A Dashboard for System Trustworthiness: Usability Evaluation and Improvements

Diego Camargo, Felipe Nunes Gaia, Tania Basso, Regina Moraes
University of Campinas - UNICAMP
Campinas, Brazil
email:{kmargod, felipegaia.comp}@gmail.com, {tbasso@cotil, regina@ft}.unicamp.br

***Abstract*—Dashboards are used to organize and display important information in a way that must be well-arranged, understandable and easy to read. Thus, the success of the dashboard depends on its usability. In this paper, we present the design and usability evaluation of a dashboard developed for the visualization of system trustworthiness properties, the relationship among them and their relevance in the composition of a trustworthiness score over usage time. The evaluation was performed to understand and support the usability regarding human perception in its operation. Security and Information Technology (IT) specialists feedback was sought throughout the process and was obtained from a usability testing and a questionnaire for user interaction satisfaction. Results revealed usability concerns regarding design and content, which led to improvements that were implemented in the dashboard.**

Keywords*—*Dashboard; Trustworthiness; User experience; Usability.

I. Introduction

A dashboard is a tool used for information management and business intelligence. Data dashboards organize, store, and display important information from multiple data sources into one, easy-to-access place.

If a dashboard is properly designed, it can help to understand the semantics of visualized information. Then, data can be easily transformed by a user to information and knowledge used for specifying tasks. On the contrary, improper design of a dashboard can lead to difficulties in its use, incomprehension of visualized data and unsuitable tasks specifications. Consequently, users avoid or quit using the dashboard [1].

In this work, we present a dashboard for trustworthiness assessment and a usability evaluation on it. The dashboard presents a trustworthiness score of a cloud application, as well as the scores of intermediate trustworthiness properties (e.g., security, privacy, dependability, isolation, scalability, among others). It allows users to interact with the monitoring and assessment of these properties and to input or modify the configuration values (e.g., weights, thresholds) to improve their scores. The dashboard was developed in the context of the ATMOSPHERE (Adaptive, Trustworthy, Manageable, Orchestrated, Secure, Privacy-assuring Hybrid, Ecosystem for Resilient Cloud Computing) project [2], a collaborative project between Europe and Brazil, whose main objective is to provide a solution for assessing the trustworthiness of cloud applications that handle large volumes of data.

The usability evaluation is an extension of a previous work [3]. We conducted a usability assessment of the dashboard with specialists through usability testing and satisfaction questionnaires in two rounds. After each round of usability testing, we revised and improved the dashboard in response to usability weakness findings before the next round of testing, until the majority of participants expressed high satisfaction. Even so, other improvements have been made to address minor usability concerns.

This paper is organized in six sections. After the first, the Introduction, Section 2 presents Background that is mandatory to understand the paper, including the dashboard importance. Related Work is presented in Section 3 and Section 4 presents the Usability Evaluation of the dashboard. Section 5 presents some Discussions and the Improvements that were performed according to the evaluation process. Conclusions and Future Works follow in Section 6.

II. Background and Related Work

This section addresses, briefly, the issues that underpin this work. It discusses the need of dashboards, as well as describes the trustworthiness dashboard used in this study and usability tests.

A. The need of dashboards

Nowadays, large volumes of information are generated by several devices connected to the Internet. Companies and even governmental organizations are interested in holding these data because it is a source of very important information that can, for example, affect the operational efficiency of that organization, increase profits, identify customers profiles, and cut unnecessary expenditures that waste the budget. In this scenario, data mining has received a lot of attention due to its strong ability of extracting meaningful information from data.

Besides mining the data, it is necessary to present the process results to users and analysts. One of the tools used for this purposes is a dashboard. It helps to summarize data obtained by the data mining process, providing a quick view of results or actual state of the activities.

Dashboards are nowadays widely used for monitoring and analysis of business processes. Numerous companies such as IBM [4], SAP [5], Tableau Software [6], to name a few well-known vendors, offer complete Business Intelligence (BI) or information visualization solutions. Nevertheless, these

