine Granularity	Coarse Granularity
- Purchase Date - Company Name - Purchase Value - Products (for e-commerce) - Purchase City	- Total of purchases Summarized by month

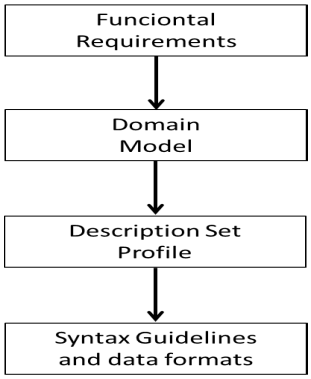

FIGURE 1. MAP development process. Adapted by DCMI.

Thus, it can be stated that data granularity is strongly related to the type of Personal Data that will be collected and, consequently, used by the controller. This is because the granularity will define the Personal Data level of detail and therefore what information can be produced about the data subject [28].

The widely use of Personal Data for several purposes made arise the concern with Transparency in order to provide software with easy and access to information about the use of personal data. Transparency is presented in next subsection.

B. PERSONAL DATA TRANSPARENCY (TRANSPARENCY)

Personal Data Transparency, shortened as Transparency in this paper, can be defined as the quality of a computer system that means the degree with which the system provides data owners privacy, security and control over their data processing, allowing the data subject to exercise his/her rights and to act on the use of data. Transparency has become a requirement for systems that collect and manage Personal Data [11].

Efforts towards Personal Data Transparency are usually concerned with making data and analytic algorithms both visible and comprehensible to data subjects [5]. Providing such information is not considered a trivial action, since certain events are technical / computational in nature and are complex to be translated into a simple information visualization mechanism [5].

Mortier *et al.* [29] highlight the fact that Transparency has a strong relationship with other qualities of Human-Computer Interaction, since providing Transparency directly depends

on techniques for mapping/translating computational processes into understandable information. The difficulty in achieving this balance is often one reason why processes with Personal Data are unclear to data subjects [30].

Aiming to provide the minimum information about data use, websites usually insert information about this usage into a knowledge document such as Privacy and Security Policy (PSP). PSPs are not efficient strategies for Transparency because they are often long and complex texts, presented to the data subject at the time of acquisition, download or execution of an action and its content refers to practices of collection, use and disclosure of Personal Data [31].

Filgueiras *et al.* [32] observed that the complexity of PSP have not encourage people to read it once, they not want to look or may not have the means to look or may not have the ability to see the truth regarding their data. Due to this reason, Data Subject are often aware of data collection and processing, but they do not know when, how or where they happen, and they do not give due importance to examining terms of use and privacy policies of digital services before consenting.

Transparency also appears as a prerequisite for a data subject to exercise his/her rights knowingly about the use of his/her Personal Data [5]. The exercise of rights refers to his/her ability to restrict or cancel a usage permission, to request a report and/or copy of the collected data. These actions are encouraged and guaranteed by the EU GDPR and other national regulations.

C. TRANSPARENCY IN THE EUROPEAN UNION GENERAL DATA PROTECTION REGULATION

The EU GDPR is the result of an effort by the European Union bodies to provide a data protection regulation for those living in the EU, while also providing greater uniformity to existing data laws [33]. The GDPR came into force in 2018 to regulate and control the use of Personal Data in any field involving the European Union [2].

The regulation is strongly based on the need to defend the freedom and privacy of the data subject who produces Personal Data frequently and has such records used by companies for a wide range of purposes. Considering that the acquisition of data is usually done in a *black box* strategy, in which data subject do not know how their data will be manipulated, the GDPR proposes a set of guidelines that aims to guarantee the Transparency of processes for the people.

The GDPR presents a set of regulations that comprise principles, rights and issues related to organizations that are responsible for protection and control, cooperation, penalties, among others features. There are 11 chapters, 99 articles and 173 considerations on the use of Personal Data [2].

It is important to highlight that the GDPR aims to protect the data subject who holds the data; it is possible to notice its focus on their freedom, privacy and rights. Thus, Transparency must be considered as a mean to ensure that Data Subject can achieve these items.

Transparency is an important concept in the GDPR. In a significant number of articles, the regulation requests that companies that use Personal Data pay attention to providing Transparency. Also, in chapters 13, 14 and 15, the GDPR presents a list of items that must be considered to provide Transparency. Among the items for Transparency in GDPR are:

- identification data of controllers, processors and control entities;
- processing purposes, data used, period of use, controllers' interests and legal basis for the use of the data;
- information on the origin of data, if obtained from third parties;
- information on information sharing with third parties;
- processing information and computer-based decision making; and
- information on the means through which data subjects may exercise their rights to request actions upon their data, such as usage restriction or copy of data.

The beneficiaries of Transparency required by the GDPR include all data subjects of companies located in the European Union or those that may have any kind of link with them, such as: subsidiary of an EU company or the ones that exchange information with any EU company. After the GDPR has been deployed, fines were established and offenders were effectively charged, the use of Personal Data tends to be controlled. Data subjects may have a clear protection from this regulation, but protection depends also on data subjects accessing data usage information and on acting if data misuse is identified.

D. METADATA APPLICATION PROFILE

Metadata is the data that represents the information about the data [34]. In essence, it describes what, who, when, where, and how about each facet of the information, assisting the organization in its publication and support [35].

Metadata articulates the context of a resource such as a picture, a text file or any digital register providing description that allow analyzing, understanding and exchanging data among different operators [36]. The concept of metadata is not new, but modern applications that are called as ''computing-oriented'' has gained space since the mid-1990s in conjunction with the growth of the Internet [36], [37]. This situation is due to factors such as the Internet placing a large number of documents which requires a structured way to provide meaningful content and context to the data subject [37].

The need to build web-based metadata that would adapt to specific domain needs gave rise, in the early 2000s, to the concept of Metadata Application Profile [38]. Metadata Application Profile (MAP) is a set of guidelines that specifies which metadata is used in a domain as well as its usage rules, data format and other specifications that allow the use of MAP [36].

The main justification for using MAPs was the ability to use a combination of different metadata patterns by extracting the best in each to meet the specific domain [24]. Still, a MAP

does not necessarily have to be a combination of several patterns, but it can have its structure designed specifically for a domain, if other patterns cannot be tapped [38].

Dublin Core Application Profile [36][1] and also Tennis [38] highlight that a MAP can contain the following document set:

- **Functional Requirements**: describes the set of actions that the MAP should allow to conduct;
- **Domain Model**: specifies the entities, agents, relations, attributes of the domain;
- **Description Set Profile**: describes the terms of the metadata and the rules to use it; and
- **Syntax Guidelines and Data Formats**: defines coding rules for computer systems to use to read / process data.

Morgana and Baptista [24] and Malta and Baptista [39] highlight several examples of MAPs that have sought to address a wide range of application domains. Examples cited by the authors are: Scholarly Work Application Profile (SWAP) [40] used to describe academic work and the Virtual Open Access Agriculture - Aquaculture Repository (VOA3R) [41] developed by the Food and Agriculture Organization (FAO).

A widely known MAP is the Dublin Core that contains fifteen metadata elements to identify and sort documents within a specific context [42]. Dublin Core usage allows its terms to be refined through the Qualified Dublin Core and this possibility allows the refinement of content specifications to improve interoperability and web semantic accuracy, considering that semantics should be the main focus [36], [42]. Interoperability is the ability to exchange information in a large and heterogeneous system environment distributed in several areas [43] while Web Semantic is the ability to classify the data based on different context and assign meaning to it in order to be better human understanding, but also in enhancing the understanding of the machines [44].

Another well-known application of metadata is to meet the demand for data sharing in a fast, reliable and quality way. So, the Darwin Core metadata standard can be used as the standard for biodiversity data sharing [45]. Darwin Core also contains a set of terms and categories that can be understood by humans or processed by machines. These terms allow to treat biodiversity occurrence data with metadata on location, biome, context, taxonomy and identification of species.[2]

The amount of existing MAPs is still relatively small as discussed by [24]. However, it is possible to identify various initiatives for using application profiling in domains such as libraries, data collections, encoding mechanisms for transmission, open data and multimedia content [24], [25].

It is believed that the development of other MAPs might make grow the possibilities for application domains that could require the correct use of metadata and their rule descriptions shall increase. In the context of this research, the use of Personal Data is considered a domain with wide possibility for using MAPs [25].

[1]http://www.dublincore.org/specifications/dublin-core/profile-guidelines/

[2]https://github.com/tdwg/dwc

III. WORK RELATED TO PERSONAL DATA TRANSPARENCY

This section presents some specific work related to this research. The related works were selected because they focus on defining some kind of strategy for identifying, classifying or modeling privacy and Personal Data usage information in meaningful features to be used in software application, formal methods or user-friendly interfaces.

In the related works, the researchers discussed specific issues similar to those presented by [46], who discussed the difficulty of working with Transparency once it usually requires presentation an excessive amount of biased, subjective and sometimes inaccessible information. Also, it may still provide unwanted disclosure of information (trade secrets) from a company to a potential competitor.

In Patrick and Kenny's research [6], the authors mapped GDPR's Personal Data usage requirements to a set of *High-level Principles*. The principles were applied to guide the development of software interfaces usable by data subjects to track the use of their Personal Data.

The principles were used to support specification of user interface requirements and also to propose possible interface solutions for Transparency. For example, for the Transparency principle, it is presented as the requirement of: *resource to data subjects access, understand and manipulate how their personal information is used.* As an example of an interface, the authors report that interfaces should exemplify and explain based on tutorials how Personal Data is used. A second example is that the interface components and content used to obtain consent may be unambiguous and obvious for data subject and controllers.

However, the authors do not highlight what information should be shown and how it should be presented to data subjects.

The paper presented by Guarda *et al.* [47] proposed a methodology and a set of techniques for integrating Personal Data security with compliance with Personal Data usage regulations. The methodology aimed to support the abstraction of complex security actions and deliver friendly information to be analyzed and support the decision about the security of their data.

Guarda *et al.* [47] presented a set of categories that organized the information to be protected in the followed categories:

1) Classes: Personal Data refers to the data that allow identifying a personal; Sensitive Data (SD) which refers to Personal Data that includes information on ethnicity, gender, religion and political opinion; and Non-Persoal Data (NPD) that was information that could not be associated with a person;
2) Legal Roles: refers to actions that could be performed by agents within the Personal Data life cycle such as:

Data Controllers, Data Processor and the Data Producing Individual;
3) Auxiliary Notions: characteristics regarding to the access features to Personal Data such as: Purpose; Consent and Data Quality.

The authors replaced the natural language specified by the presented categories in order to derive the formal rules of the template access control policies. They used standard Boolean connectives to combine them in Boolean algebra, once it were considered an expressive technique to reduce the complex privacy specification in a formal language to support software development,.

The third paper is presented by Hosseini *et al.* [46] which highlights the fact that Transparency has become a necessity for software applications. However, there are few well-defined conceptual models or strategies for support developers in understanding, specifying, and implementing Transparency requirements. Hosseini *et al.* [46] discussed that Transparency is often seen as part of other non-functional requirements such as security or privacy.

Hosseini *et al.* [46] identified a set of facets as important be used in a Transparency project. For Transparency information, the *Transparency* facets describe groups of information considered relevant in a scenario where data subject needs to know about the use of his/her data. The facet was classified into the groups:

- Transparency Stakeholders: All actors in the use of Personal Data may be identifiable in order to make possible to understand where the information originates and, which data subject produces the information, who process and/or received it;
- Transparency Meaningfulness: Stakeholders may know about information of data, actions and purposes behind the use of personal data;
- Transparency Usefulness: Information must support the stakeholders actions and their decisions-making or change their perceptions of information provider;
- Information Quality: Describes the importance of quality information features such as free-of-errors, concise, timeliness, understandability, objectivity and reputation.

We also analyzed two researches that aimed to create a GDPR-based class model for Personal Data usage: [48] and [49]. The researches focused on mapping GDPR guidelines into UML Class Diagram and thus provide an approach to support application development in accordance with the regulation. Classes such as: Purpose, Consent, Data Processing, Technical Measures among others were identified. Attributes such as: Free consent, Legal basis and public interest are part of the identified information.

Although the modeling presented by [48] and [49] did not have a specific focus on Transparency, we assumed that the class information could be used as a basis for Transparency metadata. It seems to be out of the scope of these authors concerns about data subject' ability to understand the information and user-friendly ways to present it.

The works presented in this section were initiatives to improve how to provide Personal Data Transparency. Formal languages, UML modeling and information quality facets are elements that could be used to make information more user-friendly.

However, none of the related works performed the improvements suggested in this paper. First, although Transparency metadata could be similar to the classes defined in [48] and [49] work, the authors were not concerned with information metadata standardization. Also, the cited research did not present guidelines to information presentation, although discussions about quality aspects for Transparency metadata that could be used in a presentation strategy were presented by [46].

IV. TR-MODEL - SPECIFICATION TO PROVIDE USER-FRIENDLY PERSONAL DATA TRANSPARENCY

This section presents TR-Model, a set of specifications to provide user-friendly Personal Data Transparency.

TR-Model represents three relevant concepts, which are briefly described here and analyzed in further detail in the following sections: the concept *entities* model information on actors involved in Personal Data Transparency. The concept *metadata* is associated to all information used to describe Personal Data Transparency. The concept *metaevents* represents situations which are relevant to Personal Data Transparency. These concepts are presented in three layers of abstraction: the Domain Model, the Definition and the Specification.

