Assessing the Communicability of Human-Data Interaction Mechanisms in Transparency Enhancing Tools

Patrick Barreto
Institute of Computing
Fluminense Federal University
Niterói, RJ, Brazil
Email: patrickbarreto@id.uff.br

Luciana Salgado
Institute of Computing
Fluminense Federal University
Niterói, RJ, Brazil
Email: luciana@ic.uff.br

José Viterbo
Institute of Computing
Fluminense Federal University
Niterói, RJ, Brazil
Email: viterbo@ic.uff.br

Abstract—The growing practice of accumulating personal data to generate predictions about users, leverages the need for mechanisms that allow people a more effective control of their data. An emerging field of studies called Human-Data Interaction (HDI), proposes the inclusion of human at the center of the data flow, providing mechanisms for citizens to interact explicitly with the collected data. Researches in HDI have discussed ways to offer Transparency Enhancing Tools (TETs), i.e., tools that support people on HDI issues related to privacy and personal data protection. Many works conducted about TETs focuses on usability issues, exploring aspects such as efficiency, user satisfaction and ease of learning. In this work, on the other hand, we aim to assess the communicability of HDI mechanisms in TETs. Hence, we applied the Semiotic Inspection Method (SIM) to investigate if and how HDI concepts are applied in two different TETs used for personal data management. We triangulated results from the study with findings from another investigation about communicability issues carried out in the same domain, but by observing and interviewing users.

I. INTRODUCTION

The evolution of mobile devices, such as smartphones, tablets and sensors, influenced the society lifestyle as a whole by making more flexible the access to various services on the internet, and, eventually, by bringing advances in processing capacity and agility in mobile communication. As a result, people have begun to consume and share a significant and ever-increasing amount of data on their daily lives, which encompasses, for example, social information, events, health, lifestyle and consumption habits [1].

In ubiquitous computing scenarios, data collected by monitoring the user's activities can be used in analyzes and inferences to extract information about the behavior of individuals [2], [3]. In this scenario, data is used to make predictions related, for example, to the health status of people [4] or consumption trends [5]. Thus, the growing practice of accumulating personal data and generating inferences or predictions from them, leverages the need for research and creation of mechanisms that allow people a more effective interaction in this process of data manipulation [6].

Given this scenario, an emerging field of studies called Human-Data Interaction (HDI) proposes the inclusion of human at the center of the data flow, providing mechanisms to citizens to interact explicitly with these systems and the associated data [7]. The purpose is to enable users to understand by whom and in what form their respective data is used, and how to promote desirable effects and avoid undesired consequences [1].

Research in HDI is still incipient and has been gaining strength in the last five years, although some works in the Information Systems area reflect similar questions, among which are: transparency through open data, storage and use of personal data on the daily life of individuals [8] and privacy of information [9], [10]. Other works [11], [12], however, discuss ways to offer tools that support people on HDI issues related to privacy and personal data protection, and propose Transparency Enhancing Tools (TETs).

In [11], for example, the authors consider aspects of usability for TETs, proposing an interface prototype that seeks to offer the user a comprehensive view of their data stored and made available in different online services. This prototype is based on visualization techniques, seeking to associate the personal data of a user to the service for which this data was shared. The purpose is to provide transparency to users about their personal data collected by online services.

In [12], the authors presented PrivacyInsight, a software that allows the user to access their personal data, as well as the flow of data between interested entities involved in the storage and processing of this data. In addition, the tool enables the user to perform actions on their data in an indirect manner, i.e., through requests for correction or removal, thus providing means to exercise their right granted by law. This tool is based on the European Data Protection Directive (Council of European Union. 2016. Council regulation (EU) no 679/2016 - General Data Protection Regulation 95/45/EC[1]) and on usability requirements identified in the design of this solution.

Finally, in [13] the authors propose two Privacy Design Standards to facilitate the development of applications through solutions to recurrent privacy problems, focusing on trans-

---

[1]https://eur-lex.europa.eu/eli/reg/2016/679/oj

parency as a measure to establish broad control for users about their personal data. These standards are based on a set of factors, such as objective, user profile, usage context, problem, solution and consequences.

Such works, therefore, focus on usability issues in TETs, exploring aspects such as efficiency, user satisfaction and ease of learning. Our work, on the other hand, deals with the communicability [14] of mechanisms that enable Human-Data Interaction. For this purpose, we adopt a theory that considers the interaction of human beings with the interfaces of computational systems as a particular case of metacommunication, called Semiotic Engineering [15]. Some research questions are raised in the light of this theory, among which we highlight: How concepts of HDI can be communicated to users through TETs?; How interfaces can create means for the users to act in the data manipulation process?; and Which interactive elements can be used to communicate to the user about opportunities to use the data and its associated value?

In this context, this paper aims to investigate if and how HDI concepts are efficiently communicated (communicability) in two different TETs used for personal data management. For this, we apply the Semiotic Inspection Method (SIM) for scientific purposes [15]. We triangulated results from the study with findings from another investigation about communicability issues carried out in the same domain, but by observing and interviewing 5 (five) users. This work is organized as follows. In Section 2, we set the theoretical framework for the addressed theme. In Section 3, we present the methodology used in this work. In Section 4, we discuss the results obtained. Finally, in Section 5 and 6 we present the triangulation and final considerations, respectively.

II. Theoretical Framework

In this section, we present the theoretical basis necessary for the generation of this work.

A. Human-Data Interaction

The HDI area is an emerging field of interdisciplinary studies, which aggregates elements not only from the various branches of Computer Science, but also from areas such as Law, Psychology, Behavioral Economics and Sociology [1].