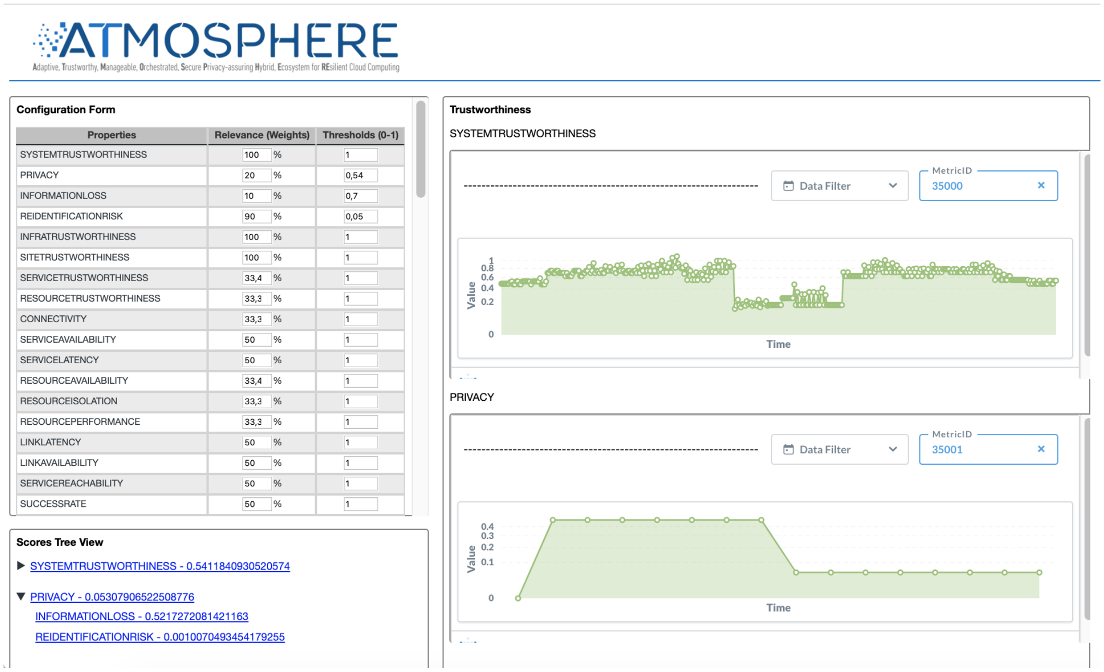

Figure 1. A first release of the trustworthiness dashboard [3]

approaches do not always integrate with specific applications, which requires a specific dashboard development.

B. The dashboard for trustworthiness assessment

As mentioned before, we developed a dashboard to present the scores of trustworthiness properties (and also intermediate scores from attributes composing the properties) for applications in cloud environment. It was developed to show the information provided by a specific solution, which could not be integrated to standards dashboards approaches.

The dashboard is based on quality model [7]. Basically, in this model, the root represents the **trustworthiness score**. The leaves represent a set of quantifiable **attributes** chosen to characterize the system (e.g., memory usage, throughput). When these attributes represent input measures, they must be normalized by applying adequate functions. For that, the definition of **thresholds** is necessary once they specify the maximum and minimum values for the inputs of the leaf-level components of the quality model.

The values for each component are influenced by an adjustable element **weight**, which specifies a preference over one or more characteristics of the system, according to established requirements (e.g., in certain contexts memory usage might be more important than throughput). The final score is computed using the aggregation of the weighted values of the attributes, starting from the leaf-level towards the root attributes, using **operators** that describe the relation between them. A first release of this dashboard was presented in [3]. More details about its requirements can be found in that reference.

It is important to mention that the input measures are provided by the *Trustworthy Data Management Services (TDMS)*. TDMS is a component of the ATMOSPHERE project similar to a database service in cloud systems dealing with mechanisms for data storage, access and management and it also considers trustworthiness properties. The trustworthiness-related information is obtained and stored according to definitions of quality models, including their weights and thresholds. More details of quality TDMS can be found in [8].

The dashboard for trustworthiness assessment was implemented using the Metabase tool [9] because it is open source and can be used for deployment on any system that is running Docker. As the dashboard requires user interaction and Metabase is not able to update information into the database, we developed a dynamic form that obtains the quality model configuration data defined by the users and saves (or updates) them in the database through REST services.

Figure 1 shows the first release of the trustworthiness dashboard. In the upper left corner, the *Configuration Form* allows users to select the attributes and configure weights, thresholds and periodicity of data collection. In the bottom left corner, the *Scores Tree View* allows the navigation through the scores. This navigation (drill down) is provided by a recursive algorithm implementation.

The calculation of the scores is made by the properties and the dashboard presents the historical scores according to the time configuration. The user evaluates the scores through the respective charts and can change the configuration (weights and thresholds) if necessary to better represent the trustworthiness composition.

 ISBN: 978-1-61208-761-0

C. Quality Models

Data privacy is one of the properties that makes up a trustworthiness system. Figure 2 illustrates the privacy quality model used in this work. Two main attributes are considered to compose data privacy: the re-identification risk and the information loss. Re-identification risk is the probability of discovering an individual by matching anonymized data with publicly available information. Information loss is the amount of information that can be obtained about the original values of variables in the input dataset. Both measurements are obtained through the use of anonymization techniques tools. More details about the privacy quality model and its attributes (thresholds, weights, normalization values) can be found in the work of Basso et al. [10].

