The Semantic Combining for Exploration of Environmental and Disease Data Dashboard for Clinician Researchers

Albert Navarro-Gallinad, Alan Meehan, and Declan O'Sullivan

ADAPT Centre for Digital Content, Trinity College Dublin, Dublin, Ireland
School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland
{albert.navarro,alan.meehan,declan.osullivan}@adaptcentre.ie

Abstract. While Semantic Web technologies facilitate the integration of heterogeneous data sources through the Resource Description Framework (RDF) and ontologies, they present an obstacle for non-technical researchers who want to access and explore the data to meet their needs. To address this problem, visual tools and analytical platforms with a user-centred approach are an emerging solution. This paper outlines the design of a dashboard called Semantic Combining for Exploration of Environmental and Disease data (SCEED), an initial visual tool designed for use by clinician researchers to explore and retrieve combined environmental and disease data for further analysis. The evaluation of SCEED consists of a combination of standard usability and effectiveness methods, using the AVERT project as a case study. In the AVERT project, clinician researchers need to address the challenges of querying specific vasculitis flare clinical data for a particular patient to retrieve linked environmental data from a triplestore, and downloading the chosen data as input for their statistical models. The initial evaluation has concluded that the SCEED dashboard is an adequate initial design to fulfil, and points towards an interface to engage clinician researchers directly with Linked Data. Furthermore, this paper helps to highlight the difficulty of conducting usability evaluations with small sample sizes and how evaluation metrics can be combined to assess the requirements for developing an effective tool.

Keywords: Semantic Web · dashboard · usability evaluation.

1 Introduction

Semantic Web technologies have a steep learning curve which can present an obstacle for non-technical researchers when trying to access and explore the data for their needs. Visualization tools and analytical platforms operating on top of Semantic Web architectures can support accessing and exploring Linked Data for non-technical or non-domain expert users by aiding query formulation

---

in an intelligible manner [5]. A dashboard approach with multiple coordinated views offers advantages for statistical data including the integration of multiple data sources, details of the underlying data, a flexible data analysis layer and a reusable framework [16, 18, 3]. Furthermore, a user-centred approach to dashboard design provides an easy and intuitive interface to be used by a focused group [16, 18, 3, 11] and the use of standard usability evaluation with standard post questionnaires enables comparison of prototype tools with later versions of the tool, as well as with other tools.

At the moment, typically clinician researchers require knowledge engineers to access, explore and retrieve data that they are interested in from datasets when implemented using standard Semantic Web technologies. Therefore, there is an opportunity to propose a semantic analytical platform following a user-centred approach. In particular, health related statisticians or a clinician with statistical experience (hereafter clinician researchers) lack tools to explore clinical and environmental linked data, which will be used as input to train their models (described further in the Section 3).

This paper outlines the design of the Semantic Combining for Exploration of Environmental and Disease data (SCEED) dashboard, an initial visual tool designed to be used by clinician researchers to explore and retrieve combined environmental and disease data for further analysis. The contribution of this paper is the SCEED dashboard itself along with an initial evaluation of the usability and effectiveness through a standard usability test.

The paper is structured as follows. Section 2 reviews related research. Section 3 overviews the AVERT project. Section 4 outlines the design and implementation of the SCEED dashboard. Section 5 describes the evaluation method, the results and analysis of the user experiment and discusses the outcome of this initial evaluation. Section 6 concludes the paper and states the future work.

2 Related Work

This section overviews the state-of-the-art in Semantic-Web based visual tools to meaningfully explore linked data for clinician researchers. We have classified the reviewed tools based on the usability evaluation for their visual techniques.

Relevant tools where non-standard usability evaluation used. The Granatum project addresses the computational challenges that genomic scientists have in analysing single-cell RNA sequencing data with a graphical analysis pipeline [25]. As part of the project, Hasnain et al. 2014 [9] evaluates the developed Liked Biomedical Dataspace for supplementing drug discovery with domain experts, following a user-centred approach for bioinformaticians and biomedical researchers. This dataspace uses ReVealD as the visual query system, which is evaluated with 'Tracking Real-time User Experience (TRUE)' methodology, and later became a platform integrated for this project. Kamdar et al. 2014 [9] evaluates this platform for biomedical researchers with metrics such as number and time taken per step to complete a task. The fact that an ad hoc usability questionnaire was used in this particular study encumbers further comparison

with other studies. Likewise, Villanueva-Rosales et al. 2015 [23] uses an ad hoc usability questionnaire to evaluate, with multidisciplinary participants, an experimental graphical user interface for The Earth, Life and Semantic Web Project [4]. In addition, Scharl et al. 2017 [21] assesses the usability of the semantic analytical platforms usability with heuristic evaluation, formative usability tests and feedback from actual users, communication professionals; non-standard usability/effectiveness metrics.