The literature presents some papers that address HDI with a focus on data analysis based on aspects of embedded interaction. In this approach, HDI is related to "Human manipulation and making sense of large complex and unstructured datasets". In [16], HDI is defined as "the customized delivery problem, creating the context of data understanding from large datasets". Thus, those proposals are based on aspects of HDI for the design of visualizations that allow to generate insights on large volumes of analyzed data.

There is a second approach, based on the proposals of Mortier *et al.* [7], which considers broader and more complex aspects, seeking mainly to address the problem of the management and use of personal data in society in general [17]. In other words, HDI is related to the manipulation of data, mainly personal, based on human factors [7].

The former approach is the one adopted in this work. It is related to the scenario in which the development of technologies and services for data generating, sharing and manipulating have, in general, allowed people to be in contact with digital tools and artifacts for consumption or production of information. Thus, people can produce data both Consciously (profile data in social networks, use of physical activity tools), and Unconsciously (robots monitoring our search history, cookies recording our browsing history, inferences of interest created from our purchase or search history) [7], [1].

Such data can be accumulated by different organizations that can perform inferences about sensitive issues related to our lives (health or emotional state, consumption habits or political preferences, for example) [7], so that these different analyzes make it possible to influence the user's behavior in a variety of ways [1]. Based on this perception, research in HDI seeks to address the new issues arising from the use of this ecosystem of personal data between different interested entities and their impacts on the actions of individuals and in society.

In [7], the authors establish three fundamental principles which address the challenges tackled in Human-Data Interaction, such as: Legibility, Agency and Negotiability. Legibility is concerned with making data acquisition and analytic algorithms more transparent and understandable to users, since, in general, interactions with data flows and processes are often obscure to people. Agency aims to provide individuals with the means to manage their data and their access by third parties, as well as to seek effective ways of acting in these systems, to the extent that individuals find it appropriate. This includes not only the ability to opt in or out of data collection and processing, but also the broader ability to engage with data collection, storage and use, and to understand and modify data and inferences. Finally, Negotiability is concerned with the various dynamic relationships that arise from data processing. This topic covers, for example, how understanding and individual attitudes change over time.

B. Semiotic Inspection Method

In this research, we used the Semiotic Inspection Method (SIM) for scientific purposes [15], a method of qualitative evaluation in Human-Computer Interaction (HCI), based on Semiotic Engineering [14]. With this method, the evaluators can analyze the communicability of the interactive artifacts [18]. The focus is to inspect the metacommunication from designer to user with the objective of identifying possible breakdowns in communication. First, in the preparation stage, the evaluation focus, the user profile and the inspection scenario are defined. In the evaluation stage, the evaluator examines the interface and classifies the signs as metalinguistic, static or dynamic.

Metalinguistic signs are the first to be analyzed, since they explicitly express and explain other parts of the metacommunication of the designer. This class of signs is usually found throughout the interface, either in instructions, explanations, warnings and error messages, with a focus on online help and user's manuals [19].

Static signs are those that communicate their meaning regardless of cause and effect relationships and can be interpreted from instant canvas pictures. Thus, they express the state of the system at a given time. They are represented by the elements present in the interface screens (or equivalents in non-visual interfaces), such as labels, images, text boxes, buttons, menus, etc., as well as layout, size, color, font and other characteristics. Its analysis should consider only the interface elements presented in each screen at an instant of time, without examining neither the behavior of the system, nor the temporal and causal relationships between interface elements [19].

The inspection of dynamic signs requires that in the analysis, the evaluator inspects the interaction process that the user can experience through the interface. These signs are perceived through changes in the interface that communicate to the user the behavior of the system as a result of user actions (clicking the mouse, pressing enter, changing the focus from one form field to another, etc.), by external events (receiving email, Internet connection failure, etc.) or over time. Dynamic signs are usually represented by animations, opening and closing dialogues, transitions between screens, or modifications to the elements of a screen (for example, activating a button, updating a text or image, modifying the layout of some interface elements, etc) [18].

To inspect the interface, SIM proposes 5 steps to be followed by the evaluator [20]. In the first three steps, the main goal is to reconstruct the metacommunication of the designer for each category of signs (metalinguistic, static and dynamic), using the following meta-model of the designer [20]: *"Here is my understanding , who you are, what I learned that you want or need to do, in what preferred ways and why. This is the system that I have therefore created for you, and this is how you can or should use it for meet a variety of purposes that fall within this version"*. The steps are:

- Step 1: Inspection of metalinguistic signs. In this step, the evaluator explores the documentation and help system.
- Step 2: Inspection of static signs. In this step, the evaluator inspects the static signs of the interface.
- Step 3: Inspection of dynamic signs. In this step, the evaluator inspects the signs that emerge from the interaction.
- Step 4: In this step, the evaluator contrasts and compares the metacommunication messages from steps 1, 2, and 3 and records possible problematic interpretations that may occur in user interaction time.
- Step 5: Appreciating the quality of metacommunication. In this step, the evaluator produces a report containing the communicability problems encountered, which may frustrate or prevent the user from understanding the message intended by the designer, affecting his productivity. In this method, the evaluator is the advocate of the user.

SIM can be used in scientific contexts and generate valid knowledge in HCI [15]. To do so, two other steps must be considered when applying the method. During the preparation phase, it is necessary to define the research question
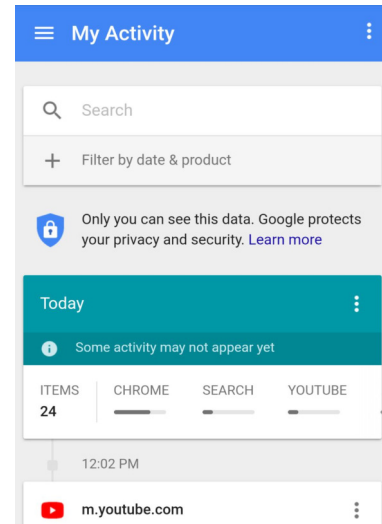

Fig. 1. MyActivity home screen.

that researchers are looking for an answer. Also, after the application, a Triangulation step is added to the analysis. Triangulation involves the generation of other results (e.g. by other specialists or through compatible methods) which valides scientifically the results obtained through SIM.