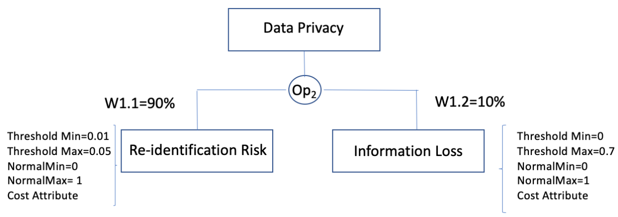

Figure 2. Privacy Quality Model Instance [10]

Figure 3 illustrates the main attributes of the other quality model used in this work. The System Trustworthiness Quality Model defines all the attributes involved in the trustworthiness score of a cloud-based application. Mainly, it is composed of the following (sub) quality models:

- Infra Trustworthiness- responsible to assess the trustworthiness of the cloud infrastructure (hardware and software resources available);
- Data Management Trustworthiness - where the trustworthiness of storage data is described;
- TDPS (Trustworthy Data Processing Services) - responsible to define the attributes of the services that are running to provide the expected results to the users.

The (sub) quality models are composed using a neutrality operator aiming at obtaining the score of trustworthiness of the system under analysis. Each one, per se, is a complex quality model composed of several attributes and sub-attributes that, for sake of simplicity, we did not represent in Figure 3. However, we briefly describe these attributes below.

The Infra Trustworthiness Quality Model is composed of the Site and the Virtual infrastructure trustworthiness. The Site is composed of Services, Resources and Connectivity and several sub-attributes, i.e., Service Availability, Service Latency, Resource Availability, Resource Performance, Resource Isolation, Link Latency and Link Availability. Some of them have their scores assessed at runtime (for example, Service Availability) and others as static metrics (for example, Resource Isolation).

Virtual Infra Trustworthiness is related to the cloud virtualisation services and resources and it is composed of used and free Central Processing Unit (CPU) and Memory, scalability capacity and CPU Isolation.

Data Management Trustworthiness is defined to assess the score related to data storage and recovery. This score is strongly influenced by the engine used and in the context of the ATMOSPHERE project, based on Vallum [11] or in a more common engine (i.e., MySQL). In any case, it has as attributes Performance (Response Time, Throughput and Bandwidth), Security (Attestation and Confidentiality), Fault Tolerance (Replication) and Data Privacy.

The TDPS is related to services provided by the application. In this context, each service has its own score and the TDPS general score is a composition of all the services executed by the user application. The attributes considered in this case are Fairness, Transparency, Stability and Data Privacy.

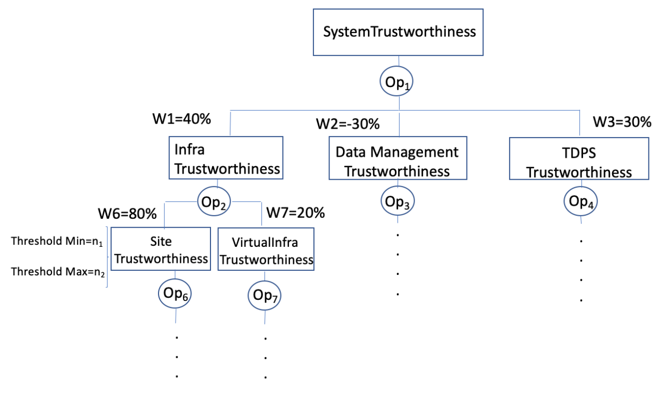

Figure 3. System Trustworthiness Quality Model Instance

Both these quality models (Privacy and System Trustworthiness) are displayed in the tree structure of the dashboard (see Figure 1 - lower left corner). An expanded view of the tree with the attributes and sub-attributes can be seen in Figure 5.

D. Usability Testing

As the dashboard is a data visualization tool, it must provide easy use and understanding of information semantics. Otherwise, the produced visualizations can be misguiding and, as a consequence, may lead to wrong conclusions. In addition, if users have difficulties in using the dashboard they can avoid or never use it. Thus, usability is an important issue that must be addressed while developing this tool.

The concept of usability is related to software quality, in terms of ease of use and learning. Brazilian Norm (NBR) 9241-11 [12] defines usability as a measure in which a product (software and hardware) can be used by specific users, in a specific context of use, to achieve specific objectives with effectiveness, efficiency and satisfaction. Nielsen [13] defines usability as a set of factors that qualify the user's interaction with the software (e.g., user control, easy to recall, efficiency, among others). These factors are related to the ease of use and learning to use the system.

According to Barnum [14], usability testing is *the activity that focuses on observing users working with a product, performing tasks that are real and meaningful for them*. There are two types of methods that can be used to assess the usability of interfaces: inspection (or analytical) methods and empirical methods. Inspection methods are those in which one or more evaluators examine the interface, judging it for usability problems, without the need to verify the interaction of real users with the system (for example, evaluation based on heuristics). Empirical methods, on the other hand, are those in which real users participate, interacting with the system being evaluated, while evaluators perform the analysis of such interaction and the problems found [15].

A *Questionnaire* is a widely used technique designed for the assessment of perceived usability, used as support for inspection and empirical methods mentioned above. Typically, a questionnaire has a specific set of questions presented in a specified order using a specified format with specific rules for producing scores based on the answers of respondents [16].