Relevant tools where standard usability evaluation was used. Sabol et al. 2014 [19] presents a toolchain to explore and visually analyse Linked Data for non-Semantic Web experts. The authors evaluate the work from a formative usability angle with a quantitative, standard NASA Task Load Index (TLX) for workload and time per task; and qualitative, think-aloud protocol metrics. Furthermore, Dafli et al. 2015 [6] evaluates the usability and efficacy of the Open Laberinth extension with specific scenarios aimed at health professionals. A System Usability Scale (SUS) questionnaire, a standard method, in combination with eViP questionnaire and expert reviews were the used metrics. This standard questionnaire is also used by Braoveanu et al. 2016 [3] in combination with the time per task and discussion of the task results for a user study with tourism researchers and practitioners; and by Zained et al. 2015 [24] with an additional custom questionnaire designed to test the FedViz interface for a researchers and engineers in Semantic Web.

In contrast to the related work outlined above, our approach is focused on providing access and exploration of linked environmental and clinical data for clinician researchers. In this paper, we present a dashboard with a user-centred approach addressing clinician researchers assessed with a standard usability/effectiveness evaluation, enabling comparison with later versions of the SCEED tool, and with other related tools.

3   AVERT Background

AVERT[1] and HELICAL[2] are two projects in the field of Healthcare Data Linkage which share the same data integration approach based on Semantic Web technologies [17], providing a scalable semantic architecture for data related to rare chronic diseases. The data model links clinical data for patients with ANCA vasculitis (a rare kidney disease) with environmental data, for the goal of predicting when flares of the disease may occur for individual patients.

In the AVERT semantic architecture [17], Semantic Web technologies are used to combine multiple diverse data sources with spatial and temporal common features between medical registries and environmental data. These datasets are converted to RDF [12] with R2RML [7], a mapping from relational databases format to RDF datasets; and stored in a triple store, allowing information retrieval through semantic queries. Then SPARQL enables the retrieval, manipulation and linkage of the stored data. Currently a knowledge engineer performs these

---

[1] https://www.tcd.ie/medicine/thkc/avert/
[2] http://helical-itn.eu/

SPARQL queries to fulfil the clinician researchers needs, a human-in-the-loop approach.

The intention going forward in such projects is to allow the clinician researchers themselves (through a dashboard) to access and explore the clinical and environmental data that are represented and linked through semantic web technologies. The data will be used as input to train their models to potentially find associations and relationships between environmental factors and the disease flares of patients.

An effective tool would thus intend to achieve the following requirements extracted from expert consensus within AVERT:

Requirement 1: enable the clinician researcher to query specific clinical patient data to retrieve linked environmental data, without the need for knowledge of the underpinning semantic web technologies;

Requirement 2: support the understanding of the clinician researcher in the use and limitations of the linked environmental data to support identification of flares for rare diseases;

Requirement 3: allow for the download of selected clinical and environmental data to be used as input in statistical models for data analysis.

The SCEED dashboard is a prototype tool aimed at satisfying these three requirements.

## 4   Design and Implementation

The development of the dashboard was motivated by the needs of clinician researchers in the AVERT project, who are not Semantic Web experts, to identify relevant environmental data that should be linked to longitudinal ANCA vasculitis patient clinical data to support spatio-temporal analysis of the data. This is done in order to support prediction of flares for individual patients and to ultimately support the discovery of environmental factors that trigger the disease in the patient cohort.

The dashboard operates on top of AVERTs semantic architecture (see Section 3), where data from multiple data sources is uplifted to RDF [17]: weather and pollution data (27.5M triples) along with infectious disease data and clinical data (2.6M triples). This relevant data is retrieved from a triplestore supporting GeoSPARQL [1] queries, which are key for the nature of our data (which has spatio-temporal components).

The initial dashboard was designed to have four tabs (see Fig. 1), each of which are described next.

Query tab. In the Query tab of the dashboard, Fig. 1 part a), the user can select options from the different flare related parameters. These selected options are then substituted into a SPARQL query template, URL encoded and executed against the data in the triplestore.This tab is aimed towards satisfying *Requirement 1*.