III. METHODOLOGY

The methodology used in this work consists, in large part, in the application of the Semiotic Inspection Method (SIM), described in the previous section, with the focus on the evaluation of the communicability considering the concepts presented by the HDI theory. We use, therefore, the predictive paradigm, making use of an interpretative and qualitative method [21]. The inspections were carried out by two evaluators together, being one a junior level evaluator and the other a senior level evaluator (specialist). Based on the application of the SIM steps, we sought to answer two research questions (RQs): (i) Which are the communication strategies that potentially enable Human-Data Interaction?; (ii) What is the relationship among the elements found in the first question and the main concepts of HDI proposed by Mortier *et al.* [7], i.e., Legibility, Agency and Negotiability?

There are few TETs tools for data management that provide ways for controlling data. In addition, there are no records that these tools were designed based on the concepts that we adopt as the foundation of HDI. Thus, we performed two studies (S1 and S2) to identify traces of the application of the fundamentals of HDI in TETs. To do this, we have selected two tools: Google MyActivity [22] and Privacy Badger [23].

MyActivity allows users to exercise greater control over the data generated by the monitoring of their activities in Google's services and products. Privacy Badger is a add-on that aims to restrain the action of third-party domains that seek to collect data through unauthorized monitoring of user activity while he is browsing the web. Figures 1 and 2 show the interface of the home screen of both tools.
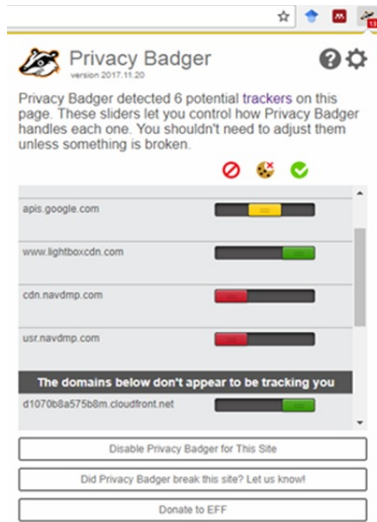
Fig. 2. Privacy Badger Home Screen.

These tools were chosen because they seek to offer the user an understanding of how their personal data are being collected or used by interested entities. Moreover, both tools provide the user with forms of control over the access and use of their data. The evaluation focus, the user profile and the reference inspection scenario in both tools were defined in the preparation stage of S1 and S2, as described below:

A. Preparation Stage for MyActivity

In what follows, we describe the evaluation focus, the user profile and the inspection scenario defined for the evaluation of MyActivity.

- User profile: The user uses mobile devices or computer to browse the internet, searching for leisure and entertainment options, and to make online purchases. He gets surprised to have the possibility of exercising control over his personal data.
- Inspection scenario: Ane performs various tasks through her smartphone. Every day she accesses social networks, conducts research of professional and personal interest and seeks to find the best path in traffic. Ane has always opted to make her personal data available to her applications and services from large companies like Google. With this, Ane seeks to experience a navigation based on their tastes, interests and types of content consumed, thus avoiding receiving unnecessary notifications or content. Recently, Ane was notified by Google about MyActivity, their Data Management tool. According to MyActivity's proposal, Ane has found that she can manage the Google's services to made them more useful. Ane was surprised to know that she can exercise control over her data or activities carried out through her smartphone. Hence, when Ane accesses Google's MyActivity, she wants to perform the following tasks: (a) To find out what types of data are being stored or monitored by Google;

(b) To exercise some intervention in the availability and access to her data.

B. Preparation Stage for Privacy Badger

In what follows, we describe the evaluation focus, the user profile and the inspection scenario defined for the evaluation of Privacy Badger.

- User profile: The user uses mobile devices or a computer to browse the internet, searching for leisure and entertainment options, or shopping online. However, he is concerned about his privacy, i.e., preserving data about his browsing history against third parties.
- Inspection scenario: Bob uses his personal computer often to read news, emails, search products, shop online, access Internet banking, interact in social networks and search about leisure options. Bob manipulates his personal information to perform a good part of these actions. Thus, concerned about the risk of invasion of privacy by tracking robots (trackers), he resorted to some tools, among them, the Privacy Badger. Hence Bob wants to accomplish the following tasks when using this tool: (a) Identify all possible trackers that can monitor his activity when using the internet; (b) Block monitoring trackers.

IV. RESULTS

In this section, first, we present the classes of signs found with the SIM inspection. Then, for both studies (S1 and S2), we tried to answer research questions (i) and (ii). For this, we analyze the main communication strategies found during the inspection, which involve the communication of mechanisms that enable Human-Data Interaction. These communication strategies were identified from traces of the application of the three main concepts of HDI theory: Legibility, Agency and Negotiability. Finally, we performed a comparative analysis between studies S1 and S2.