It is important to mention that there are some other techniques for usability testing, such as *heuristic evaluation, cognitive walkthroughs, and think aloud*. These techniques are out of the scope of this work, but they should be addressed in future work.

E. Related Work

Usability testing techniques have been applied to evaluate dashboards in the most diverse contexts. For example, Chrisna et. al [17] conducted a study that included user observations, heuristic assessment and a survey among users for a Business Intelligence (BI) application. Magdalena et al. [1] defined a strategy, based on user testing and heuristic evaluation, to provide improvements and, consequently, increase the use of the BI dashboard in their company (one of the biggest airlines in Indonesia). Lavalle et al. [18] presented an interactive dashboard to allow non-expert users to be guided towards specific data visualizations regarding tax collection. They used a questionnaire to evaluate the dashboard usability. Read [19] used the think aloud technique to develop a dashboard. Based on user behavior and comments about the system that were noted by the research team, the evaluation helped to understand the usability of the (navigational) menu layout for the design of the system.

It is obvious that each dashboard is designed and constructed according to the specific business needs from the company or organization, including data and users profiles, which requires respective specific usability evaluations. To the best of our knowledge, there is no dashboard for trustworthiness evaluation in cloud computing applications, neither a study regarding dashboard usability in this context. This work aims to help fill this gap.

III. Usability Evaluation

In order to evaluate and improve the trustworthiness dashboard, we performed two sprints of evaluation. In the first sprint, a preliminary (pilot) test was performed with three specialists on security and privacy. Based on this evaluation, some improvements were implemented in the dashboard interface to prepare a more complete validation with a larger number of users. In the second sprint, 22 IT specialists, including professionals involved in the ATMOSPHERE project, evaluated the dashboard interface. It is important to mention that, before this whole process, the validation methodology was designed and submitted to the Research Ethics Committee (*Plataforma Brasil*) for the necessary authorization.

The validation methodology is composed of (i) a user testing, which specifies a dashboard usage scenario so that users exercise the scenario and answer some essay questions; (ii) two multiple choice questionnaires. We decided to use both these evaluation techniques because they are effective for reaching a wide audience, since the professionals interviewed are from different countries. Also, their cost is low and they are quite time-saving.

Regarding the user testing, a document explaining how the dashboard works was sent to the users. It describes the quality models (the dashboard uses their structure to the hierarchical representation of scores) and defines a scenario for users to interact, in a controlled manner, with the dashboard. The goal is that, after exercising the dashboard, the users answer some questions about the experience. To perform these tests, we made available two quality models (privacy and infrastructure) and respective component attributes. These quality models have already been validated through case studies in the ATMOSPHERE project.

The scenario for exercising the dashboard suggests at least the following actions: (i) change the weight of Information Loss attribute to 0.2 (20%) and Reidentification Risk to 0.8 (80%); (ii) See and write down the score of the Link Latency attribute; (iii) See and write down the score of the Service Trustworthiness attribute. Then, the users reported their impressions about the dashboard through questions such as: (i) “Did you have any problem when using this dashboard? If so, which ones?” (ii) “Do you suggest any change to improve this dashboard? If so, which ones?” (iii) “Would you use that dashboard again? Why?”

Regarding the two multiple choice questionnaire, the first one is about the user profile. It has six questions to mainly understand the user’s experience as IT (Information Technology) professionals, as well as their experience in Human-Computer Interaction (HCI) domain. The second questionnaire is focused on the interface usability, composed of ten questions to evaluate the strengths and weaknesses of the dashboard interface. These questions were based on Nielsen’s heuristics (usability) [20] and heuristics for information visualization [21]. Each question is a statement with a rating on a four or five-point scale of “Strongly Disagree” to “Strongly Agree” or “Very Easy” to “Very Difficult” and the answers were scored as a Likert scale on strength of agreement. Some questions statements from the second questionnaire are: (i) “the use of the dashboard configuration form for setting the parameters is simple.” (ii)

"The symbols used in the tree structure make the hierarchy of attributes clear". Tables I and II present the questions and respective answers for the four and five-point scale questions, respectively.

IV. Discussions and Improvements

As mentioned before, we performed two sprints of evaluation. In the first sprint, three specialists in information security and privacy were selected. We selected these professionals because they are familiar to trustworthiness and the quality models used in the experiments (see Section II-C). Their age, in average, is 40 years; 100% work in IT more than three years; 33% never worked before in the HCI domain and 67% never worked directly but frequently use material of this domain; 33% work sometimes with system's Front-End and system's requirements while 67% work with these matters frequently; 67% work with system's testing sometimes while 33% use to work with systems testing frequently.

