

Evaluating the usability of a semantic health data framework: approach and study

Albert Navarro-Gallinad^{a,*}, Fabrizio Orlandi^a and Declan O’Sullivan^a

^a*ADAPT Centre for Digital Content, School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland*

E-mail: albert.navarro@adaptcentre.ie

Abstract. Usability testing presents an opportunity to promote collaboration between domain experts and computer scientists when developing tools and processes to address current data challenges in research and industry. Particularly to reduce the expertise required in using and benefiting from Semantic Web (SW) technologies when integrating heterogeneous data sources through semantic reasoning. We present a useful example of a usability testing approach to evaluate the usability of a framework (i.e. Knowledge Graph, Methodology and User Interface) for Health Data Researchers (HDR), when trying to link particular health events with environmental data to explore the environmental risk factors of rare diseases. The description and results of the evaluation approach are demonstrated for 17 HDRs with expertise in health data related to ANCA associated vasculitis in Ireland and Kawasaki Disease in Japan, and with no previous practical experience in using SW technologies. The usability evaluation results validate the usefulness of the framework allowing researchers themselves to link health and environmental data whilst hiding the complexities of SW technologies. Beyond the direct impact on environmental health studies, the description of the evaluation approach can guide researchers in making SW technologies more accessible to domain experts through usability studies.

Keywords: Usability Testing, Semantic Data Integration, Knowledge Graph, Environmental Health Data, Rare Disease Research

1. Introduction

Emerging technologies can address many of the data challenges that domain experts are facing, but they require knowledge and practical expertise in the technology. Collaboration between computer scientists and domain experts could facilitate overcoming the challenges [1]. For example, a doctor that needs to bring in different data types (e.g. health and environmental data) to understand if the medication prescribed for a patient needs to be reduced or increased during the treatment. In this case, a data engineer with expertise in Knowledge Graphs (KG) could collaborate with the doctor to enable the linkage between the diverse health and environmental data sources [2]. In doing so, the doctor would not only resolve the data interoperability challenge but contribute to research related to the disease or treatment, as KGs provide meaning to the data

(context), querying and reasoning capabilities to generate new insights (efficiency) and transparency of the linkage process (explainability) [3].

A common approach to facilitate the uptake and use of KGs and Semantic Web technologies is to use visualisations [4, 5]. Domain experts could link the health data through a User Interface (UI) that allows experts themselves to run queries in an intelligible manner. Then, the resulting linked data can be exported with the necessary provenance and lineage metadata facilitating the reuse and linkage in other contexts or domains. At this point, researchers will have generated in a semi-automated manner graph data represented using the Resource Description Framework (RDF) [6], which makes their data linkable by machines. Not only that but if the data is licensed, provided with a data use and deposited in a data repository; the data will already be Findable, Accessible, Interoperable and Reusable (FAIR) [7]. Therefore, the SW technologies can facil-

*Corresponding author. E-mail: albert.navarro@adaptcentre.ie.

1 itate the process of making data FAIR towards open
2 science.

3 In this paper, usability testing is used as a method
4 to promote collaboration between Health Data Re-
5 searchers (HDR) and data engineers to accomplish
6 shared goals. The usability testing approach was ap-
7 plied to identify, refine, validate and consolidate the
8 achievement of the requirements from the HDR; when
9 trying to link particular health events with environmen-
10 tal data to explore the environmental risk factors of
11 rare diseases. Previous work from the authors identi-
12 fied the researchers requirements and the limitations in
13 the state of the art [8, 9].

14 **Researcher requirements.** In this case, data linking
15 for HDRs involves querying the data linked through
16 location and time, understanding the linked data and
17 metadata generated during the linkage process, and ex-
18 porting the data to be used as input for data analysis
19 and for open data publication. The definition of link-
20 ing led to the formulation of three requirements, which
21 were extracted from expert consensus:

- 22 – *Requirement 1: Enable HDRs to query environ-*
23 *mental data associated with individual health*
24 *events through location and time, within the re-*
25 *gion of the event and a period of data before the*
26 *event.*
- 27 – *Requirement 2: Support the understanding of*
28 *event-environmental linked data and metadata,*
29 *with its use, limitations and data protection risk*
30 *for individuals.*
- 31 – *Requirement 3: Export event-environmental linked*
32 *(meta)data to be used as input in statistical mod-*
33 *els for data analysis (CSV) and for publication*
34 *(RDF).*

35 **State of the art limitation .** As part of our research,
36 we reviewed tools that facilitate the linkage between
37 health and environmental data using KGs focused in
38 the biomedical domain, and found that in summary the
39 tools required of the domain expert some expertise in
40 KGs for the tools to be practical for them and/or the
41 developers of the tools did not conduct with domain
42 expertise any usability testing using standard metrics.

43 The domain expert driven requirements and limita-
44 tions found in the state of the art, motivated the de-
45 sign of a framework called SERDIF (Semantic Envi-
46 ronmental and Rare Disease data Integration Frame-
47 work), which is a combination of a Methodology, a
48 KG and a UI, in support of HDRs [9]. Furthermore,
49 it motivated us to design a usability evaluation using
50
51

1 standard metrics, which will allow for comparison of
2 progress on usability issues between versions and also
3 as a means to deeply include researchers' perspective
4 during the development process.

5 The evaluation methods used are presented in this
6 article in a general manner to encourage other re-
7 searchers to apply in different contexts, while provid-
8 ing an example of how the approach was applied for
9 our particular scenario, that is with HDR participants,
10 use cases and sample sizes elements (Section 2). Be-
11 fore demonstrating the implementation of the evalu-
12 ation methods, the updates that have been applied to
13 SERDIF as a result of the first usability evaluation that
14 was undertaken by us [9] is described (Section 3). The
15 implementation section starts by describing the experi-
16 mental setup, execution and metrics recorded, and fol-
17 lows with the demonstration of the findings and rec-
18 ommendations (Section 4 and 5). Furthermore, the im-
19 plementation exemplifies how to conduct the analysis
20 of the metrics, providing the data and processing steps
21 to guide researchers when applying the methods, in a
22 transparent manner.

23 The main contribution of this article is the results
24 of the evaluation approach taken to validate the use-
25 fulness of a framework aimed at facilitating the use
26 by health data researchers of Semantic Web technolo-
27 gies in support of data linkage. Two minor contribu-
28 tions, which are a consequence of the main contribu-
29 tion, are the potential implementation of the evaluation
30 approach to other health data linkage studies; and the
31 open source artefacts (i.e. data and processing) to pro-
32 mote reuse, reproducibility and credibility of the ap-
33 proach taken.

34 The relevance of the contributions is highlighted by
35 comparing our evaluation approach with other evalu-
36 ation approaches applied in the development of other
37 tools who aim to promote the use of Semantic Web
38 technologies for health data researchers (Section 6).
39 The comparison includes the limitations and advan-
40 tages of the reviewed studies, compared with the ap-
41 proach presented in this article. Finally, the article con-
42 cludes with the future work and potential impact of
43 collaborations between health experts and computer
44 scientists when trying to address data challenges (Sec-
45 tion 7).

46 2. Evaluation approach

47
48 The methods to evaluate the usefulness of the
49 SERDIF framework through usability testing include
50
51

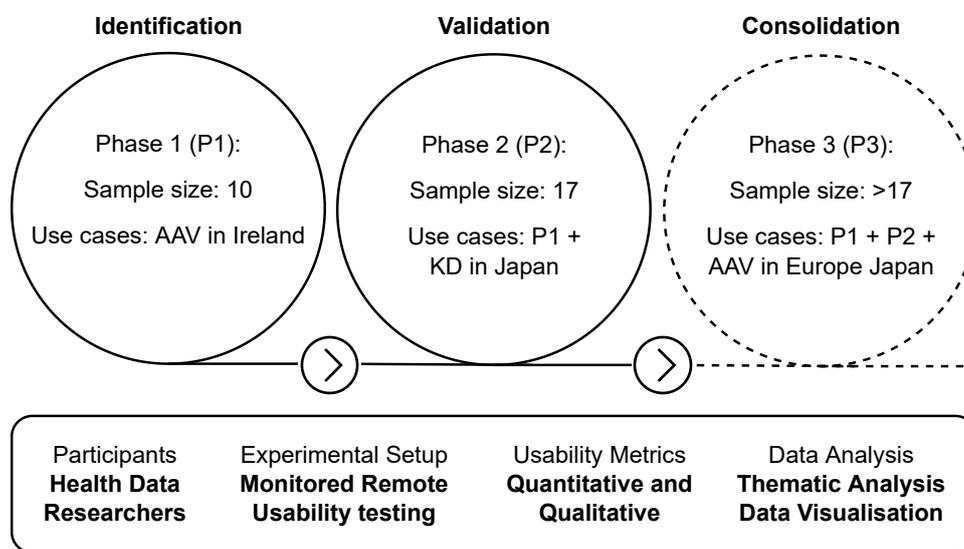


Fig. 1. Overview of the usability testing approach. AAV: Anti-neutrophil cytoplasmic autoantibody (ANCA)-associated vasculitis (AAV). KD: Kawasaki Disease.

static elements: the strategy, participants, experimental setup, metrics and data analysis; and dynamic elements for environmental health studies of rare diseases including the use cases and sample size. We define static elements as the elements that remain the same throughout the usability study while dynamic elements are progressively incorporated in each phase of the usability study. The evaluation approach was submitted to and gained approval from the relevant Ethics Committee in Trinity College Dublin.

