Usability evaluation of Kadi4Mat*

Marco Joost[1,2][7364514] and Imke Schwenke[1,2][7351430]

[1] University of Cologne, Cologne, Germany
https://www.uni-koeln.de/
[2] Deutsches Zentrum für Luft- und Raumfahrt, Linder Höhe, 51147 Köln, Germany
https://www.dlr.de/de

Keywords: Kadi4Mat · Usability Evaluation · Electronic laboratory notebook.

1 Introduction

Electronic laboratory notebooks play a vital role in modern scientific research, offering researchers an efficient and organized platform for documenting experiments and managing data. As the scientific community transitions from traditional analog laboratory notebooks to electronic alternatives, the usability of these digital tools becomes a critical factor in their adoption and effectiveness (Higgins, 2022).
This research presents a formative evaluation of the usability of the Kadi4Mat software, a digital laboratory notebook designed to support researchers in managing research data and collaboratively working on experiments (Brandt & Selzer, 2021). Guided by the principles of interaction, we chose Suitability, Learnability, Customizability, and Error Tolerance. By testing these principles we aim to assess the software's strengths and areas for improvement (ISO 9241-110, 2006).
The significance of this research lies in its contribution to the broader understanding of electronic laboratory notebooks usability, in the context of Kadi4Mat. By identifying specific challenges and proposing practical solutions, our findings provide a way forward for both academia and practice. The insights gained from this study can guide the development of user-friendly features and act as a baseline for future research.
We proceed as follows. First, we explain concepts in the theoretical background necessary to understand this research. Second, we present the research methodology, detailing the goals, metrics, and usability test plan design. Next, the results are presented. Then, the discussion outlines the specific findings of our usability evaluation, highlighting challenges and proposing recommendations for enhancing the usability of the Kadi4Mat software as well as presenting the limitations of our work. Finally, the conclusion completes this research.

* Supported by organization University of Cologne and Deutsches Zentrum für Luft- und Raumfahrt.

2 Background and Related Work

In this section, we break down the main concepts our research is based on. First off, we give a clear definition of Usability. Then, we explain the principles of interaction. Next, we present the System Usability Scale used to calculate a general usability score. Finally, we explain the Kadi4Mat software on which the usability test procedure is carried out.

2.1 Usability Definition

There have been many attempts to define usability. Often the result has remained rather vague. The ISO 9241 Part 11 standard, however, has succeeded in making the term more concrete. It defines usability according to three criteria: effectiveness, efficiency and satisfaction.
Effectiveness refers to whether the user is able to achieve a specified goal completely and accurately. Efficiency, on the other hand, is concerned with the resources used to achieve the user's goal. Satisfaction refers to the user's comfort and acceptance after using the software (ISO 9241-11, 1997; Theis, 2023).

2.2 Interaction Principles

The interaction principles were set down the first time in the DIN EN ISO 9241-110 standard in 2006. They are intended to provide guidance on how a user interface should be designed. They are made up of the following principles:

- Suitability for the user's task
- Self-descriptiveness
- Conformity of the user expectation
- Controllability
- Error tolerance
- Learnability

Suitability refers to whether the design allows the user to perform their task effectively and efficiently. A system is considered self-descriptive if it can be understood through feedback from the dialogue system or if it provides an explanation to the user upon requesting relevant information. To reach conformity of the user expectation, the system should be consistent and implement familiar concepts and conventions. Controllability is achieved if the user controls the entire dialogue until they completed the respective task. The principle error tolerance is implemented when user errors are reduced to a minimum with the design and the design allows the user to easily correct errors. A system is customizable if the interface can be adapted to the user's preferences and tasks. Lastly, learnability is reached when users can understand the system quickly and will use it more efficiently over time (ISO 9241-110, 2006; Theis, 2023).

2.3 System Usability Scale

The System Usability Scale (SUS) is a widely used questionnaire for assessing the perceived usability of a system or product, developed by John Brooke in 1986. The SUS consists of a 10-item questionnaire, each item representing a statement about the usability of the system. Respondents rate their agreement with each statement on a Likert scale from 1 (strongly disagree) to 5 (strongly agree). The scores are then translated to a single usability score on a scale of 0 to 100, with a higher score indicating better usability. General guidelines suggest a score of 68 or higher to be above average. The SUS provides a simple, quick and reliable measure of subjective usability. We use this scale to capture the test users perception of the Kadi4Mat software and substantiate the findings discovered in the usability test (Brooke, 1995).

2.4 Kadi4Mat

The Kadi4Mat (Karlsruhe Data Infrastructure for Materials Science) is an open-source electronic laboratory notebook designed for managing research data. Kadi4Mat aims to create a virtual research environment for seamless collaboration primarily in material science, by electronically recording the entire scientific workflow. In the long term Kadi4Mat aims to support the unique workflows of scientists of various research disciplines and to be widely adapted as a electronic laboratory notebook. This is enabled by its modular architecture, that accommodates the unique workflows of scientists. However, the successful adaptation of Kadi4Mat highly depends on the usability it offers underlining the need for a usability evaluation (Brandt & Selzer, 2021).

3 Method

This section outlines our methodology for assessing the usability of Kadi4Mat. We establish goals derived from the interaction principles and define corresponding measurements to gauge the achievement of these objectives. We then present the design of the usability test plan, detailing the structure and procedures employed. Furthermore, the text provides relevant information about the participants, giving us insights into the demographics and backgrounds that influenced our usability evaluation.

