dHealth 2020 – Biomedical Informatics for Health and Care
G. Schreier et al. (Eds.)

doi:10.3233/SHTI200097

Assessing and Improving the Usability of the Medical Data Models Portal

Daniel REICHENPFADER[a,1], Robin GLAUSER[b,1], Martin DUGAS[c] and Kerstin DENECKE[b,2]
[a] *FH Joanneum, Graz, Austria*
[b] *Bern University of Applied Sciences, Biel, Switzerland*
[c] *University of Münster, Münster, Germany*

Abstract. Case report forms (CRF) specify data definitions and encodings for data to be collected in clinical trials. To enable exchange of data definitions and in this way to avoid creation of variants of CRF for similar study designs, the Medical Data Model portal (MDM) has been developed since 2011. This work aims at studying the usability of the MDM portal. We identify issues that hamper its adoption by researchers in order to derive measurements for improving it. We selected relevant tools (e.g. Nibbler, Hotjar, SUPR-Q) for usability testing and generated a structured test protocol. More specifically, the portal was assessed by means of a static analysis, user analysis (n=10), a usability test (n=10) and statistical evaluations. Regarding accessibility and technology, the static code analysis resulted in high scores. Presentation of information and functions as well as interaction with the portal still has to be improved: The results show that only limited functions of the webpage are used regularly and some user navigation errors occur due to the portal's design. In total, six major problems were identified which will be addressed in future. A continuous evaluation using the same structured test protocol allows to continuously measure the website quality, to compare it after changes have been implemented and in this way, to realise a continuous improvement. The effort for a repeated evaluation of the same evaluation with 10 persons is estimated with 10 hours.

Keywords. metadata, usability, semantic annotation, medical data models

1. Introduction

Case report forms (CRF) [1] are questionnaires tailored to a specific study design in which the necessary examination data of patients are captured. The data definitions and specifications in CRFs should enable to collect data in sufficient detail without ambiguity, unnecessary details and should avoid redundancy. CRFs are usually only accessible to the respective institution who generated it. Due to lack of accepted standards for data capturing, missing metadata and varying data definitions in CRFs, sharing and reusing them is complicated and results in avoidable time and resource consumption [2]. To address this issue, there have been efforts to offer a centralised data dictionary, where data definitions and corresponding metadata can be shared among researchers. One of

[1] Contributed equally

[2] Corresponding Author: Kerstin Denecke, Bern University of Applied Sciences, Quellgasse 21, Biel, Switzerland, E-Mail: kerstin.denecke@bfh.ch

those efforts is the portal for Medical Data Models, or MDM portal for short, which is a metadata registry for the creation, analysis, release and reuse of medical forms [2], [3]. It has been the largest freely accessible database for medical data models and is a registered European information infrastructure [2],[3]. It provides a platform for structured collection, discussion, export and cross-reference of medical forms and data models. MDM contains forms in the CDISC Operational Data Model (ODM) format with ~520,000 data elements (as of 2020). Among those, core data sets, common data elements or data standards, code lists and value sets are provided. It is possible to view, discuss, and export forms in various technical formats (PDF, CSV, Excel, SQL, SPSS, R, etc.). Specifically, it provides a multilingual platform for exchange and discussion of data models in medicine, both for medical research and healthcare. The main functionalities are uploading and editing ODM-XML files, searching for data models and creating new data models. As of 2020, there are ~2600 registered users of the portal; it is intended to further enlarge this user base which requires a high usability of the platform. However, the usability of the portal has not been assessed so far. Therefore, this work aims at studying the usability of the MDM portal and at identifying issues that hamper its adoption by researchers in order to derive measurements for improving it and in this way, increase the number of users.