In order to present TR-Model, subsection A present the methodology used to develop the model and the basis upon which the model lays its foundations. Subsection B presents the Domain Model, that defines the entities that compose the model. Subsection C contains the definition of metadata and metaevents for TR-Model entities, similar to attributes and operations in object orientation modelling. Section D ellaborate on metadata and metaevents representation in software tools, focusing on the model's quality in use. Sections E, F, G, H and I respectively bring detailed description of metadata and metaevents (where applicable) for entities Actors, Personal Data, Purpose of Use, Access and Agency.

In next section we present the requirements definition for TR-Model.

A. REQUIREMENTS DEFINITION FOR TR-MODEL

The TR-Model represents a concept in which Transparency metadata are information with about events, agents and relationship involved in the use of Personal Data. Therefore, we considered the Personal Data Transparency as an information about data usage organized in a domain model with entities, metadata and descriptions.

One concern in the construction of the TR-Model was its standardization, consistency, use, understanding and the need to avoid ambiguity and subjectivity to assist data subjects in analyzing the use of his/her Personal Data and act to ensure their rights. To avoid the problems, it was decided to use the Metadata Application Profiling (MAP) approach to support

the development of TR-Model as, according to [25], the MAP strategy makes it possible to identify functional and technical requirements for metadata, address issues of ambiguity and generalization, and facilitate testing.

Thus, TR-Model was created considering the Dublin Core Metadata Initiative (DCMI)[3] specification once, we intend to provide a group of specifications to support the deployment of software tools to provide Transparency for data subjects. DCMI provides a process to be followed in order to created a MAP to support a specific domain. The Figure 1 base on DCMI presents the TR-Model was organized in a set of entities, metadata, events and usage specifications. Events are occurrences that interfere in the use of Personal Data and may be presented for Data Subject. Thus, in this paper we will refer for event as *metaevents*. Details about the meaning and reason to use the metaevents are presented in Section IV-C.

The TR-Model was defined to be a **pattern** for Personal Data Transparency (Transparency) focused on the Data Subjects' interests concerning information and its comprehension capacity. To achieve this goal, in the development process, we focused on the needs of non-experienced/novice users regarding data transparency, that is, those who do not have advanced skills in computing or information science to understand the implications of personal data processing, even though they are regular users of a computer application. This model was designed to be viewed in a high level in which technical information is considered to be of little interest to the data subject and consequently must be supplied by metadata that relevant information in an understandable language and at the same time conveying the full concept of Transparency efficiently.

In the development of the TR-Model, the understanding of the Transparency domain was developed using the Pressman's Domain Analysis [50]. This technique does not consider the active presence of the end user, but uses documents, articles and other means to obtain details of domain requirements. The use of sources other than the user was necessary since users' incipient knowledge about Transparency provided few and sometimes divergent inputs to the model. Thus, the following resources were used to understand the Transparency domain:

- **The GDPR** was the main source of knowledge. It presents a well-defined set of Transparency requirements that must be applied to computer systems. The requirements are clear and understandable to support the model development, although some of them still require some interpretation by the reader;
- **Technical and scientific articles** provided information on the context of usage of Transparency, in a number of distinct areas. This enabled us to identify applicable forms of Transparency in tools such as TETs or initiatives to implement Transparency and ensure what was called *data subject-centric transparency*. Also, it allowed the comparison of the concepts of GDPR with the applications that occur in the TETs;
- **Scientific Papers** have several contributions to Transparency. The articles showed that the concept of Transparency is being discussed in several research centers, but its concepts, structures and application forms still have many open points for discussion, including the definition of a set of structured and described information about agents and events involved in the use of Personal Data. We highlight the work in [5], [11], [29] and [6].
- **Websites** leaded to technical content and research groups on HDI and Transparency studies that provided information on concerns, challenges, TETs and perspectives regarding the implementation of Transparency. Accessing the http://hdiresearch.org/ website we identified materials that encouraged the study about the relationship between IHC areas and Personal Data.
- **Users**: even though they do not have advanced Transparency knowledge to contribute in the requirements elicitation activity, they significantly contributed with the model validation once the information was subjected to critical evaluation of the data subjects through scenarios simulations. Such analysis has fostered discussions to refine TR-Model in order to provide more understandable and interesting Transparency.

In the first stage of the TR-Model conceptualization and development the needs of Transparency were highlighted. So, we tried to understand what kind of information would be relevant to Transparency and how software tools would provide this information in such a way that they could be understandable. For this, the regulations and scientific technical articles that analyzed the main challenges arising from the use of Personal Data and mainly (how) about Personal Data can generate problems of privacy, security and freedom were analyzed.

People who presented themselves as potential data producers were also interviewed and consulted. We interacted with participants through workshops. Workshops were conducted in universities and educational institutes which the researcher presented the concepts, challenges and features of the Personal Data usage. Participants with different expertise such as computing, law, marketing, administration, education and logistic attended the workshops. The workshop's tasks were: presentation about Personal Data usage; resolution of questionnaires; and PSP analysis and discussion. It was common some participants required a talk face-to-face. Although this kind of discussion was informal, it provided relevant data that support the TR-Model development as well as the result of questionnaires and PSP analysis.

Based on the cited previous work, we define the Functional Requirement for TR-Model as: *Support the development of software tools to provide Transparency about the Personal Data usage with information and events about actors, purpose of use, personal data, transfer/disclose and agency in a way that information may be readable, relevant and*

[3]https://www.dublincore.org/specifications/dublin-core/profile-guidelines/

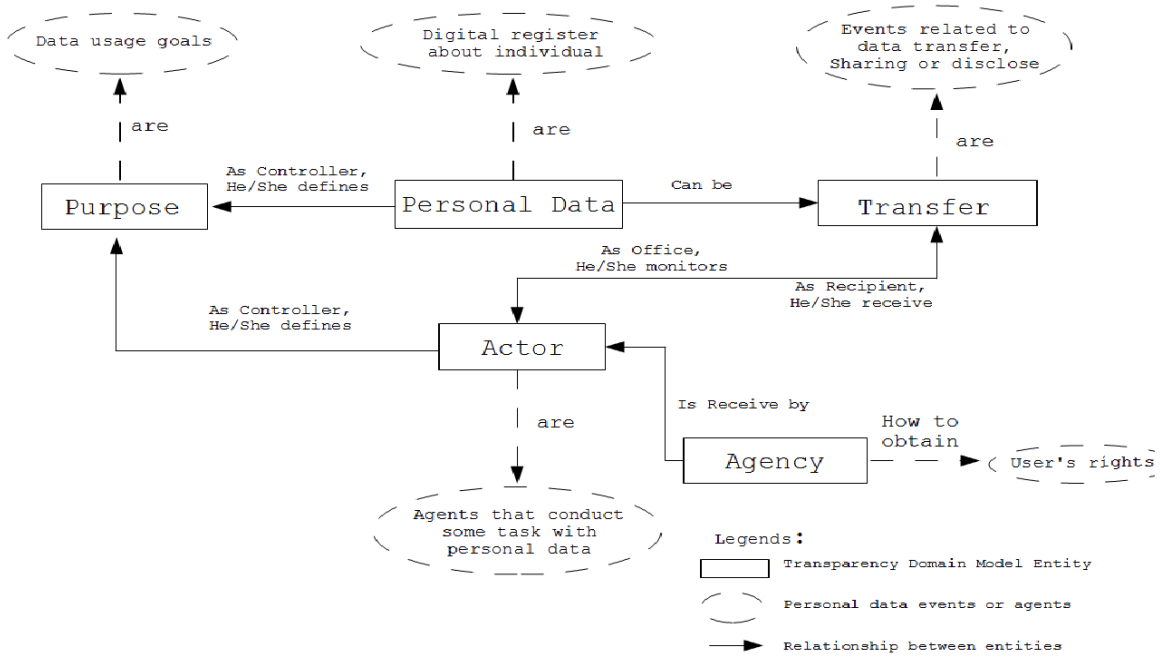

FIGURE 2. Personal data transparency domain model.

understandable for data subject in order to insert him/her in the flow of data usage.

To accomplish the functional requirement, we assume that TR-Model considers the following individuals' needs:

1) the individual must know what data is collected and how they are collected. Thus, a minimally necessary Transparency information to provide understanding about what Personal Data are collected and about Personal Data collection strategies. Also, data origin and data security preservation strategies must be considered;
2) Data subject must be informed about the Stakeholders involved in the data usage, their roles and how to access / identify them. This information is required in Personal Data regulations in order to ensure to provide knowledge about stakeholders identification to support any kind of individual analysis or action;
3) Data subject must be informed on the purpose of using the Personal Data and on the legal document that legitimates this usage;
4) Data subject must know the means to exercise their rights in case the usage turns to be against their interests or principles;
5) Data subject should be provided minimum processing and events information; and
6) Data subject should be informed on Personal Data transfers, sharing and / or disclosure.

Defined the Functional Requirement, the next step in building the TR-Model is to organize the entities, attributes and relationships to assemble the TR-Model's Domain Model. Domain Model is presented in the next subsection.

B. DOMAIN MODEL

The TR-Model domain was structured as presented in Figure 2.

The TR-Model domain was proposed based on GPDR's Articles 13 and 14 and respective sections and paragraphs. The domain model presents the follow entities:

Entity **Actors (Controllers, Processors, Recipients and Protection Office)**: this entity represents persons and legally established companies that participate in the production, collection, processing, sharing or using of Personal Data. Actors is an important entity because every other entity is related to at least one Actor. The focus of GDPR regarding to actors is related with identified/contact details for individuals. The actors are classified in GDPR as:

- Data Subject: Person who uses any device and who produces a Personal Data;
- Controller: A private or public person or company that determines the purposes of using the data and must comply with laws and regulations regarding the use of Personal Data;
- Processor: Private or public person or company that conduct events under controller control;
- Protection Office: An independent public entity established in a country or specific area that is responsible for oversee and enforcing regulations on the use of Personal Data; and
- Recipient: Person or company that receives any Personal Data for use or that is the subject of any claim, complaint or action of the individual. Recipient can also be any of other type of actors that receive a request of actions to ensure data subject rights and freedom.

Entity **Purpose**: This entity is aimed to support transparency information about the purpose of use for Personal Data. The purpose is always defined by the controller or by a group of controllers called as ''Jointly''. Information on purpose is relevant for transparency because it is the basis for legitimisation of collection and usage of personal data.

Entity **Personal Data**: This entity represents the description of the pieces of information that are collected from the data subject and used by Controllers. The Personal Data entity is justified by the fact that the variety and data of interest of the controllers is very large and the data owner may not be properly informed on what data need to be collected. Depending on what data are collected, the generation of information about the data subject can seriously affect their privacy, security and/or freedom.

Also, Personal Data Transparency Information allow data subjects to exercise their rights on canceling, restricting or denouncing the unwanted or misuse of their data. For such actions, it is necessary to have information about the collected data.

Entity **Transfer**: This entity is responsible to support Transparency about the data transference, sharing or disclosing. So, the Recipient information must be enforced, by a data protection regulation, that can be the GDPR or any local other one, to verify and ensure the correct use of the obtained data.

The transfer of Personal Data is a primary concern of data subjects and it is strongly related to privacy and data security. There are specific techniques, such as the use of log files, maps and graphs seek to show the path taken by the data after being produced and / or collected, but the information provided may not meet the data subject's needs or regulations for using Personal Data.

The GDPR is emphatic in affirming the need and concern to observe the legality of actions aimed at the disclosure and sharing of Personal Data. It requires that information on what data are shared and with which recipient(s), what is the legal basis for sharing and what is the justification of the transfer.

Entity **Agency**: This entity is responsible to support software tools for presenting information on how the data subject may exercise his or her rights to report irregular use of the data, cancellation and restriction of use and / or request for a copy of the data.

This entity of Transparency aims to make clear to the data subject how he/she can guarantee their rights to actions such as: Request access to the data, report an irregular use of data, restrict access to the data or others data subjects' rights.

Considering that each controller can provide different mechanisms to guarantee data subject' rights, a set of metadata focused on presenting the existing path(s) to carry out the action was proposed and will be presented in Table 7. The mechanism that the controller must present to allow data subjects to argue about their data use can be a website, a phone number, an email address or any other type of resource that allows the insertion of the complains.

With the presented entities, we assumed that TR-Model is able to provide a set of agent information and events related to the use of Personal Data.

As TR-Model was created to support software tools to provide Transparency, a set of descriptive metadata was also developed to explain how each entity should be used as well as its attributes and events. This description can be applied to both controllers at the time of filling Transparency data as the data subject can use them to understand the information transmitted.

Next section presents the Description Set Profile for TR-Model.

C. METADATA AND METAEVENTS DEFINITION

In this section we are presenting the definition of the metadata and metaevents for TR-Model's entities. These specifications were created to be used as a guideline for providing the Transparency Information Visualization.

The concept used to create metadata and metaevents was similar to object oriented modeling proposed by Pressman [50] that establishes a set of attributes and operations for each domain entity. In this approach, metadata are equivalent to attributes, describing the entities' characteristics and their meaning in the domain. Metaevents are then elementary actions that take place upon metadata in the context of an entity, and thus are similar to operations. Metaevents are also meaningful information for Data Subjects to understand how their data are used.

All TR-Model entities have metadata to describe their characteristics in the context of Personal Data usage. All TR-Model entities have elementary metaevents of creation, retrieval, update and deletion, however, they are not significant from the Data Subject perspective. Thus, only the entities Purpose, Personal Data and Transfer had their metaevents represented in the model.