3.1 Goals and Metrics

The goal of this research is to formatively evaluate the usability of the Kadi4Mat software. The usability is assessed by adhering to the principles of interaction defined by the DIN EN ISO 9241-110 (ISO 9241-110, 2006). For a meaningful usability evaluation we are choosing a subset of the principles that are most suitable for the evaluation. This is done by analyzing the user base of the Kadi4Mat software, environmental factors, and the general goals set for this software project.

First, Kadi4Mat needs to mimic the basic features of an analog laboratory notebook, such as independence of technology and easily changeable entries that can be adapted to the experiment. However, as an ELN, it also has to enable better organization and data search as well as easy collaboration with other researchers. Therefore, the suitability for the task is vital and will be assessed by analyzing the efficiency and effectiveness of Kadi4Mat. Effectiveness is assessed by the success rate. This can be calculated by dividing the number of successfully completed tasks by the number of all tasks. Our goal is to have an overall success rate of 80%. To determine how efficient the software is, the time for performing a task was taken. The time for all tasks should be less than 30 minutes. This is a reasonable amount of time for taking notes of an experiment, especially with a new software.
Second, the users of Kadi4Mat software consist of experienced researchers seeking to transition from traditional analog laboratory notebooks to ELNs. These users are tech-savvy but lack experience with electronic laboratory notebooks. Their openness to new technologies suggests a willingness to adapt, provided the software is intuitive and user-friendly. This suggests that the principle learnability is important to evaluate the usability of Kadi4Mat. Learnability is assessed by comparing the time per task of the participants with the time an expert needed. The expert is a person who is already familiar with the software. It is examined whether the time of the participants approximates the time of the expert.
Third, accuracy and reliability in documenting experiments are crucial for scientific work, highlighting a need for a system that not only facilitates efficient data collection but also ensures precise and error-free data. Thus, error tolerance needs to be analyzed in this evaluation. We developed specific tasks to see how the software reacts to errors users perform (see Task 6). Of this task the success rate and the time is analyzed.
Last, one goal of Kadi4Mat is to be an electronic laboratory notebook for various research disciplines and given the high level of education and specialized knowledge of these researchers, a flexible digital laboratory notebook capable of catering to various scientific disciplines and needs is essential. This makes customizability an important principle to assess the usability of Kadi4Mat. For this principle, two tasks were created (see Tasks 8 and 9). The time per task and the success rate is again evaluated.
In addition, the thinking aloud method was used to gain insight into the user's thinking. In order to collect all the thoughts, the user was recorded as he or she performed the usability test tasks. This method can help to find elements in the design that lead to misunderstandings and errors. Note that the Simplified Thinking Aloud method was used. The traditional method is carried out by psychologists or user interface experts, but for our purpose computer scientists as experimenters are able to use this method to evaluate interfaces.(Nielsen, 1994)

3.2 Usability Test Plan

According to Hamborg (1996), the usability of a system can only be evaluated if three units are fixed: the software under investigation, the to be performed

tasks and the users who are supposed to asses the software. These units are set out in the usability test plan. Each part of the plan, which can be found in the appendix, is described and explained below.
The first part of an usability test plan is the overall goal of the evaluation. There are two types of evaluation in usability testing: summative and formative. Summative evaluation compares the software to a certain standard. This happens after the system has been designed and is usually used to determine how competitive the software is. Formative evaluation focuses on finding weaknesses of the software during system design and thus finding possible improvements (Gediga, 1999; Theis, 2023). In our case, the software is already in production and we were trying to see if it can compete with analog notebooks. However, we did this by analysing weaknesses and then making suggestions for improvement. So we had a certain summative character, but we focused on formative evaluation.
The experiment was conducted multiple times, with each iteration overseen by an experimenter, and involving the participation of a single participant. They took place in a quiet room with only the experimenter, the participant and the necessary resources present. We estimated the duration of each experiment to be around 30 minutes. The resources needed were a computer with web access to perform the usability tasks, a copy of the task, a recording device, a time tracking device and a notepad and pencil. The experimenter presented and explained the tasks. Then showed the starting page of Kadi4Mat. After that, he or she informed the participant that he or she will be recorded and asked him or her to speak out his or her thought during the experiment. Moreover, the participant should speak out when a task or sub task is finished. In this way every thought of the participant could be recorded and used for further analysis. From that moment on, the only help the experimenter should give is to remind the participants of the Kadi4Mat help page. At the start of the experiment, the experimenter switched on the recording and time tracking device. During the experiment, the experimenter took notes on a notepad, which was structured as shown in Table 1. He or she was supposed to record whether a task was completed and how long it took to complete a task. These measures were needed to calculate our metrics.

Table 1: Experimenter's notes

Tasks	Task finished	Finishing Time
Task 1	Yes	1:39

The chosen set of tasks was designed to cover a typical workflow in the software, thus testing how quickly a user could learn the software and carry out their work. He or she should register, create a group, a project and two experiments, one from scratch and one using a template. To understand how the software

behaves when users make mistakes, the participant was then asked to modify one experiment. After that, he or she was supposed to find entries and then perform more general tasks such as changing the language, customizing the software and then exploring it freely. We decided that the participant should not do the tutorial that Kadi4Mat offers when you register. In this way we wanted to understand how intuitive the software is. The only help the participant should use was the Kadi4Mat help page.
After the experiment was carried out, the participant then answered the questions of the usability scale. The survey can be found in the appendix.