A. Classes of Signs

In this section, we present some visual design options that were identified through SIM. These options represent visual cues used to interact with systems and have been adopted by the designers of these applications. The MyActivity tool uses all of the identified types, while the Privacy Badger tool does not offer the 'Cards' and 'Modal' classes, as follows:

- **Cards** are the registered activities of the user. Each activity is represented by a title (showing the service used), a link for the activity performed preceded by a keyword that characterizes the type of activity recorded, such as 'Watched' (Videos), 'Visited' (Web Pages), 'Searched' (Google Search) or 'Viewed Area' (Use of Maps), for example. In addition, 'details' or 'delete' options are displayed in each activity log. Generally, a figure can be associated with an activity, and activity groupings can be done automatically to summarize the display of records. Thus, an option is displayed in the card footer if the user wants to view items that have been deleted. It is interesting to note that all activities compose a
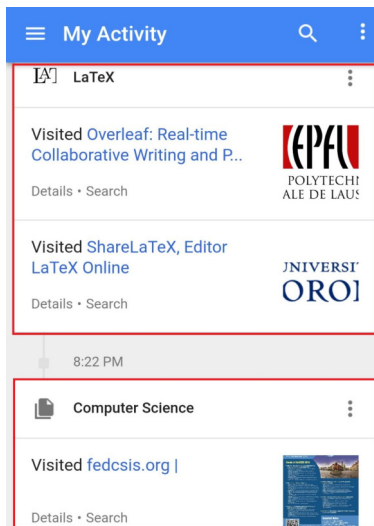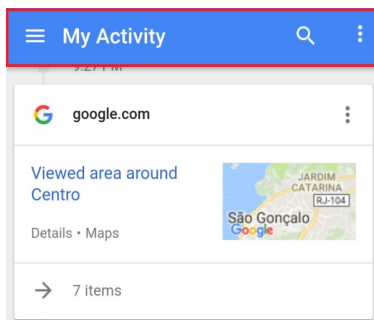
Fig. 3. Example of MyActivity 'Cards'.

Fig. 4. Example of MyActivity Persistent Menu.

record history, so that these records are distributed in their respective days. Each card, which may be an activity or a grouping of records, is linked by a timeline with an associated timestamp. Figure 3 shows an example of cards.

- **Persistent Menu** allows the user to access a set of options at any time during their interaction with the tool. See the example in Figure 4.
- **Modals** are alerts usually displayed to confirm user actions, with texts that briefly describe their consequences. Figure 6 shows an example of a Modal in MyActivity.
- **Search Filters** allow the user to refine the view of records by context. For example, in MyActivity, you can select service types, date range, and more. In Privacy Badger, you can choose the status associated with the domain, for example. Figure 5 shows the example of a filter used by the Privacy Badger.
- **Sliding Buttons** represent user preferences regarding his privacy, i.e., what can be seen about him by entities interested in his personal data. Figure 7 shows examples of Sliding Buttons in Privacy Badger.

B. Communication Strategies in MyActivity

Based on the mapping of metalinguistic, static and dynamic signs, and their respective metamessages, the communication strategies identified in the MyActivity tool were:

- **CS1:** Provide different categorizations of the collected data;
- **CS2:** Show monitored activities and the level of use of products/services;
- **CS3:** Offer forms of intervention on data collection;
- **CS4:** Provide alerts of actions performed by the user;
- **CS5:** Provide means to report problems or collaborate with ideas.

About Communication Strategy (CS1), MyActivity, linked to MyAccount, gives the user a list of categories of data that can be collected about him, such as 'Location History' or 'Device Information'. For each category, a brief description of the purpose of the collection is presented. If he would like more information on this, the 'learn more' link will direct him to the 'Help' page. From this categorization, MyActivity also provides more specific categories of user activity on Google services, such as "Feedback no interest in YouTube" and "Location responses". In doing so, we believe that design intent was to provide a means for the user to have a comprehensive view on what types of data can be collected by Google, as well as to understand the company's objectives in acquiring the data of its users. In our interpretation we evaluated that this strategy points to some aspects of the concept of Legibility in HDI, since the tool makes available to the user information about who is monitoring him, the means used to perform the data collection, what types of data will be collected, and the intended interest in that process. However, we did not find explanations about the algorithms and methods used to generate inferences from the user data.

About Communication Strategy (CS2), MyActivity allows the user to access and review the history of activities he performed. By default, the most recent activities are displayed at the beginning of the history, i.e., in reverse chronological order. In addition, the user has access to the 'details' option that provides explanations on how monitoring is performed. The search engine helps the user find a specific set or activity in their activity history, giving the user better navigability, since a large number of records are expected. MyActivity also offers two types of activity view: packet-based (records are listed individually), or product/service-based (groups a sequence of records by their respective product/service). He can also check how often he uses the services each day. The activity history query should be based on its category. However, initially this may not be communicated appropriately to the user, since, by default, when accessing MyActivity, activities related to a category, usually 'Web and Apps Activity', are listed. Hence, we can observe indications of the concept of Legibility in this strategy, since the tool provides resources that inform the user about the capture of his activities in Google services and products.

Fig. 5. Menu filter in Privacy Badger.

About Communication Strategy (CS3), MyActivity, linked to MyAccount, offers the user ways to intervene on the data collected about him. For each data collection category, he can define whether monitoring is enabled or not. By default, some monitoring types are already enabled and others are disabled (paused). However, the user has the right to intervene, at any time, on the type of monitoring that he wants to pause or enable. The tool uses color to distinguish monitoring status, applying grayscale to 'paused' monitoring, and color to 'enabled' monitoring. In this case, we understand a possible intention to use colors is to communicate to the user about possible benefits from collecting their data. MyActivity also allows the user to delete activity logs. In this case, excluding records implies disregarding such data in the aggregation and processing made by Google. Therefore, we can perceive in this strategy the perspective addressed by the concept of Agency, so that the user has mechanisms to determine what types of data can be accessed and collected, as well as how his records are generated and deleted.

About Communication Strategy (CS4), MyActivity provides alerts when the user wants to take actions on his data, such as pausing/enabling a type of monitoring or deleting an activity, for example. These notifications tell the user about the implications of the intended action, generally seeking to encourage him to make his data available to Google, showing how this reflects benefits in delivering services or products. In addition, clarifications on how to collect and store such data are also communicated, thus proposing clarity and transparency, aiming at user confidence. From this, we can notice a link with the concept of Legibility, because this strategy seeks to deal with the user's concerns about his data and the processing that is performed from them.