According to ISO definition 9241-11:2018, 3.1.1, usability is defined as a measure of effectiveness, efficiency and user satisfaction when using a certain service, product or system [4]. The duration users stay on a website depends directly on implicitly received feedback as part of usability (e.g. when clicking, scrolling, bookmarking). In case of poor presentation and lack of feedback, users often stay between 37 and 47 seconds on a webpage when using a desktop application [5]. There are several quantitative and qualitative methods available for assessing usability, navigability of websites and to identify usability problems [6]. There are generic usability methods originating from the field of human–computer interaction that can be divided into expert-focused and user-focused methods [7]. Expert-focused methods, such as heuristic evaluations [8], rely on the quality judgments of experts. User-focused methods (think-aloud usability testing, user page reviews, user surveys) try to collect relevant data among (potential) users of the website. The objective of website evaluation is to make a website useful, profitable, and accessible to a user [9]. Depending on the content, target user group and tasks to be supported, different aspects are of interest and have to be studied [10]. Due to the specific tasks and content of the MDM portal, we have to develop a usability assessment framework tailored to the MDM portal. In this paper, we present our test protocol and study results.

2. Methods

Based on the work of Thielsch [6], we identified aspects and tools that are of interest when studying the usability of the MDM portal. We focused on tools and methods that allow to assess the quality of a website (and not its content). A limiting factor to be considered in method selection is the project duration of 15 weeks. Further, the evaluation method is expected to be repeated later on to continuously improve the website. Therefore, conducting the tests shoul not be too time consuming. As a result, we decided for four methods: Static code analysis, user analysis, usability testing combined with a statistical evaluation. They aim at providing information on technical improvements of the website, details on user behaviour, subjective impressions of users

and information on usability. The results of the four methods are analysed and interpreted. Conclusions are drawn with respect to possible improvements of the website. Details on the single methods are provided in the following.

By means of a static code analysis, the code of a website is analysed. For selecting a code analysis tool, we based on the work of Kumar and Hasteer [11]. The web tool "Nibbler" (https://nibbler.silktide.com) was finally chosen because of its ease of use and wide range of functions. The tool automatically examines the website http://www.medical-data-models.org and related subdomains for aspects such as accessibility, SEO (search engine optimisation), readability, etc. A rating on a ten-level scale for different categories is provided as results accompanied by comments and suggestions for improvement.

User analysis analyses user behaviour while interacting with a website. This can be realised non-obtrusively in the background by recording mouse movements and click behaviour to identify to what extent functions and website areas are accessed. The user analysis of the MDM portal is carried out with scientific employees of the Institute for Medical Informatics of the University of Münster that use the platform on a regular basis. They have a background in life sciences, medical informatics or computer sciences. 10 persons were requested to join the test. We decided to use the tool “Hotjar” (http://www.hotjar.com) since it has already been used successfully in other studies [12]–[14]. It is used to collect data on user interaction behaviour. More specifically, an extension for the web browser “Chrome” is developed and provided to the participants who installed it on their computers. Once installed, the behaviour of the users is recorded on three specific areas of the webpage: form search, detailed view of a form, and the editor for creating a new form. The data is collected over a period of one week or five working days. The results are provided by Hotjar as three heat maps, which visualise the users' click and scroll behaviour and mouse movements.

Usability testing is combined with a statistical evaluation using a questionnaire. We decided to use SUPR-Q (Standardised User Experience Percentile Rank Questionnaire, [15]). SUPR-Q comprises eight questions concerning usability and a final optional question "comments/feedback” [15]. Answers are provided on a 5-item Likert scale. 10 test persons were recruited from the clinical trial units (CTU) of the universities of Bern and Basel. These test persons have different professional backgrounds and as members of the CTU staff they offer the scientific, technical and computing expertise needed to support patient-oriented clinical research at all stages. The participants have to solve three tasks: to search (task 1), compare (task 2) and create (task 3) medical forms followed by filling the SUPR-Q. These are the main tasks to be supported by the MDM portal. We measure task completeness, time needed and occurrence of errors. A test is stopped for a user when a serious error occurred. They are only allowed to solve the three tasks – there is no time to get familiar with the portal or test other functionalities. The usability test is conducted in two runs. The participants have to use the portal on a provided test device. Based on the results of the first test run, the website is adapted. The changes to be realised are selected depending on the feasibility of the changes as well as the number and impact of an error. The changes made are applied directly to the productive version of the portal via a self-implemented browser extension, but only available locally on the test device. The test persons of the second run thus work with a modified and improved version of the website.