Link data tab. This is the first tab the user sees when submitting a query, Fig. 1 part b). The aim of this tab is to provide the environmental data linked
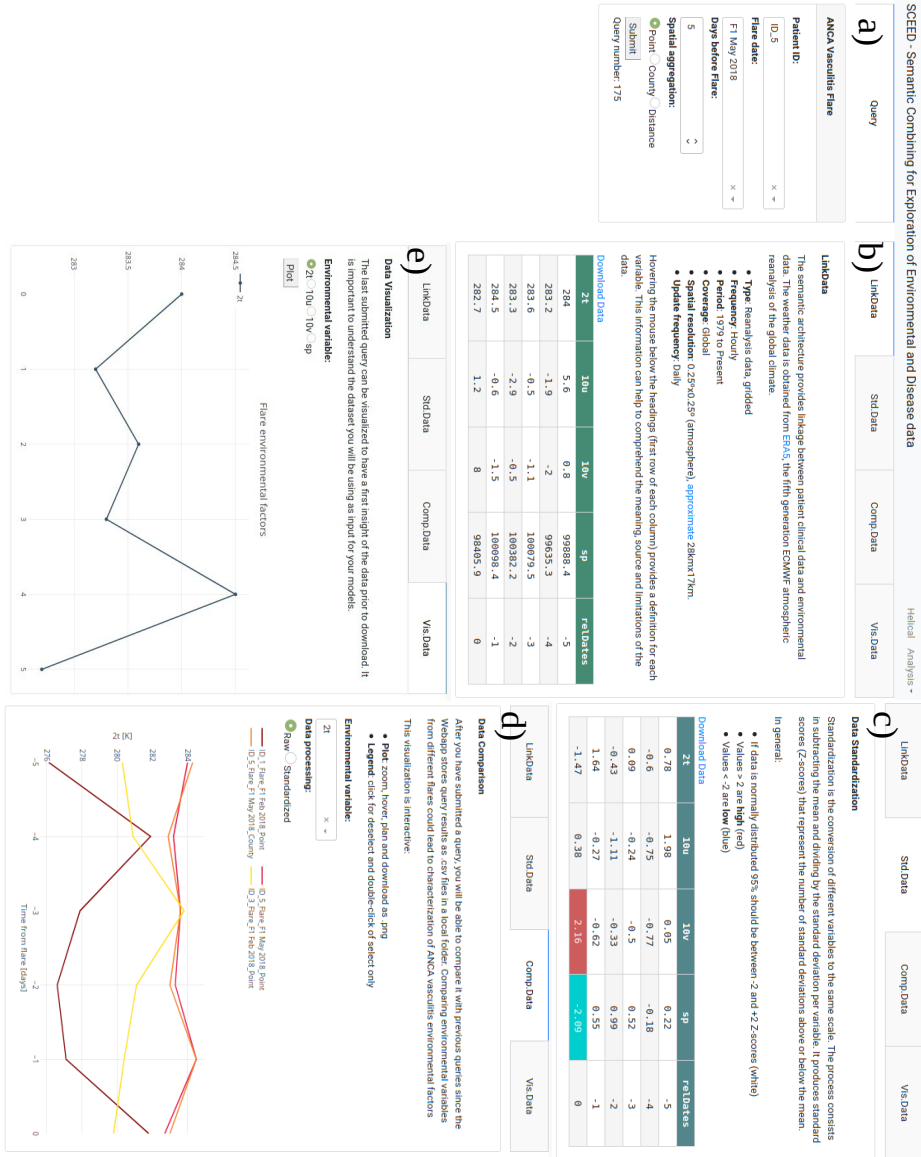
Fig. 1. SCEED dashboard multiple view after finishing the tasks from the study. a) Query section. The user can select an option from a dropdown display list for Patient ID and Flare date, a numeric value for Days before Flare and an input from the radio buttons in Spatial aggregation. Data tabs: b) Link Data, c) Std.Data, d) Comp.Data and e) Vis.Data; the different tabs allow the user to navigate, compare, visualize and download meaningful information. Each tab starts with an introductory text information to guide the user and ends with a data visualization as table or graph.

to the clinical patient queried in the Query tab. This data is displayed as a table with a hovering feature that displays the data description, gathered from the climate data store provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) parameter database[3]. The table displayed after the submission of the query can be downloaded as a CSV file (in support of *Requirement 3*).