About Communication Strategy (CS5), MyActivity presents a space that offers the user a feature to report usage problems or errors in the activity log. In this way, however, it is necessary to wait for the analysis of the request before it can be acknowledged. This mechanism also acts as a channel for sharing ideas or suggestions. Thus, we can establish a link with the concept of Agency, because this strategy allows the user means to inform and correct the data provided.
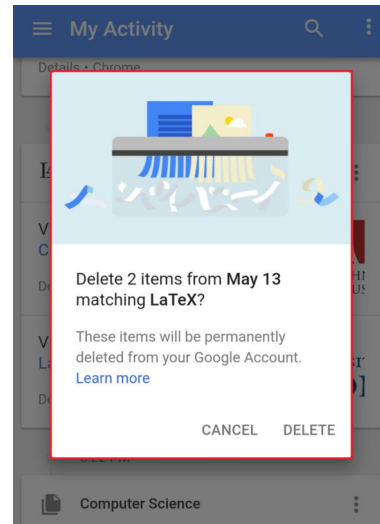

Fig. 6. Modal Example in MyActivity.

C. Communication Strategies in Privacy Badger

Based on the mapping of metalinguistic, static and dynamic signs, and their respective metamessages, the communication strategies identified in the Privacy Badger tool were:

- **CS6:** Offer forms of intervention regarding the use of data;
- **CS7:** Show the third-party domains identified.

About Communication Strategy (CS6): In general, the content of a web page can come from a number of different sources, i.e., third-party domains. On an e-commerce page, for example, the display of products will be carried out by a virtual store, the search engine can be from a company hired to make this service available and the ads will be from an advertising company. In this way, Privacy Badger analyzes the behavior of all identified third-party servers as the user navigates through different domains, applying one of the following statuses: 'Blocked', 'Partially Blocked' and 'Allowed'. If any domain is attempting to monitor the user's browsing record via cookies without permission, then PB will automatically block the content coming from that server. If this domain is providing an important content type for the page to work, then Privacy Badger will allow connections to this server, but will block its cookies injection in the browser. Finally, Privacy Badger will allow third-party content to be injected if no monitoring activity is detected.

The tool also allows the user to modify the applied status, thus offering the power to intervene in the result of the classification algorithm. In this way, this domain will be classified according to the decision of the end user. In addition, PB offers a local white-list where the user can add domains he trusts, so that they are out of the Privacy Badger analysis. Thus, we can observe a correlation with the concept of Agency in supporting the user to exercise control over access to their browsing data. However, the tool does not allow the user to
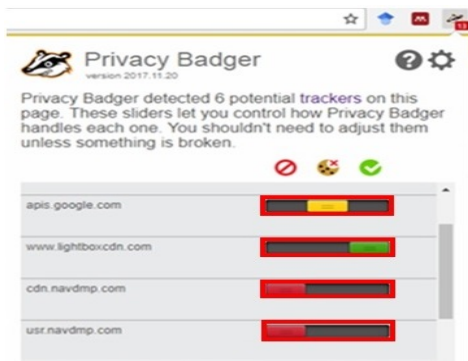
Fig. 7. Sliding Button in Privacy Badger.

define which types of data he wants to make available, for example.

About Communication Strategy (CS7): The Privacy Badger provides a list with the address and status of each identified third party domain, allowing the user to be aware of which third-party domains are active, as well as those who have attempted to perform some kind of hidden monitoring. However, the tool failed to communicate information about the identified domains, such as the type of content or what functionality a particular server was attempting to enter, which could assist user understanding of how third-party domains act. Thus, we can point out, in this strategy, a reference to the concept of Legibility based on the identification of which domains have attempted to collect the user data. However, the user will not be informed about some important aspects of the data collection, storage and processing, such as the purpose of the data collection or what types of data, in fact, can be collected about him.

D. Comparative Analysis between S1 and S2

One of the main challenges observed during the inspections was the partial application of the aspects advocated by the concepts of HDI, pointing to possible barriers or limitations in the adoption of human factors in relation to the use and storage of personal data by third parties. On the other hand, we have identified the possibility of enabling Human-Data Interaction without all traces of its concepts being present.

Regarding the concept of Legibility in HDI, we identified some of its aspects present in Communication Strategies CS1, CS2 and CS4 (Study S1) and in Communication Strategy CS6 (Study S2). We understand that this concept represents the first step in enabling Human-Data Interaction, setting guidelines that support people's understanding of the actions of third-party that are interested in their personal data. The communication problems encountered may point to resistance in making more transparent the algorithms used to infer new knowledge about people. This corroborates Mortier's observation in [7], when mentioning the conflict in making public such algorithms that are intellectual property of companies. However, it was possible to identify important mechanisms related to Legibility aspects, such as the classification of

monitored data, the identification of those who want to access personal data, the forms of collection used and the intended goals of the interested parties. However, from the possible communication problems reported in S2, we noticed the more timid use of this concept by the Privacy Badger tool.

In relation to the concept of Agency in HDI, traces of this concept were also perceived in both tools. We can see that the comprehensive use of Legibility offers a more propitious context for the adoption of Agency aspects. In other words, if people are not aware of who they are interested in collecting their data, what their intended data are, their collection methods and desired goals, then the ability of people to act on their data is limited, so that there will be no important information available to users in order to support them in their decisions or to assist in the creation of more adequate criteria and controls on the access and use of their data by third parties. Therefore, we note that the MyActivity tool communicates these aspects more clearly, reflecting the concept of the Agency more comprehensively.