2.1. Static evaluation elements

Strategy. The usability testing approach consists of three phases (Fig. 1): (P1) identifying and refining the initial user requirements, (P2) validation of the usefulness of the framework for HDRs and (P3) consolidation of the requirements. Each phase represents an iteration that comprises the development or refinement of the framework based on the user requirements and the evaluation that provides the basis for the refinement of the framework and requirements.

Participants. The participants are HDRs, meaning researchers with a health background and statistical or data analysis experience (e.g. clinicians, health informaticians and epidemiologists), or statisticians and data analysts that are studying health related outcomes. In addition, only HDRs without practical expertise in using KGs are eligible for the study. A minimum of 10 participants is recommended for the first phase of the

study [10], progressively increasing in the following phases. The eligible researchers are recruited by sending two personal emails asking if they would like to participate in the usability study. The first email contains a brief introduction and the rationale of the experiment, and expert participants are asked to complete an online poll to select a time slot that works for them. The second email provides the participants with the details of the experiment: virtual meeting details, participant information sheet and consent form to be read and signed, the UI link and login credentials.

Experimental setup. Participants are asked to complete a series of tasks using the UI while thinking aloud [11]. The tasks were derived from consensus among HDRs with real workflows in mind. Then, the participants are asked to fill out a post usability test survey. The experiment is conducted remotely using a video conferencing platform where participants are asked to share their screen and audio while observed and assisted, if necessary, by a usability moderator.

Metrics. The metrics gathered to assess the usefulness of the framework combine quantitative and qualitative metrics to support the findings.

Quantitative. Time spent to complete each task during the usability testing session (time per task); the task being completed or abandoned with/out assistance from the usability session moderator (task completion or failure); the Post-Study System Usability Questionnaire (PSSUQ) [12], which is a standard questionnaire meant to assess the usability evolution during the de-

velopment of a system (usability test survey scores and scales).

Qualitative. The session transcripts with the think aloud comments of the participants (usability session transcripts); the open answers the user survey, which allowed users to record text statements for further clarification of the scores (usability test survey open comments); the usability moderator recorded participant thoughts in a summarised manner for an initial impression on the usability, which were not used in the analysis (moderator notes).

Data analysis. A Thematic Analysis [13] is conducted on the usability testing session transcripts, supported by the quantitative metrics and Thematic Analysis of the PSSUQ open comments, to identify the findings and recommendations for each phase. Therefore, triangulating participant comments, questionnaire responses and observations to demonstrate the findings from multiple perspectives. The PSSUQ scores and scales are compared with previous phases or similar tools comparing the distributions with box plots.

2.2. Dynamic evaluation elements

Use cases. The use case refers to ongoing rare disease research involved in environmental health studies that would benefit from applying KG to integrate the diverse sources of data. In other words, these are studies that require meaningful data linkage to validate their hypothesis of environmental risk factors, which can contribute in providing new insights to the aetiology of the disease. For example, vasculitis such as AAV and KD (Fig. 1), where the current theory for the aetiology involves a complex interaction between environmental and epigenetic factors, in a genetically susceptible individual [14–17]. Researchers in projects such as HELICAL¹, PARADISE², WIND-BIOME³ and FAIRVASC⁴ provide ideal participants for this particular study as they are international professors, researchers and PhD students with fluent English and no previous practical experience in using KGs.

Sample size. The sample sizes presented in Fig. 1 cover the requirement of a specialised tool for linking environmental and health data [10]. Specific num-

bers are given for Phase 1 and 2 as they have been conducted already, but Phase 3 will be conducted in the coming months after the publication of this article, but is expected to be larger given the numbers involved in the FAIRVASC project who will be the target participants. The study advances in a progressive manner through each phase, including researchers and the use case from the previous phase. This iterative approach improves the chances to find errors, ambiguous information and confusing features while generalising the health data input capabilities of the framework and also expanding the environmental data coverage to different countries.

3. Framework updates after Phase 1

The SERDIF components have been updated after the results from Phase 1 of the usability study and the refined requirements, which have been published as previous work from the authors [9]. The updates are presented here for each of the components of the framework (KG, Methodology and UI) and for the usability testing execution, to provide context for the reader of this article so that the evaluation results of Phase 2 presented in this article can be interpreted.

3.1. Framework components

KG component. The KG component is where environmental data and health data is linked together through location and time using RDF and SPARQL queries. Environmental data is described using the RDF Data Cube vocabulary (QB) [18] instead of the Semantic Sensor Network Ontology (SOSA/SSN) [19] that was used in the KG at the start of P1. QB focuses on describing statistical and multi-dimensional datasets, which presents an advantage as it provides a general data structure and flexibility to represent other types of environmental data besides sensors such as occupational exposures or environments. The KG now includes metadata information beyond the weather and air pollution variables descriptions with the dataset descriptors [20], and the origin and processing of the datasets [21]. While keeping the spatial dimension of the data (i.e. locations and regions) represented as GeoSPARQL [22] geometries, and the time dimension as xsd:dateTime. Enabling spatial and temporal reasoning features, which are key to answer HDR research questions.

¹<https://helical-itn.eu/>

²<https://www.tcd.ie/medicine/thkc/avert/>

³<https://www.isglobal.org/en/-/clima-y-salud>

⁴<https://fairvasc.eu/>

1 Regarding the data sources uplifted to RDF, the air
2 pollution data source was replaced with the European
3 Environmental Agency (EEA)⁵ while the weather data
4 remains the same from MetEireann⁶. The uplifted data
5 includes 25 weather and 2000 air quality individual
6 datasets with hourly data from the 2000-2021 period
7 and for the Republic of Ireland. The total amount of
8 triples from environmental data has increased from
9 30M to 80M triples. The KG data has been published
10 as open data in Zenodo [23] and a SPARQL endpoint
11 is available at the following URL for public use:

12 <https://serdif-example.adaptcentre.ie/>
13
14

15 **Methodology component.** The methodology is a
16 series of steps that guides the researcher in linking par-
17 ticular events with environmental data using SW tech-
18 nologies. Two new steps were added to the method-
19 ology regarding data protection and privacy and open
20 data publication. The data protection and privacy step
21 provides the user with relevant information to make
22 sure the processing of the personal data (i.e. health
23 data) meets the requirements of the contract signed to
24 use the data under a specific purpose (e.g. a data shar-
25 ing agreement or consent form) and complies with the
26 General Data Protection Regulation (GDPR)⁷ in data
27 processing. The open data publication step includes the
28 semi-automatic means to make the data FAIR; for the
29 later publication in an open data repository. The pub-
30 lication of the data will only be possible after explicit
31 permission from the data controller, and when it does
32 not provide means for re-identification of the data sub-
33 ject.

34 **UI component.** The UI component is designed from
35 a user-centric perspective to support HDRs access,
36 explore and export the linked health-environmental
37 data with appropriate visualisations, and by facilitat-
38 ing the query formulation for non-SW experts. The
39 data is linked using a SPARQL query [24] reason-
40 ing over location and time. The GEOSPARQL func-
41 tion `geof:sfWithin` is used to select the environmental
42 datasets within the region of the event (e.g. a county or
43 country). The `xsd:dateTime` datatype allows to select a
44 certain period before the health events. Therefore, en-
45 vironmental data is associated to health data for a par-
46 ticular region and period before the health event, as
47 the researchers are trying to understand the risk factors
48

49 ⁵<https://discomap.eea.europa.eu/map/fme/AirQualityExport.htm>

50 ⁶<https://www.met.ie/climate/available-data/historical-data>

51 ⁷<https://gdpr-info.eu/>

1 that led to the event. The UI was modified to present
2 environmental data linked to individual health events
3 instead of aggregated data for an event type. The mod-
4 ification affected existing features of the UI like the
5 data table (Fig. 2D) and the time series plot, which be-
6 came a temporal heat map (Fig. 2E). The comparative
7 tab was removed as it was confusing to the users and
8 the relevance was not clear at the end of Phase 1. In ad-
9 dition, a check button was added for the user to check
10 if environmental data is available for options selected
11 from the previous dropdowns (Fig. 2B); and a login
12 step was added to avoid multiple participants running
13 queries at the same time, which could affect the exper-
14 imental metrics (Fig. 2A).

15 Then, information about the origin of the data
16 and processing steps was made available through the
17 data provenance and lineage buttons as RDF graphs
18 (Fig. 2C). The user can also explore the full meta-
19 data generated for the linkage process that includes
20 dataset descriptors (e.g. licence, distribution, temporal
21 and spatial information and structure of the dataset),
22 data provenance and lineage information (e.g. datasets
23 used and SPARQL query to link the data), data protec-
24 tion and privacy aspects of the linked data (e.g. data
25 controller, processing risks and purpose of the linkage)
26 and the data use (e.g. for studying a specific disease
27 by researchers with a signed data sharing agreement to
28 process the health data). Furthermore, the results from
29 a query can be exported as FAIR data with the poten-
30 tial for researchers to further link the data with other
31 studies and for open data publication.