3.3 Participants

The goal in selecting test subjects is to align the participant group with the potential user base as much as possible. The user base of the Kadi4Mat software mostly includes researchers from various disciplines. The experiment involved five people, three males and two females. The participants' ages range from 25 to 30, and each participant has at least a bachelor's degree along with some experience in scientific research. Additionally, all members show a willingness to familiarize themselves with new software. None of the participants has any prior experience using an electronic laboratory notebook. Additionally, an expert familiar with the software participated in the test. This allows us to gain insights into the performance of someone who is already used to the Kadi4Mat software.

4 Results

This section is designed to present the findings gathered by conducting the methods specified in the research methodology. First, the results of the usability test are presented. Therefore, the time per task and success rate are reported and analysed. Then, the perceived usability by the participants is presented by analysing the SUS questionnaire and determining a usability score for the Kadi4Mat software. Finally, this section concludes with an overview of the analysis of the notes collected using the thinking aloud method, which gives us a more subjective assessment of the software's usability.

4.1 Usability Test

Throughout the usability test, we tracked the success of each task seen in Figure 1 and the time required by each participant to complete them displayed in Figure 2. Then, we computed the mean success rate for each task and the overall mean success rate across all tasks. Additionally, we calculated the mean time needed per task and the total average time taken by all participants to complete all tasks.
In the usability testing, each participant completed 17 tasks, making a total of 85 tasks. Of these, 67 tasks were successfully completed, resulting in an overall

mean success rate of 78.8%. Notably, nine tasks achieved a 100% success rate, meaning that more than half of the tasks could be solved by all participants. Two tasks were completed by four participants, four tasks were completed by three participants, one task could only be completed by two participants and the last task in the usability test could not be completed by any participant.
On average, participants spent 43 minutes and 23 seconds completing the entire test plan. In contrast, the expert only required 11 minutes and 32 seconds. Task 4a emerged as the lengthiest, with an average of 5 minutes and 40 seconds needed by each participant. Task 7b had the shortest mean time per task at 26.4 seconds, making it the task completed in the least amount of time by the participants.

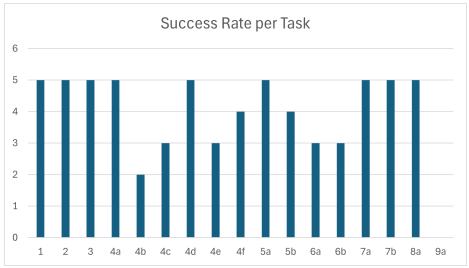

Fig. 1: Successrate per Task

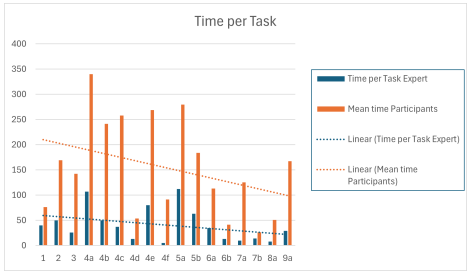

Fig. 2: Time per Task with Linear Regression

4.2 System Usability Scale

Each participant conducted the SUS questionnaire. Consequently, we had a SUS score of each participant and could therefore compute a mean SUS score seen in Table 2 The mean SUS score for the test group was calculated to be 40,5. It is important to note that, despite being displayed on a 0-100 scale, these scores do not represent percentages and should be interpreted in relation to their percentile ranking (Brooke, 1995).

Table 2: SUS Score

Participant	SUS Score
P1	37,5
P2	47,5
P3	40
P4	42,5
P5	35
Mean SUS Score	40,5

4.3 Thinking aloud notes and Feedback

Besides the quantitative results which were gathered during the testing and the SUS questionnaire, we wrote down the thought process and active feedback provided by the participants during the testing. After analyzing the notes of the participants we found six statements that have been expressed by three or more participants during the testing as seen in Table 3. All the participants had difficulties to understand the structure of Kadi4Mat and at some point of the testing were confused at what part of he software they were currently at. Four participants were not satisfied with the appereance of the software. Three out of five participants complained about the color range used in Kadi4Mat, had problems to understand the meta data, and complained about the arrangements of parts of the software. In addition to the quantitative data obtained from the testing and the SUS questionnaire, we documented the participants' thought processes and received active feedback throughout the testing. After analyzing these participant notes, we identified six statements that have been expressed by three or more participants, as shown in Table 3. Observed among all participants was a difficulty in understanding the structure of Kadi4Mat and as well as a confusion about their current location within the software. Furthermore, dissatisfaction with the software's appearance was expressed by four participants. Three out of five participants complained about the color range used in Kadi4Mat, encountered difficulties understanding metadata, and were opposed regarding the arrangement of software components.

5 Discussion

In the following, the meaning and possible interpretations of the presented results are discussed. For this purpose, we look at each of the defined goals and analyze whether they have been achieved. If not, recommendations are made to improve the Kadi4Mat in relation to the goal analyzed.