Standard data tab. In the Standard data tab, Fig. 1 part c), the user can compare different environmental variables for a better understanding of their variability (in support of *Requirement 2*), since they have been (statistically) standardized, hence they have been converted to the same scale, producing standard scores (Z-scores). The table displayed has some highlighted values with colour encoding depending on the category of the value, available to download as a CSV file.

Comparative data tab. A CSV file is stored with data from the previous submitted query. These are then compared in an interactive plot with legend features, allowing for selecting and deselecting of options key for multiple flare environmental related data comparison (see Fig. 1 d). The multiline plot allows a user to discover/present/identify trends, seasonality, comparison and check for outliers previously discovered in the standardization tab; to improve their comprehension of the environmental data previous to the patient's flare event (in support of *Requirement 2*).

Visualization tab. In this tab Fig. 1 part e), the user can visualize the last submitted query to have a cleaner view of each variable. The tab is aimed to provide a quick insight of the data prior to download to make sure it has accomplished the statisticians needs.

The SCEED dashboard (shown in Fig. 1) is coded in Python (3.6), using the *Dash 1.7.0)*[4] package as a framework that facilitates building cross-platform analytical platforms. This dashboard is coded dynamically, displaying the drop down options for each parameter reacting to available data in the triplestore endpoint. Therefore, if new data is added to the triplestore, according to the same data model, the dashboard will react accordingly showing the new available data. This approach is ideal when managing both clinical data (since data collection is an ongoing process), as well as environmental time series data which is constantly being updated

5 Evaluation

An experiment was undertaken to evaluate the usability of the initial SCEED dashboard in accessing environmental data linked with clinical de-identified patient records, by clinician researchers who have no practical experience with in Semantic Web technologies. The user experiment was structured by a brief introduction to the dashboards background, the tasks to be completed by the participants and a follow up post-questionnaire.

---

[3] https://apps.ecmwf.int/codes/grib/param-db
[4] https://dash.plotly.com/

5.1   Experimental Setup and Execution

The target group was clinician researchers that are not Semantic Web experts, who would be users of the analytical platform and who had data exploration needs in the health domain similar to those of the AVERT project. This targeted selection criteria resulted in the recruitment of seven participants: PhD students (3) and professors (4), who were experienced in analysing clinical data with statistical models. These participants are within the 30-50 age range, with a female to male ratio of 2:5 and international with fluent English. This sample size is covering the requirements, evaluating a prototype of a novel user-interface design in the first stages that has a specialised nature (exploration of clinical and linked environmental data) [15].

The experiment started with the participants signing the informed consent document, which commenced with a short explanation about the purpose of the dashboard, its main target contribution to research and a mention of the semantic technologies operating in the back-end. This was the first contact with the SCEED dashboard and the participant had not interacted with or seen the dashboard previously. Furthermore, each participant was asked to follow a role while testing SCEED: that is a researcher with access to clinical patient data would like to extract and comprehend environmental data related to patient flares.

Clinical data was simulated from the AVERT data model tailored to support the chosen tasks. Environmental data was obtained from ERA5, the fifth generation ECMWF atmospheric reanalysis of the global climate [8], and reduced to four variables (columns in the data table of Fig. 1), again to support timely exploration of the dashboard for the given tasks.

Each participant was asked to follow a concurrent think-aloud protocol (CTA) [2] and participants' think-aloud statements were recorded by hand by the experiment designer during the evaluation. The think-aloud protocol requires listening to the participant process while completing the tasks as well as encouraging the think-aloud action. The participants think-aloud statements and extra feedback were recorded by hand for a qualitative analysis with Grounded Theory [10, 22].

As the experiment was conducted during COVID-19 restrictions, synchronous remote testing was the method used through a video conferencing platform with remote control functions. Interestingly, we observed that the remote testing nature of the study reinforced the ideal spectator role within the participant-observer interaction, an optimal testing environment for the CTA method. An hour was allocated to each participant to explore the dashboard, complete the tasks and fill out the post-questionnaire.