Based on the questionnaire results, all the evaluators agree that the dashboard interface is nice, presents adequate volume, intuitive configuration form and results are ease to understand through the charts. On the other hand, based on the user testing, i.e., through the exercise of the predefined scenario, all evaluators had some problems to navigate in the Quality Model tree structure. One of them was really not able to realize how to navigate, i.e., to open the structure and see the scores. So, navigation seems to be the most significant interface problem. One of them commented that "The tree did not appear on her/his screen, making it difficult for him/her to find the tree". A second comment is "The need to click the arrow to open the tree is not intuitive". Both comments complain about the navigation through the tree, pointing the need for improvements in that specific part of the dashboard. Some improvements were implemented trying to make the tree navigation more intuitive before continuing with a more wide evaluation (the second sprint).

In the second sprint, 25 IT professionals acted as evaluators. It includes professionals working in IT companies (32%), research (36%), professors (4%) and undergraduate IT students (28%). Their age, in average, is 32 years; 68% work in IT more than three years; 23% have already worked in the HCI domain; 68% work with system's Front-End and 73% work with system's requirements while 64% work with these matters frequently.

Based on questionnaires results, the majority of the evaluators (approximately 73%) agree that the dashboard interface is nice, presents adequate volume of data (approximately 55%), intuitive configuration form (50%) and results are ease to understand through the charts (64%). However, similarly to the sprint 1, the evaluators still had some problems to navigate in the Quality Model tree structure.

Although a minority (approximately 30%) of the respondents stated that the tree structure navigation is difficult and the symbols used in this structure are not intuitive, some comments and suggestions were made as a result of the user testing. Three evaluators suggested that the submit button could be fixed in the display field, avoiding the scroll, which would improve the usability. At least four evaluators commented about the difficulty on navigating through the tree and suggested improving the symbols to make them more intuitive. Two of them mentioned that the quality model should be clearer in the tree structure.

The majority of the comments were about the large volume of data displayed on the dashboard, which at least 10 evaluators found excessive. They suggested to improve the charts reducing (i.e., grouping) information and removing the data markers. Also, a considerable number of evaluators commented about the difficulty in understanding the meaning of the information displayed on the dashboard. Although the evaluators are familiar with IT context and technologies, the information about trustworthiness is quite specific. Even though we sent a document explaining how the dashboard works, at least 9 respondents stated that they needed a lot of effort to understand the properties, weights and thresholds. Two of them suggested the use of chart legends.

A summary of the results from the two sprints is presented in Tables I to IV.

Table I shows the questionnaire statements with a rating on a four-point scale. The statements refer to the evaluation of the interface and the volume of data presented in the dashboard. About 76% of the evaluators agree/strongly agree that the dashboard has a friendly interface and 60% agree/strongly agree that the volume of data is adequate. However, we considered that 40% of disagree/strongly disagree answers is a considerable percentage to indicate that improvements regarding the volume of data must be done.

Table II shows the statements with a rating on a five-point scale, also from the questionnaire. The statements refer mostly to navigation through the dashboard, the tree structure and the presentation of the charts. 48% of the respondents found that navigation through the dashboard was easy/very easy, while 24% found it difficult and 28% found that the level of difficulty is medium. Regarding the configuration form, 52% found that it was easy/very easy to use, while 28% found that it is difficult/very difficult to use and 20% classified the level of difficulty as medium.

When asked for the navigation in the tree structure, 52% found it easy/very easy, 32% found it difficult/very difficult and 16% found the level of difficulty as medium. However, about the symbols used, 32% classified them as easy/very easy to use, the same percentage for medium classification. And the majority of the evaluators (36%) found it difficult/very difficulty to use. This is a strong indication that these symbols must be improved.

Finally, regarding the charts evaluation, 56% found them easy/very easy to understand, while 32% found them difficult/very difficult. 12% found that the level of understanding is medium.

Table III presents the results from the two-point scale questions, which were answered by the evaluators as part of the user testing. In this evaluation, 68% stated that they would use the dashboard again; 44% stated that they

TABLE I
ASSESSMENT RESULTS (FOUR-POINT QUESTIONS)

Statement	Strongly Agree	Agree	Disagree	Strongly Disagree
The dashboard interface is friendly (for example, colors, easy viewing)	16%	60%	16%	8%
The volume of data displayed is adequate (i.e., the dashboard does not have excessive information)	16%	44%	36%	4%

TABLE II
ASSESSMENT RESULTS (FIVE-POINT QUESTIONS)

Statement	Very Easy	Easy	Medium	Difficult	Very Difficult
Navigation through the dashboard was ...	20%	28%	28%	24%	0%
The use of the configuration form for setting the parameters was ...	20%	32%	20%	24%	4%
The navigation in the tree structure to view the scores of the attributes was ...	20%	32%	16%	24%	8%
The symbols used in the tree structure made the hierarchy of attributes.... to use	20%	12%	32%	16%	20%
The results presented in the form of charts based on the history of the scores let the understanding ...	12%	44%	12%	20%	12%

TABLE III
ASSESSMENT RESULTS (USER TESTING TWO-POINT QUESTIONS)

Question	Yes	No
Would you use the dashboard again?	68%	32%
Did you have any problem when using the dashboard?	44%	56%
Do you suggest any change to improve the dashboard?	64%	36%

TABLE IV
ASSESSMENT RESULTS (USER TESTING ESSAY COMMENTS)

Category	Comments (%)
Understanding	37%
Volume of Data	37%
Navigation	44%
Charts	26%

had problems when using the dashboard and we considered this a high percentage and indicates that the problems must be investigated; 64% suggested improvements to the dashboard. We considered most of these suggestions and the improvements are described in the next subsection. Table IV summarizes the problems and suggestions pointed out by the evaluators through essay questions.