32 3.2. Improving the experimental setup 33

34
35 The participants experienced some delays and control
36 malfunctions during the P1 of the usability study.
37 That was because the moderator had to open the UI
38 in a local computer, share the screen during the video
39 conference call and then, give remote control to the
40 participant. In P2, an example UI was made available
41 for the participants under the following URL:

42 <https://serdif-example-dash.herokuapp.com/>
43
44

45 By example UI, we mean that the health data ac-
46 cessed and integrated with the environmental data is
47 simulated health data. The health data is available in
48 a different repository within the triplestore (i.e. inter-
49 nal repository) to work as a prove of concept for a fed-
50 erated scenario, where health data does not leave the
51 storage location but it is only consulted. The improve-

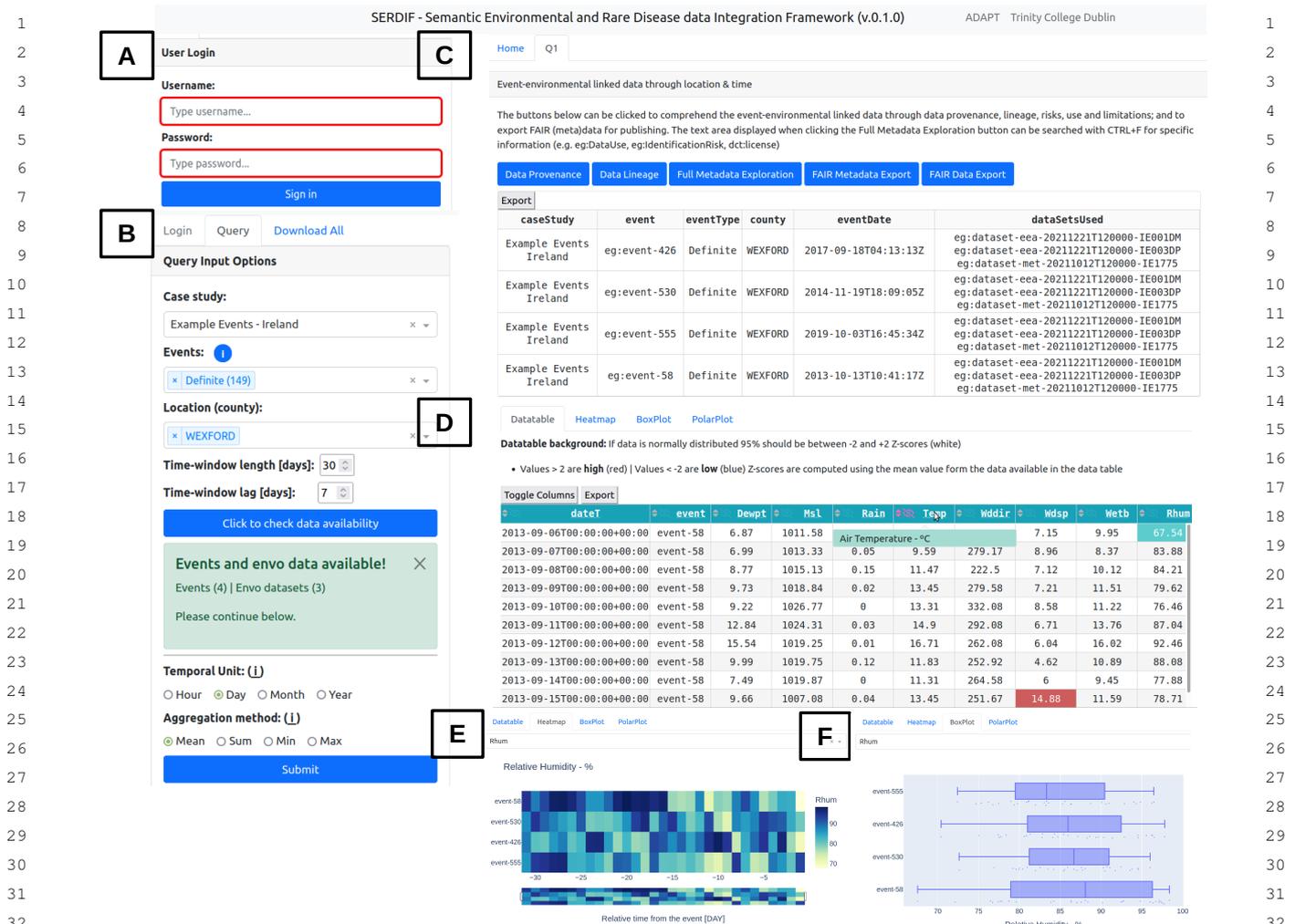


Fig. 2. Screenshot of the SERDIF UI displaying (A) the login panel, (B) the query input panel with selected options, (C) the tabs generated after submitting a query, which includes metadata section with the FAIR export button, (D) a data table and three different visualisations, (E) heatmap, (F) box plot and polar plot.

ment also granted the possibility to include researchers that had not signed a data sharing agreement to use and process the health data (Section 4).

In addition, the tasks were rephrased using simpler language and dividing the tasks in subtasks to improve the readability and avoid having to go back and forth between the PDF document with the tasks and the browser window with the UI (Table 1).

4. Evaluation Execution: Phase 2

The purpose of the Phase 2 (P2) of the evaluation was to evaluate the usefulness of the SERDIF frame-

work, after the refinements from P1, for HDRs in their hypothesis validation of environmental factors for rare disease research. The evaluation was conducted with 17 HDRs with expertise in health data related to AAV in Ireland and KD in Japan during the fourth quarter of 2021. In particular, the researchers were international (associate) professors, clinicians, research assistants, technicians and PhD students with fluent English, and without practical experience in using KGs. The participants were recruited by email, which was facilitated by the involvement of the authors in the HELICAL, PARADISE and WINDBIOME projects. The participants were asked to complete the tasks in Table 1 while thinking aloud and using the SERDIF UI component.

Table 1

Tasks that the participants are asked to complete during the usability testing associated with the requirements from Section 1.

Task	Requirement	Description
T1	R1	Read the 'Home' tab information. (a) Was the information presented enough for you to get an overall understanding of the SERDIF framework? (b) If not, what information would you add?
T2	R1	Login with the credentials provided in the invitation email.
T3	R1	Submit a query using the input options available in the 'Query' tab on the left panel. (a) Please explain and justify aloud the specific choices that you make. (b) Are the query input options clear to you? (c) Would you add any extra information to get a better understanding? (d) After completing all the inputs, please click the 'Submit' button once.
T4	R2	Explore the 'Q1' tab generated after submitting a query. (a) Read the information provided in the introductory paragraph. (b) Explore and discuss aloud the usefulness of the information displayed for each of the three buttons below, in order to understand the event-environmental linked data. (c) Click the following buttons: Data Provenance, Data Lineage and Full Metadata Exploration. (d) Can you say aloud how many data sets were used for each event in the Data Provenance table? (e) Can you identify aloud the data use comment (eg:DataUse), identification risk comment (eg:IdentificationRisk) and license (dct:license) in the Full Metadata Exploration text?
T5	R2	Explore the data table displayed underneath the metadata buttons. (a) Can you understand all the column headings when hovering over the abbreviations? (b) Do you understand the meaning of the cell background colours? (c) Can you hide columns by using the eye icon next to the column heading or by using the 'Toggle Columns' button? (d) After coming to conclusion on the previous questions, discuss aloud whether you found the data table useful to comprehend the event-environmental linked data? (e) Say aloud if you would add any feature to the current display
T6	R2	Explore the Heat map, Box Plot and Polar Plot tabs. (a) Select an environmental variable from the inputs provided. (b) Please explain aloud the usefulness and clarity of each plot. (c) Say aloud if you would add any more visualisations to the current ones.
T7	R2	Submit one more query with different input options from the ones used before. (a) Which individual event uses the most data sets between Q1 and Q2? (b) Do both data sets (Q1 and Q2) have the same license? (c) Which of the data tables (Q1 and Q2) seem to have more extreme values (i.e., more colored cell backgrounds)? (d) Choose one of the following plots from Heat map, Box Plot and Polar Plot tabs and compare aloud Q1 and Q2 plots. (e) Say aloud if you would add any other feature to help you comprehend better the data at this point.
T8	R3	Export/download the metadata and data from the first query (Q1) (a) FAIR data and metadata. (b) Data provenance and Data table as CSV files. (c) Export all data tables at once in the 'Download All' tab in the left panel. (d) Say aloud whether such event-environmental linked data and metadata would be useful and re-usable as input for environmental research.
T9	All	Could you summarise verbally your overall experience in completing these tasks using the SERDIF dashboard?

The metrics recorded throughout the P2 of the usability evaluation (Section 2) and the subsequent data analysis leading to the findings and recommendations are presented in this section, following the six steps towards a trustworthy Thematic Analysis by Nowell et al. 2017 [13].

4.1. Step 1: Familiarising yourself with your data

While in a typical Thematic Analysis only qualitative data (in many formats) is considered, quantitative metrics were also included as they provided further insight in the analysis to support the findings. The recorded metrics (data) include the previous quantitative and qualitative metrics mentioned in Section 2.1, which are available in the data folder of the paper's

1 GitHub repository:

2
3 <https://github.com/navarral/swj2022-usability-paper>

4
5 Data was collected by different interactive means
6 by the usability moderator: stopwatch (time per task),
7 manual typing during the session (moderator notes),
8 automatic process using an online version of the ques-
9 tionnaire (usability test survey scores, scales and open
10 comments), manual transcription of the audio record-
11 ings of the sessions (usability session transcripts), and
12 manual input after the usability session transcription
13 (task completion or failure).