Each participant was asked to complete a series of tasks carefully selected to assess the three core requirements of the dashboard stated previously. These tasks were written and given together with the informed consent document at the start of the video conference. The observer tracked manually the time spent per task with a stopwatch when the participant explicitly made a comment that the task had been completed. The tasks were set out as follows. First the

instruction, and then the criteria for when the task will have been completed by the participant. The tasks were sequenced and numbered as follows:

1. **Submit a query for a specific patients flare.** The task will be complete when an environmental data table is displayed in the LinkData tab.
2. **Explain the meaning of each column in the environmental table.** The task will be complete when the participant has hovered over the columns headings of the main table and read the description of the environmental variables.
3. **Try different aggregation approaches.**The task will be complete after the participant has explored the different spatial aggregations available in the Query section and the main table reacts/changes accordingly.
4. **Compare variables for the same flare in the standard data tab.** The participant will have finished this task after selecting the Std.Data tab.
5. **Compare environmental data from different flares in the comparison tab.** The participant will have finished this task when they have successfully compared two flares in the Comp.Data tab.
6. **Visualize Link Data variables in the visualization tab.** The task will be completed when the participant successfully visualizes the environmental data prior to download in the Vis.Data tab.
7. **Download useful raw data for the researchers needs.** The participant will have finished this task after downloading the data from either the LinkData or Std.Data tabs.

After the completion of the tasks, the participants were asked to complete a Post-Study System Usability Questionnaire (PSSUQ) (described further in the next section) to evaluate the user experience in a quantitative metric.

The methods described above include a CTA protocol, successful completion of the tasks and time on task to support the PSSUQ standard questionnaire. The CTA protocol grants feedback to understand the effective task completion and time on task in a meaningful way. These methods combine of quantitative and qualitative metrics to evaluate the usability of the SCEED tool.

5.2 Results

Quantitative results *Time on task.* The box plots in Fig. 2 compare participants times spent on task, which all the participants completed successfully. The spread of the time values per task, the length of the boxes (IQR), is below 1 min for the simplest tasks of submitting a query and downloading the data (T1 and T7); between 1-2 min for the tasks of selecting different query parameters and tabs (T3, T4 and T6) and around 3 min for the more complex tasks of explaining the meaning and comparing the data (T2 and T5). Furthermore, the median follows a similar pattern than the spread with most of the tasks below 3.5 minutes, increased by 1 min for T2 and doubled for the most complex task of comparing patients flares (T5). The box plots in Fig. 2 also identified 3 outliers and proved to be suitable in studying the patterns of this data with a sample size of 7.
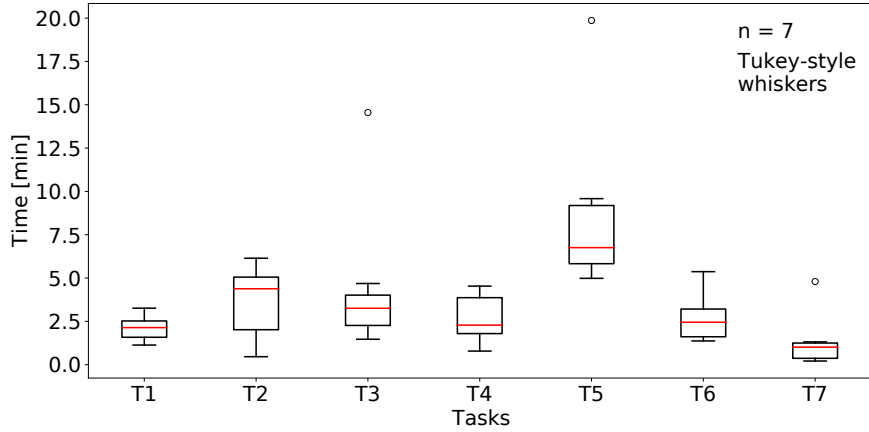
Fig. 2. Time spent to complete each task during the experiment.

PSSUQ: The Post-Study System Usability Questionnaire. The PSSUQ is a general questionnaire meant to assess the usability evolution during the development of a system with 19 questions [13], second version of the questionnaire was used in this study. The PSSUQ follows a 7-point Likert Scale and assesses four different metrics: system usefulness (SysUse), information quality (InfoQual), interface quality (IntQual)and overall, averaged from 1-8, 9-15, 16-18, 1-19 questions. The questionnaire results and aggregations per group for the SCEED dashboard are presented as box plots in Fig. 3. This visualization allows us to compare the distributions without any assumptions, again adequate for our sample size. Most of the PSSUQ scores have a median of 2 and a spread between 1-1.5 points, reduced to ≤ 1 for the four averaged metrics with larger sample sizes. Moreover, Q7 and Q16, which indicate the learning easiness and the interface pleasantness, got the best scores. However, Q12 and Q18, regarding easy finding of information and having the needed functions, have an increased median of 3, the higher the worse in this scale, and Q9 has a spread of 3.5 points indicating a diversity in opinions on the error messages. The identified outliers for the individual questions are from 2 participants, which were not satisfied with the system use and features available. Furthermore, some participants provided qualitative comments through the PSSUQ open comment section coherent with the previous results.