We received in total 27 essay comments. We classified them in 4 categories: *Understanding*, which refers to the understanding of the meaning of the information displayed in the dashboard; *Volume of Data*, which refers to the amount of information displayed in the dashboard; *Navigation*, referring to the navigation through the tree structure and the pages of the dashboard; and *Charts*, which refers to the visualization of the charts. As we are dealing with essay comments, some of them were classified in more than one category.

A. Usability improvements for the dashboard

Based on the evaluation, we provided the improvements to the dashboard. Figure 4 shows the latest release.

The first improvement is about the submit button, in the Configuration Form (upper left corner). We removed the scroll and now it is fixed, facilitating its visualization. The scroll is only for the properties. Also, in the Configuration Form, we added a help system where the user can position the mouse over the "?" symbol or the property itself and a popup will open with information and details about that property (see Figure 4, where a popup example is shown with the "The name of trustworthiness metric" message). This would facilitate the use of the dashboard by less expert users.

Regarding the excessive information on chart visualization, first, we removed the data markers, which left the chart cleaner. We also optimized some queries to group information. One example of query optimization is presented in Table V, where it is possible to observe the use of DISTINCT, AVG and GROUP BY clauses, which better organize and reduce the amount of information.

Another improvement for chart visualization is the addition of caption for the x-axis of the chart. In Figure 4, we can observe the dates throughout the metrics (December 1, 2019, December 8, 2019 and December 15, 2019 for System Trustworthiness property and August 1, 2019, September 1, 2019, October 1, 2019, November 1, 2019 and December 1, 2019 for Privacy property). These dates can be configured dynamically by the user, which can specify a period for visualization.

To improve the tree structure navigation, we decided to replace the symbols. It is possible to observe that, in Figure 4, bottom left corner, we used the plus and minus ("+", "-")