14 It was found that the accuracy of the automatic tran-
15 scription feature of the video conferencing platform
16 used was not enough for the later coding. Many sen-
17 tences and words could not be understood without the
18 context and previous involvement in moderating the
19 experiment. As a consequence, the transcription pro-
20 cess was performed manually from the audio record-
21 ings of the sessions and the text was organised by
22 each task heading, improving the readability of the
23 documents. The manual transcription was a time con-
24 suming process as the sessions lasted between 30 and
25 60min but led to rich documents facilitating the fol-
26 lowing steps. The audios and transcripts were stored
27 locally and then, the audios were deleted and the tran-
28 scriptions without any personal information uploaded into
29 a Taguette [25] local database for the subsequent cod-
30 ing. The 17 documents are named Participant ID_date
31 time (e.g. P1_GMT20211118) for easier management
32 and interpretation of the audit trail.

33 34 4.2. Step 2: Generating initial codes

35
36 The interactive data collection influenced the gener-
37 ation of the codes as the moderator and data analyst
38 were the same person, but two other researchers, au-
39 thors of this paper, were involved in the coding pro-
40 cess. Besides the initial influence mentioned above,
41 the researcher did not use an initial codebook at the
42 start of the process, addressing the data with an open
43 mind. The initial generation of the codes was per-
44 formed line by line and assigning specific codes. Then,
45 redundant codes were merged resulting into 120 dif-
46 ferent initial codes for the 17 documents (see ‘code-
47 book_20220223_0.csv’ in the paper’s Github). The
48 number of highlights per code started to show some
49 tendency for some of the codes. In addition, the code
50 ‘open tag’ stored sentences which seemed to not fit

1 into any of the other codes but that could be useful in
2 the later iterations of the coding process.

3 The Taguette software proved to be a great asset
4 for the tagging and categorisation of the sentences in
5 the transcripts, facilitating the researcher complex task
6 to highlight the sentence, assign a code and write a
7 memo as a description of the code for the debriefing
8 with the other two researchers in Step 3. The memo
9 was useful as some of the codes’ names would have
10 been hard to be understood by themselves for the other
11 two researchers involved in the analysis and future re-
12 searchers trying to reproduce this work that are experts
13 in the content (Appendix A).

14 The researchers met three times within 8 days for
15 a debriefing of the coding process. In each meeting,
16 the researchers tried to reduce the number of codes
17 by merging codes discussing similar concepts. The
18 descriptions associated to the merging codes were
19 rephrased to include all the relevant information from
20 the predeceasing codes. After 8 iterations, the final
21 codebook was agreed by consensus between the three
22 researchers. The final codebook included 22 codes
23 with at least 30 references of each, which were gener-
24 ated from the 120 initial ones (Appendix A). The code-
25 book generated at each iteration and a document sum-
26 marising the 1930 highlights are available in the pa-
27 per’s GitHub.

28 29 4.3. Step 3: Searching for themes

30
31 The searching for themes process was conducted
32 by displaying the number of references per partici-
33 pant for each code as an annotated heat map (Fig. 4E).
34 The heat map was generated using one of the exports
35 supported in Taguette’s software. The export included
36 the transcription highlights per document (i.e. per par-
37 ticipant ID) with the associated code as a csv file
38 (all_tags_date.csv documents in the paper’s GitHub).

39 The heat map rows were reordered by number of
40 references but also by their importance in assessing the
41 objective of the usability test: ‘*Validate the usefulness*
42 *of the framework for the users*’. The reordering natu-
43 rally led to codes that identified with the achievement
44 of the requirements. However, some of the codes did
45 not fit into the requirements, which were left as ‘open
46 theme’ and not discarded at this step.

47 48 4.4. Step 4: Reviewing themes

49
50 The document highlights were read again two times
51 to validate the assignment of each code to a particular

theme. During the reviewing, the naming and descriptions of the codes had to be adjusted. In addition, the ‘open theme’ codes started to shape into two different themes regarding new features or functionalities not present in the framework and references to the overall usability experience when conducting the experiment. The participants’ experience was then included in the five themes.

4.5. Step 5: Defining and naming themes

The five themes identified are described in detail as the findings of the usability study, supported by the quantitative metrics and Thematic Analysis of the PSSUQ open comments (Section 5). Therefore, triangulating participant comments, questionnaire responses and observations to demonstrate the findings from multiple perspectives. The combination of the metrics is visualised in Fig. 4 to serve as reference when providing the evidence for each finding. Showcasing the strength of the proposed approach of combining quantitative and qualitative data within the definition of the themes. In addition, the comparison between the PSSUQ scales between the first and second phase of the usability study is represented in Fig. 3.

4.6. Step 6: Producing the report

The thematic analysis report was written as a usability report since the aim of the analysis was to study the metrics recorded during the usability evaluation. Therefore, the structure of the document was as follows: executive summary, methods, metrics, results, analysis, findings and recommendations, and next steps. The content of the report is included in this paper together with the recorded metrics in the paper GitHub to ensure the transparency of the methods and findings. The study did not finalise by checking the findings with all the participants as it was not viable to organise a meeting where the 17 expert participants were present and could participate in a discussion. The participants were sent the usability report by email and they had the chance to come back to the authors with further comments or feedback.

5. Evaluation results: Phase 2

The results, findings and recommendations sections of the usability report are presented in this section.

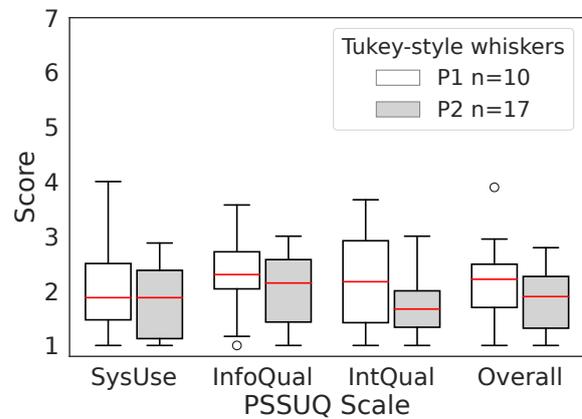


Fig. 3. Comparative PSSUQ scales box plots for System Usefulness (SysUse), Information Quality (InfoQual), Interface Quality (IntQual) and Overall for the P1 (white) and P2 (grey) phases of the usability testing. The sample sizes for each of the metrics are 80,70,30 and 190 (P1) and 136, 119, 51 and 323 (P2) respectively. The scores are in a Likert 7 points scale where the lower the value the higher the satisfaction.

The themes name, description, number of references, progress towards achieving the requirements and scope of the finding in terms of the framework artefacts (KG, Methodology and UI) from the participants are summarised in Table 2. The themes clearly capture the progress towards achieving the user requirements important for the overall research question, while also including emerging requirements and the experience when completing the tasks using the user interface.

The findings supported by the other metrics together with the recommendations are described in the section below.

5.1. Requirement 1: Querying

Finding. “While querying environmental data associated with particular events was possible, the event concept and approach were complex, confusing the query process.”

Evidence. The 17 expert participants were able to complete the first three tasks (T1, T2 and T3), which were formulated based on the querying requirement (Fig. 4A). The time per task variability is below 2 minutes for the first three tasks (Fig. 4C), supporting that querying was possible for all participants in a similar manner. The majority of the code references for the first theme indicate clarity and usefulness of the query inputs, elements and processes (253 out of 386),