Finally, no traces of the application of the aspects related to the Negotiability concept were found in the strategies identified. This may point to the difficulty in defining mechanisms that allow identifying characteristics that are expressed in data and are likely to change over time, such as individual attitudes and interests, for example. However, these issues are relevant in the context of processing and generating inference from personal data. The Table 1 summarizes the answers by each research question.

TABLE I
SUMMARY OF THE ANSWERS TO OUR RESEARCH QUESTIONS
OBTAINED IN S1 AND S2.

RQ(i)	RQ(ii)
CS1: Provide different categorizations of the collected data.	Legibility
CS2: Show monitored activities and the level of use of products/services.	Legibility
CS3: Offer forms of intervention on data collection.	Agency
CS4: Provide alerts of actions performed by the user	Legibility
CS5: Provide means to report problems or collaborate with ideas.	Agency
CS6: Offer forms of intervention regarding the use of data.	Agency
CS7: Show the third-party domains identified.	Legibility

V. Triangulation

Triangulation is a standard procedure in the validation of qualitative research results [24]. In this case, we validated the results obtained from applying the SIM (S1 and S2), comparing them with the results obtained from a set of interviews and observation sessions with users, thus seeking to identify convergences and divergences, ensuring the scientific validity of the results.

The triangulation step counted on 5 (five) participants. Before starting the tests, a consent form was presented to the

participants, with information about the data collection and use conditions performed in this research. With due acceptance, the next step was to make a brief presentation of HDI for each participant, exploring the three main concepts proposed by Mortier *et al.* [7] . Then the tests were started.

We also used the inspection scenario described in sections III-A and III-B, with some adaptations in their tasks to carry out this study with participants, in order to generate a good comparability of the results obtained in S1 and S2. After completing each task, participants responded verbally to a post-test questionnaire. With the participants' speech, it was possible to identify and highlight the convergences and divergences, regarding research questions proposed in this paper. The convergences and divergences identified will be presented in the following subsections A and B. The Table 2 summarizes these results obtained in the triangulation step.

TABLE II
Summary of the convergent and divergent results identified in the triangulation step.

CSs	Convergences		Divergences	
	RQ(i)	RQ(ii)	RQ(i)	RQ(ii)
1	P1, P2, P3, P4, P5	P1, P2, P3, P4	-	P5
2	P1, P2, P3, P4, P5	P1, P2, P3, P4, P5	-	-
3	P1, P2, P3, P4, P5	P1, P2, P3, P4, P5	-	-
4	P2, P3, P4	P2, P3, P4	P1, P5	P1, P5
5	P2, P3, P4, P5	P3, P4, P5	P1	P1, P2
6	P2, P3, P4	P2, P3, P4	P1, P5	P1, P5
7	P2, P3, P4	P3, P4	P5	P1, P2, P5

A. Convergences

In this section will be presented, for each communication strategy, the discourse excerpts that show the convergences identified, related to the answers questions RQ(i) and RQ(ii).

As evidence for the CS1, provide different categorizations of the collected data, the following discourse excerpts were collected:

P1: "From my point of view, Google is keeping an eye on my interests, especially those that move my everyday life. So it seeks to monitor my clicks, texts, videos, comments and where I went. This is related to the concept of Legibility."

P2: "... I realize that Google wants to know your location, the places you've visited, what you search for (search engines, Youtube and Play Music) because this will help them in their recommendations. This is related to the concept of Legibility because the tool shows what Google is interested in doing with this data, but not in depth. That is, it tells what to do and with what, but does not say how it will do."

P3: "You can easily see what data Google has about you. In this case, it is the concept of Legibility involved."

P4: "Google has pretty much everything about me, like location, places I've visited and the time that happened, what I watched or did on Youtube, my vocal signature... The Legibility is not completely applied because I realize that there is still a certain lack of transparency on forms of using my data and its purposes."

P5: "Yes, for example the location, types of songs, what I see on Youtube."

As evidence for the CS2, show monitored activities and the level of use of products/services, the following discourse excerpts were collected:

P1: "I can see what Google has recorded about me. This is associated with Legibility, since there are texts or words like "Learn More" that help you understand what has been recorded."

P2: "It is possible to visualize, including the circuit that the user performed in a locality, a kind of history of how you visited a place. Thus, it is possible to associate with the concept of Legibility because you are aware about what it is monitoring."

P3: "It is possible to consult the records by means of histories, such as the one of location, that allows to identify even the route accomplished, in an easy and organized way for the user. So, this view fits into the concept of Legibility, because it is very clear what was recorded i.e., a query."

P4: "MyActivity shows the history of all these types of activities ... so it's Legibility."

P5: "MyActivity logs (in the case of Chrome), not only which site I've visited, but which sections of the site I passed. Here I think it's Legibility because it shows exactly what I did."

As evidence for the CS3, offer forms of intervention on data collection, the following discourse excerpts were collected:

P1: "I didn't trust too much in controlling my data because we don't have a policy or something that makes that control more present in our everyday lives. However, with the option of downloading my data, I was more reassured. This is tied to the concept of Agency."

P2: "I realized that in some cases these options are hidden. But you can control what they can and can't access your data. It's related to the Agency, by allowing action on the data."

P3: "It is possible, for example, to both exclude and prevent them from continuing to monitor you. So there are Agency these options that give you control over your data."

P4: "I can erase history, intervene in some things ... the minimum exists, which is the case of being able to delete, modify, allow or deny. Agency, but in this case, I can not manage for third parties, but only for myself."

P5: "It does provide an option of what I can release or not. I think it's Agency, because it's a way to conduct management over my data, it's you showing what you want to happen."

As evidence for the CS4, provide alerts of actions performed by the user, the following discourse excerpts were collected:

P2: "The pros and cons of letting them monitor their activities are clear. These alerts reinforce what your action will cause, not allowing the user to simply take action, forcing the user to heed it. Thus, the concept of Legibility appears again."