3. Results

In this section, we summarise the results achieved by the four methods (see Table 1) and derive suggestions for improving the website based on these results.

Table 1. Methods and tools used in the test together with a summary of results.

Method	Tool	Result
Static code analysis	Nibbler	Specify image size, no tool for user analysis implemented, replace some complex URLs
User analysis	Hotjar browser extension	Relevant areas of the website, login errors, identification of inaccessible functions, pages are too comprehensive
Usability test	Task-based usability test	Some functions cannot be found, login errors
Statistic evaluation	SUPR-Q	Relevance and reliability of the provided information is good, usability can be improved

3.1. Results of static code analysis

The overall score for static code analysis amounts to 6.7 out of 10 points. This overall score is split into four aspects: Accessibility (8.6/10), Experience (7.4/10), Marketing (3.6/10) and Technology (8.0/10). Accessibility and technology are ranked with a high score. The code analysis showed that internal links are specified which helps SEO and aid accessibility. The webpage is optimised for printing and for access from mobile devices. All pages have defined headings. URLs for changing language are complex. The experience aspect is ranked lower, but still good. The main reasons are images without specified size and missing links to social media such as Twitter or Facebook. The latter aspect and missing implementation of a tool for user analysis impact on the very low marketing score.

3.2. Results of the user analysis

Within the defined test period of seven days, the Hotjar tool registered a total of 238 user interactions. The visualisation of the mouse movements did not provide interpretable results and was therefore not evaluated. Ten calls of the page "search form" were registered. From recorded click behaviour, it becomes evident that the filter function is not used when searching for a form; the name of the form is entered and selected directly from the list. The scrolling record also shows that the searched form was always listed within the first five search results, as more than 80% of users on the page did not scroll down. The detailed view of a form shows 32 hits within the period of investigation. On this page, too, all actions performed took place in the upper quarter of the page. Actions frequently registered on this page include logging on to the website, downloading the form, changing the form language and leaving comments.

The "Create form" page was accessed 196 times during the study period, i.e. six and nineteen times more often than the other two pages. Half of the users remain on the top 50% of the page length. Records of clicks show very clearly that when a new form is created, three of the thirteen input fields are used: Field name, encoding and data type. Existing support functions for finding data fields and codes are not used. In addition, 24 video recordings of the users' surfing behaviour were automatically taken, lasting between nine seconds and 230 minutes. The recordings show that users must log in again several times during a session because the log-in data is regularly reset automatically.

3.3. Usability test results

The first test run was successfully conducted with five employees of the CTU of the University of Bern. Only task 1 (find form) could be completed successfully by all participants (see Figure 1). Task 2 (compare forms) could not be completed by any of the test subjects and task 3 by 40% of the participants. Time on task differed among the tasks with 10 minutes in average for task 3 to around 4 minutes spent on task 1. A total of 13 different errors were observed in the first test run. Problems occurred in identifying some interaction elements such as adding a data model to a selection, create a form or the filter function. Further, the users became frustrated because they had to log in several times and there was confusion between the language setting functions for the website and the form. Accordingly, after completion of the first test run and analysis of the results, the following adjustments are made to the portal and applied directly to the changed elements of the website using a browser extension: 1) added entry in the navigation bar to create a new form, 2) adjusted layout of the "Create form" page, 3) changed colour and icon of the "Add form to selection" button.

The second test run was also successfully conducted with five employees of the CTU of the University of Basel (see Figure 1). The success of improvements becomes obvious in an improved average time on task which was shorter for all tasks. The completion rate for two tasks increased. However, the completion rate dropped for task 1 compared to the results from the first test run, as one user experienced insuperable, recurring login problems. Unexpected problems were caused using a German keyboard on the test device, which caused problems entering special characters (@, +). In these cases, the test leader intervened. These errors were not further considered.