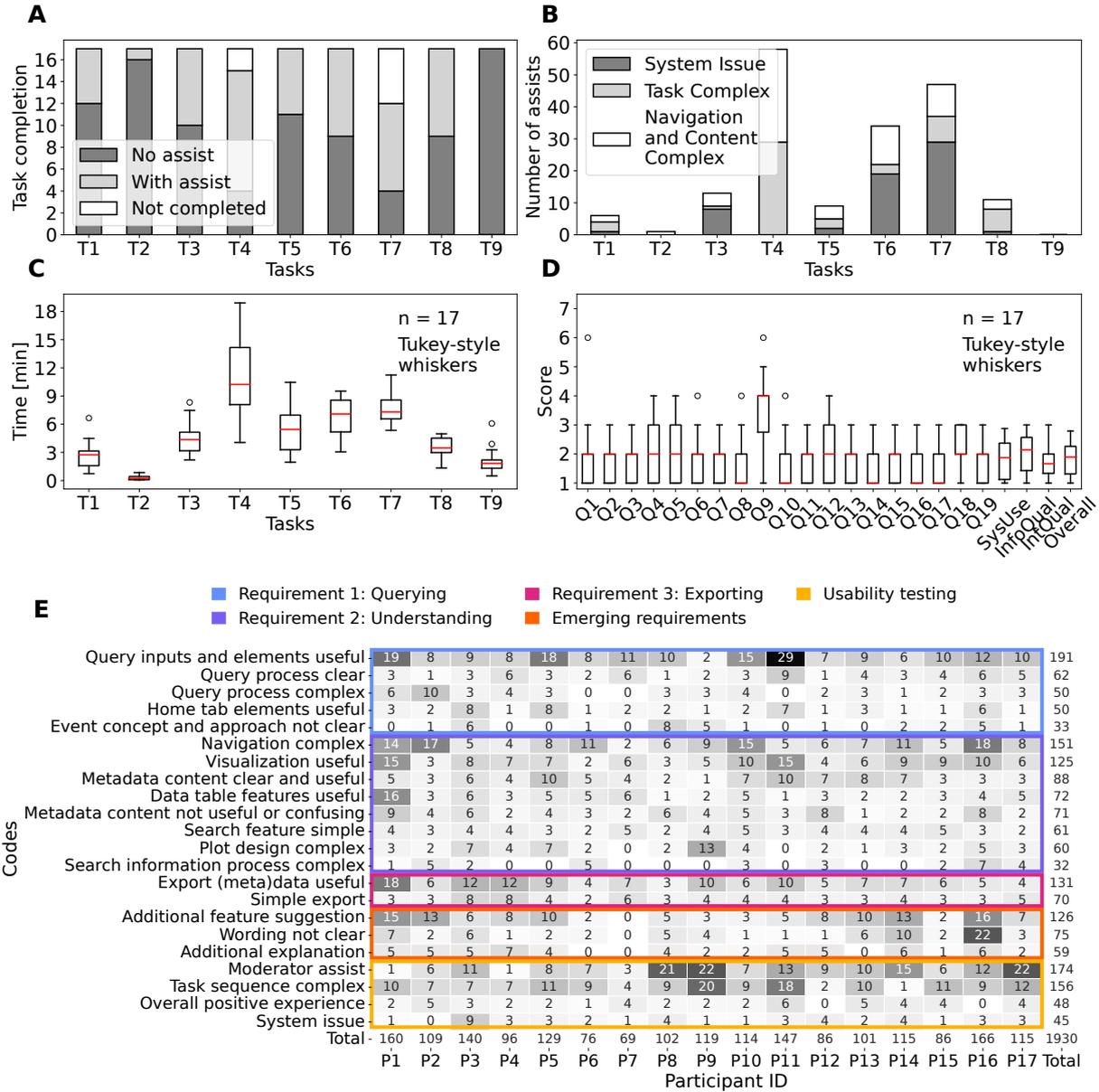


Fig. 4. Combination of the quantitative and qualitative metrics collected in the Phase 2 of the evaluation. **A** – Task completion bar plot including the if assistance was needed from the participants (n=17): no assistance required (no assist, dark grey), with at least one intervention to assist (with assist, light grey) and task or subtasks not completed (not completed, white). **B** – Number of assists from the usability testing moderator to the participants (n=17) including the type of assistance needed: the system is crashed or is not responsive (System Issue, dark grey), the task is confusing or not completed in a sequential manner (Task Complex, light grey) and the design and/or content of the user interface complicates the completion of the tasks (white). **C** – Time spent magnitude and variability to complete each task during the usability session represented as a box plot. **D** – PSSUQ scores box plots with the four averaged metrics (SysUse, InfoQual, IntQual and Overall) on the right end with a sample sizes of 136, 119, 51 and 323. The scores are in a Likert 7 points scale where the lower the value the higher the satisfaction. **E** – Codes and themes that emerged from the thematic analysis of the usability sessions transcripts as a heatmap for traceability.

Table 2

Thematic Analysis summary of the usability sessions transcripts colour coded as in Fig. 4E heat map.

Themes	Findings	References	Progress	Scope
Requirement 1: Querying	While querying environmental data associated with particular events was possible, the event concept and approach were complex, confusing the query process.	386	In progress	Framework
Requirement 2: Understanding	The metadata is enough to understand the provenance and lineage of the linked data and the visualisations are useful to explore the data but the content is complex and hard to navigate making it not user friendly.	660	In progress	UI
Requirement 3: Querying	Exporting the (meta)data is simple and useful to be used as input for analysis.	201	Completed	Framework
Emerging Requirements	Additional features and explanations together with simpler words would increase the usability of the framework.	260	New	Framework
Usability Testing	Overall positive experience when using the dashboard but moderator interventions were needed due to system technical issues and task design.	423	In progress	UI, task design

and the codes are referenced across all participants (Fig. 4E).

However, the tasks that involved actual querying (T3 and T7) required the most assistance due to system issues related to a memory exceeding issue due to the hosting service used during the usability session. The memory limit was noticed three times more in T7 since the free memory was even less than at the start (T3). The system issue is also reflected in the high PSSUQ score in Q9 (Fig. 4D).

Furthermore, 14 out of 17 participants found the query process and query inputs complex at least once (Fig. 4E). That was mostly due to the time window length and lag and the event query inputs, which together made the query process less straightforward. The PSSUQ open comments reinforce the complexity of the query process and the understanding of the event approach (10 out of the 13 comments, Appendix B). An event is something that occurs in a certain place during a particular time. While it was understood that environmental data was being gathered for events that happen within a spatial region, the temporal part of the event was less clear. Our hypothesis is that the events exemplified a particular use case but the researchers were thinking of answering their own research question. Therefore, limiting the events to a particular use case may be the source of the confusion in the query process together with the terminology of the time window parameters.

Recommendations:

1. P2.R.Qu.1: Host the dashboard in a site, service or virtual machine where memory capacity should not be a problem for a regular query.

2. P2.R.Qu.2: Provide an example to visualise the time window parameters and the outcome of the query.
3. P2.R.Qu.3: Allow users to input their own events to answer their research questions instead of limiting to the available example event data in the triplestore.

5.2. Requirement 2: Understanding

Finding. “The metadata is enough to understand the provenance and lineage of the linked data and the visualisations are useful to explore the data but the content is complex and hard to navigate making it not user friendly.”

Evidence. Most of the expert participants needed assistance when completing T4, T5, T6 and T7, which are tasks formulated towards the understanding of the event-environmental linked data (Fig. 4A). The task completion pattern together with the time per task magnitude and variability of tasks T4, T5, T6 and T7 indicate that these were complicated tasks compared to the rest (Fig. 4A,C). The results are also coherent with the number of subtasks within each task.

Furthermore, the most complicated task, T4, was associated with the understanding where data came from (i.e. data provenance) and the processing steps from the initial data sources to the aggregated version (i.e. data lineage). Therefore, learning about the data from understanding the content of the metadata. While the metadata was useful to learn about the data, the content was complex and searching for the information was not straightforward. This statement is supported by the codes emerged regarding the metadata from the

1 usability transcripts (Fig. 4E) and PSSUQ open com-
 2 ments (Appendix B) together with the Q12 PSSUQ
 3 score (Fig. 4D).

4 The visualisation of the data as a data table and with
 5 plots have been useful for the participants. The fea-
 6 tures in the data table were appreciated when trying to
 7 have a first insight of the queried data. While the plots
 8 were also useful to explore the data from different per-
 9 spectives, they need an additional explanation to clar-
 10 ify some of the elements and guide the participants in
 11 what they should be looking for (Fig. 4E).

12 Nevertheless, the most negative finding of the us-
 13 ability study is that the navigation was complex as sup-
 14 ported by the type of assists for tasks T4, T5, T6 and
 15 T7 (Fig. 4B), and the code with the most references
 16 from the *Requirement 2: Understanding* theme in the
 17 usability transcripts (151 out of 660, Fig. 4E) and the
 18 PSSUQ open comments (12 out of 28, Appendix B).
 19 Therefore, the layout of the dashboard needs to be im-
 20 proved to be more inline with the workflows and story
 21 lines from the researchers.

22 **Recommendations:**

- 23 1. P2.R.Un.1: Metadata content needs to be simpli-
 24 fied and summarised in a user friendly way.
- 25 2. P2.R.Un.2: Visualisations need explanations to
 26 guide researchers in the interpretation of the re-
 27 sults.
- 28 3. P2.R.Un.3: The dashboard approach with multi-
 29 ple tabs will be switched to a simplified user in-
 30 terface with 3 main sequential steps: (i) upload
 31 the event data, (ii) select query options and (iii)
 32 exportable and explorable output.

33 *5.3. Requirement 3: Exporting*

34 **Finding.** *“Exporting the (meta)data is simple and
 35 useful to be used as input for analysis.”*

36 **Evidence.** All of the expet participants completed
 37 T8 (Fig. 4A), which was about exporting the data and
 38 metadata, but 8 participants needed a one time inter-
 39 vention due to the complexity of the task (5) or be-
 40 cause the navigation was complex (3) (Fig. 4B). Only
 41 one participant needed 3 interventions, 2 related to the
 42 task complexity and 1 to the complex navigation (Ap-
 43 pendix B). The short time spent in T8 and the low
 44 variability below 2 min support the simplicity of the
 45 (meta)data export (Fig. 4C).

46 The number of references for the codes related to
 47 the usefulness (131) and simplicity (70) to export the

1 (meta)data (Fig. 4E). Exporting the data for a subse-
 2 quent analysis is adamant for researchers. The meta-
 3 data is also useful to understand and contextualise
 4 the data for the researcher but also to allow other re-
 5 searchers to reuse the data after publication. However,
 6 some of the comments included in the navigation com-
 7 plex code referred to the confusion around the graph
 8 formats (i.e. rdf and ttl) of the exports (Fig. 4E).

9 **Recommendations:**

- 10 1. P2.R.Ex.1: Allow users to choose the data export
 11 format between data table and graph.