P3: "They provide more in-depth information about your data. Soon, it becomes Legibility."

P4: "It gives me more in-depth information when I click activate or when I want to pause. Legibility."

As evidence for the CS5, provide means to report problems or collaborate with ideas, the following discourse excerpts were collected:

P2: "It is possible to report problems, but it is not known whether this will be answered or not."

P3: "MyActivity even allows screenshots to be sent to help identify the problem. They even allow modifications for legal reasons, and as it is a global company, then it is important to have that same option. It is related to Agency."

P4: "It offers a way to report legal issues, but no issues with third parties. However, the screen passes generic information so I guess it could be for any kind of problem as well. On the other hand, the system provides a 'learn more', so maybe it is the case to go and read more about it. It would be the Agency."

P5: "In case it is the feedback option. This option allows me to participate, either to complain or to suggest or make some kind of contribution. So I think it's more related to the Agency than to the other two concepts."

As evidence for the CS6, offer forms of intervention regarding the use of data, the following discourse excerpts were collected:

P2: "I've been able to distinguish the intentions of third-party domains by colors and by some domain names, i.e. those that are interested in my behavior or not. This identification can relate to the concept of Legibility."

P3: "The total he gives you easily before you click. And when you click, it gives you the names, so it's very easy to identify these things in it. This relates to Legibility (a bit), because it gives you just that this domain is trying to monitor you, but you do not know what it's registered for, i.e., it's the minimum level of Legibility."

P4: "I found it very good to be able to identify because he showed me many domains that may be monitoring me, but on the other hand lacked more transparency because he does not describe very well the purpose of these possible trackers are acting. I think if I knew that, I could allow monitoring if it was to benefit people. It is related to the principle of Legibility."

As evidence for the CS7, show the third-party domains identified, the following discourse excerpts were collected:

P2: "I can enable and disable, for example, third-party domains. This is related to the concept of Agency. But the concepts applied here are weak, because it could provide more information about what each third-party domain wants to do, or what behavior the third-party domain has presented to the PB to block."

P3: "I managed to block some domains, but since it does not provide information on the consequences of this action, it might impact the functioning of the page. It is related to the concept of Agency, but only to the part of managing access."

P4: "The Privacy Badger allows me to perform the blocking or the release, so it is the minimum of management. Therefore, it is associated with the Agency concept."

B. Divergences

The divergences identified in the triangulation step can be classified into two cases: The first case comprises the participants who didn't identify some communicative strategy and, therefore, couldn't answer the RQ(i) and RQ(ii). The second case comprises the participants who were able to identify communication strategies as mentioned in the SIM, but associating them with other concepts of HDI, different from those pointed out during the SIM. In this case, the perceived divergences apply only to RQ(ii).

Regarding the first case, some communicative strategies were not perceived by a few participants, for many reasons. The examples identified were: P1 and P5 had contact with CS4. However, they did not consider it as an alert, but rather as another textual information presented in the view. P1 was also unable to locate CS5 nor understand CS7, as follows: *"The identified names of third-party domains made no sense to me."*. In addition to these examples, P5 couldn't to evaluate the Privacy Badger, because this tool did not offer the language desired by the user, as evidenced in his report: *"The interface does not make sense to me, because I do not know English, so I would not manipulate this program."*

Regarding the second case, for example, from the textual description presented for each activity category in MyActivity through CS1, P5 understood that activity monitoring allows users to better understand themselves over time through a possible processing of their historical records. This new understanding may be relevant in subsequent data exchanges, allowing the sharing of new reassessments in relation to their behavior or interests expressed in data, with others interested in their data. So, P5 linked CS1 with Negotiability concept, as reported: *"I think it's Negotiability, because we may want to be monitored for certain types of activities at some point, maybe for benefits. By knowing what types of activity I am monitored, it allows me to set up your profile and perceive, over time, changes in behavior or interests about me. It may be interesting that other people might know about these changes."*

In a second example, P2 considers that CS5 is related to the concept of Negotiability, because it understood that reporting problems can serve not only to report system failures, but also to express its considerations about the use of its data by third parties, as reported: *"It may have to do with Negotiability, because it allows you to negotiate about your data, saying what you disapprove of the use of data."*

There were two other divergences over CS7. P1 considers it to be related to the concept of Legibility because Privacy Badger automatically performs a possible block on a third party domain. Thus, the tool suggests that it is not necessary for the user to modify such controls, but only to observe the type of constraint applied to a particular third-party domain, as reported: *"The option to block a third-party domain may refer to the concept of Legibility because it is in my discretion to block it or not."* Finally, P2 considers that CS7 is linked to the Negotiability concept because *"it allows me to not only block or enable, but make a middle ground by blocking*

only the cookie and letting the third party domain perform its functionality on the page."

VI. CONCLUSION

This work aimed to identify and evaluate the communicability of strategies that enable Human-Data Interaction, through the application of the main HDI concepts proposed by Mortier *et al.*[7], in the context of personal data management. For this, the Semiotic Inspection Method was used to inspect the MyActivity and Privacy Badger tools. A methodology was established based on Semiotic Engineering theory [14], contributing to the consistency of the evaluation and results obtained.

The tools were inspected by two evaluators and, through the results obtained, it was possible to answer the research questions raised in this work. We found out evidences (see section IV) of high and low communicability in some Strategies to communicate Legibility, Agency and Negotiability. For instance, in CS1, because they are not explaining the algorithms and methods used in generating inferences from the user data (low legibility). In CS3, because the user has mechanisms to determine what types of data can be accessed and collected (high agency). The results also allowed us to present problems related to the partial application of the inherent aspects to the concepts of Legibility and Agency. However, it was observed that, even without the identification of all concepts related to HDI, it was possible to observe traits that allow Human-Data Interaction.