Qualitative results The experiment also followed a concurrent think-aloud protocol providing the study with qualitative data, analysed by means of Grounded Theory. The observer/note-taker coded and categorized the annotations manually and with the note-taker's criteria alone. The categories of the annotations were recurrent and natural, commenting upon the important features of the dashboard. All the participants discussed the usefulness and understanding of the patient IDs exploration, the Z-scores, the comparative plot visualization
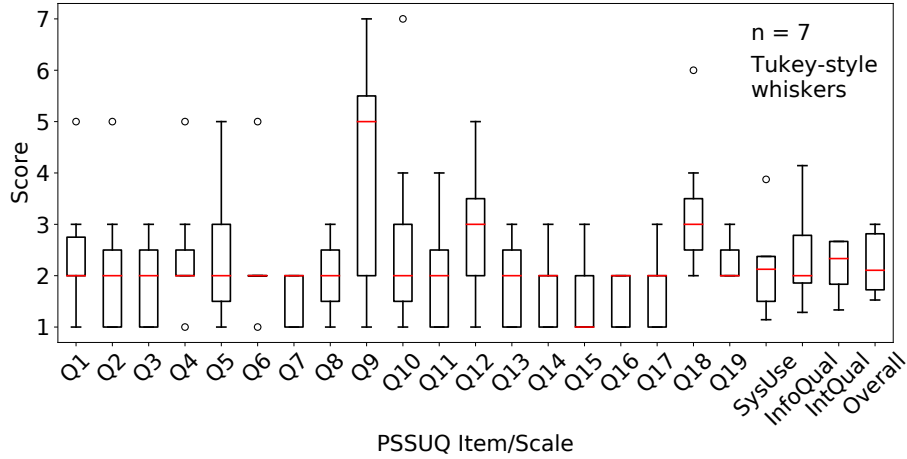
Fig. 3. PSSUQ scores box plot with the four averaged metrics (SysUse, InfoQual, IntQual and Overall) on the right end with a sample sizes of 56, 49, 21 and 133.

and the downloading approach of patient flare linked environmental data in the dashboard. However, the rest of the features, including the dates selection (days before flare), the spatial aggregation feature, hovering for the variable information display and the usefulness of the Vis tab; presented a variety of patterns. Participants expressed that the linked environmental data would be more useful if provided as a time series for a specific period instead of only dates before the flare. Moreover, clinician researchers wanted the possibility to explore and download all data in a summarized way.

The emerging themes from the Grounded Theory analysis were (1) accessing flare related environmental data, (2) associating multiple patients and (3) exploring longitudinal data. These themes acknowledged the perception associated to the complex topic of comprehending linked environmental data to support rare disease research (*Requirement 2*).

5.3 Discussion

The SCEED dashboard was developed in order to support clinician researchers exploring clinical data linked with environmental data by querying the data, visualizing these datasets as tables and visualizations for comprehension and downloading the data for their models; without previous knowledge of Semantic Web technologies. We conducted a user experiment to evaluate the SCEED dashboard by the completion of 7 tasks using standard methodologies: time on task, CTA and PSSUQ; to assess the usability. These tasks were selected to assess the three core requirements.

First, all the participants were successful in completing the tasks displaying a similar pattern (Fig. 2). This pattern suggests that the time spent to complete each task increases with the complexity of the tasks. This complexity directly

relates to the difficulty of fulfilling the requirements by the clinician researchers, supporting the achievement of *Requirements 1* and *3*.

Second, the analysis of the PSSUQ responses leads to a better understanding of the SCEED dashboard specific features. The PSSUQ aggregated results in Fig. 3 show the known consistent pattern of poor ratings for InfoQual relative to IntQual and for Q9 [14], supporting the robustness of the questionnaire with less than 15 participants [20]. Moreover, these aggregated results are lower than the norm defined for the PSSUQ version 2 [13], the lower the value the higher the satisfaction; and provide a reference for the next versions of the dashboard. The open comments of this questionnaire indicated that the system was easy to use and had good features; while requiring additional ones, explained by Q18 score (see Fig. 3), to fulfil *Requirement 2*.