12 *5.4. Emerging Requirements*

13 **Finding.** *“Additional features and explanations to-
 14 gether with simpler words would increase the usability
 15 of the framework.”*

16 **Evidence.** The overall framework is useful for the
 17 participants, which is represented by the Q19 and
 18 SysUse scores, but the framework needs additional
 19 features denoted by the higher score in Q18 (Fig. 4D).
 20 Most of the references for the forth theme, emerging
 21 requirements, are related towards extending the cur-
 22 rent features of the framework in the usability tran-
 23 scripts (126 out of 260, Fig. 4E) and PSSUQ open
 24 comments (21 out of 29, Appendix B) Thematic Anal-
 25 ysis. All participants but one suggested additional fea-
 26 tures, which would increase the usability of the frame-
 27 work. The additional features include: (i) adding an ad-
 28 vanced aggregation method and selecting all query fea-
 29 tures, (ii) summarising the metadata in a user-friendly
 30 way, (iii) facilitating the understanding of the data ta-
 31 ble and increasing the amount of data, (iv) defining, se-
 32 lecting and grouping events feature for exploration, (v)
 33 adding a time series, scatter and histogram plots, and
 34 (vi) distinguishing outliers based on historical data.

35 Even though the framework is specialised for envi-
 36 ronmental health research, some of the technical terms
 37 and paragraph wording were hard to understand. All
 38 but one participant referred to the confusing wording
 39 at least twice with an overall reference number of 75
 40 out of 260 (Fig. 4E). In particular, the terms in the
 41 metadata content, which require additional explana-
 42 tion. Another element that required further explanation
 43 was the plots to explore the data. The number of ref-
 44 erences in the usability sessions transcript (59 out of
 45 260, Fig. 4E) and the PSSUQ open comments (7 out
 46 of 29, Appendix B) support the need for additional ex-
 47
 48
 49
 50
 51

planations to facilitate the use of the framework.

Recommendations:

1. P2.ER.1: Include the additional features suggested by the users.
2. P2.ER.2: Simplify the wording when possible and include a tooltip providing a definition for complicated terms.
3. P2.ER.3: Include an explanation for all the visualisations in the dashboard.

5.5. Usability Testing

Finding. “Overall positive experience when using the dashboard but moderator interventions were needed due to system technical issues and task design.”

Evidence. The expert participants had an overall positive experience when using the dashboard to complete the tasks as supported by the number of references in the usability sessions transcript (48 out of 423, Fig. 4E) and the PSSUQ open comments (16 out of 26, Appendix B). The Overall scale of the PSSUQ displays a distribution of values between 1 and 2 (Fig. 3), which reinforces the positive results from the Thematic Analysis. In addition, the Interface Quality and Overall PSSUQ scales have improved in respect to the P1 usability evaluation (Fig. 3).

However, the participants required assistance to complete most of the tasks related to system issues and the task being complex to understand (Fig. 4B). The system issues lead to the non completion of T7 for 5 out of 17 participants (Fig. 4A) due to the memory limit explained in 5.1. Following, some of the plots did not work for certain environmental variables, mostly for pollutants, leading to confusion.

The participants also needed assistance in understanding the tasks or found the tasks complex (Fig. 4E), which could mean that the sequence of tasks do not represent a real workflow or the wording of the tasks is not intelligible. In addition, the PSSUQ score distribution for Q4 and Q5, which are related to the quickness and efficiency when completing the tasks, are higher than the rest of the questions (Fig. 4D). Most of the assists due to a task being complex (29 out of 54) are for T4 (Fig. 4B), which represent a particular step in the workflow that the participants had trouble reinforcing the discussion about metadata not being clear in the second finding.

Recommendations:

1. P2.UT.1: Check why some of the environmental variables do not display a plot.
2. P2.UT.2: The wording and story line of the tasks will be improved and simplified.

6. Related work

This section overviews the state-of-the-art in empirical usability evaluations of user interfaces (UI) to facilitate the use of Semantic Web technologies (i.e. RDF, OWL or SPARQL) for health researchers without practical experience in using the technologies. It focuses on studies where users interact with the UI in the health domain, and not use-case studies that present how the tool can answer competency or domain specific questions (e.g. ontology evaluations or rank search systems). The review is centred on the metrics used, the evaluation strategy, the availability of the raw and processed data and the sufficient description to reproduce the results of the study.

A common strategy when evaluating the usability of a UI is the combination of quantitative and qualitative metrics, identifying the issues and the underlying reason why they are an issue for the users respectively. The qualitative metrics concur with the ones described in Section 2.1, transcriptions of the usability sessions and the open comments from the usability surveys. However, some of the reviewed studies only included interviews with invited doctors [26, 27] or conducted a structured interview encouraging users to think aloud [28] without a usability questionnaire. Including a standard usability questionnaire would have improved the understanding of the UI’s usability as it would have provided the study with the possibility to compare the results with further versions or similar tools (reliable). In addition, the questionnaires have been validated to be sensible to small sample sizes (<10), which tends to be the most common size for the reviewed studies, as well as, including an objective and quantitative view to the study.

In other studies, the System Usability Scale (SUS) questionnaire [29] was used alone to evaluate the usability and efficacy of a tool for a specific purpose, repurposing tool [30]. The SUS questionnaire was also combined with the Thematic analysis of the experiment transcription obtained from completing the tasks while thinking aloud [31–35]. Only one of the six studies reviewed using a usability questionnaire used the PSSUQ in combination with the thematic analysis [36]. The choice between the use of SUS or PSSUQ

1 depends on each case. For example, the evaluation approach
2 presented in this article decided to adopt the
3 PSSUQ since the questionnaire was designed specifically
4 for scenario-based usability studies, and it provides
5 scores for the System Usefulness, Information
6 Quality, Interface Quality scales for a more detailed
7 view of how the HDR perceived the usability of the
8 SERDIF framework.

9 Regarding the evaluation strategy, most of the reviewed
10 studies included health experts from professional and
11 researcher profiles. Therefore, the purpose of the studies
12 was to facilitate the use of SW technologies to domain
13 experts. However, a single usability evaluation was
14 presented without the reporting of further evaluation
15 comparing the results to track the progress. Researchers
16 mentioned the intention to do further evaluations in
17 the strategy or plan as future work since the studies
18 were not at that stage. Only two studies [34, 35] provided
19 follow-up evaluations comparing the SUS results of
20 college students, consumers or workers when using a
21 UI. The authors argue that the SUS results did not
22 follow an increasing trend from phase to phase (i.e.
23 improving the usability) since new functionalities or
24 features requested by the users were added, making
25 the UI more complex.

26 Coinciding with the evaluation results of the SERDIF
27 framework, they concluded that dividing complex
28 tasks in subtasks and choosing simplicity over including
29 all functionalities will improve the usability of the
30 UI. Nonetheless, only [36] provided information about
31 the assistance required by the usability moderator for
32 the participants when completing the tasks. The
33 assistance combined with the completion or failure
34 assessment and the time per task can be a key
35 combination of metrics when trying to understand the
36 adequacy of a task in the experiment and the complexity
37 for the user.

38 The reviewed studies did not provide either the raw
39 and processed data or sufficient description to reproduce
40 the results of the study as supplementary information
41 or deposited in an open data repository generating a
42 persistent Digital Object Identifier (DOI). The additional
43 information provided was limited to tables and figures
44 linking the themes, usability findings and recommendations.
45 Therefore, the complicating reproducibility of the work
46 by other researchers and making it hard for novice
47 researchers in usability studies but knowledgeable in
48 SW technologies to learn how to analyse the results.

49 In contrast to the related work outlined above, this
50 article presents the result of the evaluation approach to
51 validate the usefulness of a SW-based detailing each

1 of the steps of thematic analysis and provides the
2 necessary data to fully reproduce the study. The manual
3 transcription process allows for the removal of any
4 personal information (e.g. references to people, experiences
5 or locations) carefully, making the data safe to be
6 published together with the methods on GitHub. The
7 results and the description of the methods can guide
8 and promote future usability studies of tools to make
9 the SW technologies more accessible to domain
10 experts, and researchers seeking to learn about SW
11 technologies. Therefore, answering the call for more
12 research related to reducing the expertise required to
13 benefit from the technologies and engaging new
14 researchers from other domains [37].

15 Our evaluation approach combines the usability
16 metrics discussing the quantitative results within the
17 Thematic analysis process (Section 5) supporting the
18 identification and description of the themes as the
19 findings of the usability test. The results are collated
20 together in a multi-panel picture reducing the cognitive
21 load compared to presenting the information as tables
22 and text as in the literature reviewed. The display
23 also offers the multi angle perspective needed to face
24 the complex process of evaluating the usability of a UI.

25 The PSSUQ scales improved or remained the same
26 from P1 to P2 (Fig. 3), even though new functionalities
27 and features have been added to the framework
28 required by the users. The increase in usability
29 between the two phases validates the usefulness of the
30 framework but also the adequacy of the usability
31 evaluation approach. In addition, the involvement of
32 the authors in the research project mentioned above,
33 facilitated the recruitment of participants. Addressing
34 one of the main issues for the lack of user studies
35 in the SW domain [4]. The increase in the sample
36 size between P1 and P2 together with the positive
37 results support the generalisation of the framework
38 beyond a single case study.

39 However, our study presents a few limitations. The
40 design of the framework and tasks are focused on
41 simplifying the formulation of SPARQL queries,
42 exploration of the retrieved data (CSV) and the
43 generation of FAIR data (RDF). Therefore, the design
44 does not provide additional tools beyond the
45 GraphDB triplestore⁸ functionalities to explore the
46 KG, which limits the property and class exploration
47 of the KG. Following, the data in the KG needs to
48 be relevant to each use case, which did not happen
49 for researchers study-

⁸<https://www.ontotext.com/products/graphdb/>

ing KD in Japan. That was because only Irish health events and environmental data were available in the KG. Therefore, HDRs need to be familiar with the health events.