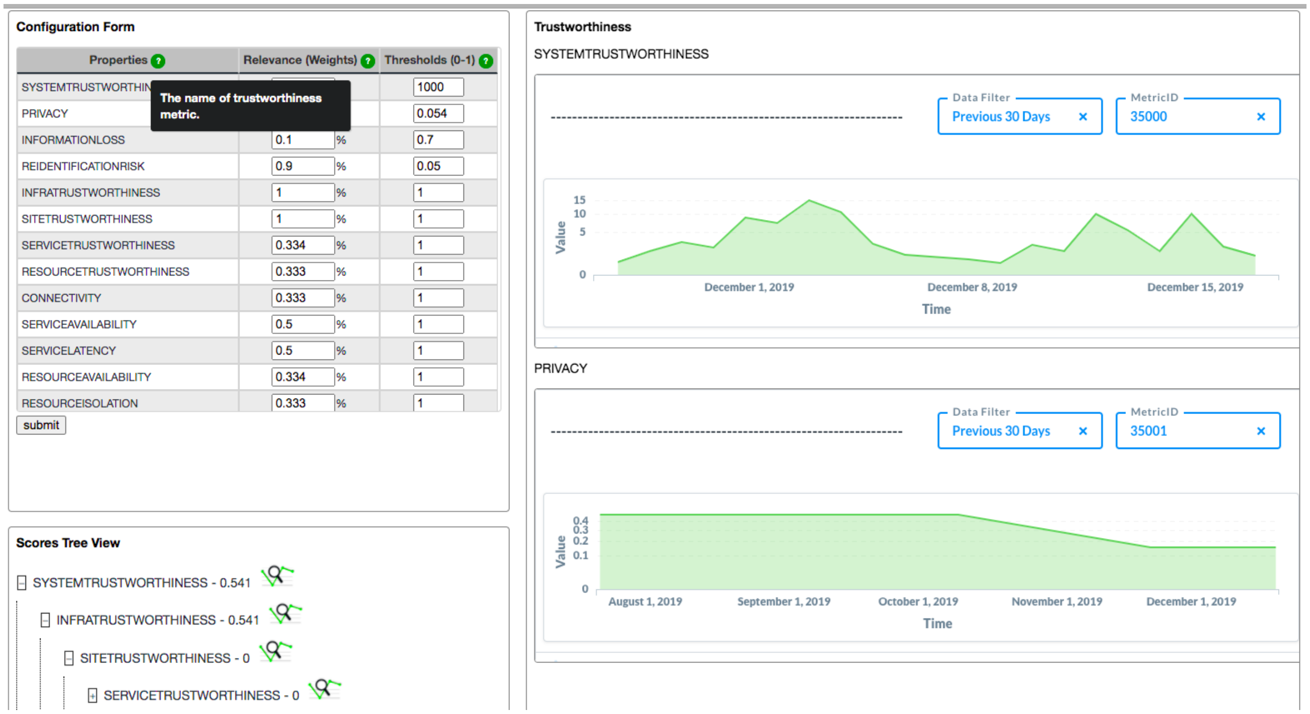

Figure 4. Improved dashboard based on the usability evaluation

TABLE V
EXAMPLE OF QUERY OPTIMIZATION FOR CHART VISUALIZATION IMPROVEMENT

Before	After (optimized)
SELECT ‘MetricData‘.‘metricId‘ AS ‘metricId‘, ‘MetricData‘.‘valueTime‘ AS ‘valueTime‘, ‘MetricData‘.‘value‘ AS ‘value‘ FROM ‘MetricData‘ WHERE ‘metricId‘ IS NOT NULL AND {{filter}} LIMIT 2000	SELECT distinct ‘MetricData‘.‘metricId‘ AS ‘metricId‘, CAST(‘MetricData‘.‘valueTime‘ as date) AS ‘valueTime‘, AVG(cast(‘MetricData‘.‘value‘ as decimal(10, 2))) AS ‘value‘ FROM ‘MetricData‘ WHERE ‘metricId‘ IS NOT NULL AND {{filter}} GROUP BY ‘valueTime‘, ‘metricId‘ LIMIT 2000

control symbols to expand or collapse the branch, i.e., to show and hide subgroup properties when navigating through the hierarchy. We believe that these are more universal symbols and the users are more familiar with them.

We also introduced a shortcut icon for charts (a magnifying glass with a chart) on the right side of each property in the scores tree view. This allows users to select a specific chart to be visualized while navigating or having a general view of the tree.

Finally, we reduced the number of decimal places in the score in order to make the tree visualization clearer. Figure 5 shows an example of expanded tree, where each property has its own shortcut icon.

B. Impacts on ATMOSPHERE project

It is important to mention that the dashboard and respective usability evaluation provided some improvements in the ATMOSPHERE project regarding the maturity level of some requirements and adaptation scenarios.

The ATMOSPHERE project produced a realtime platform that self-adapts when the score threshold is reached [8] and the dashboard was used as a front-end of it. The use of the dashboard with adequate usability allowed to validate adaptation scenarios in a visual way and, consequently, in a faster way too. For example, when the value of the privacy property is greater than the threshold assigned in the configuration form, the platform can use an adaptation plan that reduces the value of re-identification risk and/or information loss sub-properties and, consequently, the value of privacy too. Using the dashboard, these values are updated automatically in the interface, identifying to user the adaptation. Previously, the validation was performed by scripts and analysis of log files, which required a lot of time.

With respect to the project requirements, they required a experimental evaluation of different interfaces during the development and the evaluation of each one helped to improve the development of the components, as they were better designed and tested. At this point, we can say that the

 ISBN: 978-1-61208-761-0

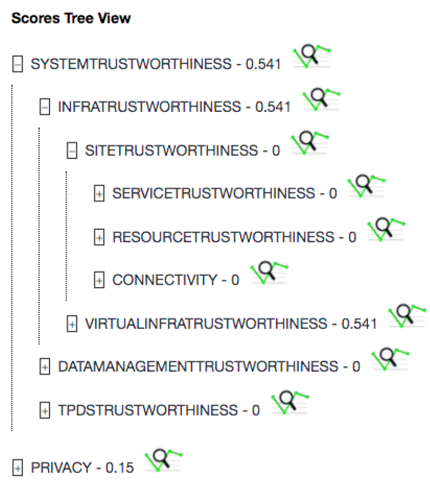

Figure 5. Expanded tree and shortcuts to the respective properties charts

dashboard and the usability evaluation process performed in this work helped the development of the project, accelerating the validation and integration of components.

V. Conclusions and Future Work

This work presented a usability evaluation of a dashboard for assessing the trustworthiness of cloud applications that handle large volume of data. We applied questionnaires and usability tests to perform this evaluation.

The results and the improvements highlighted the importance of mixed-methods evaluation of usability as a part of the design of the dashboard. The different profiles of users (but all of them active in the IT universe) offered an efficient way to assess the needs of users, generate ideas and develop a more viable product for use. This could be done iteratively, through two sprints. When we talk about product viability, it is worth mentioning that the usability evaluation process can help improve the system requirements identification and maturity, as well as help define testing scenarios.

