Evaluating the Potential of a Data-Driven Approach in Digital Service (Re)Design

Tea Mijač, Mario Jadrić, Maja Ćukušić
Faculty of Economics, Business and Tourism, University of Split
Cvite Fiskovica 5, Split, Croatia
{tea.mijac, mario.jadric, maja.cukusic}@efst.hr

***Abstract**. The amount of data is exponentially growing each day. With every interaction with digital services, users create their digital footprint. It is not unusual that usage data can demonstrate the need more objectively than users themselves leading to the fact that huge potential is hidden in data-driven development approaches. In order to investigate the reasoning behind using objective versus subjective data to changes in design of digital services, a pilot research study was conducted. The study was performed using A/B testing and a questionnaire. Original website was used as Design A, and with same functionalities kept, another webpage was launched (Design B). Among other findings, the analysis confirms that there is demonstrated need and rationale behind using objective users' data rather than subjective in testing the changes in design.*

Keywords. Digital service design, data-driven approach

1 Introduction

Number of digital artefacts is growing rapidly each day (Tomitsch, 2018). Since the amount of data produced by users nowadays exponentially increases, and by going through the more and more available usage data (generated through interaction with digital services), the development teams got the chance to understand what users are really doing and how they react better (King, Churchill, & Tan, 2017; Spiess, Joens, Dragnea, & Spencer, 2014). Since the user habits and interests are changing rapidly and new trends emerge daily, and having in mind that user experience (UX) is ultimately subjectively, dynamically and contextually dependent (Halvorsrud, Kvale, & Følstad, 2016), designers and developers have no choice but to take into consideration the data generated by different user actions and feedback collected from the overall user experience (Lee, Smith, Calvert, & Snajdr, 2016; Lemon & Verhoef, 2016).

Consequently, tracking the objective user data is extremely important and this data is a key component when evaluating the user experience (Sengers, Boehner, Mateas, & Gay, 2008). Although many studies confirm that direct contact with users is the key in user-oriented approaches and in fulfilling their expectations (Kujala, Kauppinen, & Rekola, 2001), the process of redesign is still often based on intuition rather than actual data (Havice, 2017) when, in effect, the designers could use e.g. the mouse movement, keyboard clicks, and so on, as the best input for the improvement (King et al., 2017). There is nothing more direct than the data that users produce themselves. With more and more data becoming available, the greater the chances are of understanding the users' needs (Anderson, 2015; Spiess et al., 2014).

The importance of using objective data in the context of improving the user experience has been a popular topic recently. For example, several authors studied data-driven development in telecommunications where they collected data to foresee potential customer complaints and tried to improve their user experience (Bao, Wu, & Liu, 2017). Similarly, Lee et al. (2016) investigated the approach in improving a library website. Despite the fact they highlighted the importance of using the objective data to improve the user experience, the main limitation of their research was using only the external source data – Google Analytics. Generally speaking, in the Human-Computer Interaction (HCI) field there is a lack of research about user experience metrics based on behavioural objective data produced by users themselves (Rodden, Hutchinson, & Fu, 2010).

In one of our recent studies (in the process of publication), we propose a methodological framework for user-oriented data-driven information systems modelling devised around the well-known IS development phases. The difference is that the proposed framework aims to emphasise the user experience and fosters the data-driven approach. The data can be used in order to either improve the user experience by way of eliminating critical errors, if these are detected, or to improve the whole users' journey while interacting with the system. In order to use the objective data in this whole process, metrics have to be devised and incorporated in the development phase. The proposed framework highlights these aspects as an important step/phase before any redesign.

Why we propose this methodological framework? Three major reasons why IT projects fail are the lack of information from end-users, unfinished specifications and frequent changes in specifications (Geogy & Dharani, 2016). Data-driven approach makes it easier to collect and understand the user needs and to provide much better quality of interaction – which takes us a step closer to collecting and analysing the actual needs of users as the basis for (re)design and development of user-oriented digital services. Data Driven Development (DDD) assumes that the development teams must base their decisions regarding new versions of digital services based on collected data (King et al, 2017; Maalej, Nayebi, Johann, & Ruhe, 2015). The development teams should be able to consider the requirements of the mass users when deciding on what needs to be developed (Spiess et al., 2014; Maalej et al., 2015). By adopting the user-oriented and data-driven approach - user experience should be improved. In general, the cycle of developing a digital service should never end.

To support the development of our framework, we need to inspect the data that is to be used in the process. It can come from different sources, external (tools used for passive tracking) or internal (server logs produced by users and by system itself). Despite the several divisions of objective data types, authors agree on two main types: passive (implicit) and active (explicit) data (Liikkanen, 2016; Maalej et al., 2015). Difference is that passive data comes from passive tracking such as session recordings of server logs and active data is collected via surveys, chats etc. (Liikkanen, 2016; Maalej et al., 2015). Authors (Rodden et al, 2010) also bring up types of metrics: (a) PULSE metrics (number of visits, activities per visitor etc.) and (b) HEART metrics: (1) Happiness – aesthetics, ease of use (2) Engagement – frequency of using, (3) Adoption – number of unique users (4) Retention – giving up using the service and (5) Task success – efficiency, effectiveness or error rate. Based on this, three types of (relevant) objective data (passive) are inspected further in the paper:

- Server logs – can contain all relevant information about former and current state of the system (digital service); there are of course many types of logs, depending on the defined settings.
- Visitors' metrics – most likely it is data collected via external sources such as Google Analytics.
- Visual metrics – heat maps and click maps.