7. Future work and conclusions

From the results obtained from the evaluations performed to date (P1 and P2), the usability testing approach was able to validate the usefulness of the framework for HDR. However, while researchers can link particular health events with environmental data to explore environmental risk factors of rare diseases, the framework needs to be refined to enhance the linkage process.

A third usability study (P3) will be conducted on the updated version of the framework towards consolidating the requirements. The sample size will be increased one more time including researchers studying ANCA associated vasculitis across Europe. The KG will be expanded accordingly with environmental data from the necessary countries and with geometry data representing those countries. Therefore, generalising once more the results to support the usefulness of the framework for environmental health studies.

Besides the direct impact on environmental health studies, the framework has the potential to become a component to base an early warning system for public health that could support doctors making informed decisions.

Acknowledgements

This research was conducted with the financial support of HELICAL as part of the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 813545 and No. 713567 at the ADAPT Centre for Digital Content Technology (grant number 13/RC/2106_P2) at Trinity College Dublin.

References

- [1] Y. Mao, D. Wang, M. Muller, K.R. Varshney, I. Baldini, C. Dugan and A. Mojsilović, How Data Scientists Work Together With Domain Experts in Scientific Collaborations: To Find The Right Answer Or To Ask The Right Question?, *Proceedings of the ACM on Human-Computer Interaction* **3**(GROUP) (2019), 237:1–237:23. doi:10.1145/3361118.
- [2] M. Roos, E. López Martin and M.D. Wilkinson, Preparing Data at the Source to Foster Interoperability across Rare Disease Resources, in: *Rare Diseases Epidemiology: Update and Overview*, M. Posada de la Paz, D. Taruscio and S.C. Groft, eds, Advances in Experimental Medicine and Biology, Springer International Publishing, Cham, 2017, pp. 165–179. ISBN 978-3-319-67144-4. doi:10.1007/978-3-319-67144-4_9.
- [3] A. Hogan, E. Blomqvist, M. Cochez, C. D'amato, G.D. Melo, C. Gutierrez, S. Kirrane, J.E.L. Gayo, R. Navigli, S. Neumaier, A.-C.N. Ngomo, A. Polleres, S.M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab and A. Zimmermann, Knowledge Graphs, *ACM Computing Surveys* **54**(4) (2022), 1–37. doi:10.1145/3447772.
- [4] C. Pesquita, V. Ivanova, S. Lohmann and P. Lambrix, A Framework to Conduct and Report on Empirical User Studies in Semantic Web Contexts, in: *Knowledge Engineering and Knowledge Management*, C. Faron Zucker, C. Ghidini, A. Napoli and Y. Toussaint, eds, Lecture Notes in Computer Science, Springer International Publishing, Cham, 2018, pp. 567–583. ISBN 978-3-030-03667-6.
- [5] M. Al-Tawil, V. Dimitrova and D. Thakker, Using knowledge anchors to facilitate user exploration of data graphs, *Semantic Web* **11**(2) (2020), 205–234. doi:10.3233/SW-190347.
- [6] O. Lassila and R.R. Swick, Resource Description Framework (RDF) Model and Syntax Specification. <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>.
- [7] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg and et al., The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* **3**(1) (2016), 160018, Bandiera_abtest: a Cg_type: Nature Research Journals Number: 1 Primary_atype: Comments & Opinion Publisher: Nature Publishing Group Subject_term: Publication characteristics;Research data Subject_term_id: publication-characteristics;research-data. doi:10.1038/sdata.2016.18. <https://www.nature.com/articles/sdata201618>.
- [8] A. Navarro-Gallinad, A. Meehan and D. O'Sullivan, The Semantic Combining for Exploration of Environmental and Disease Data Dashboard for Clinician Researchers **2778** (2020), 13. <http://ceur-ws.org/Vol-2778/paper7.pdf>.
- [9] A. Navarro-Gallinad, F. Orlandi and D. O'Sullivan, Enhancing Rare Disease Research with Semantic Integration of Environmental and Health Data, in: *The 10th International Joint Conference on Knowledge Graphs, IJCKG'21*, Association for Computing Machinery, New York, NY, USA, 2021, pp. 19–27. ISBN 978-1-4503-9565-6. doi:10.1145/3502223.3502226.
- [10] R. Macefield, How To Specify the Participant Group Size for Usability Studies: A Practitioner's Guide **5**(1) (2009), 12.
- [11] T. Boren and J. Ramey, Thinking aloud: reconciling theory and practice, *IEEE Transactions on Professional Communication* **43**(3) (2000), 261–278. doi:10.1109/47.867942.
- [12] J. Lewis, Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies, *Int. J. Hum. Comput. Interaction* **14** (2002), 463–488. doi:10.1080/10447318.2002.9669130.
- [13] L.S. Nowell, J.M. Norris, D.E. White and N.J. Moules, Thematic Analysis: Striving to Meet the Trustworthiness Criteria, *International Journal of Qualitative Methods* **16**(1) (2017), 160940691773384. doi:10.1177/1609406917733847.

- [14] A.R. Kitching, H.-J. Anders, N. Basu, E. Brouwer, J. Gordon, D.R. Jayne, J. Kullman, P.A. Lyons, P.A. Merkel, C.O.S. Savage, U. Specks and R. Kain, ANCA-associated vasculitis, *Nature Reviews Disease Primers* **6**(1) (2020), 1–27, Number: 1 Publisher: Nature Publishing Group. doi:10.1038/s41572-020-0204-y. <https://www.nature.com/articles/s41572-020-0204-y>.
- [15] J. Scott, J. Hartnett, D. Mockler and M.A. Little, Environmental risk factors associated with ANCA associated vasculitis: A systematic mapping review, *Autoimmunity Reviews* **19**(11) (2020), 102660. doi:10.1016/j.autrev.2020.102660. <https://www.sciencedirect.com/science/article/pii/S1568997220302354>.
- [16] X. Rodó, R. Curcoll, M. Robinson, J. Ballester, J.C. Burns, D.R. Cayano, W.I. Lipkin, B.L. Williams, M. Couto-Rodriguez, Y. Nakamura, R. Uehara, H. Tanimoto and J.-A. Morguá, Tropospheric winds from northeastern China carry the etiologic agent of Kawasaki disease from its source to Japan, *Proceedings of the National Academy of Sciences* **111**(22) (2014), 7952–7957. doi:10.1073/pnas.1400380111. <https://www.pnas.org/content/111/22/7952>.
- [17] X. Rodó, J. Ballester, R. Curcoll, J. Boyard-Micheau, S. Borràs and J.-A. Morguá, Revisiting the role of environmental and climate factors on the epidemiology of Kawasaki disease, *Annals of the New York Academy of Sciences* **1382**(1) (2016), 84–98. doi:10.1111/nyas.13201.
- [18] R. Cyganiak and D. Reynolds, The RDF Data Cube Vocabulary, 2014. <https://www.w3.org/TR/vocab-data-cube/>.
- [19] R. Catherine, B. Stephan, A. Géraldine and B. Daniel, Weather Data Publication on the LOD using SOSA/SSN Ontology, 10.
- [20] R. Albertoni, D. Browning, S.J.D. Cox, A. Gonzalez-Beltran, A. Perego and P. Winstanley, Data Catalog Vocabulary (DCAT) - Version 2, 2020. <https://www.w3.org/TR/vocab-dcat-2/>.
- [21] T. Lebo, S.S. Sahoo and D.L. McGuinness, PROV-O: The PROV Ontology, 2013. <https://www.w3.org/TR/prov-o/>.
- [22] M. Perry, J. Herring, N.J. Car, T. Homburg, J. Abhayaratna, S.J.D. Cox, M. Bonduel and F. Knibbe, GeoSPARQL - A Geographic Query Language for RDF Data | OGC, 2012. <https://www.ogc.org/standards/geosparql>.
- [23] A. Navarro-Gallinad, Weather and Air Quality data for Ireland as RDF data cube, Zenodo, 2021, Type: dataset. doi:10.5281/zenodo.5668287. <https://zenodo.org/record/5668287>.
- [24] E. Prud'hommeaux and A. Seaborne, SPARQL Query Language for RDF. W3C recommendation. <https://www.w3.org/TR/rdf-sparql-query/>.
- [25] R. Rampin and V. Rampin, Taguette: open-source qualitative data analysis, *Journal of Open Source Software* **6**(68) (2021), 3522. doi:10.21105/joss.03522.
- [26] R. Li, C. Yin, S. Yang, B. Qian and P. Zhang, Marrying Medical Domain Knowledge With Deep Learning on Electronic Health Records: A Deep Visual Analytics Approach, *Journal of Medical Internet Research* **22**(9) (2020), e20645, Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada. doi:10.2196/20645. <https://www.jmir.org/2020/9/e20645>.
- [27] M. Sheng, Y. Shao, Y. Zhang, C. Li, C. Xing, H. Zhang, J. Wang and F. Gao, DEKGB: An Extensible Framework for Health Knowledge Graph, in: *Smart Health*, H. Chen, D. Zeng, X. Yan and C. Xing, eds, Lecture Notes in Computer Science, Springer International Publishing, Cham, 2019, pp. 27–38. ISBN 978-3-030-34482-5. doi:10.1007/978-3-030-34482-5_3.
- [28] B. Rose-Davis, W. Van Woensel, S. Raza Abidi, E. Stringer and S. Sibte Raza Abidi, Semantic knowledge modeling and evaluation of argument Theory to develop dialogue based patient education systems for chronic disease Self-Management, *International Journal of Medical Informatics* **160** (2022), 104693. doi:10.1016/j.ijmedinf.2022.104693. <https://www.sciencedirect.com/science/article/pii/S1386505622000077>.
- [29] J. Brooke, SUS: A quick and dirty usability scale, *Usability Eval. Ind.* **189** (1995).
- [30] E. Daffi, P. Antoniou, L. Ioannidis, N. Dombros, D. Topps and P.D. Bamidis, Virtual Patients on the Semantic Web: A Proof-of-Application Study, *Journal of Medical Internet Research* **17**(1) (2015), e16. doi:10.2196/jmir.3933. <http://www.jmir.org/2015/1/e16/>.
- [31] N. Ammar, J.E. Bailey, R.L. Davis and A. Shaban-Nejad, Using a Personal Health Library-Enabled mHealth Recommender System for Self-Management of Diabetes Among Underserved Populations: Use Case for Knowledge Graphs and Linked Data, *JMIR Formative Research* **5**(3) (2021), e24738. doi:10.2196/24738. <http://formative.jmir.org/2021/3/e24738/>.
- [32] M.R. Stöhr, A. Günther and R.W. Majeed, The Collaborative Metadata Repository (CoMetaR) Web App: Quantitative and Qualitative Usability Evaluation, *JMIR Medical Informatics* **9**(11) (2021), e30308, Company: JMIR Medical Informatics Distributor: JMIR Medical Informatics Institution: JMIR Medical Informatics Label: JMIR Medical Informatics Publisher: JMIR Publications Inc., Toronto, Canada. doi:10.2196/30308. <https://medinform.jmir.org/2021/11/e30308>.
- [33] R. Marcilly, L. Douze, S. Ferré, B. Audeh, C. Bobed, A. Lillo-Le Louët, J.-B. Lamy and C. Bousquet, How to interact with medical terminologies? Formative usability evaluations comparing three approaches for supporting the use of MedDRA by pharmacovigilance specialists, *BMC Medical Informatics and Decision Making* **20**(1) (2020), 261. doi:10.1186/s12911-020-01280-1.
- [34] X. He, R. Zhang, R. Rizvi, J. Vasilakes, X. Yang, Y. Guo, Z. He, M. Prosperi, J. Huo, J. Alpert and J. Bian, ALOHA: developing an interactive graph-based visualization for dietary supplement knowledge graph through user-centered design, *BMC Medical Informatics and Decision Making* **19**(4) (2019), 150. doi:10.1186/s12911-019-0857-1.
- [35] X. He, H. Zhang and J. Bian, User-centered design of a web-based crowdsourcing-integrated semantic text annotation tool for building a mental health knowledge base, *Journal of Biomedical Informatics* **110** (2020), 103571. doi:10.1016/j.jbi.2020.103571. <https://www.sciencedirect.com/science/article/pii/S1532046420302008>.
- [36] R. Hanlon, M. Barry, F. Marrinan and D. O'Sullivan, Towards an effective user interface for data exploration, data quality assessment and data integration, in: *2021 IEEE 15th International Conference on Semantic Computing (ICSC)*, 2021, pp. 431–436, ISSN: 2325-6516. doi:10.1109/ICSC50631.2021.00077.
- [37] A. Hogan, The Semantic Web: Two decades on, *Semantic Web* **11**(1) (2020), 169–185, Publisher: IOS Press. doi:10.3233/SW-190387. <https://content.iospress.com/articles/semantic-web/sw190387>.