The Domain model presented in the preceding sections was created according to the DCMI strategy for MAP creation. Because the TR-Model's main goal is to be a guide for development or software tool for Transparency with user-friendly and information quality content, other particular strategies that are considered common in a MAP creation were not adopted: (1) reusing attributes and definitions of other MAPs once we did not identified metadata from other MAPs that could be reused; and (2) create and specify usage syntax in order to support interoperability once TR-Model did not aimed to work with it.

TR-Model metadata and metaevents are presented in Figure 3.

The next subsection presents the metadata, metaevents and descriptions for TR-Model Entity.

D. METADATA/METAEVENTS TRANSPARENCY DESCRIPTION

The metadata and metaevents descriptions show how metadata or metaevents must be deployed in software tools to presented Transparency to Data Subject.

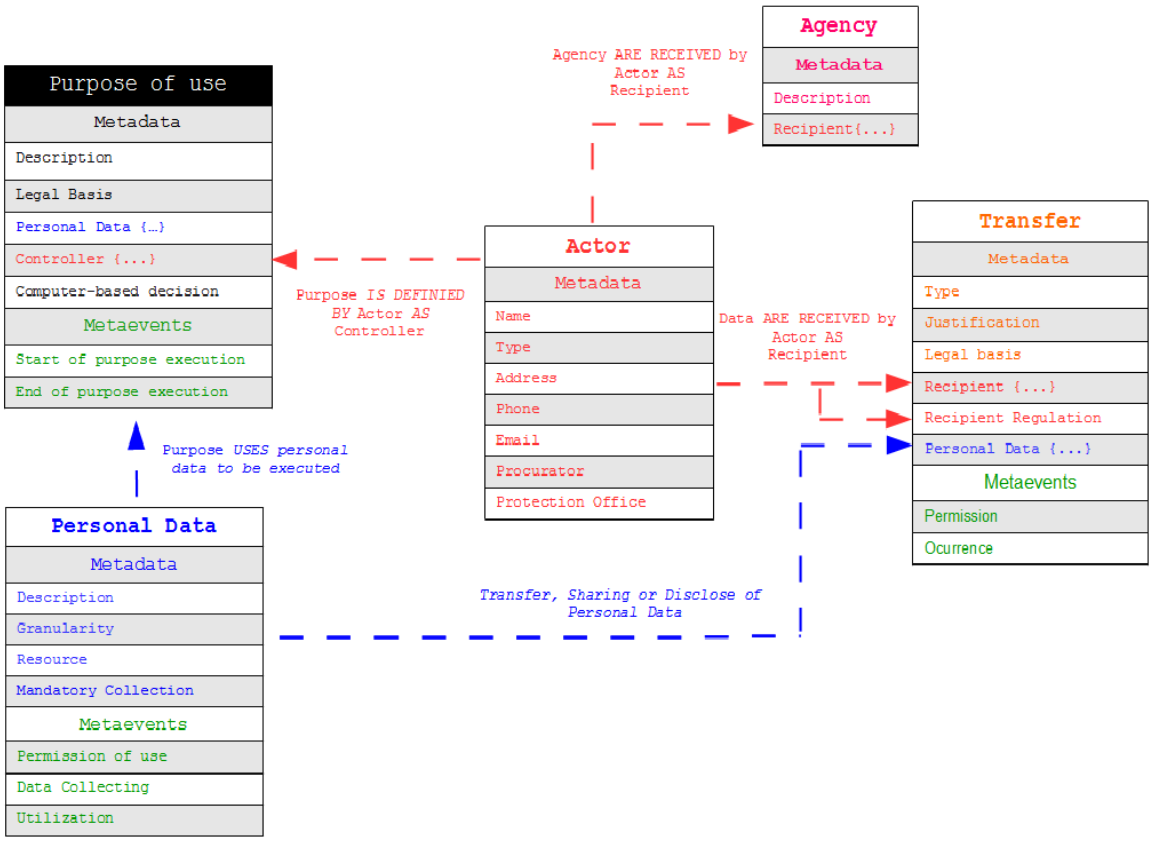

FIGURE 3. TR-model entities metadata and metaevents.

The descriptions were created based on HCI Features [51] and Information Quality Dimensions [52] as we considered techniques used to provide appropriated information visualization for the data subject. Concepts as Readability, Presentation Format and Information Quality were also applied.

The concept of Readability refers to the ability to produce more readable and understandable texts to the data subject and considers aspects such as text size, number of words and focus on the audience [53], [54]. It must be applied to information which is believed to be best presented in textual format. The readability features selected for Transparency were: (1) use short sentences to be more attractive to readers; (2) focus on the readers considering that the information's users are not experts in computing or data analysis; and (3) avoid complex words making the use of words with few syllables and that are popular for readers.

Also, the concept of InfoVis (Presentation Format) is the study/application of visual resources such as graphs, info-graphics or interactive resources for abstract data representation [55]. This approach was considered to propose a Transparency mechanism that could be better applied with visual elements, as they would represent the information in a more interactive and easy to assimilate manner.

The Information Quality (IQ) is a multidimensional concept that aims to support the development and delivery of appropriated information for data subjects [56]. Concern about IQ has grown due to the widespread use of the Internet by people who use information presented for their decision making [56]. In this way, people's decision is increasingly tied to their ability to understand, analyze and also in information accessed confidence [52]. According to [57] information with quality is the information that fits the concept of *fit for user*.

IQ is related to information delivery considering dimensions such as Contextual Quality (relevance, completeness and value added) and Representational Quality (ease of understanding, ability to interpret and concise representation) [58]. It is also considered for Transparency specification since data subjects access, analyzed and understand the information provided in order to analyze how his/her Personal Data is used (and by whom) and, thus, decide on any intervention. The dimensions considered for metadata and metaevents specification in TR-Model are shown in Table 2 based on [52], [58], [59].

In the following subsections we will present the Metadata and Metaevents Description Set for each one of the Personal Data Transparency Entity in TR-Model.

V. METADATA, METAEVENTS AND DESCRIPTIONS FOR ENTITIES

In this section, Metadata and metaevents list is presented as well as it descriptions.

TABLE 2. IQ dimensions for TR-Model. Adapted from [52], [58], [59].

Dimension	Description
Understandability	Transparency should provide simple, clear and understandable information.
Objectivity	Transparency should be presented for Data Subject as a unbiased, unprejudiced and impartial information.
Relevance	Transparency should be relevant for the Data Subject to support the analysis of his/her Personal Data usage.
Completeness	Transparency information must be complete. Data Subject must access all necessary information without having to search for information elsewhere.

A. METADATA AND DESCRIPTION FOR ACTORS

The definition of attributes of this entity was strongly based on the GDPR section 1 paragraphs a, b, and f for both articles 13 and 14. These sections require that systems must present a list of items for the actors identification. Among them must be a contact information. Considering that the term ''contact data'' allows a certain subjectivity, we decided to establish some metadata/description that allows the data owner to locate the actor by maps, phones or internet.

Actors are all individuals or organizations involved in the Personal Data life cycle. The people involved can present at least the following responsibilities: determining data usage rules, performing processing, sharing data, receiving shared data, and monitoring / supervising actions.

The actors metadata and description are presented in Table 3.

TABLE 3. Metadata description set for actors entity.

Metadata	
Name	**Transparency Information Description**
Actor's Name	An identifiable name.
Type	Specify a single task that the actor can perform using the Personal Data. It may provide one of the following data: (1) Controller; (2) Protection Office; and (3) Recipient
Address	List of following items: (1) Street Name; (2) Number; (3) City; (4) State/Province; (5) Postal Code; and (6) Country
Phone	Phone Country Code + Phone Number
Email	Internet mailbox associated with exactly one.
Procurator	Presents the name of a person that answers for control management. This field is mandatory for Controllers or Recipients.
Protection Office	An instance of entity Actor with Metadata Type set as: Protection Office. This field may not be filled if the actor is a "Protection Office".

The next subsection displays the metadata description for Personal Data.

B. METADATA, METAEVENTS AND DESCRIPTION FOR PERSONAL DATA

The Personal Data metadata was developed to present information about which data was collected, how it was collected, and how it would be stored / protected. This entity aims to comply with the Transparency principles discussed by Bellamy and Alonso [10] that states that the use of Personal Data should be open to the Data Subject.

The Personal Data metadata and metaevents aimed to specify issues related to which Personal Data are collected. It includes data structure, collecting features and devices used to collected data. Three events describe since the moment when data subject grant the use of Personal Data until the tasks conducted with a data. The Personal Data metadata are described and specified in Table 4.

We assumed that, with the metadata and metaevents presented, the data subject will be able to know and understand which of their data is collected, how it is collected and how it is used. This information can be used by the data producer for better understanding of interactions made by the software such as: suggesting a product or service, anticipating a person's action or presenting certain knowledge about people's actions.

In the next subsection the Purpose usage metadata and metaevents will be presented.

C. METADATA AND METAEVENTS AND DESCRIPTIONS FOR PURPOSE OF USE

Purpose-of-use metadata and metaevents intend to support software to provide for Data Subjects Transparency information to answer one of the most common questions regarding to the use of Personal Data: ''*What do you want my Personal Data for*?''. The entity's metadata is focused on the reason of use, responsible for the use and the legality. The data subject will also be able to understand what data will be used for the purpose and whether any decision will be made solely by computer.

A relation with Actors entity provide for Purpose of use information about the actor as a Controller. The Purpose of use metadata named as *Controller* has information that allow Data Subject to access contact data about who define how Personal Data may be used.

Regarding to metaevents, it support to understand the purpose of use by providing information about the beginning and the end (period of time or event trigger) of the purpose's execution. Thus, the Data Subject can identify whether the execution of the purpose of use (and the consequent use of his data) is momentary, temporary or while he is using an application or service.

And finally, knowing about the purpose of use for Personal Data, we believe that the data subject can develop more confidence in the data controller organization, besides having information that can support an eventual complaint or denunciation of abuse in the use of the data.

The Purpose of use metadata and its description set is presented in Table 5.

The next subsection will display the Transfer/Sharing/Disclose metadata and metaevents

D. METADATA, METAEVENTS AND DESCRIPTIONS FOR TRANSFER/WWWWW/SHARING/WWWWW/DISCLOSE PERSONAL DATA

Information about how Personal Data are disclose, sharing or transfer were identified as the most worrying Transparency

TABLE 4. Metadata, Metaevents and descriptions for personal data entity.

Metadata	
Name	**Transparency Information Description**
Description	Identificable and short description given to Personal Data. **Example**: *(1) Location; (2) Credit Card purchase value; or (3) Performance of a physical activity*.
Granularity	A combination of items separated by "+" combined with the sentence: "Personal Data {Personal Data Name} is composed by: {granularity/combination of items}. **Example:** *Personal Data Location is composed by: Latitude + Longitude + Date + Time*.
Resource	If the Personal Data is collected by the controller that will use the data directly from data subject, the system may present a minimal amount of sentences describing the resources used to create the data items. If the resource is unusual for people, the system should supplement the information with a brief description of the resource. **Example:** *The latitude and longitude data are registered by smartphone's GPS system. GSP is an internal resource of hardware that cannot be visualized (just accessed via software) by users*.
Mandatory Collection	It must present whether the data collection is mandatory in the related purpose of usage. This information must be followed by a justification sentence case the mandatory value is YES. Case the information is NO, the software can provide a text describing the consequences of the lack of data collecting.
Metaevents	
Permission of use	This event must consider the moment or the action which the data owner granted the use of his/her data. To present it for data subjects, the system must use a set of images presenting the interface used to give the consent in order to remember/explain to user about the moment when he/she allowed (or will allow) the use of his/her Personal Data. The images can be followed by the text such as: *View when/how you did consent the use of your data* or *View when/how you will allow the use of your data*
Utilization	This metaevent may provide Transparency about how the data will be used and what information about the data subject can be produced through processing (or what questions about the data subject will be answered). The description of the information must be in a language in which the complexity of the use of data is abstracted for the data subject. The description of the Personal Data should be based on the structure: *The Personal Data will be used as follows {{describe form of use }* and the produced information or answered questions are: {list of information produced or answered questions about the person}. Cases with several different usage of the Personal Data, the system should repeat the specifications for each use.

TABLE 5. Metadata, Metaevents and descriptions for purpose of use entity.