Table 3: Thinking Aloud Notes and Feedback

Statements	Number of Participants
Monotonous colors	3
Problems understanding meta data	3
Problems to understand the structure of application	5
Didn't understand at where in the software participant is	5
Complained about arrangement of parts in software	3
Not satisfied with appearance of software	4

5.1 Suitability of the task

In our usability test of the Kadi4Mat digital laboratory notebook, we aimed to evaluate its suitability for the task by assessing both efficiency and effectiveness. Effectiveness, in particular, was measured through the success rate. Our goal was to achieve an 80% success rate, but the observed overall one was 78.8%. This result indicates a reasonable level of effectiveness in terms of task completion. However, several factors may contribute to the success rate falling short of the 80% goal. A closer look shows that tasks 9a and 4b had the lowest success rates of 0% and 40% respectively. In both tasks the participants were asked to edit the description of a record. The description text field in Kadi4Mat is based on Markdown[3]. This is a lightweight markup language for formatting text. However, 3 out of 5 participants were not familiar with the language. There is no information about Markdown on the help page of Kadi4Mat. We recommend providing a cheat sheet with the basic syntax of Markdown.
Tasks 4c and 6 had also a low success rate of 60%. This is due to the fact that the users had problems displaying the content of a markdown file. It is only possible to copy the content into the description text file or to insert a link to the file if it has been uploaded to Kadi4Mat. However, we suggest adding the possibility to import the content once the file has been uploaded. These changes could increase the individual success rates and therefore, the overall one.
In our evaluation of the Kadi4Mat digital laboratory notebook's efficiency, we measured participants' task completion times, revealing an average of 43 minutes and 23 seconds across all tasks — surpassing our 30-minute goal, which reflects the assumed typical time for notebook usage. As note-taking in a laboratory notebook is a secondary task for an experimenter, an optimal system should not consume too much time. Task 4a, as highlighted in Figure 2, emerged as the most time-consuming. Our analysis indicates that participants encountered difficulties with adding metadata to the experiment, particularly in selecting the appropriate type from a drop down box. Notably, the Kadi4Mat help page contains information about these types, but participants often neglected this

[3] `https://markdown.de/`

resource. To address this, we recommend incorporating a button that opens a pop-up window offering information about each type and thus, streamlining the understanding of metadata selection. In addition, the 'Key' and 'Value' fields were not clear to the participants. Some only used one of these fields to enter metadata. Proposing a term change from 'Key' to 'Attribute' could potentially enhance clarity and distinguish these fields more effectively.

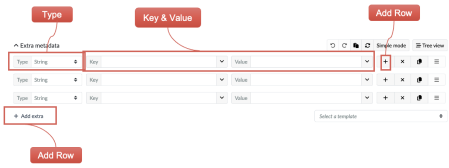

Fig. 3: Metadata Field

Moreover, participants invested more time than required on the starting page, according to our observations. This extended duration can be attributed to the presence of a header that shares a similar size and color with the navigation bar. Consequently, the attention shift from the navigation bar resulted in a lengthened search. Since the header lacks any functional significance, our recommendation is to remove it, streamlining the user experience.

5.2 Learnability

To assess the learnability of Kadi4Mat, we employed an analysis of the time per task, comparing the performance of participants against an expert. As depicted in Figure 2, linear regressions were applied to both sets of data, showcasing the time per task trends. The resulting lines on the graph reveal an interesting dynamic. The expert's line indicates a decreasing time per task, suggesting that the tasks become, in general, shorter. Conversely, the participants' line also experiences a slightly higher decline, indicating a small learning curve among users. While the trend suggests that participants become more efficient with the course of tasks, the degree of improvement is not substantial.
A contributing factor to the observed low learnability is the challenge users faced in navigating the system, particularly when performing tasks with similar structures. Notably, the pages for collections, records, and groups share identical structures and color schemes. While the intent may have been to leverage familiarity for enhanced learnability, our findings suggest that this approach led to user confusion, with individuals occasionally forgetting the specific page they were on. To address this issue, we recommend implementing a distinct color scheme for each page, preferably selected from a sequential color palette. This design choice would enable users to associate each entity (collections, records, and groups) with a unique color, facilitating quicker recognition and reducing the likelihood of confusion. Moreover, introducing a hierarchical display of these

entities through color association could enhance the understanding of their relationships, further aiding users in distinguishing between collections, records, and groups. Ultimately, these adjustments contribute to a more learnable software environment.

5.3 Customizability

In evaluating the customizability of the Kadi4Mat software, two tasks were employed – switching the language and adapting the experiment description (task 8 and 9). The results highlight notable challenges in these aspects.
Switching the language proved to be a time-consuming task, with an average duration of 51 seconds. The primary challenge encountered by users was the placement of the language switch button in the bottom container, positioned at the end of the page rather than within the user's immediate view. This design choice contributed to users struggling to locate the language switch option promptly. To enhance customizability, i.e. adapting the software to the user's needs, we suggest to place the bottom container at the end of the view, so no scrolling is needed to find the language switch.
Adapting the experiment description presented a more significant challenge, as evidenced by a 0% success rate. The difficulty arose from the fact that the description field uses Markdown but does not support coloring. This limitation hindered users' ability to customize the appearance of the text. To address this issue, incorporating support for color customization within the Markdown-based description field could significantly improve the customizability of experiment descriptions. Enabling users to personalize the visual aspects of text enhances the flexibility and adaptability of the software, aligning with the principle of customizability.