Appendix A. Code descriptions for the Thematic Analysis (P2)

Table 3

Description of the codes and references from the participants in the Thematic Analysis of the usability sessions transcripts (P2)

Code	Description	References
Additional explanation	The participant requests additional explanation for a feature, visualisation, text description or tool tip.	59
Additional feature suggestion	Distinguish outliers based on historical data. Add time series, scatter and histogram plots. Add an advanced aggregation method and select all features. Define, select and group events feature for exploration. Summarise the metadata in a user friendly way. Facilitate the understanding of the data table and increase the amount of data.	126
Data table features useful	Hiding, restoring, ordering and colouring the columns and values in the data table improves the usability of the table.	72
Event concept and approach not clear	The purpose to gather data for particular events complicates the understanding of the event concept and terminology.	33
Export (meta)data useful	Exporting the data, once it is understood, is useful for a subsequent analysis and the metadata for the provenance and reusability.	131
Home tab elements useful	The text, diagram and links in the home tab facilitate the overall understanding of processes underpinned by the UI.	50
Metadata content clear and useful	The content from the data provenance, data lineage and/or full metadata windows is clear and/or useful.	88
Metadata content not useful or confusing	The information provided in the data provenance table, data lineage and/or full metadata exploration windows is not useful or confusing.	71
Moderator assist	The moderator intervenes because the participant struggles to complete a task, wanders off task or goes too deep into a task or the system crashes.	174
Navigation complex	The design of the export and metadata buttons together with the multiple tab approach, the functionality of the pop ups and some of the data table features complicate the navigation.	151
Overall positive experience	Generic comment on the positive experience when using the tool.	48
Plot design complex	The content, axis ticks and labels of the plots are not clear.	60
Query inputs and elements useful	The query input options are clear in general and the tool tips, text and drop downs (multi)inputs help the user in understanding the query process.	191
Query process clear	The query process and execution is clear.	62
Query process complex	The query process and the sequence and meaning of the query inputs are not understood.	50
Search feature simple	Control+F function to look for specific elements in the metadata is simple and easy for the user.	61
Search information process complex	Finding relevant information in the metadata windows can be complex.	32
Simple export	Data and metadata generated are simple and easy to export.	70
System issue	The user identifies a system issue as taking long to load, crash or not responsive.	45
Task sequence complex	The task was confusing, not completed in the required order or had to be read again.	156
Visualisation useful	Visualising the data as a heat map, box plot and/or polar plot is useful, simple, interactivity adds value and the plots can be compared.	125
Wording not clear	A word or expression that appears in a paragraph is not clear or used appropriately.	75

Appendix B. PSSUQ open comments Thematic Analysis

Table 4

Theme references of the PSSUQ open comments colour coded as in Fig. 4E heat map.

Themes	References
Requirement 1: Querying	13
Requirement 2: Understanding	28
Requirement 3: Querying	1
Emerging Requirements	29
Usability Testing	26

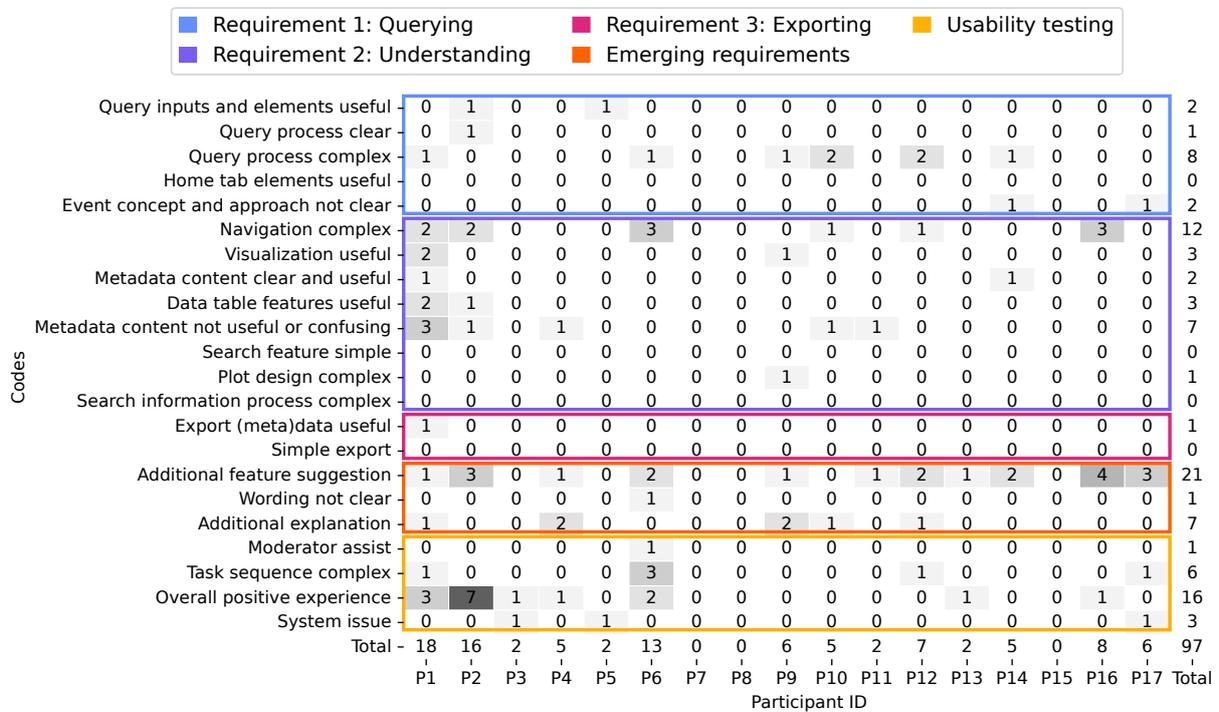


Fig. 5. Categorization of the PSSUQ open comments using the same codes as in the usability sessions transcripts displaying coherence between both sources.