Since the improvements required by the respondents users did not demand huge changes and significant software development skills, we can state that, for the experiments in this work, the use of user-centered evaluation to mitigate potential usability challenges can easily help increasing user satisfaction and adoption of the dashboard. However, it is important to mention that the dashboard is in the early stage of its development. For late stages, it is recommended the use of other complementary techniques because the end stages of dashboard development can mask potential functional problems that will prevent proper usage and lead to misinterpretation of results.

So, as future work we intend to identify and apply different usability evaluation techniques together with usability testing to identify specific usability issues and room for improvement.

Acknowledgment

This work has been partially supported by the projects ATMOSPHERE (https://www.atmosphere-eubrazil.eu/ - Horizon 2020 No 777154 - MCTIC/RNP) and ADVANCE (http://advance-rise.eu/ - Horizon 2020-MSCA-RISE No 2018-823788).

References

[1] R. Magdalena, Y. Ruldeviyani, D. I. Sensuse, and C. Bernando, "Methods to enhance the utilization of business intelligence dashboard by integration of evaluation and user testing," in *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)*. IEEE, 2019, pp. 1–6.

[2] ATMOSPHERE, "Adaptive, trustworthy, manageable, orchestrated, secure, privacy-assuring hybrid, ecosystem for resilient cloud computing," 2018, URL: https://www.atmosphere-eubrazil.eu/ [accessed October, 2020].

[3] D. Camargo, F. N. Gaia, T. Basso, and R. L. Moraes, "A dashboard for system trustworthiness properties evaluation," in *Anais do XXI Workshop de Testes e Tolerância a Falhas [Testing and Fault Tolerance Workshop]*. SBC, 2020, pp. 1–14.

[4] IBM, "Accelerate your journey to ai with a prescriptive approach," 2020, URL: https://www.ibm.com/br-pt/analytics [accessed October, 2020].

[5] SAP, "Introducing sap customer data platform," 2020, URL: https://www.sap.com/index.html [accessed October, 2020].

[6] Tableau, "Get a full picture of your business, inside and out," 2020, URL: https://www.tableau.com/ [accessed October, 2020].

[7] ISO, "International organization for standardization. systems and software engineering — systems and software quality requirements and evaluation (square) — guide to square," 2014, URL: https://www.iso.org/standard/64764.html [accessed October, 2020].

[8] T. Basso, H. Silva, L. Montecchi, B. B. N. de França, and R. Moraes, "Towards trustworthy cloud service selection: monitoring and assessing data privacy," in *XX Workshop de Testes e Tolerância a Falhas (WTF 2019)[Testing and Fault Tolerance Workshop]*, Gramado, RS, Brazil, 2019, pp. 7–20.

[9] Metabase, "Have questions about your data? metabase has answers," 2020, URL: https://metabase.com [accessed October, 2020].

[10] T. Basso, H. Silva, and R. Moraes, "On the use of quality models to characterize trustworthiness properties," in *International Workshop on Software Engineering for Resilient Systems*. Springer, 2019, pp. 147–155.

[11] Vallum Software, "A nextgen network monitoring & management solution without the complexity and cost," 2020, URL: https://vallumsoftware.com [accessed October, 2020].

[12] ABNT - Brazilian Association of Technical Standards, "Ergonomic requirements for office work with visual display terminals (vdts) part 11:guidance on usability (abnt nbr iso 9241-11:2011)," 2011, URL: https://www.abntcatalogo.com.br/norma.aspx?ID=86090 [accessed October, 2020].

[13] J. Nielsen, *Usability engineering*. Morgan Kaufmann, 1994.

[14] C. M. Barnum, *Usability testing essentials: ready, set... test!* Morgan Kaufmann, 2020.

[15] J. Nielsen, "Usability inspection methods," in *Conference companion on Human factors in computing systems*, 1994, pp. 413–414.

[16] J. Sauro and J. R. Lewis, "Standardized usability questionnaires," in *Quantifying the user experience*, vol. 8. Morgan Kaufmann, 2012, pp. 198–212.

[17] C. Jooste, J. Van Biljon, and J. Mentz, "Usability evaluation for business intelligence applications: A user support perspective," in *South African Computer Journal*, vol. 53, no. Special issue 1. South African Computer Society (SAICSIT), 2014, pp. 32–44.

[18] A. Lavalle, A. Maté, J. Trujillo, and S. Rizzi, "Visualization requirements for business intelligence analytics: A goal-based, iterative framework," in *2019 IEEE 27th International Requirements Engineering Conference (RE)*. IEEE, 2019, pp. 109–119.

[19] A. Read, A. Tarrell, and A. Fruhling, "Exploring user preference for the dashboard menu design," in *2009 42nd Hawaii International Conference on System Sciences*. IEEE, 2009, pp. 1–10.

[20] J. Nielsen and R. L. Mack, *Usability Inspection Methods*. John Wiley & Sons, Inc., 1994.

[21] T. Zuk, L. Schlesier, P. Neumann, M. S. Hancock, and S. Carpendale, "Heuristics for information visualization evaluation," in *the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, 2006, pp. 55–60.