Third, the CTA allowed us to understand participants thoughts as they occurred while completing the tasks. The categorization for the think-aloud statements made clear that the dashboard needs improvements in a number of areas which will be addressed in the next versions. Furthermore, the emerging themes of the Grounded Theory analysis endorse the previous statements made along fulfilling *Requirements 1* and *3*, and acknowledging alternatives on how *Requirement 2* could be addressed in the following versions. These next versions will be updated with a more focused multiple patients approach on a selcted date range to improve the environmental data exploration.

When the results of the various evaluation methods described above are examined together, we were able to achieve some more insights. A number of examples of these insights are worth presenting. The emotional responses noted during the CTA provides an explanation for the three outliers in Fig. 2. These participants were curious and wanted to explore all the functionalities during the tasks. On the other hand, task 2 dispersive values can be explained by the inefficient formulation of the task, since some participants discovered the hover functionality while performing task 1, resulting in quicker times.

Having selected only clinician researchers provides more relevant results than enlarging the sample size for this first version. However, this supposes a challenge when making statements around the quantitative results which is why we combine different metrics in the evaluation. Another limitation of this work is that the manual evaluation for the qualitative data with a limited number of participants, lacking depth of the qualitative results. These limitations will be addressed in later evaluations with a more comprehensive and automatic design, and increasing the sample size with the involvement of more real end users including different interests for a wider acceptance of the prototype.

However, the evaluation conducted on this paper, which combines different standard metrics, could be beneficial for assessing other tools with low sample sizes. Finally, the results on this initial evaluation hold promise for producing an interface that will engage clinician researchers directly with Linked Data.

6   Conclusions and Future Work

From the design and the results obtained from the evaluation performed, the SCEED dashboard is an adequate initial design to fulfil the clinician researchers requirements of querying specific clinical patient data to retrieve environmental data linked to vasculitis patient flare clinical data from the triplestore and downloading meaningful data to be used as input for statistical models. However, new features are necessary for comprehending the use and limitations of the environmental data for rare disease flare discovery.

The combination of measuring the time per task, CTA protocol and PSSUQ provided enough data to assess the usability of the dashboard, highlighting the successful aspects, identifying the items that need to be improved and the new features to be added. In the future, this methodology will be used as a baseline to track the evolution of the dashboard.

References

1. Battle, R., Kolas, D.: Enabling the geospatial Semantic Web with Parliament and GeoSPARQL. Semantic Web **3**(4), 355–370 (2012). https://doi.org/10.3233/SW-2012-0065
2. Boren, T., Ramey, J.: Thinking aloud: Reconciling theory and practice. IEEE Transactions on Professional Communication **43**(3), 261–278 (Sep 2000). https://doi.org/10.1109/47.867942
3. Braşoveanu, A.M.P., Sabou, M., Scharl, A., Hubmann-Haidvogel, A., Fischl, D.: Visualizing statistical linked knowledge for decision support. Semantic Web **8**(1), 113–137 (Jan 2017). https://doi.org/10.3233/SW-160225
4. Chavira, L.A.G.: The Earth Life and Semantic Web Project Experiment GUI. Tech. rep., U.S. Department of Health & Human Services (Jul 2015)
5. Dadzie, A.S., Rowe, M.: Approaches to visualising Linked Data: A survey. Semantic Web 2 **2**(2), 89–124 (Jan 2011). https://doi.org/10.3233/SW-2011-0037
6. Dafli, E., Antoniou, P., Ioannidis, L., Dombros, N., Topps, D., Bamidis, P.D.: Virtual Patients on the Semantic Web: A Proof-of-Application Study. Journal of Medical Internet Research **17**(1), e16 (2015). https://doi.org/10.2196/jmir.3933
7. Das, S., Sundara, S., Atkinson, R.: R2RML: RDB to RDF Mapping Language. https://www.w3.org/TR/r2rml/ (2012)
8. ERA5, C.C.C.S.C..: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate. Copernicus Climate Change Service Climate Data Store (CDS). ECMWF. https://cds.climate.copernicus.eu/cdsapp#!/home
9. Hasnain, A., Kamdar, M.R., Hasapis, P., Zeginis, D., Warren, C.N., Deus, H.F., Ntalaperas, D., Tarabanis, K., Mehdi, M., Decker, S.: Linked Biomedical Dataspace: Lessons Learned Integrating Data for Drug Discovery. In: International Semantic Web Conference 2014. pp. 114–130. Lecture Notes in Computer Science, Springer International Publishing (2014). https://doi.org/10.1007/978-3-319-11964-9_8