5.4 Error tolerance

The evaluation of error tolerance in the Kadi4Mat software unveiled insights into challenges users faced during specific tasks, impacting the success rates.
Task 6 involved changing values in the description field and had a success rate of 60%. As mentioned before, implementing a feature that allows users to import the content directly could provide a more flexible and error-tolerant approach, allowing users to modify values seamlessly.
Furthermore, user errors were identified in the metadata field. Users tended to enter all information into the Key field, neglecting the Value field. The software lacks explicit guidance to use both fields, resulting in an error-prone situation. Implementing clear instructions or visual hints to encourage the use of both Key and Value fields can enhance user understanding and errors in data entry.
Similarly, challenges emerged when creating dashboards, where users misunderstood the concept of panels. The button for adding a panel is presented as a dropdown menu, leading to users consolidating all images of graphs into a single panel instead of creating individual panels. To address this, simplifying the interface by providing a straightforward "Add" button and allowing users to choose

between graph and key performance indicator (KPI) panels could mitigate confusion and errors during dashboard creation.

5.5 Limitations

Although our recommendations show promise in enhancing Kadi4Mat's suitability for the task, learnability, customizability, and error tolerance, it is important to acknowledge the limitations of our study. The scope of our project was constrained, which limited the number of experiments we could conduct and resulted in a smaller sample size that may affect the generalizability of our findings. In addition, it should be noted that the participants in this study have research backgrounds and at least a bachelor's degree, but not the same experience conducting experiments as the user base of Kadi4Mat has. This may also limit the generalizability of our findings. The principles chosen for evaluation were based on our assumptions about their importance, which may have excluded other critical aspects. Recognizing these limitations, future studies with a larger participant pool and a more comprehensive exploration of additional usability principles may provide a more nuanced understanding of Kadi4Mat's user experience.

6 Conclusion

In conclusion, our evaluation of the Kadi4Mat software's usability, guided by the principles of interaction—Suitability, Learnability, Customizability, and Error Tolerance- has provided valuable insights into its strengths and areas for improvement. The findings reveal specific challenges users face and offer recommendations for enhancing the overall user experience.
Reflecting on the key contributions of this research, we've identified areas where Kadi4Mat can be changed or adapted to better align with user expectations and needs. The examination of suitability highlighted task-specific as well as Markdown language issues, while the assessment of learnability revealed opportunities for optimizing system navigation. Evaluating customizability uncovered challenges in language switching and experiment description adaptation, and insights into error tolerance provided suggestions for enhancing various user-friendly features.
In terms of significance, this research contributes to the broader understanding of usability in the Kadi4Mat software. By highlighting specific challenges and proposing practical solutions, it provides a way forward for both academia and practice. The insights gained from this study can guide the development of user-friendly features. Additionally, future research can complement this work by conducting a usability test with more subjects or conduct an heuristic usability evaluation of the Kadi4Mat software.

In essence, this research not only identifies opportunities for improvement in the Kadi4Mat software but also underlines the importance of user-centric design in the development of digital tools for scientific tasks. If the Kadi4Mat software

can offer a high usability to the potential user base, it can be adapted as an electrical laboratory notebook. This is crucial to effectively support researchers in their work and ultimately advancing scientific research.

A Appendix

Table A.1: Usability Test Plan

Usability Test Plan	
Attribute	Value
Goal	Identify strengths and weaknesses of Kadi4Mat, an electronic notebook
Place	Room with one participant and one experimenter
Duration	approximately 30 minutes
Resources	– computer with web access for performing usability tasks – recording device – time tracking device – notepad and pencil for experimenter
Software State	Demo Version of Kadi4Mat, web-based application
Users	Insert list of users

Usability Test Plan	
Attribute	Value
Usability Test Tasks (Part 1)	1. **Login:** Register for the demo version of *Kadi4Mat* and log in. If necessary, change the language settings during that step. 2. **Create a group:** Create a group 'Study Group' and add the users 'ischwenk' and 'mjoost'. All participants should have at least editor rights in the group. 3. **Create a project:** You want to create a series of related experiments. Create a new collection, 'Study Project', for these experiments. Ensure that the collection is private and that only the members of the previously created group can access the collection. 4. **Create experiment 1:** (a) The first experiment has been carried out. Create a new entry for the experiment. Create a new data set (record) based on the following metadata: – Title: Beam quality of the diamond laser and mode composition – Type: experiment – Visibility: private – Tags: beam and laser – Extra metadata: Device = Diamond Laser, Power level = 45%, Location = Zugspitze – Collection: Study Project – Access rights: Type: Group - Study Group - Editor role (b) Add the experimental setup (versuchsaufbau1.png) to the experiment. The picture should be visible in the description of the experiment described with a heading 'Setup'. (c) Add the notes and measurements to the experiment (Experiment1.md). These should be also visible in the description of the experiment. (d) Add the files plot1.png, plot2.png and plot3.png to the experiment. (e) Create a new dashboard with the three plots in the dashboard tab of the experiment. (f) Add the experiment to the collection you have created. 5. **Create a template:** (a) Create a template for further measurements (Records) with the laser from the experiment already carried out. The template should have the same metadata as experiment 1. The parameter "power level" should not yet be defined, and there should be no measurement data in the template. (b) Use the template to create experiment 2 (power level 70%) and 3 (power level 100%). You can copy the content from Experiment2+3.md. Both experiments should be part of the project.