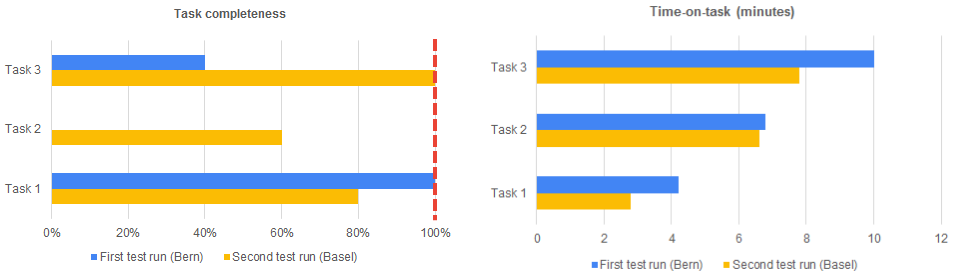

Figure 1. Results from the usability test: Task completeness (left), time on task (right). The tasks are: search (task 1), compare (task 2) and create (task 3) medical forms.

3.4. Result of SUPR-Q

Table 2 shows the results of the SUPR-Q questionnaire. The results show that the users recognise the relevance and reliability of the information provided in the portal (statements 3 and 4). 80% of the participants agree or strongly agree with these statements. There is room for improvement in the presentation of the information and functions as well as in the interaction. 50% of the participants disagree with the statement that the portal provides a simple and clear presentation. The participants criticised the evaluation setup, since they had only limited time to handle a predefined task and in this

way, only limited time to interact with the system to get an overall impression. They confirmed that operating and navigation elements can often not be recognised easily. The target user group of the portal does not seem to be clearly specified. A positive statement was that the creation of forms is easy. Meta information about menu items, fields and options is missing. 4 out of 10 persons would rather recommend the portal; two claim they will not recommend and 4 persons were rather neutral regarding recommendation.

Table 2. SUPR-Q results (n=10)

	1 (strongly disagree)	**2**	**3**	**4**	**5 (strongly agree)**
1. The system is easy to interact with.	0	2 (20%)	5 (50%)	3 (30%)	0
2. The system is easy to navigate.	0	3 (30%)	3 (30%)	4 (40%)	0
3. Provided information is credible and plausible	0	1 (10%)	1 (10%)	6 (60%)	2 (20%)
4. Provided information is trustworthy and reliable	0	1 (10%)	1 (10%)	5 (50%)	3 (30%)
5. Portal is well designed	0	3 (30%)	4 (40%)	3 (30%)	0
6. Portal provides a simple and clear presentation	1 (10%)	4 (40%)	2 (20%)	3 (30%)	0
7. I will use the system in future.	1 (10%)	2 (20%)	4 (40%)	2 (20%)	1 (10%)

3.5. Suggestions for improving the Website

By combining the results from the various test methods, suggestions for improvement were made. They are shown in Table 3, sorted by severity of problem on a scale from one to four. In the static test, the website already achieved very good results, only in the area of "social media" the portal received a larger deduction, as neither a Twitter nor Facebook page was created and linked for the website. It has to be clarified whether links to social media will help to make the platform more popular among researchers.

Table 3. Errors and solution proposals

Severity	**Error**	**Proposed solution**
3	Re-Login required	This problem is due to a caching problem and should be addressed accordingly.
3	Buttons for adding forms are unobtrusive	The buttons for adding forms should become more obvious.
3	"Create Form" functionality is not obvious.	The link to create a new form should be directly accessible via the navigation bar, as this is one of the most used functions on the website.
2	Comparing forms: unclear output function	Within the tool to compare forms, it would be easier for users if another label like “Web output” would be used instead of the label “HTML Output”.
2	Unclear submission of the request in the field search when creating forms	The submission of the form in the field search context has to be facilitated; the key “Enter” key should not be used to submit the form.
2	Search field is not sufficiently highlighted	The search field should be highlighted as it is a powerful function of the website and currently users miss this function.