Metadata	
Description	Transparency Specification
Purpose description	Must describe the purpose-of-use's name and the the legitimate interests pursued by the controller or by a third party. To do this, a single sentence specifying the puporse (or purposes) of use as followec "The purpose of use for your Personal Data is {purpose description}". **Example:** *The purpose for your Personal Data is to know your top destinations and common routes during the week to offer the best routes* or *The purpose of use of your Personal Data is to know your preferences*.
Legal Basis	Presents the regulation that guarantees the use of the Personal Data for the specific purpose. Information about the law/regulation must be as detailed as possible (Article, Paragraph, Topic, Recital, etc.) to support the data subject to locate and confirm it content. **Example:** *The use of your Personal Data is to know your top destinations during the week to offer the best routes is according to General Data Protection Regulation (GDPR), Article number 89 item A available in https://gdpr-info.eu/art-89-gdpr/*
Personal Data	This metadata must present a sample of Personal Data entity/metadata/metaevents for each Personal Data used. Transparency metadata specifications must be followed by the Personal Data Metadata/Metaevent Specification.
Controller	This metadata must present a sample of the Actor entity/metadata which the metadata named Type has the value *Controller*. Transparency metadata specifications must follow the Actor entity metadata specifications.
Computer-based Decision	It must highlight whether any kind of decision is made just based on algorithms/computer results without human supervision. The event must also provide a justification sentence if the computer-based decision value is YES.
Metaevents	
Start of purpose execution	This event must present information about the moment or trigger when the data usage start. A moment refers to a specific date, time or period, for example: *Start of use purpose execution: 01/01/2020 at 00:00*. An action is a system or user interaction. **For example**: *Start of use purpose execution: The agreement with the Privacy and Security Policy.*
End of purpose execution	This event must present information about the moment or trigger when data usage ends. A moment refers to a specific date, time or period, for example: *End of use purpose execution: 31/01/2020 at 00:00*. An action is a system or user interaction. **For example**: *End of use purpose execution: When to uninstall the app.*

users' concern. Due to this concern we decided to create a set of Transparency information that could support the knowledge about which data are being disclose and who are accessing Personal Data besides the app/website controller.

The information about data are provided by Personal Data entity and the relation with Transfer entity. Who access Personal Data are presented through two relations between entities Transfer and Actors. Two different actors are involved in the Personal Data disclose: Recipient: Other controller that receive Personal Data; and Recipient regulation: Protection Office that supervise the Recipient.

As complementary information for identification of irregularities in data disclosure, we proposed a metadata set related to laws/regulation description as well as metadata about how and who are regulating the data usage in the recipient context. The proposed metadata are in compliance with GDPR Article 13 and 14 Item 1(f) and with LGPD's V that discuss international data sharing/transfer.

Regarding to metaevents, TR-Model proposes two information related to transfer events: (1) the moment when the Data Subject authorize the data transfer, disclosure or sharing; and (2) which occurrences trigger the distribution of the data, once a controller can send data in different context of use in different interval or packages.

Based in this needs the Transfer metadata and description were defined as presented in the table 6.

TABLE 6. Metadata, Metaevents and descriptions for personal data transfer entity.

Metadata	
Description	Transparency Information Description
Title	Short text to be used as a title for Transfer, Share or Disclosure action. **For example:** *Improving your health*
Type	Specify which type of transfer is done following the list: **(1) Transfer** :(Personal Data are transferred to third parties. The third controller may use the data for different purposes even it is unrelated to the purpose of use presented by the controller who collected the data.; **(2) Sharing**: Personal Data are transferred to a recipient who will work with the controller to improve, supplement, supply or otherwise need related to the purpose of use presented by the controller who collected the data.; or (3) **Disclose**: The controller makes the Personal Data open for public access.
Justification	Sentence that describes the reason to support Personal Data transfer. The sentence may follow the structure: The Personal Data {name of Personal Data} is/are {transfer/shared/disclose} due to the reason {transfer reason}. **Example:** *The Personal Data: Location is disclose for policy and other authorities (if required) to support any kind of investigation*. The sentence can reference all data as long as the action is done this way. **For example:** *The transfer of Personal Data cited here is due to computing resource sharing*.
Legal basis	A sentence reaffirming the name of the Personal Data and its objective followed by information of the law / regulation that guarantees the Transfer, Share or Disclose of the Personal Data is legal. Law/regulation information desired: (1) - Name of the law/regulation; (2) - Number of the article or section; and (3) - Name/number of specific item. **Example:** *The Disclose of Personal Data Location is according to General Data Protection Regulation (GDPR), Article number 48. https://gdpr-info.eu/art-48-gdpr/*
Recipient	This metadata must present a sample of the Actor entity/metadata which the metadata named Type has the value *Recipient*. Transparency metadata specifications must follow the Actor entity metadata specifications.
Recipient regulation	This metadata should present information about how Personal Data will be protected at the data recipient. A combination of two pieces of information may be displayed: (1) Recipient Personal Data law/regulation: State which law/regulation the recipient responds to; (2) Data Protection Office at Recipient: An instance of Actor entity with data with metadata type as *Protection Office*.
Personal Data	It must provide a sample Personal Data entity metadata for each Personal Data used. Transparency guidelines should follow the Personal Data entity metadata descriptions.
Metaevents	
Permission	This event must consider the moment or the action which the Data Subject authorized the transfer of his/her Personal Data. In order to delivery the information for Data Subject, the system must use a set of images presenting the interface when he/she give the consent in or to remember/explain to data subject about the moment when he/she allowed (or will allow) the transfer of his/her Personal Data. The images can be followed by the text such as: *View when/how you did consent the transfer of your data* or *View when/how you will allow the sharing of your data*. Images may also have highlights for texts and interface components read and accessed by Data Subject.
Ocurrence	This metaevent should provide information about how Personal Data are transfer/ sharing/ disclose. Personal Data can be transferred, shared or disclosed over a period of time or based on some trigger. Structure recommended: *Personal Data {can present data description} will be { transfer/shared/ disclosed} { data dissemination strategy}*. **For example:** *Personal Data will be transferred every time you use the app* or *A copy of the Personal Data will be shared every thirty minutes*.

It is also important to highlight the metadata named Type that shows which type of data distribution action with third parties. The definitions of each word were created after analyze how Personal Data could be used by several controllers in partnership and searching in dictionary which would be most appropriate to represent it.

This metadata can help the data subject to identify whether data distribution is beneficial to the him/her, because in partnership with other controllers, the primary controller can improve the quality of information to achieve the purpose which Personal Data was collected. The Type metadata can also support the identification of actions that is not in compliance with data subject authorization since the data recipient could use data for some action for which the data subject has no interest.

Thus, TR-Model Transfer entity proposed Transparency that may allow the data subject to identify the destination of their Personal Data as well as the legality of the actions. With this knowledge, he/she can make decisions about continuing to use the service or application and consequently having your data distributed to third parties.

Next section presents the metadata for Action and Negotiation (Agency).

E. ACTIONS AND NEGOTIATION METADATA (AGENCY)

To provide Agency metadata, the followed features must be considered: 1) This entity's metadata should work as a tutorial/informative-style to guide Data Subject how he/she may acts to intervene in the use of their Personal Data; (2) Each controller can provide different mechanisms to ensure data subjects' rights; and (3) TR-Model is not aimed to make any kind of pattern for these mechanisms.

The name *Agency* refers to any action or negation that can (may) be performed by Data Subjects to ensure their rights in the Personal Data usage.

To meet these factors we assumed that this entity would need just three metadata: (1) Agency title; (2) The recipient of the action (Who will receive the message, report or request) that may be an actor with type as Controller, Processor or Protection Office; and (3) A metadata with the flexibility specification to allow the controller to present several kinds of information on how the Data Subject may acts. Thus, the tutorial information presented can be according to Controller preferences since it guide the Data Subjects in their actions.

The metadata for agency are shown in the Table 7.

Next section presents the TR-Model validation.

VI. VALIDATION

This section presents the TR-Model validation carried to verify whether it could support software development in order to provide Transparency information focused on Data Subjects. TR-Model must support software tools to assist in the visualization, understanding and analysis of the Personal

TABLE 7. Metadata for agency.

Label	Transparency Information Description
Agency's Title	Simple sentence describing the agency's name. **For example**: *(1) Request a copy of Personal Data; or (2) Report Incorrect use of your data.*
Description	This metadata should provide information the tasks that Data Subject may do in order to ensure their rights. The application can display a list of events or set of images or an email address. It is not the obligation of the Transparency tool to provide the resource to the data subject, but to guide him/her to the interface or path provided by Controller or Protection Office website.
Recipient	This metadata must present a sample of the Actor entity/metadata which the metadata named Type has the value *Recipient*. Transparency metadata specifications must follow the Actor entity metadata specifications.

Data usage by Data Subject. Also, it should support the compliance with the Personal Data regulations such as GDPR.

Due to the fact that TR-Model cover different aspects, we decided to perform the validation based on the following approaches:

1) **Transparency coverage:** Personal Data Transparency is one the main request provided by GDPR and LGPD in order to ensure the correct Personal Data usage. Considering this request, we decided to verify whether the proposed information about data usage were in compliance with these regulations. The choice of GDPR was due to the fact that it is a regulation already in force and with its well-defined texts. The LGPD was chosen because the origin of this research takes place in Brazil and there is the expectation of use of this model in the country;
2) **User Evaluation:** TR-Model was created to support Transparency information for the data subject usage. It aimed to abstract the complexity of events involved in the use of Personal Data in user-friendly and relevant information. Therefore, testing with data subject was necessary to verify the effectiveness of information from the user's point of view. Validation with participants was performed using a combination of controlled tests and questionnaires resolution. The data collected were analyzed on two approaches: (1) quality of HCI elements; and (2) IQ dimensions.

The next subsection presents the coverage validation of the activities performed during this research.

A. INFORMATION COVERAGE VALIDATION

In this validation task we analyzed whether TR-Model covered Transparency requirements proposed by regulations are attended with its metadata and metaevents. For GDPR, we considered the Articles 13, 14 and 15 because these articles present the data subject' right regarding to access to information about the use of their Personal Data. For LGDP we did not considered a specific article or chapter, because

TABLE 8. Relation between GDPR transparency requirements and TR-Model Metadata/Metaevents.

GDPR Article, Section and Item	TR-Model Entity, Metadata or Metaevent Information
Art. 13 and 14/1/a.	Actor Entity metadata.
Art. 13 and 14/1/b.	Actor Entity metadata.
Art. 13 and 14/1/c.	Description e Legal Basis metadata of *Purpose of use* Entity.
Art. 14/1/d.	Description and Granularity metadata of *Personal Data* Entity.
Art. 13/1/d and Art. 14/2/b	The combination of *Purpose of use* and *Personal Data* Entities can provide more Transparency details.
Art. 13/1/e and Art. 14/1/d .	Metadata Controller or Recipient provided in *Personal Data* and *Transfer* Entities. *Actors* metadata complement the information.
Art. 13 and 14/1/f.	Recipient Regulation metadata in the *Transfer* Entity.
Art. 13 and 14/2/a.	Utilization metaevent in *Personal Data* Entity.
Art. 13/2/b and Art. 14/2/c .	*Agency* entity.
Art. 13/2/e	Mandatory Collection metadata in *Personal Data* Entity.
Art. 13/2/f and Art. 14/2/g .	Computer-based decision metadata in *Purpose of use* Entity.
Art. 14/2/f	Resource metada in *Personal Data* Entity.

Transparency is not discussed in a specific article as GPDR, but the Transparency requirements are discussed during in text topics that need to address Transparency.

The inspection was performed listing the Transparency requirements presented by the regulations and for each requirement (or group of requirements) it was indicated how the information could be made available by TR-Model entities, metadata and metaevents.

The GDPR analysis is presented in Table 8. This table presents the Articles, section and items that request Transparency information and the related contribution made by TR-Model.

Important to highlight that, there are TR-Model's contribution that are applied in more than one Transparency request.

The LGPD analysis were consucted similarly and is presented in Table 9.

TABLE 9. Relation between LGPD transparency requirements and TR-Model Metadata/Metaevents.

LGPD Chapter, Section, Art. and Item	TR-Model Entity, Metadata or Metaevent Information
2/1/9/IV.	Controller metadata of *Puporse of use* Entity and Recipient/Protection Office metadata of *Transfer* Entity.
3/NA/18/I - IX	*Agency* Entity, *Transfer* Entity and Permission of use metaevent in *Personal Data* Entity
5/NA/34/III	*Transfer* Entity.
2/3/14/2	*Personal Data* Entity and *Agency* Entity.

The information presented in Tables 8 and 9 allow to conclude that TR-Model provide compliance with GDPR and LGPD in the feature: *"what information to provide as Transparency.* In fact, these regulations were the main requirement

TABLE 10. TR-Model validation scenario.

N	Description
1	You are starting running, walking, of performing any other gym training activities. Following the instructions provided by your Personal Trainer, you download an application to register your activities. Before installing, you are warned that your Personal Data will be collected and used.**Considering this information, you should access the Transparency interface provided by the application and analyze the purpose of using your Personal Data and which Personal Data will be used.**
2	Suppose you already know the application usage of your Personal Data. However, you want to find out if the application shares or discloses your data for other companies. So you decide to: **check whether your data will be shared, who is supposed to have access to this data, and how you should report any data misuse**
3	You are inputting a specific information about your training activities. A feedback message displays that your Personal Data will be used. You decide to: **check for what purpose your Personal Data will be used and verify whether it in compliance with the previously authorization.**
4	A medical clinic sends you an email offering their services. You notice that the services are related to your physical activities and you suspect the clinic may have accessed information about your activities from the application. You then decide to: **verify if you have authorized the application to share your Personal Data for that particular purpose and the moment or event in which that permission occurred. In addition, you check when your Personal Data was collected.**
5	You have identified an unauthorized usage of your data. Thus, **you decide to contact the Controller, who is responsible for the use of the data to negotiate the situation. Also, you decide to report the fact to the Data Protection Agency**;

artifact used to support the construction of TR-Model, and, in a certain way, we expected that it suitability were achieved.

Although regulations have some differences in Transparency requirements such as: (1) greater emphasis on one Transparency requirement in one regulation than in another; or (2) certain Transparency information may be required with more details in one regulation than another, all the requirements are accomplished by the TR-Model.

Next subsections presents the validations conducted with users participation.