10. Khan, S.: Qualitative Research Method: Grounded Theory. International Journal of Business and Management **9** (Oct 2014). https://doi.org/10.5539/ijbm.v9n11p224

11. Koopman, R.J., Kochendorfer, K.M., Moore, J.L., Mehr, D.R., Wakefield, D.S., Yadamsuren, B., Coberly, J.S., Kruse, R.L., Wakefield, B.J., Belden, J.L.: A Diabetes Dashboard and Physician Efficiency and Accuracy in Accessing Data Needed for High-Quality Diabetes Care. The Annals of Family Medicine **9**(5), 398–405 (Sep 2011). https://doi.org/10.1370/afm.1286

12. Lassila, O., Swick, R.R.: Resource Description Framework (RDF) Model and Syntax Specification. https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/ (1999)

13. Lewis, J.: Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. Int. J. Hum. Comput. Interaction **14**, 463–488 (Sep 2002). https://doi.org/10.1080/10447318.2002.9669130

14. Lewis, J.R.: IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. International Journal of Human–Computer Interaction **7**(1), 57–78 (Jan 1995). https://doi.org/10.1080/10447319509526110

15. Macefield, R.: How To Specify the Participant Group Size for Usability Studies: A Practitioner's Guide. Journal of Usability Studies **5**(1), 12 (2009)

16. McKenna, S., Staheli, D., Fulcher, C., Meyer, M.: BubbleNet: A Cyber Security Dashboard for Visualizing Patterns. Computer Graphics Forum **35**(3), 281–290 (2016). https://doi.org/10.1111/cgf.12904

17. Reddy, B.P., Houlding, B., Hederman, L., Canney, M., Debruyne, C., O'Brien, C., Meehan, A., O'Sullivan, D., Little, M.A.: Data linkage in medical science using the resource description framework: The AVERT model. HRB Open Research **1**, 20 (Mar 2019). https://doi.org/10.12688/hrbopenres.12851.2

18. Reynolds, D., Cyganiak, R.: The RDF Data Cube Vocabulary. https://www.w3.org/TR/vocab-data-cube/ (2014)

19. Sabol, V., Tschinkel, G., Veas, E., Hoefler, P., Mutlu, B., Granitzer, M.: Discovery and Visual Analysis of Linked Data for Humans. The Semantic WebISWC 2014, Lecture Notes in Computer Science **8796**, 309324 (Oct 2014). https://doi.org/10.13140/2.1.3744.2566

20. Sauro, J., Lewis, J.R.: Quantifying the User Experience: Practical Statistics for User Research. Elsevier, Cambridge, 2nd edition edn. (2016)

21. Scharl, A., Herring, D., Rafelsberger, W., Hubmann-Haidvogel, A., Kamolov, R., Fischl, D., Föls, M., Weichselbraun, A.: Semantic Systems and Visual Tools to Support Environmental Communication. IEEE Systems Journal **11**(2), 762–771 (Jun 2017). https://doi.org/10.1109/JSYST.2015.2466439

22. Tie, Y.C., Birks, M., Francis, K.: Grounded theory research: A design framework for novice researchers:. SAGE Open Medicine (Jan 2019). https://doi.org/10.1177/2050312118822927

23. Villanueva-Rosales, N., Chavira, L.G., del Rio, N., Pennington, D.: eScience through the Integration of Data and Models: A Biodiversity Scenario. In: 2015 IEEE 11th International Conference on E-Science. pp. 171–176 (Aug 2015). https://doi.org/10.1109/eScience.2015.77

24. Zainab, S., Saleem, M., Mehmood, Q., Zehra, D., Decker, S., Hasnain, A.: FedViz: A Visual Interface for SPARQL Queries Formulation and Execution. In: VOILA@ISWC2015 (2015)

25. Zhu, X., Wolfgruber, T.K., Tasato, A., Arisdakessian, C., Garmire, D.G., Garmire, L.X.: Granatum: A graphical single-cell RNA-Seq analysis pipeline for genomics scientists. Genome Medicine **9**(1), 108 (Dec 2017). https://doi.org/10.1186/s13073-017-0492-3