4. Discussion and conclusion

There are numerous guidelines and tools available for studying the usability of websites. However, some of them are not directly applicable to the MDM portal. The portal’s

specific content (medical data models), target groups (mainly medical professionals) and numerous supported functions (search, compare, export, create, encode, comment medical data models) require an innovative approach. The proposed test strategy using several methods and tools is based on current literature but has neither been carried out with the four selected methods nor applied on health-related websites. The analysis with the tool Nibbler basically replaces the manual check of the website for conformity with best-practice guidelines regarding aspects of marketing, technology, accessibility and user experience. However, level of detail and number of examined elements are, in contrast to manual investigation, limited. Functional limitations of the tool occurred during our tests since the free version of the software was used.

In our evaluation, we used SUPR-Q to collect data on the user experience. Its advantages, including the reduced number of questions while maintaining high internal coherence, differentiates the SUPR-Q from other test procedures such as the System Usability Scale. However, the questions do not specifically address aspects of health-related websites. In future, it could be assessed whether the Health Information Technology Usability Evaluation Scale is better suited [16]. The number of test persons was limited to 5 persons per test run. Turner et al. claim that the most serious usability problems can be revealed with only three subjects [17]. So, we believe that the main problems could have been detected. A limiting factor for answering the SUPR-Q questionnaire is that the participants only had time to solve the given tasks and not to get familiar with the portal and study other functionalities. Therefore, we assume, that additional usability issues might be revealed when other tasks have to be solved. Regarding the statistical evaluation using the SUPR-Q questionnaire, a direct approach of users should be considered, as investigated on a library portal by Chiranov [18]. The use of additional qualitative methods may yield additional insights. Thielsch for example recommends combining qualitative inspection methods (e.g. expert review) with qualitative in-depth interviews or focus groups [6]. Consideration of psychometric quality criteria, namely objectivity, reliability and validity are still to be clarified [19]. The legal requirements of the European General Data Protection Regulation, GDPR, complicated conducting the user analysis: integrating the selected tool Hotjar into the website was legally imposing. Instead, we used staff of the University of Muenster. The users were sufficiently informed about the data to be collected. Installation of the necessary tracking code could be done by staff members themselves. The created browser extension including the installation instructions can be regarded as an innovation of this work, since neither the analysis tool itself provides a convenient way to add required tracking code to the source code, nor such or similar extensions could be found online. Similar to static analysis, a free test version was used for user analysis, entailing constraints in the scope of functions.

In conclusion, we introduced a test method for assessing health-related websites with comprehensive functionalities. In particular, when time and resources are limited, we suggest quantitative methods: Static code analysis, user analysis, usability testing combined with a statistical evaluation. It could be shown that serious usability problems can be revealed. The approach can be easily applied to other websites. The set of methods could be extended by qualitative methods such as expert interviews or focus groups. Alternative commercial tools for static (e.g. Silktide) and user analysis (e.g. Matomo) offer an extended range of functions and could also be considered. For the comprehensive investigation of user behaviour, technical, organisational and legal aspects must be considered in order to comply with the general data protection regulation [20]. Further usability tests with other tasks can easily be carried out based on the defined

test procedure. To realise this, new tasks and indicators for measuring task success have to be defined. The user analysis with installed Hotjar browser extension can be used without additional effort, just requires analysis of the results.

References

[1] S. Bellary, B. Krishnankutty, and M. S. Latha, Basics of case report form designing in clinical research, *Perspect. Clin. Res.*, vol. 5, no. 4, p. 159, Jan. 2014, doi: 10.4103/2229-3485.140555.

[2] M. Dugas et al., Portal of medical data models: information infrastructure for medical research and healthcare, *Database*, vol. 2016, p. bav121, 2016, doi: 10.1093/database/bav121.

[3] M. Dugas, Portal of Medical Data Models (MDM-Portal). [Online]. Available: https://medical-data-models.org/. [Accessed: 29-Oct-2019].

[4] International Organization for Standardization, ISO 9241-11:2018(en), Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en. [Accessed: 29-Oct-2019].