B. TR-MODEL VALIDATION WITH USERS

In order to verify whether the TR-Model guidelines are effective in guiding the production of satisfactory Transparency information for data subjects, we have performed some user tests. User tests were based on a physical-activity data-collection application, which collects and processes data about running, walking and gym activities. Using this application, we developed five scenarios, in which the user was presented to usual situations of Personal Data Transparency relevance. In this application, we simulated information using TR-Model metadata and metaevents specifications.

Additional to user testing, we applied attitudinal verification through surveys. Results are presented using descriptive statistics and qualitative analysis.

We designed a questionnaire to gather each participant's opinion about the quality of transparency presented by the TR-Model. The validation of this model was based on the Qualitative Approach [60] in order to identify each participant's experience with Transparency and also to identify issues that could be improved with the TR-Model. The instrument is presented in Table 11. The participant answered all questions for each scenario.

Closed questions were used to assess interface elements features as Readability and presentation format, and Information Quality features as discussed in the previous sections.

Some questions depended on previous answers. For example: the question *If in the previous question you answered that you could not understand Transparency or understood only part of the information.* was presented if the answer to the question *On Regarding the Ease of Understanding and Interpretation of Transparency of the evaluated scenario, what is your opinion?* was: Bad or Fair.

Those dependent questions were descriptive and aimed to collect participants' suggestions or justification regarding answers of other questions. The answers of these questions provided insights for corrections and/or improvements in TR-Model.

C. VALIDATION PARTICIPANTS PROFILE

A total of 121 (one hundred and twenty-one) participants participated in the validation tests. The demographic profile of the participants is shown in Table 12.

Participants were invited to participate voluntarily in the activity. Invitations were distributed via email and social networks. Before starting the testing, they were advised that the purpose of validation was the TR-Model and that their performance was not being evaluated at all. They were also informed that no personal data would be collected, and their actions and opinions would be used solely and exclusively for the purpose of this research.

Also, the participants were told that they could give up anytime if they wanted to or felt embarrassed. They received researchers' phones and contact emails for any questions or complaints. Tests only started after the participant agreed to participate.

We assumed that the participants' profile represented a significant sample of users of websites and mobile applications. Limitations due to the sample were due to a larger number of participants from the STEM areas due to the fact that many participants were students, teachers and / or employees of educational institutions in the area of computing, mathematics and physics. Also, we had no participants over 50 years old.

Before the evaluation, participants answered a questionnaire intended to assess their previous experience with this subject, their awareness about the use of their Personal Data by applications, and the level of concern associated.

Regarding to prior knowledge about the use of Personal Data by applications, the result was almost unanimous: 97 % of the participants indicated previous knowledge about the subject and only 3 % (4 participants) answered that they did not know about the use of Personal Data. For identification of

TABLE 11. Questions and alternative answers in validation questionnaire.

Question	Alternatives
Q1 - Use the range of 1 - Very Bad to 5 - Very Good. How do you rate the ease of finding and/or identifying information about the use of your Personal Data?	Value range
Q2 - In your opinion, the amount of information presented in the Transparency of the evaluated scenario. Your opinion is:	(a) Excessive; (b) Insufficient; or (c)Appropriate
Q2a - If your answer for the amount of information was Excessive or Insufficient, please describe what information should be added or removed.	Open answer.
Q3 - Regarding the ease of understanding and interpretation aspects of the evaluated scenario, what is your opinion?	(a) Bad; (b) Regular; (c) Appropriate.
Q3a - If in the previous question you answered Bad or Regular, please describe the reason why you could not understand Transparency or understood only part of the information.	Open answer.
Q4 - Regarding the objectivity of Personal Data Transparency in the evaluated scenario, what is your opinion?	(a) Objective or (b) Not objective
Q4a - If you considered Transparency information in the evaluated scenario as NOT OBJECTIVE, please explain or exemplify what led you to this opinion.	Open answer
Q5 - Regarding the completeness of Transparency in the evaluated scenario, what is your opinion?	(a) The information is complete or (b) There is not sufficient information.
Q5a - If the completeness of Transparency information in the scenario is poor, please inform what you think could be added to make this Transparency information complete	Open answer.
Q6 - Some data are better visualized by text, others by images (photos, videos, etc.), colors, formats, etc. Regarding the format in which the Transparency information is displayed (texts, colors, images etc.) what is your opinion?	(a) Appropriate or (b) Inadequate ou (c) Appropriate, but need improvements
Q6a - If you classified the display format of any information Inadequate, please describe what format would improve the visualization.	Open answer
Q8 - In a range of 1 - Not Relevant to 5 - Most Relevant, what is your rate about the relevance of the Transparency information of your Personal Data usage?	Value range

TABLE 12. Participants' profile.

Profile	Options	Results (%)
Age	Under 18 y-o	3.3
	18 - 35 y-o	95.0
	36 - 50 y-o	1.7
	+50 y-o	0.0
Gender	Male	71.1
	Female	28.9
	Not informed	0.0
Educational Level	Elementary school	0.0
	High School	4.1
	University Education	95.1
	Postgraduate studies	0.8
Ocupation	STEM	82.6
	Social and Human Sciences	5.8
	Communication and Information	1.7
	Environmental Sciences	4.,1
	Others	5.8
App download	Rarely	15.7
	Frequently	84.3

participants in this section we will name participants who had prior knowledge as **Profile A** and those who did not know as **Profile B**.

Profile B participants also showed little or no concern about mechanisms that could provide Transparency. These participants answered that they never read the PSP; 75 % of them never worried about the use of their Personal Data and 25 % had only some concern. This behavior was expected because if they ignore the use of their data, there is no reason to be concerned and no reason to read PSPs in order to find information about this. We expect concern to rise after knowledge.

For Profile A participants, the results presented subtle change as 64 % of participants also answered that they had never read the PSP and 36 % who had done some reading. Similar to Profile B, no participants indicated that they read PPS frequently.

The Profile A answers regarding to the considerations about the use of Personal Data are presented in Figure 4.

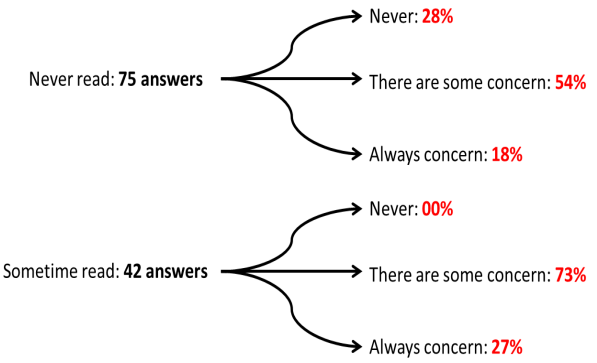

FIGURE 4. Responses related to concern with the use of personal data.

Another question asked the participants was: *how often they seek information about the use of their Personal Data*. This question will be discussed only for the respondents of Profile A, because the others naturally answered that they never seek such information. The answers to this question were: (1) Never: 34.2 %; (2) Rarely: 54.7%; Often: 10.3 %; and Always: 0.8 %.

For participants who responded that *Never* or *Rarely* search for information, they were asked to describe why they provided these answers. The essay question was answered

by 104 respondents who presented answers that included elements such as:

- Lack of interest in information;
- Lack of time, for one must read a lot of information;
- Difficulty in identifying and / or finding information;
- Lack of choice, since denying access to data implies in no access to the service;
- Confidence in application providers;
- Little knowledge about the subject; and
- Long, technical, legal and subjective texts;

For participants that answered that they do search for information about the use of Personal Data, they were asked to describe the strengths (S) and weaknesses (W) of the resources / information provided by the applications. Thirteen participants marked one of these options and their answers described the issues such as:

- (S) Targeting content of interest;
- (S) Information about collected data;
- (S) Existence of information about the use of Personal Data;
- (W) Information is not restrictive and gives scope for interpretation;
- (W) Inaccurate information on how and for what data will be used;
- (W) Too much content to review; and
- (W) Complexity of terms and words used to present information.

The presented information corroborates a previous discussion in which the lack of a *user-friendly* pattern as well as the use of inaccessible means make Transparency information difficult to be accessed, found or understood by the data subject. Consequent data subjects' detachment in verifying the use of Personal Data leaves a clear path for companies to promote a *black box* approach in which either the user accepts the company's terms or the user does not use the application.

When companies provide information, the content is too complex to be analyzed. Information can be written to provide a margin of interpretation favorable to the company and to hinder possible actions by data subjects.

Thus, even when the participant has a concerned profile about the use of Personal Data and searches for information about it, problems related to information quality affect the content and lead to serious difficulties for data subjects. Although much of the use of data may be positive and beneficial to the data subject, difficulties reported by the data subject can create ways for the misuse.

Finally, we concluded that people are not uninterested or silent about the use of Personal Data, but instead, the way information is disposed leads to lenient behaviors as people usually respond better to quick, objective, simplified, simple-language content.

The next subsection analyzes participants' expectations regarding the Transparency information *versus* the information proposed by the TR-Model.

D. PARTICIPANTS' TRANSPARENCY EXPECTATIONS VERSUS TR-MODEL INFORMATION

A mandatory question was proposed to participants, which aimed to verify whether TR-Model was according their expectations regarding Transparency information: *If you could choose any information to require in order to learn about the usage of your Personal Data, which information would you require?*. The question required an essay answer.

The 119 responses received were analyzed using textual analysis and grouped in 14 categories that express users' expectations. Results are shown in Figure 5.

This chart shows that the participants' biggest concerns were *what data is used*, *by whom they are accessed* and *purpose of using the data*. Some concerns has also been identified such as *how the data will be used* which we assumed that would be a more technical information, for example, that the data are compared or summarized or classified by a specific algorithm.

The expectation on *who will access the data* was also evident. Among it, concerns about sharing or disclosing data among several organizations were considered by participants. They are more focused on the knowledge about who will receive the data (or whom the data will be shared with) than the concern about its usage by the Controller who collected the data.

More technical expectations such as *resources used to collect data*, *frequency of collection*, *safety information*, *storage* and details on *how to cancel the permission to use data* were presented, but in lower frequency. Security and storage information may be suggestions from people with computer knowledge, once this information is technical in nature. Expectations for information on how to cancel, change or restrict the use of data may be the need for those knowledgeable about the use of Personal Data and who are concerned about its consequences, as in the event of disagreement or disagreement they could cancel. permission to use the data.

Regarding the analysis of the suitability of the TR-Model to the participants' expectations, we concluded that the model supported directly 12 expectations, that is, participants' expectations were described in the metadata and metaevents description in the entities. The following expectations were not met or were not directly addressed:

- *List of processes conducted in the Personal Data*: with just one indication, we assumed that the participant expected technical information about Personal Data processing history, something close to the concept of Data Provenance. Considering that this is a requirement that needs a specific support model, that information was not included in the present release of TR-Model;
- *Information about storage and data security*: Although we considered including this information in TR-Model during its development, it was dropped because we assumed it to be too technical, and possibly might not be relevant for data subjects. However, the availability of this information will be considered in future

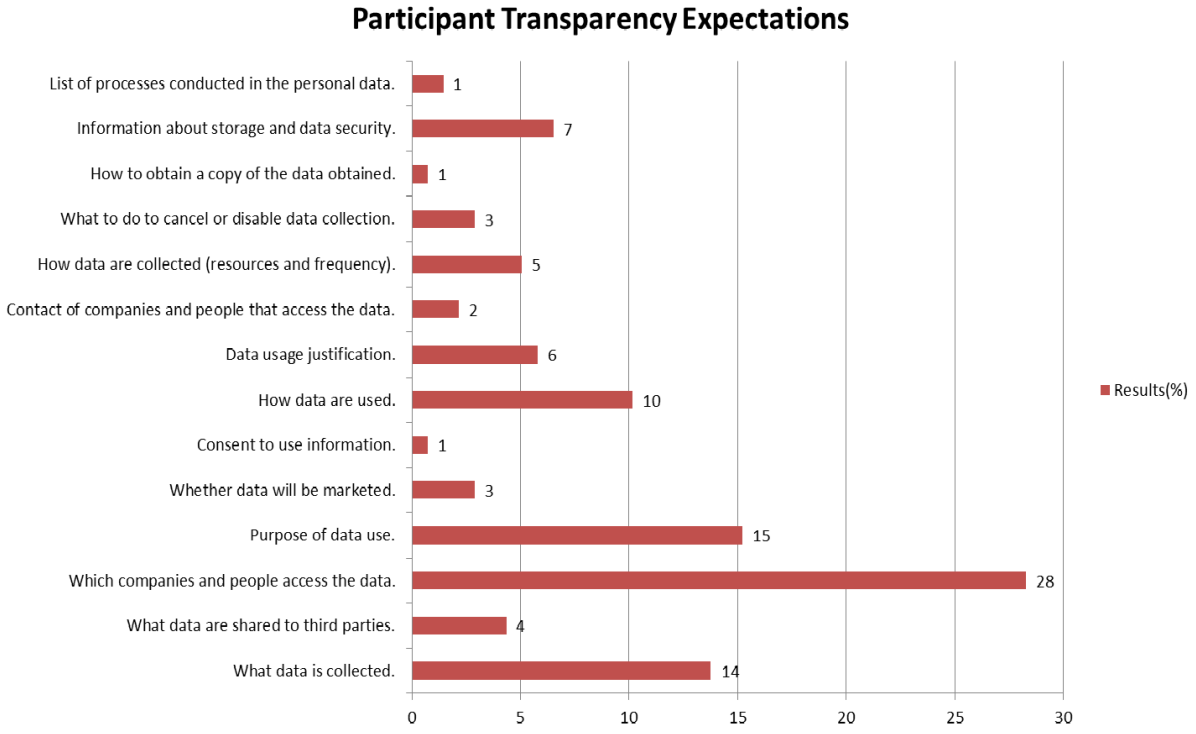

FIGURE 5. Participants' expectations about transparency.

releases of TR-Model once is necessary more studies to abstract storage and security technique in user-friendly information;

- *Usage justification*: This information is included in the *Transfer* entity, but has not been entered in the *Purpose of use* entity. In *Purpose of use* entity, this information can be suited in *Description* metadata.