This paper brings three main contributions. The first is the application of SIM in an yet unexplored context of HDI related to the use of TETs, showing that the application of the method was relevant in that context. The second contribution is the identification of a set of communicative strategies and the classes of signs used by designers to make the HDI feasible. Such strategies may support designers of other TETs in their decisions about which strategies to use. Finally, the research results of the model proposed by Mortier *et al.* [7] in association with the application of the SIM, i.e., the inspection and evaluation by model, has shown to be promising in the HCI evaluation of applications that seek to provide means for Human-Data Interaction. The results presented here were validated (as is typical in validation of qualitative research) through an endogenous triangulation [25]. This motivates us to carry out new empirical studies with HDI designers to explore the practical effects of designing TETs with the communication strategies.

VII. ACKNOWLEDGMENTS

REFERENCES

[1] H. Hornung, R. Pereira, M. Baranauskas, and K. Liu, "Challenges for human-data interaction–a semiotic perspective," in *International Conference on Human-Computer Interaction*. Springer, 2015, pp. 37–48.

[2] J. Han, H. Ding, C. Qian, W. Xi, Z. Wang, Z. Jiang, L. Shangguan, and J. Zhao, "Cbid: A customer behavior identification system using passive tags," *IEEE/ACM Transactions on Networking*, vol. 24, no. 5, pp. 2885–2898, 2016.

[3] C. Meurisch, U. Naeem, M. A. Azam, F. Janssen, B. Schmidt, and M. Mühlhäuser, "Smarticipation: intelligent personal guidance of human behavior utilizing anticipatory models," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, 2016, pp. 1227–1230.

[4] A. Doryab, M. Frost, M. Faurholt-Jepsen, L. V. Kessing, and J. E. Bardram, "Impact factor analysis: combining prediction with parameter ranking to reveal the impact of behavior on health outcome," *Personal and Ubiquitous Computing*, vol. 19, no. 2, pp. 355–365, 2015.

[5] F. Zhang, N. J. Yuan, K. Zheng, D. Lian, X. Xie, and Y. Rui, "Mining consumer impulsivity from offline and online behavior," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 1281–1292.

[6] R. Mortier, H. Haddadi, T. Henderson, D. McAuley, and J. Crowcroft, "Challenges & opportunities in human-data interaction," *University of Cambridge, Computer Laboratory*, 2013.

[7] R. Mortier, J. Crowcroft, D. McAuley, H. Haddadi, and T. Henderson, "Human-data interaction: The human face of the data-drivem society," 2014.

[8] E. W. Ritter and S. J. Rigo, "Fitdata: A system for monitoring physical activity based on mobile devices," in *Proceedings of the XII Brazilian Symposium - Volume 1*, ser. SBSI 2016. Porto Alegre, Brazil: Brazilian Computer Society, 2016. ISBN 978-85-7669-317-8 pp. 72:550–72:557.

[9] C. Buck and S. Burster, "App information privacy concerns," *AIS Electronic Library - Americas Conference on Information Systems*, 2017.

[10] C. Buck, "Stop disclosing personal data about your future self," *AIS Electronic Library - Americas Conference on Information Systems*, 2017.

[11] J. Angulo, S. Fischer-Hübner, T. Pulls, and E. Wästlund, "Usable transparency with the data track: a tool for visualizing data disclosures," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2015, pp. 1803–1808.

[12] C. Bier, K. Kühne, and J. Beyerer, "Privacyinsight: the next generation privacy dashboard," in *Annual Privacy Forum*. Springer, 2016, pp. 135–152.

[13] J. Siljee, "Privacy transparency patterns," in *Proceedings of the 20th European Conference on Pattern Languages of Programs*. ACM, 2015, p. 52.

[14] C. S. De Souza, *The semiotic engineering of human-computer interaction*. MIT press, 2005.

[15] C. F. and Leitão and C. S. De Souza, "Semiotic engineering methods for scientific research in hci," *Synthesis Lectures on Human-Centered Informatics*, vol. 2, no. 1, pp. 1–122, 2009.

[16] F. Cafaro, "Using embodied allegories to design gesture suites for human-data interaction," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 560–563.

[17] A. Crabtree and R. Mortier, "Human data interaction: historical lessons from social studies and cscw," in *Proceedings of the 14th European Conference on Computer Supported Cooperative Work, 19-23 September 2015, Oslo, Norway*. Springer, 2015, pp. 3–21.

[18] C. S. de Souza, C. F. Leitão, R. O. Prates, and E. J. da Silva, "The semiotic inspection method," in *Proceedings of VII Brazilian symposium on Human factors in computing systems*. ACM, 2006, pp. 148–157.

[19] C. S. de Souza, C. F. Leitão, R. O. Prates, S. A. Bim, and E. J. da Silva, "Can inspection methods generate valid new knowledge in hci? the case of semiotic inspection," *International Journal of Human-Computer Studies*, vol. 68, no. 1-2, pp. 22–40, 2010.

[20] R. O. Prates, C. S. de Souza, and S. D. Barbosa, "Methods and tools: a method for evaluating the communicability of user interfaces," *interactions*, vol. 7, no. 1, pp. 31–38, 2000.

[21] S. Lewis, "Qualitative inquiry and research design: Choosing among five approaches," *Health promotion practice*, vol. 16, no. 4, pp. 473–475, 2015.

[22] "Google myactivity," https://myactivity.google.com/, acessado em 12/11/2017.

[23] "Privacy badger," https://www.eff.org/privacybadger, acessado em 10/01/2018.

[24] N. K. Denzin and Y. S. Lincoln, *The landscape of qualitative research*. Sage, 2008, vol. 1.

[25] J. W. Creswell and J. D. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications, 2009.