[5] R. Homma, K. Soejima, M. Yoshida and K. Umemura, Analysis of User Dwell Time on Non-News Pages' 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4333-4338..

[6] M. Thielsch, Expertise Website-Evaluation: Übersicht über bestehende Evaluationsmethoden und Entscheidungshilfe für die Evaluation bestehender sowie neu geschaffener Websites. Arbeitsbericht, Köln: Bundeszentrale für gesundheitliche Aufklärung (BZgA). https://doi.org/10.17623/BZGA:225-EWE-1.0, 2018.

[7] W. Tan, D. Liu, and R. Bishu, Web evaluation: Heuristic evaluation vs. user testing, *Int. J. Ind. Ergon.*, vol. 39, no. 4, pp. 621–627, Jul. 2009, doi: 10.1016/j.ergon.2008.02.012.

[8] J. Nielsen, 10 Usabilty Heuristics for User Interface Design. *Nielsen Norman Group*, 1995.

[9] O. Signore, A Comprehensive Model for Web Sites Quality, in *Seventh IEEE International Symposium on Web Site Evolution*, Budapest, Hungary, 2005, pp. 30–38, doi: 10.1109/WSE.2005.1.

[10] L. Hasan and E. Abuelrub, Assessing the quality of web sites, *Appl. Comput. Inform.*, vol. 9, no. 1, pp. 11–29, Jan. 2011, doi: 10.1016/j.aci.2009.03.001.

[11] R. Kumar and N. Hasteer, Evaluating usability of a web application: A comparative analysis of open-source tools, in *2017 2nd International Conference on Communication and Electronics Systems (ICCES)*, 2017, pp. 350–354, doi: 10.1109/CESYS.2017.8321296.

[12] J. Grigera, A. Garrido, J. M. Rivero, and G. Rossi, Automatic detection of usability smells in web applications, *Int. J. Hum.-Comput. Stud.*, vol. 97, pp. 129–148, Jan. 2017, doi: 10.1016/j.ijhcs.2016.09.009.

[13] B. Amaro, E. Mira, L. Dominguez, and J. P. D'Amato, A Usability Analysis of a Serious Game for Teaching Stock Market Concepts in Secondary Schools, in *New Knowledge in Information Systems and Technologies*, Cham, 2019, pp. 839–849, doi: 10.1007/978-3-030-16184-2_80.

[14] M. Eaton and C. Argüelles, Usability Study for a Community College Library Website: A Methodology for Large-Scale Data Gathering, *Community Jr. Coll. Libr.*, vol. 23, no. 3–4, pp. 99–113, Oct. 2017, doi: 10.1080/02763915.2019.1645552.

[15] J. Sauro, SUPR-Q: A Comprehensive Measure of the Quality of the Website User Experience, vol. 10, no. 2, pp. 68–86, 2015.

[16] R. Schnall, H. Cho, and J. Liu, Health Information Technology Usability Evaluation Scale (Health-ITUES) for Usability Assessment of Mobile Health Technology: Validation Study, *JMIR MHealth UHealth*, vol. 6, no. 1, p. e4, Jan. 2018, doi: 10.2196/mhealth.8851.

[17] C. Turner, J. Lewis, and J. Nielsen, Determining Usability Test Sample Size, in *International Encyclopedia of Ergonomics and Human Factors*, vol. 3, 2006.

[18] M. Chiranov, Applying pop-up survey software to incorporate users' feedback into public library computing service management, *Perform. Meas. Metr.*, vol. 12, no. 1, pp. 50–65, Jan. 2011, doi: 10.1108/14678041111124298.

[19] K. Kubinger. Gütekriterien. In M. A. Wirtz (Hrsg.), Dorsch – Lexikon der Psychologie (18. Aufl., S. 664). Bern: Verlag Hogrefe Verlag, 2014, p. 664.

[20] M. Muth, GDPR - Required Changes By Website Operators And Technical Ways To Check For Compliance, in *Proceedings zu den Seminaren Future Internet (FI) und Innovative Internet Technologien und Mobilkommunikation (IITM)*, München, 2018, pp. 57–66.