Usability Test Plan	
Attribute	Value
Usability Test Tasks (Part 2)	6. **Changes:** (a) You made a mistake when creating the data for experiment 1. Please change the values in column 2 of the measurement to the following values: 1200, 1045, 800, 540, 170, 150 (b) Take a look at your changes under Revisions. 7. **Find entries:** (a) Go to the overview page of all recoirds and display all entries with the tag 'beam'. (b) Limit the results to the entries you have created yourself. 8. **Language Settings:** (a) Switch the language. 9. **Customize:** (a) Your team reminded you that the power level of your first experience is of importance. Add it to the beginning of your experiment description in bold and red. 10. **Free exploration:** (a) Experiment with the functions that Kadi4Mat makes available to you.
End of Table	

B Appendix

Table B.1: Usability Survey

Question	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. I think that I would like to use this system frequently.					
2. I found the system unnecessarily complex.					
3. I thought the system was easy to use.					
4. I think that I would need the support of a technical person to be able to use this system.					
5. I found the various functions in this system were well integrated.					
6. I thought there was too much inconsistency in this system.					
7. I would imagine that most people would learn to use this system very quickly.					
8. I found the system very cumbersome to use.					
9. I felt very confident using the system.					
10. I needed to learn a lot of things before I could get going with this system.					

C Appendix

Table C.1: Experiment with Participant Florian Wolfrum.

Task	Task Finished	Time To Finish
1.	Yes	1:27
	does not like colors: only grey and white	
2.	Yes	3:00
	– navigation bar appeared to be helpful – did not recognize red warnings	
3.	Yes	2:48
	– tried first the Collections field of the Groups navigation bar, not of the overall one – needed to click on pencil button to change identifier – Workaround: instead of adding the group, participant added users individually	
4.a)	Yes	3:17
	– does not create new type 'experiment' but chooses already existing type 'experimental' – did not understand how the meta data is filled in – Workaround: added metadata to description – that time participant added group not users	
4.b)	Yes	4:39

Table C.1: Experiment with Participant Florian Wolfrum.

Task	Task Finished	Time To Finish
	– most time was needed to find the file on the computer	
4.c)	Yes	2:25
	– opened md file and added content to description	
4.d)	Yes	0:40
4.e)	Yes	5:12
	– found quickly the dashboard tab – did not understand structure – tried out the button 'Add panel' but did not know what it would mean – added all pictures in one panel – received error message when saving it – tried again and it worked	
4.f)	Yes	1:33
	– tried first Permissions tab in the record and then went back to Collection – was automatically part of it	
5.a)	Yes	4:44

Table C.1: Experiment with Participant Florian Wolfrum.

Task	Task Finished	Time To Finish
	– searched a bit for where to create templates – Workaround: added Metadata now in the respective field, but puts all information in the key-field	
5.b)	No	3:34
	– just changes template again – did not see, he was still in the template field	
6.a)	Yes	1:20
	– clicked first in the field – then realized he had to click the edit button first	
6.b)	Yes	1:06
	– time specification is not precise	
7.a)	Yes	0:36
7.b)	Yes	0:30
8.	No	0:39
	– could not change the color	
9.	Yes	1:31
	– tried first the general settings – complained that language switch is at the very bottom	

Table C.1: Experiment with Participant Florian Wolfrum.

Task	Task Finished	Time To Finish
10.	No	/
	– did not want to explore the web page any more	

Table C.2: Experiment with Participant Annika Schwenke.

Task	Task Finished	Time To Finish
1.	Yes	1:03
2.	Yes	1:36
	– quickly found field for creating group	
3.	Yes	1:41
	– Added Group, not users	
4.a)	Yes	6:25
	– creates record via button in Collections ('Link record' > 'New record') – does not create new type 'experiment' but chooses already existing type 'experimental' – Workaround: confused by input field metadata, first skips it and then enters key and value but ignores type	
4.b)	Yes	4:38
	– quickly uploaded file – needed time to find description of experiment – called the process laborious	
4.c)	Yes	1:44

Table C.2: Experiment with Participant Annika Schwenke.

Task	Task Finished	Time To Finish
	– uploaded the file and added the link to the file – could not see the values directly but after clicking on the link	
4.d)	Yes	0:41
4.e)	No	4:24
	– found quickly the dashboard tab – did not understand panel but clicked the button anyways – complained about how small the picture is – created extra dashboards for the other plots	
4.f)	Yes	0:49
	– was automatically part of it	
5.a)	Yes	3:32
	– tried out records first – Metadata same as before	
5.b)	Yes	2:26
6.a)	No	0:34
	– could not edit file – after clicking 'Edit file' the content is empty	
6.b)	No	1:32
7.a)	Yes	2:04

Table C.2: Experiment with Participant Annika Schwenke.

Task	Task Finished	Time To Finish
7.b)	Yes	0:33
8.	No	2:19
	– did not figure out how to color the text or make it bold	
9.	Yes	0:59
	– tried first the general settings – tried help page and found no information – uses general search to find Language	
10.	No	/
	– did not want to explore the web page any more	

Table C.3: Experiment with Participant Nikita Baklanov.