Therefore, considering the information provided by the participants, results allow us to conclude that the TR-Model usage provides 85% of the information expected by this sample of data subject. For information that presented higher frequency of expectations, TR-Model was effective in meeting 100% of participants' expectations in its entities, metadata and metaevents.

This positive result is due to the participation of data subjects (users) in the entire TR-Model development cycle. The engagement of users not only provided good user experience but an efficient set of specifications and metadata.

The next subsection presents the Transparency user experience and information quality evaluation.

E. USER EXPERIENCE AND INFORMATION QUALITY EVALUATION

This subsection presents the data analysis and results of the questionnaires answered by participants during the evaluation related to the Transparency *User Experience* and *Information Quality*.

Each Transparency scenarios were evaluated considering HCI features such as Readability, Infovis and also considering IQ elements. The analysis and discussions will be present first by HCI features and after, by IQ elements. For the discussions we analyzed the results using descriptive statistics to provide information considering all the evaluated scenarios. After, we discussed eventual situations that drew attention for their results. Also, we discussed the the analysis of essay questions answered to justify or describe a specific answers in a question.

Participants conducted 320 individual evaluations in all available scenarios. The amount of evaluation for each scenario is:

- Scenerio 01: 71 evaluations;
- Scenario 02: 68 evaluations;
- Scenario 03: 62 evaluations;
- Scenario 04: 60 evaluations; and
- Scenario 05: 59 evaluations.

Next, we discuss HCI features analysis.

1) HCI FEATURES ANALYSIS

The HCI features' evaluation aimed to verify the effectiveness of metadata descriptions that focused on the interface *design*. Although TR-Model was not designed to be an interface *pattern*, the Transparency metadata and metaevents also describe some HCI and Infovis specification to produce appropriate Transparency. In a certain way, TR-Model tried to support the *friendly* information *design* to data owners.

The questions used to collect the data for this analysis were: Q1, Q2, Q6, Q2a and Q6a presented in Table 11. The Figure 6 presents the graph with percentages for the answers to these questions, considering all scenarios evaluated.

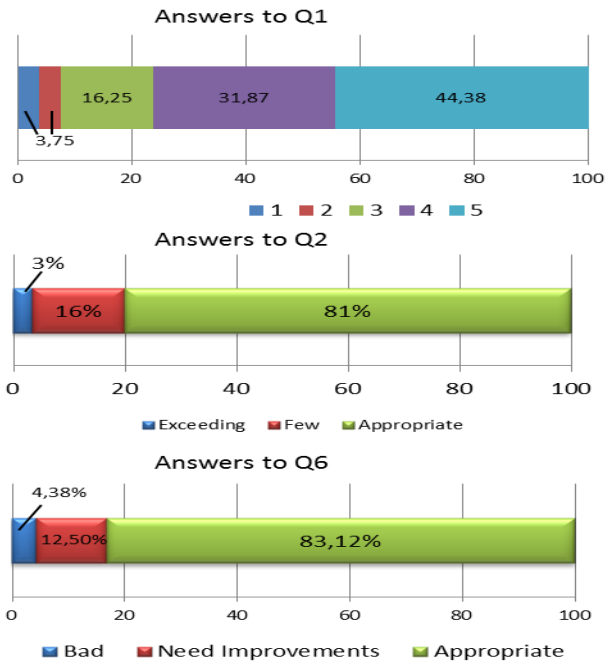

FIGURE 6. General results for transparency HCI features.

Figure 6 presents that HCI features were satisfactorily evaluated by the users. The Q1 results were that 76 % of them considered it easy to find or identify Transparency information. We assumed that this result is due to the selection of the information presented and the labeling through the metadata titles. In contrast with texts in a Privacy and Security Policies (PSP) where the data subject needs to search information within large amount of texts usually without good labels, information created based on the TR-Model is labeled and displayed in order to make it easier to distinguish it from other content. The use of simple, small and straightforward labels (metadata titles) helped in identifying the information.

The answers related to *amount of information* (Q2) presented a similar result to the previous question, once 81% answered that the amount of information is appropriate. This result is justified by the fact that descriptions of metadata and metaevents stipulate a minimal set of information, with limited amount of words and that aimed to present only that can be used and manipulated by data subject in order to avoid access and analysis of unnecessary content.

Regarding to negative answers, 19 % (64 answers) indicated that there is too much or too little information. For these answers, the suggestions of respondents (question Q2a) were analyzed in order to apply future improvements in the amount of information. An initial analysis of the Q2a answers identified that the difficulty was in locating additional desired information in a specific scenario. The difficult occurred because the participant search for a kind of information that was not available in the evaluated scenario.

Assumed that the problems indicated in Q2a were caused by some respondents misunderstanding the tested scenarios that limited Transparency to information strongly related to the scenario context, once there was no access components for other Transparency beyond that contemplated in the scenario

Q6: **Some information are better visualized by text, others by images (photos, videos, etc.), colors, formats, etc. Regarding the format in which the information is displayed (texts, colors, images etc.) what is your opinion?** that focused on Infovis had a considerable percentage of positive responses once 86 % considered the display format appropriate. Although the TR-Model does not clearly specify what interface components to use and which decision may have a direct bearing on *design*, we have considered some standards that were proposed in [61]. These standards were validated by several users in some usability testing, showing effectiveness in displaying information. The display format is an important feature as it aims to abstract the appropriate Transparency information to the data subject so that they can use it for decision making. Regarding negative answers or need for improvement, considerations made by respondents in Q6a provided some suggestions, such as: better use of colors; content organization (following a kind of sequence), improvements in images quality; better grouping and distinction of information; better use of icons; and use just one webpage. We concluded that the suggestions are all related to the information display adopted for the experiments in this research. No suggestions or criticism related to the metadata descriptions and meta-events existing in TR-Model were identified.

Analyzing the results, scenario by scenario, we noticed that there was a pattern of responses and that the numbers remained relatively similar to the values obtained in the general evaluation, as shown in Figure 7.

So, we could define some assumptions based on the related results: **(1) The result of Q1 (Use the range of 1 - Very Bad to 5 - Very Good. How do you rate the ease of finding and/or identify information about the use of your Personal Data) in scenario 5**: The scenario had a mostly appropriate opinion by the participants regarding the identification of the information. we considered that this situation occurred due to the scenario presenting a set of registration forms of the actors and such information is commonly used in the day to day personal and is simple.

This result is due to the reason that the Transparency presented is based on popular information design for respondents. This scenario presented registration information of actors involved in the use of Personal Data. A design of registration information is already very common on websites and the information presented is equally routine for personal information. In addition, the TR-Model specifications did not present severe changes in the way of informing the registration data, but instead tried to make it as simple as possible.

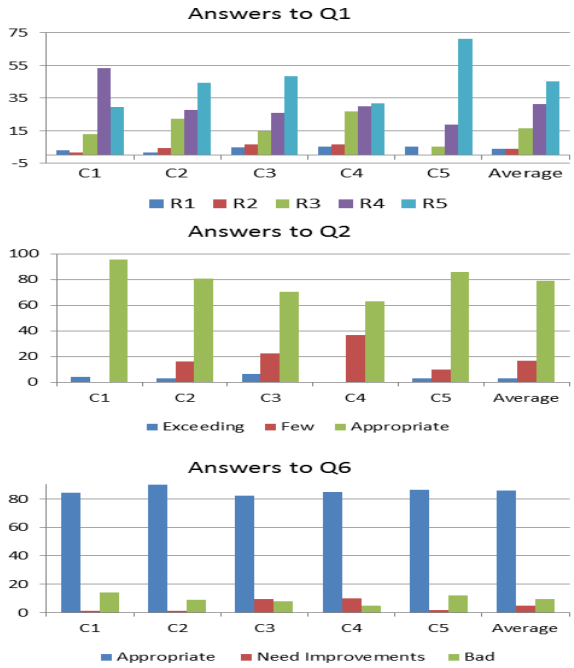

FIGURE 7. Results for transparency HCI features by scenarios.

(2) The answers for Q2 (Regarding the amount of information presented in the Transparency of the evaluated scenario. Your opinion is:) in scenario 4: This assessment presented a result that had some divergence of opinions (little - 40 %; appropriate - 60 %). This scenario used a concept of *photo album* to display a sequence of images extracted from the application. The images highlighted with different colors the moment when the data subject granted permission to use or share his/her Personal Data. This scenario aimed to remind or indicate to the participant when the permission to use their Personal Data is granted.

The Transparency information was created following the specifications of the *Permission of use* metaevent from *Personal Data* entity and the meta-event *Permission* from *Transfer* entity. The index that 40 % pointing to little information was alarming, as it indicates that the application of the TR-Model did not supported the Transparency.

Analyzing the answers from the question Q2a (If your answer for the amount of information was Few or Exceeding, please describe what kind of information could be added or should be removed whether you thought there is too much information), we can assume that participants search for complementary information. For example: in the interface that highlighted the permission to share data, respondents also looked for information about who would be the recipient of the data. They also look for information about the purpose and the processor contact data in the image that presented the moment of granting permission to use the data.

In certain a way, this information could be available by the Controller in the software grant interface. The TR-Model may be improved to expand the access and sharing permission information to provide more information to data subjects, especially the information available from other TR-Model entities and related to grant information. Other forms of displaying this information than displaying information with images may also be considered even though it is not the purpose of TR-Model.

Finally, the results allow to conclude that the use of Readability and InfoVis were effective to support the development of Transparency for scenarios. We concluded this because the most of the respondents answered that they did not have difficulties to find information, the presentation formats were friendly and the amount of information was sufficient to provide understanding about the use of data.

The evaluations that presented results that can be considered as negative for Transparency disclosure were analyzed with the support of comments described by the respondents' essay questions. The information allowed the identification of points that must require improvement in the elements of HCI, mainly in the amount of information used for some types of Transparency. Specific improvements were also pointed out for a better distribution of interface information, specially for denser information such as Purpose of Use and Personal Data entities.

The next subsection presents the results and discussions of the Information Quality dimensions.

2) INFORMATION QUALITY (IQ) DIMENSIONS ANALYSIS

The IQ dimension evaluation aimed to verify the TR-Model effectiveness to provide appropriate content that could be used by data subject to analyze and understand the how his/her Personal Data are used.

The questions used to collect data (Table 11) for this analysis were: Q3, Q4, Q5, Q8, Q3a, Q4a and Q5a. The Figure 8 presents the graph with percentages for the answers to these questions, considering all scenarios evaluated.

The results presented in Figure 8 allowed to conclude that TR-Model was effective regarding to the quality of the content presented, since all responses had above 80% of responses for positive aspects. We highlight the dimension of the *Objectivity* had a positive evaluation above 90%.

The results that consider all scenarios allowed the assumption that the TR-Model was effective in supporting the IQ dimensions. Considerable percentages present that respondents considered the content of the information appropriated to use as information for analyzing the use of Personal Data.

Thus, considering the appropriated/positive answers for the evaluated dimensions we can consider: (1) That the Transparency in the scenarios was **easy to understand** which confirms that the metadata and metaevents present non-technical information and thus become known and understandable to data subjects in general; (2) that the Transparency is **objective** and presented the content without bias or unnecessary text and that it focused solely on providing the Transparency for the data subject to know the use of their data; (3) that the Transparency was **complete**, as it allowed the participants' to solve their possible doubts and so dispensed the need

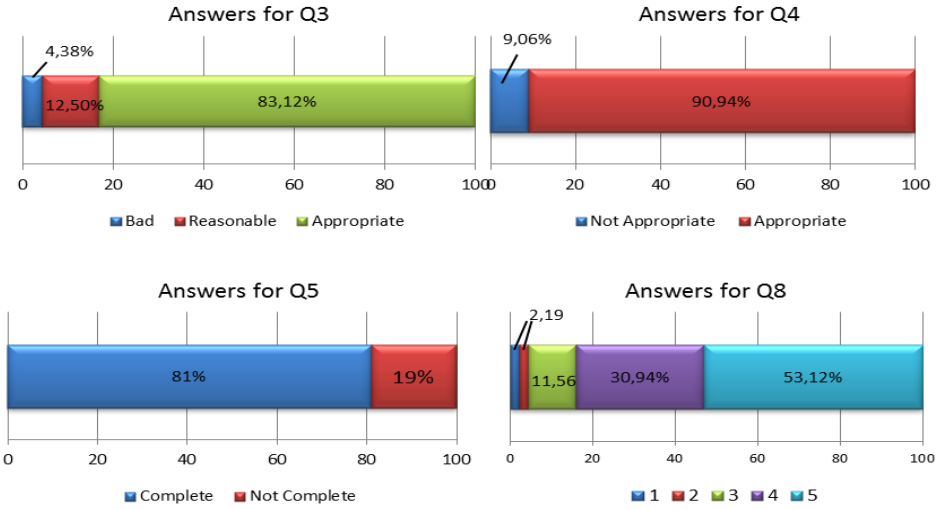

FIGURE 8. Results for transparency QI features.

to search additional information from other sources; and (4) Transparency was **relevant** as presented information of their interests thus avoiding the analysis of unnecessary or non-value added information.

Although the TR-Model had positive evaluations, participants also presented not appropriated answers in the evaluations. Questions Q3, Q4 and Q5 were complemented with essay alternatives (Q3a, Q4a, Q5a) to justify / explain inadequate model evaluations. Suggestions, justifications, and criticisms had some similarity in the responses of questions Q3a (Understanding and Interpretation) and Q4a (Objectivity).

The negative assessment regarding the questions understanding and interpretation was justified by 54 respondents, while the negative assessment on Objectivity was explained by 29 respondents. The presented texts were not well detailed and their analysis required some degree of inference of the researchers. Thus, the following justifications were identified: (1) Design, (2) Lack of information or excess of information; (3) Information presentation; and (4) Extremely technical content.

The items 1 and 3 presented above are particular HCI features that may have interfered in the use of Transparency. The analysis of these elements was discussed in the previous section and we assumed the same considerations for IQ. The lack and excess of information may indicate the data subject's need for a specific information or the design of the information presented that may have left the environment ''polluted'' giving excess information the idea. Technical content is an issue to be reviewed in the TR-Model since this model aims to avoid this issue and, so, facilitate the information understanding.

The completeness dimension (Q5a) had negative evaluations justified by factors as: lack of information; and poorly presented information, which made it more difficult to understand. However, respondents did not explain whether they tried to find information in other sources. Another factor that influenced completeness was the personal need for TDP that led to the search for specific information that were not available in the evaluated scenario.

The lack of details about which scenario was informed by the respondents and the anonymization of the questionnaire makes it difficult to identify the specific scenario in which the evaluation occurred, but indicates the need for a general review, specially in two features: (1) identify whether the technical information was a design issue or metadata / metaevent description; and (2) verify Transparency needs that were not found by participants.

We also conducted an analysis to identify whether the overall results indicated are basically the same for each evaluations' specific situation. With this analysis, we tried to conclude whether there were any unique advantages or problems in any scenario that may not have been identified in the overall analysis and that may indicate any specific improvement in the TR-Model.

The results of the scenario IQ assessment are shown in Figure 9.

The chart in Figure 9 shows that respondents' opinions for each scenario and for each question had similar values when compared to the overall assessment in the scenario assessment. Thus, we assumed that the considerations already discussed can be applied to all evaluated scenarios.

Considering IQ evaluation, Scenario 4 presented a slightly unfavorable behavior in relation to the other scenarios in question Q5. Thus, the use of the photo album with static images may require improvements in the amount and list of information presented to avoid completeness issues.

Next section presents the considerations about TR-Model validation.

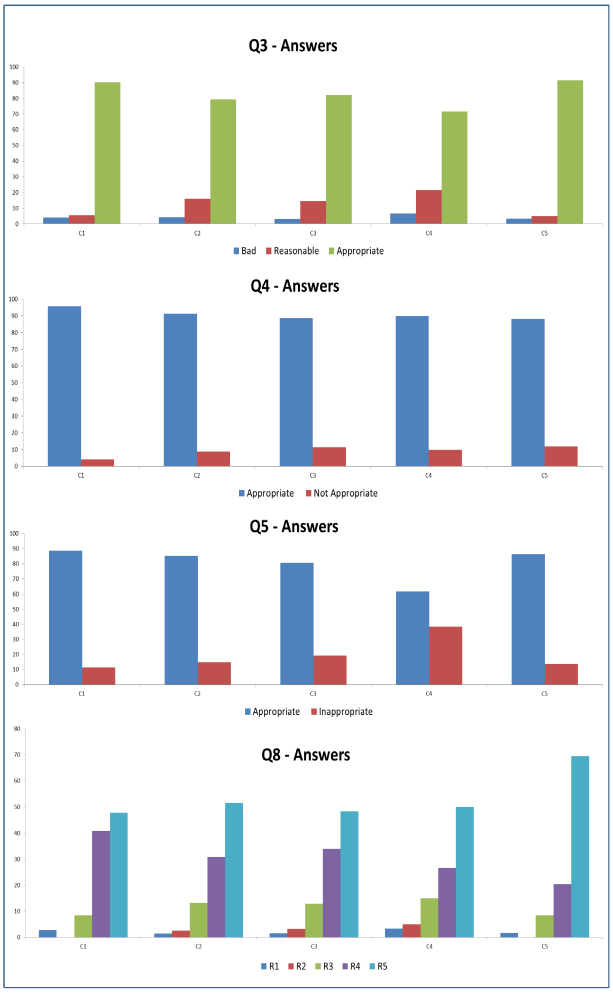

FIGURE 9. Results for transparency HCI features by scenarios.

F. TR-MODEL VALIDATION FINAL CONSIDERATIONS

The validation of the TR-Model aimed to verify the effectiveness of this model in the context of **information coverage** and **friendly and quality Transparency information for the data subject**. The TR-Model was developed concerning about the use of information by data subjects, but to provide such information, companies using Personal Data would need to use TR-Model to support their systems.

Regarding to TR-Model coverage, we assume that the TR-Model may be appropriate for users and therefore:

- Comply with the Transparency requirements set in the GDPR, and LGPD texts. So, that the software application using the template can assure the data subject and Controller that the regulations are being met;
- Meeting 90 % of validation participants' Transparency expectations informed by respondents in pre questionnaire through metadata and metaevents descriptions, thus ensuring that the user is effectively met by Transparency.

The results related to the information coverage characteristics were very good. This was already expected as the regulations were observed during the requirements stage. Coverage validation was a way of checking whether it was possible to classify the TR-Model as a model in *compliance* with regulations.

The validation with the users had no pre-established expectations, because it was the first time the model, metadata, metaevents and descriptions were used. However, the scenario-based validation allowed to simulate interesting usage situations such as Scenario 01 - Purpose of use Transparency; and Scenario 02 - Personal Data Distribution, as well as Personal Data Transparency information that was created based on the TR-Model specifications.

The results showed that positive evaluations of HCI and IQ characteristics were predominant and allowed us to conclude that the TR-Model metadata, metaevents and descriptions were effective in providing information about the use of Personal Data focused on data subjects needs. The numbers, analyzing both the general and each one of the particular scenarios, show that a large majority of respondents considered the information presented as appropriate as they were able to understand, use and analyze it.

Some few suggestions and criticisms were made but they were not significant when compared to the amount of positive evaluations. These information focused on the need for more information than the information provided by the scenarios. This needs may be personal or unanticipated in test scenarios, but should be considered in future versions of TR-Model, as such needs may occur in other cases.

Next section presents the conclusions.

VII. CONCLUSION

This article presented the TR-Model, a Metadata Profile application intended to support the implementation of Personal Data Transparency information and to provide information about the use of Personal Data to data subjects. The main objective of this work was to provide data subjects with relevant, understandable, accessible and regulation-compliant information that allows them to understand how their Personal Data are used.

TR-Model's main contribution is a set of specifications that determine *what should be displayed* for presenting Transparency Information and *how/when it should be displayed* in order to avoid ambiguity, misunderstandings or bias and present ease of understanding and relevant information to assure the user its rights for the Transparency. Thus, data subjects' main concerns regarding the use of their Personal Data are addressed as well as companies accomplishes their duties with regard to regulations.

The main objective of this research was then considered to be effectively achieved. Metadata Application Profile approach allowed the construction of a Domain Model: a set of entities, metadata, and metaevents. Specific descriptions of the metadata and metaevents use were proposed considering elements such as *Readability*, *Infovis* and Information Quality and mainly by the participation of data subjects in activities

such as *workshops*, lectures and interviews in which Data Transparency issues were discussed.

Validations conducted through scenarios evaluation and user surveys showed that the TR-Model was very effective in supporting Transparency. This conclusion was done as: (1) the TR-Model met 90 % of the Transparency expectations presented by the data subjects in the surveys; and (2) evaluations performed related to the presentation form and the quality of the content information were evaluated positively by 85 % of the validation participants.

TR-Model was considered a model focused on the data subject, that can also be used by software development companies that use personal data.

From a software developer's point of view, it can also be assumed that applications and websites that adopt the TR-Model as the basis for their Transparency can provide a set of information of interest to their data subjects. Applications using TR-Model will also be compliant with the GDPR, as shown in Table 8, presented in Section V. We can also assure that the model presents, with its metadata and metaevents, the information required by the regulation.

With regard to data subjects, we believe that, by using TR-Model based Transparency software, they will be able to access a set of user-friendly and appropriate information on content, quantity and presentation. Consequently, data subjects' cognitive load usually required to analyze complex content, understand technical terms, classify information that might be considered as important or seek information from other sources will be decreased. It is also assumed that the data subjects will have more confidence in applications that disclose information about how their data will be used and, embedded in the data usage flow, can act to ensure the textit fair use of their personal data.

As the primary objective, we assumed that guidelines can be used effectively in applications and websites to provide data subjects with greater knowledge and trust.

In future works we intend to revise and improve the TR-Model metadata and meta-events to meet the needs and suggestions presented in the questionnaires and also to improve support to design patterns. We also intend to extend the model by including Personal Data traceability features so that TR-Model can also meet requirements of Personal Data Provenance.

REFERENCES

[1] A. S. Bataineh, R. Mizouni, M. E. Barachi, and J. Bentahar, "Monetizing personal data: A two-sided market approach," *Procedia Comput. Sci.*, vol. 83, pp. 472–479, Jun. 2016, doi: 10.1016/j.procs.2016.04.211.

[2] P. Voigt and A. V. D. Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide*, 1st ed. Cham, Switzerland: Springer, 2017.

[3] M. S. Ackerman and S. D. Mainwaring, *Privacy Issues and Human-Computer Interaction*. Newton, MA, USA: O'Reilly, 2005, pp. 1–19.

[4] P. Bonatti, S. Kirrane, A. Polleres, and R. Wenning, "Transparent personal data processing: The road ahead," in *Proc. Int. Conf. Comput. Saf., Rel. Secutity*, in Lecture Notes in Computer Science, vol. 10489, Sep. 2017, pp. 337–349.

[5] R. Mortier, H. Haddadi, T. Henderson, D. Mcauley, J. Crowcroft, and A. Crabtree, "Human-data interaction: The encyclopedia of human-computer interaction," *Encyclopedia Hum.-Comput. Interact.*, vol. 2, pp. 1–48, 2016. [Online]. Available: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed

[6] A. S. Patrick, *From Privacy Legislation to Interface Design: Implementing Information Privacy in Human-Computer Interactions* (Lecture Notes in Computer Science). Berlin, Germany: NRC Publications Archive, Mar. 2003.

[7] M. Jayabalan and V. Thiruchelvam, "A design of patients data transparency in electronic health records," in *Proc. IEEE Int. Symp. Consum. Electron. (ISCE)*, Nov. 2017, pp. 9–10.

[8] A. Poikola, K. Kuikkaniemi, and H. Honko, "Mydata a nordic model for human-centered personal data management and processing," Ministry Transp. Commun., Helsinki, Finland, Tech. Rep., 2014. [Online]. Available: http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78439/MyData-nordic-model.pdf

[9] G. Iachello and J. Hong, *End-User Privacy in Human-Computer Interaction*, vol. 1. Boston, MA, USA: NOW, 2007, pp. 1–137. [Online]. Available: https://www.nowpublishers.com/article/Details/HCI-004

[10] B. Bellamy and C. Alonso, "Reframing data transparency. Centre for information policy leadership and telefónica senior roundtable," Tech. Rep., Jun. 2016, vol. 1, pp. 1–20. [Online]. Available: https://www.telefonica.com/documents/341171/2445513/CIPL+and+Telefonica+-+Reframing+Data+Transparency.pdf/9c007899-451c-4a5b-854d-784082e37bf7

[11] P. Murmann and S. Fischer-Hubner, "Tools for achieving usable ex post transparency: A survey," *IEEE Access*, vol. 5, pp. 22965–22991, 2017.

[12] P. Murmann and S. Fischer-Hubner, *Usable Transparency Enhancing Tools*. [Online]. Available: http://www.diva-portal.org/smash/get/diva2:1119515/FULLTEXT02.pdf

[13] P. Santos, L. Salgado, and J. Viterbo, "Assessing the communicability of human-data interaction mechanisms in transparency enhancing tools," in *Proc. Federated Conf. Comput. Sci. Inf. Syst.*, vol. 15, Sep. 2018, pp. 897–906.

[14] Y. Li, F. Chen, T. J.-J. Li, Y. Guo, G. Huang, M. Fredrikson, Y. Agarwal, and J. I. Hong, "PrivacyStreams: Enabling transparency in personal data processing for mobile apps," in *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, Sep. 2017, vol. 1, no. 3, pp. 1–26, doi: 10.1145/3130941.

[15] "Cartilha LGPD," Camara Brasileira de Comércio Eletrônico, São Paulo, Brazil, Tech. Rep. 5, 2019.

[16] PIPEDA, "Canadian personal information protection and electronic documents act," Canadian Bar Assoc., Otawwa, ON, Canada, Tech. Rep., Mar. 2000. [Online]. Available: https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/

[17] M. Janic, J. P. Wijbenga, and T. Veugen, "Transparency enhancing tools (TETs): An overview," in *Proc. 3rd Workshop Socio-Tech. Aspects Secur. Trust*, Jun. 2013, pp. 18–25.