Task	Task Finished	Time To Finish
1.	Yes	1:08
2.	Yes	3:35
	– took some time to find groups, started searching in the main panel – identifier is already in use, changes name instead of identifier – when trying to add members, first clicked 'add Members' button – added users individually – was confused that the add Button for a new user is on the side of the old user	
3.	Yes	2:02
	– changed again identifier via the name – started adding users individually, but then saw that type could be changed and whole groups could be added – liked the 'recently visited' field	
4.a)	Yes	4:18
	– does not create new type 'experiment' but chooses already existing type 'experimental' – Workaround: confused by input field metadata, enters key and value but ignores type – liked the automated filtering during search for users	
4.b)	No	3:58

Table C.3: Experiment with Participant Nikita Baklanov.

Task	Task Finished	Time To Finish
	– quickly uploaded file – liked Drag and Drop – the description field of the record had no header in the preview mode, thus participant had problems finding that field – did not knwo how to create a header	
4.c)	No	7:58
	– did not know how to add a markdown file – tried dragging it directly in the description text field – searched for term 'Markdown' in the help page – uploaded as a link	
4.d)	Yes	1:06
4.e)	Yes	4:24
	– tried clicking in the grid – When new Panel is activated it contains same content as previous panel – did not understand panel but clicked the button anyways – first, added all pictures to one panel – then, one for each – complains about small picture size and preselection of title and subtitle	
4.f)	Yes	0:44

Table C.3: Experiment with Participant Nikita Baklanov.

Task	Task Finished	Time To Finish
	– was automatically part of it	
5.a)	Yes	5:44
	– problems with distinguishing template description and template data description – went back to the record and clicked button 'as template' – deleted md file	
5.b)	Yes	3:10
	– that time, copied content of md file – liked preview of md text	
6.a)	Yes	2:40
	– added content of md file now	
6.b)	Yes	0:40
	– Time not specific – liked that changes could be seen in red and green	
7.a)	Yes	3:00
	– tried search bar – complained about single filters and positions of it	
7.b)	Yes	0:19
8.	Yes	0:09

Table C.3: Experiment with Participant Nikita Baklanov.

Task	Task Finished	Time To Finish
9.	No	5:03
	– bold worked – googled how to color the text, html-tags and R Markdown did not work	
10.	No	/
	– did not want to explore the web page any more	

Table C.4: Experiment with Participant Luisa Hideg.

Task	Task Finished	Finishing Time
1.	Yes	1:50
	colors are monotonous	
2.	Yes	3:10
	– identifier are not self-explanatory. – Afraid to click the button "Weiter" without adding group members. – Search for group member not intuitive. The buttons beside the search bar are not structured logically.	
3.	Yes	3:12
	– The project looks exactly like a group. It is hard for me to keep track.	
4.a)	Yes	9:46
	– did not find "records" in the application – Identifier was automatically filled by the name of the Experiment. Then the identifier was to long and could not be changed. Only when clicking on "Weiter" and a error message occurred, it could be changed.	
4.b)	No	2:50
	– The picture is not seeable	
4.c)	No	1:45

Table C.4: Experiment with Participant Luisa Hideg.

Task	Task Finished	Finishing Time
	– The notes are not visible	
4.d)	Yes	0:35
4.e)	No	4:35
	– couldn't find a way to add the plots	
4.f)	No	3:30
	– didn't really understand the concepts of experiments/collection and what the difference is	
5.a)	Yes	4:34
	– cannot copy metadata from last experiment	
5.b)	Yes	4:22
	– difficulties to find the project – concepts are hard to understand and distinguish. Project, Group, Collection, Experiment	
6.a)	No	2:49
	– by clicking on the .md file, it should be changeable in the browser.	
6.b)	No	0:32

Table C.4: Experiment with Participant Luisa Hideg.

Task	Task Finished	Finishing Time
	– there are no changes	
7.a)	Yes	1:12
7.b)	Yes	0:18
8.	Yes	0:40
	– to much at the bottom of the settings	
9.	No	1:20
	– could not change the color	
10.	No	/
	– did not want to explore the web page any more	

Table C.5: Experiment with Participant Tim Joost.

Task	Task Finished	Finishing Time
1.	Yes	0:54
2.	Yes	2:45
	– Asks why he is seeing these users – Search process for other users is not intuitive	
3.	Yes	2:10
	– User asks what the difference between a collection and experiments are	
4.a)	Yes	4:34
	– User ask what are keys, and how to specify them – Did not correctly understand difference of int, String, etc	
4.b)	No	4:04
	– Picture is very small, hard to see it	
4.c)	Yes	7:27
	– measurements, notes and picture are not really seeable	
4.d)	Yes	1:27
4.e)	Yes	3:50

Table C.5: Experiment with Participant Tim Joost.

Task	Task Finished	Finishing Time
	– The plots are very small, would like to see the plots as big pictures in a pop up	
4.f)	Yes	1:02
5.a)	Yes	4:38
	– The metadata has to be added manually again, despite already entering it before	
5.b)	Yes	2:47
	– This was good	
6.a)	Yes	2:02
	– hard to find where to change the .md data.	
6.b)	Yes	0:38
7.a)	Yes	3:34
7.b)	Yes	0:32
8.	Yes	0:27
	– the language setting should be at the navbar	
9.	No	5:03

Table C.5: Experiment with Participant Tim Joost.

Task	Task Finished	Finishing Time
	– couldn't find where to change colour to red, didn't like the restrictions for taking notes. Ask, if possible to have free-hand notes	
10.	No	/
	– did not want to explore the web page any more	

Table C.6: Experiment with Expert

Task	Task Finished	Time To Finish
1.	Yes	0:40
2.	Yes	0:50
3.	Yes	0:26
4.a)	Yes	1:47
4.b)	Yes	0:50
4.c)	Yes	0:37
4.d)	Yes	0:13
4.e)	Yes	1:20
4.f)	Yes	0:05
5.a)	Yes	1:52
5.b)	No	1:03
6.a)	Yes	0:35
6.b)	Yes	0:13
7.a)	Yes	0:10
7.b)	Yes	0:14
8.	Yes	0:08
9.	No	0:29
	– could not change the color	
10.	No	/
	– no exploration done	

D Appendix

Table D.1: Usability Survey with Florian Wolfrum

Question	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. I think that I would like to use this system frequently.		X			
2. I found the system unnecessarily complex.			X		
3. I thought the system was easy to use.		X			
4. I think that I would need the support of a technical person to be able to use this system.			X		
5. I found the various functions in this system were well integrated.			X		
6. I thought there was too much inconsistency in this system.			X		
7. I would imagine that most people would learn to use this system very quickly.				X	
8. I found the system very cumbersome to use.			X		
9. I felt very confident using the system.		X			
10. I needed to learn a lot of things before I could get going with this system.		X			

Table D.2: Usability Survey with Annika Schwenke

Question	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. I think that I would like to use this system frequently.	X				
2. I found the system unnecessarily complex.				X	
3. I thought the system was easy to use.		X			
4. I think that I would need the support of a technical person to be able to use this system.		X			
5. I found the various functions in this system were well integrated.		X			
6. I thought there was too much inconsistency in this system.			X		
7. I would imagine that most people would learn to use this system very quickly.		X			
8. I found the system very cumbersome to use.		X			
9. I felt very confident using the system.		X			
10. I needed to learn a lot of things before I could get going with this system.		X			

Table D.3: Usability Survey with Nikita Baklanov

Question	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. I think that I would like to use this system frequently.			X		
2. I found the system unnecessarily complex.		X			
3. I thought the system was easy to use.		X			
4. I think that I would need the support of a technical person to be able to use this system.		X			
5. I found the various functions in this system were well integrated.			X		
6. I thought there was too much inconsistency in this system.				X	
7. I would imagine that most people would learn to use this system very quickly.	X				
8. I found the system very cumbersome to use.				X	
9. I felt very confident using the system.		X			
10. I needed to learn a lot of things before I could get going with this system.				X	

Table D.4: Usability Survey with Luisa Hideg

Question	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. I think that I would like to use this system frequently.			X		
2. I found the system unnecessarily complex.				X	
3. I thought the system was easy to use.		X			
4. I think that I would need the support of a technical person to be able to use this system.		X			
5. I found the various functions in this system were well integrated.			X		
6. I thought there was too much inconsistency in this system.		X			
7. I would imagine that most people would learn to use this system very quickly.		X			
8. I found the system very cumbersome to use.				X	
9. I felt very confident using the system.		X			
10. I needed to learn a lot of things before I could get going with this system.			X		

Table D.5: Usability Survey with Tim Joost

Question	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. I think that I would like to use this system frequently.		X			
2. I found the system unnecessarily complex.		X			
3. I thought the system was easy to use.		X			
4. I think that I would need the support of a technical person to be able to use this system.			X		
5. I found the various functions in this system were well integrated.			X		
6. I thought there was too much inconsistency in this system.			X		
7. I would imagine that most people would learn to use this system very quickly.		X			
8. I found the system very cumbersome to use.				X	
9. I felt very confident using the system.	X				
10. I needed to learn a lot of things before I could get going with this system.				X	

References

Brandt, G.-L.-H. C. S. E. T. G. Z. Y. Z. P., N., & Selzer, M. (2021). *Kadi4mat: A research data infrastructure for materials science.* doi: https://doi.org/10.5334/dsj-2021-008

Brooke, J. (1995, 11). *Sus: A quick and dirty usability scale* (Vol. 189).

Gediga, H.-K.-C. D. I., G. (1999). The isometrics usability inventory: An operationalization of iso 9241-10 supporting summative and formative evaluation of software systems. *Behaviour Information Technology*, *18*(3), 151-164. doi: https://doi.org/10.1080/014492999119057

Hamborg, K.-C. (1996). Zum einfluss der komplexität von software-systemen auf fehler bei computernovizen und experten. *Zeitschrift für Arbeits- und Organisationspsychologie*, *40*(1), 30-11.

Higgins, N.-V.-A. . S. M., S.G. (2022). *Considerations for implementing electronic laboratory notebooks in an academic research environment.* doi: https://doi.org/10.1038/s41596-021-00645-8

ISO 9241-11. (1997). *Ergonomic requirements for office work with visual display terminals (vdt's), part 12: Presentation of information.*

ISO 9241-110. (2006). *Ergonomics of human-system interaction, part 10: Dialogue principles.*

Nielsen, J. (1994). *Usability engineering.* Morgan Kaufmann.

Theis, S. (2023, November). *Usability engineering and human factors.* Institute for Software Technology.