[18] I. Kabanov, "Effective frameworks for delivering compliance with personal data privacy regulatory requirements," in *Proc. 14th Annu. Conf. Privacy, Secur. Trust (PST)*, Dec. 2016, pp. 551–554.

[19] C. Riederer, D. Echickson, S. Huang, and A. Chaintreau, "FindYou: A personal location privacy auditing tool," in *Proc. 25th Int. Conf. Companion World Wide Web (WWW Companion)*, 2016, pp. 243–246.

[20] A. Mashhadi, F. Kawsar, and U. G. Acer, "Human data interaction in IoT: The ownership aspect," in *Proc. IEEE World Forum Internet Things (WF-IoT)*, Mar. 2014, pp. 159–162.

[21] R. Mortier, H. Haddadi, and T. Henderson, "Challenges & opportunities in human-data interaction," in *Proc. Open Digital 4th Annu. Digit. Economy All Hands Meeting*, 2013, pp. 4–6. [Online]. Available: http://de2013.org/wp-content/uploads/2013/09/de2013_submission_15.pdf

[22] C. Moiso and R. Minerva, "Towards a user-centric personal data ecosystem the role of the bank of individuals' data," in *Proc. 16th Int. Conf. Intell. Next Gener. Netw. (ICIN)*, Oct. 2012, pp. 202–209.

[23] C. Gurrin, A. F. Smeaton, and A. R. Doherty, "LifeLogging: Personal big data," *Found. Trends Inf. Retr.*, vol. 8, no. 1, pp. 1–125, 2014.

[24] A. Morgana and A. A. Baptista, "The use of application profiles and metadata schemas by digital repositories: Findings from a survey," in *Proc. Int. Conf. Dublin Core Metadata Appl.*, 2015, pp. 146–157.
[25] D. G. Sampson, P. Zervas, and G. Chloros, "Supporting the process of developing and managing LOM application profiles: The ASK-LOM-AP tool," *IEEE Trans. Learn. Technol.*, vol. 5, no. 3, pp. 238–250, Jul. 2012.
[26] S. A. M. Mostafa, S. R. H. Noori, and S. Jafreen, "Transparency—A key feature integration in existing privacy frameworks for online user," in *Proc. Int. Workshop Comput. Intell. (IWCI)*, Dec. 2016, pp. 74–78.
[27] B. Schneier, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*. New York, NY, USA: Norton, 2015.
[28] M. Turilli and L. Floridi, "The ethics of information transparency," *Ethics Inf. Technol.*, vol. 11, no. 2, pp. 105–112, Jun. 2009.
[29] R. Mortier, D. M. Hamed, and J. C. Haddadi, *Human-Data Interaction*, no. 837. Cambridge, U.K.: Univ. of Cambridge, 2013, pp. 1–9. [Online]. Available: http://128.232.0.20/techreports/UCAM-CL-TR-837.pdf%5Cnpapers2://publication/uuid/20A8C3B6-3820-4ECB-A8C2-402BC7EB7899%5Cnhttp://www-ipv4.cl.cam.ac.uk/techreports/UCAMCL-TR-837.pdf
[30] A. Crabtree and R. Mortier, "Human data interaction: Historical lessons from social studies and CSCW," in *Proc. 14 Eur. Conf. Comput. Supported Cooperat. Word*, 2015, pp. 19–23.
[31] J. B. Earp, A. I. Anton, L. Aiman-Smith, and W. H. Stufflebeam, "Examining Internet privacy policies within the context of user privacy values," *IEEE Trans. Eng. Manag.*, vol. 52, no. 2, pp. 227–237, May 2005.
[32] L. V. L. Filgueiras, A. S. F. Leal, T. A. Coleti, M. Morandini, P. L. Correa, and S. N. Alves-Souza, "Keep system status visible: Impact of notifications on the perception of personal data transparency," in *Proc. Hum.-Comput. Interact. Perspect. Design*, vol. 1, 2019, pp. 513–530.
[33] J. Cusick, "The general data protection regulation (GDPR)," *Irish Med. J.*, vol. 111, no. 5, p. 747, 2018.
[34] (Oct. 2014). *Metadata and Privacy: A Technical and Legal Overview*. [Online]. Available: https://www.priv.gc.ca/information/research-recherche/2014/md_201410e.pdf
[35] G. Moran, "Understanding metadata concepts and properties," in *Proc. Metadata*, 2008, pp. 11–13.
[36] DublinCore. *Dublin Core Metadata Initiative*. Accessed: Dec. 2014. [Online]. Available: http://dublincore.org/metadata-basics/
[37] J. Mori, Y. Matsuo, M. Ishizuka, and B. Faltings, "Keyword extraction from the Web for personal metadata annotation," in *Proc. CEUR Workshop*, vol. 184, May 2004, pp. 51–60.
[38] J. T. Tennis, "Metadata Application Profiles," in *Encyclopedia Archival Concepts, Princess, Practices*. Lanham, MD, USA: Rowman & Littlefield, Jan. 2015, pp. 2–5.
[39] M. C. Malta and A. A. Baptista, "A panoramic view on metadata application profiles of the last decade," *Int. J. Metadata, Semantics Ontologies*, vol. 9, no. 1, pp. 58–73, 2014.
[40] (2009). *Scholarly Works Application Profile*. [Online]. Available: http://www.ukoln.ac.uk/repositories/digirep/index/Scholarly_Works_Application_Profile
[41] (2017). *Virtual Open Access Agriculture & Aquaculture Repository*. [Online]. Available: https://cordis.europa.eu/project/rcn/204632/en
[42] A. Apps and R. MacIntyre, "Dublin core metadata for electronic journals," in *Proc. Int. Conf. Pract. Digit. Libraries*, in Lecture Notes in Computer Science, vol. 1923, 2000, pp. 93–102. [Online]. Available: http://eprints.rclis.org/12183/2/appsmacecdl2000.pdf
[43] S. Y. Diallo, H. Herencia-Zapana, J. J. Padilla, and A. Tolk, "Understanding interoperability," in *Proc. Emerg. MS Appl. Ind. Acad. Symp. (EAIA)*, May 2014, pp. 84–91.
[44] S. Ismail and T. Shaikh, "A literature review on semantic Web–understanding the Pioneers' perspective," in *Proc. Comput. Sci. Inf. Technol. (CS IT)*, Sep. 2016, pp. 15–28.
[45] J. L. McKillip, L. A. Jaykus, and M. Drake, "rRNA stability in heat-killed and UV-irradiated enterotoxigenicStaphylococcus aureus and escherichia coliO157: H7," *Appl. Environ. Microbiol.*, vol. 64, no. 11, pp. 4264–4268, 1998.
[46] M. Hosseini, A. Shahri, K. Phalp, and R. Ali, "Foundations for transparency requirements engineering," in *Proc. Int. Work. Conf. Requirements Eng., Found. Softw. Qual.*, 2016, pp. 225–231. [Online]. Available: http://link.springer.com/10.1007/978-3-319-30282-9_15
[47] P. Guarda, S. Ranise, and H. Siswantoro, "Security analysis and legal compliance checking for the design of privacy-friendly information systems," in *Proc. 22nd ACM Symp. Access Control Models Technol. (SACMAT Abstr.)*, 2017, pp. 247–254.
[48] J. Tom, E. Sing, and R. Matulevičius, "Conceptual representation of the GDPR: Model and application directions," in *Proc. Int. Conf. Bus. Inform. Res.*, in Lecture Notes in Business Information Processing, vol. 330, Jan. 2018, pp. 18–28.
[49] M. Robol, M. Salnitri, and P. Giorgini, "Toward GDPR-compliant socio-technical systems: Modeling language and reasoning framework," in *Proc. Work. Conf. Pract. Enterprise Modeling*, in Lecture Notes in Business Information Processing, vol. 10, 2017, pp. 236–250.
[50] R. S. Pressman, *Software Engineering: A Practitioner's Approach*, 8th ed. New York, NY, USA: McGraw-Hill, 2014.
[51] Y. Rogers, H. Sharp, and J. Preece, *Design de Interação: Além da Interação Humano-Computador*, 3rd ed. Porto Alegre, Brazil: Bookman, 2013.
[52] G. Abib, "A qualidade da informação para a tomada de decisão sob a perspectiva do sensemaking: Uma ampliação do campo," *Ciencia da Informacao*, vol. 39, no. 3, pp. 73–82, 2010.
[53] L. Vieira. (2017). *A 'Readability' Vai Mudar a Sua Forma de Produzir Conteúdo*. [Online]. Available: https://blog.apiki.com/readability/
[54] M. Zamanian and P. Heydari, "Readability of texts: State of the art," *Theory Pract. Lang. Stud.*, vol. 2, no. 1, pp. 43–53, 2012.
[55] O. Sindiy, K. Litomisky, S. Davidoff, and F. Dekens, "Introduction to information visualization (InfoVis) techniques for model-based systems engineering," *Procedia Comput. Sci.*, vol. 16, pp. 49–58, Dec. 2013, doi: 10.1016/j.procs.2013.01.006.
[56] J. S. Kandari, E. C. Jones, F. F. Nah, and R. R. Bishu, "Information quality on the world wide Web: A framework to measure and its validation," in *Proc. 2nd Int. Multi-Conf. Complex., Inform. Cybern. (IMCIC)*, vol. 2, 2011, pp. 204–209.
[57] J. Kandari, E. C. Jones, F. F. H. Nah, and R. R. Bishu, "Information quality on the world wide Web: Development of a framework," *Int. J. Inf. Qual.*, vol. 2, no. 4, pp. 324–343, 2011.
[58] R. Y. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *J. Manage. Inf. Syst.*, vol. 12, no. 4, pp. 5–33, Mar. 1996.
[59] A. T. S. Calazans, "Qualidade da informação: Conceitos e aplicações," *Transinformação*, vol. 1, no. 1, pp. 29–45, 2008.
[60] R. Budiu. (2017). *Quantitative vs. Qualitative Usability Testing*. [Online]. Available: https://www.nngroup.com/articles/quant-vs-qual/
[61] T. A. Coleti, M. Morandini, L. V. L. Filgueiras, P. L. Correa, I. G. Oliveira, and C. R. C. S. C. Barbosa, "Design patterns to support personal data transparency visualization in mobile applications," in *Proc. Design Patterns Support Pers. Data Transparency Vis. Mobile Appl. (HCII)*, vol. 1, 2019, pp. 46–62.

THIAGO ADRIANO COLETI graduated in data processing from the Technology College of Sao Paulo, in 2006. He received the master's degree in system information from EACH-USP, in 2013, and the Specialization Degree in software engineering from UML, UNIFIL, Londrina, in 2009. He is currently pursuing the Ph.D. degree in electrical engineering from EPUSP, in 2014. He is also an Auxiliary Professor at the State University of North of Parana, in the system information and computer science courses. He is the Manager of JoinSoft Corporate Systems, where he also works as a Software Engineer for supporting the analysis and requirements elicitation process. His research areas are human–computer interaction, human-data interaction, big data, software engineering, and information systems.

PEDRO LUIZ PIZZIGATTI CORRÊA (Member, IEEE) received the bachelor's degree in computer science from ICMC/USP, in 1987, and the Ph.D. degree in electrical engineering from EPUSP, in 2002. He has been an Associate Professor at the Department of Computer and Digital Systems, Escola Politécnica da USP (EPUSP), since 2003. He was a Visiting Scholar with The University of Tennessee, Knoxville (UTK), School of Information Science, Climate Change Science Institute of Oak Ridge National Laboratory (CCSI/ORNL), and United States Geological Survey (USGS). He was a Consultant on Integration of Information Systems and Data Base of the Ministry of Environment (MMA), from 2012 to 2013. He was a Fellow of the United Nations Development Programme (UNDP) (BRA/97/001) at the Modernization Program and a Coordinator of Information Systems at São Paulo State Finance (UNDP), from 1998 to 2002. He was a Software Engineering at Accurate Software, responsible for development of electronic data interchange (EDI) software, from 2002 to 2007. He was a Researcher (Level II) from the Centro Nacional de Pesquisa em Informatica - CNPTIA – from Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), from 1999 to 2002. His areas of interests are distributed database, information science, scientific data management, and e-government. His other areas of interests are modeling and simulation computer systems, distributed systems, and software engineering. He was a Fellow of the National Council for Scientific and Technological Development (CNPq), Brazil.

LUCIA VILELA LEITE FILGUEIRAS received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Escola Politécnica, Universidade de São Paulo, in 1983, 1989, and 1996, respectively. She has been an Assistant Professor with the Computer Engineering Department, Escola Politécnica, Universidade de São Paulo, since 1990. Her research interests are in the areas of human–computer interaction, human-agent interaction, human-data interaction, and information visualization.

MARCELO MORANDINI received the master's degree in computer science from USP, in 1996, and the Ph.D. degree in systems engineering from the Federal University of Florianopolis, in 2002. He holds a postdoctoral position at The University of Tennessee, Knoxville, USA. He also works as a Ph.D. and master's degrees Supervisor in researches of the human–computer interaction in big data, data science and climate changes at the University of São Paulo. He is currently an Assistant Professor at the School of Arts, Sciences and Humanities (EACH), USP. His research interests include software engineering, requirements engineering, software architectures, usability, design, evaluation, and accessibility.