The paper continues with a brief theoretical background addressing the methods relevant to data-driven approaches to support our study on the use of objective over subjective data in the digital services redesign projects. After the brief outline in section 2, the specific objectives and method is presented in section 3. The results of the survey and A/B experiment are presented in section 4 while the results are discussed and implications and limitations of the study are offered in section 5 and concluded in section 6.

2 Theoretical bases for the study

2.1 SUS and TAM as subjective measures

Standard ISO 9241-11 defines usability as an "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO, 1998). Usability is one of the main factors that influences the increase in the level of use of digital services (Huang & Benyoucef, 2014). The research has confirmed that end users prefer a website with a higher usability rating, although it is important to keep in mind that user's usability and design requirements depend on the type of digital service, users themselves, and the very purpose (Ilbahar & Cebi, 2017). As can be expected, some objective features of digital services such as loading speed, enhances user experience, and even speed perception itself has a positive impact on user experience (King et al, 2017). If for an example the task requires a large number of clicks, it is very likely that users will perceive that digital service as a complex and less usable (Venkatesh, Chan, & Thong, 2012).

Even though, there are a number of methods and theories in the literature for understanding, predicting, and assessing personal factors, behaviour, and the environment while interacting with software, the two most popular are System Usability Scale and Technology Acceptance Model (Harrati, Bouchrika, Tari, & Ladjailia, 2016). The System Usability Scale (SUS) is a well-researched and widely used questionnaire for assessing the usability of mostly web applications. It is considered that SUS is the simplest method which achieves the most reliable results according to sample size. With the number of 8 respondents used in SUS method, the expected accuracy of the results is over 75%, with the higher number of respondents the reliability increases, and the relevant conclusion can be deduced from the survey even if the sample size is 8 to 12 respondents (Brooke, 1996, 2013). The method was developed at DEC in 1986. It is a Likert scale where the respondent indicates his/her agreement or disagreement with the statement. By analysing each questionnaire one gets a result in the range from 0 to 100 – an indicator of the overall usability of the system being observed (ibid). Due to the lack of measurement for technology usage and acceptance, the TAM - Technology Acceptance Model has been developed in 1989 (Davis, 1989). To assess the user acceptance for technological products, it is one of the most well established models (Harrati et al, 2016). The first version of the model consisted of two variables that affect the acceptance or use of technology; the two variables are perceived usefulness and perceived ease of use. TAM has quickly become a dominant model for researching factors that affect user acceptance of technology (Marangunić & Granić, 2015). Variables of TAM model are also measured via a Likert scale.

2.2 Using server-side data metrics

Many authors explored the potential of server-side data-driven approaches in different contexts some of which are presented in table 1. As mentioned before, there are several types of server logs, which can be predefined and very useful. In addition to server logs, time metrics and error logs stand out as valuable data.

Table 1. Server-side data metrics

Server logs	Andrica & Candea, 2011; Garrido, Firmenich, Grigera, & Rossi, 2017; Gordillo, Barra, Aguirre, & Quemada, 2014; Grigera, Garrido, Rivero, & Rossi, 2017a; Harrati, Bouchrika, Tari, & Ladjailia, 2015; Inversini, Cantoni, & Bolchini, 2011; Maalej et al., 2015; Rodden et al., 2010a; Rodriguez, 2002
Time metrics	(Grigera et al, 2017)
Error rate	Au, Baker, Warren, & Dobbie, 2008; Rodden et al., 2010a

Regarding the external data sources, such as Google analytics, there are also several types of metrics, which can be used for this purpose. In one of our previous researches, we used only data from Google Analytics to develop a data-driven web persona (Mijač, Jadrić, & Ćukušić, 2018). Typical metrics are listed in table 2.

Table 2. Google Analytics metrics

High organic click-through rates for keyword(s)	(Bakaev, Bakaev, & Mamysheva, 2016; Lee et al, 2016; Rodden et al, 2010)
Bounce rate	(Bakaev et al, 2016; Lee et al, 2016)
Visitor traffic	(Bakaev et al, 2016; Lee et al, 2016; Rodden et al, 2010)
User activity	(Bakaev et al, 2016; Lee et al, 2016; Rodden et al, 2010)

2.3 Mouse metrics

Even though mouse metrics can also be obtained through server logs, these are usually separated from typical data server logs as it could be collected by using special software tools such as Mousotron. It is a mouse and keyboard activity monitor which enables tracking of different activities performed using a computer mouse, such as total number of left, right and double clicks, speed achieved and so on. (Blacksunsoftware, 2018). Table 3 below lists most common mouse metrics, used by different authors in their papers.

Table 3. Mouse metrics

Mouse clicks	(Andrica & Candea, 2011; Frantz, 2018; Garcia & Paiva, 2016; Harrati et al, 2015; Oertel & Hein, 2003)
Amount of scrolling and speed	(Au et al, 2008)

3 Research method

3.1 Research instrument and procedure

Drawing on the potential of generated usage data (presented in section 1) and based on the typical methods and metrics (presented in section 2), we devised a research study in order to substantiate the use of objective over subjective data in the digital services redesign projects. To be more precise, the purpose of the study was to investigate the sensitivity of objective and subjective variables to changes in design. An experiment using the A/B testing method was first performed. A/B testing is essentially an online experiment used when changes are made to a product/service to measure the effect (King et al, 2017; Lee et al, 2016; Rodden et al, 2010). In order to evaluate different versions of a digital product, in addition to the original website (design A), another one was made (design B). An experiment was then conducted with two different groups of participants (half of the classroom group A and the other half group B). With A/B testing it was possible to quantify the results and to compare the two versions of the design.

Besides automatically collecting usage data for two designs, data was also obtained using an online survey tool. The participants (second year undergraduate students of business studies) completed the questionnaires voluntarily and anonymously after the compulsory classes finished so as not to impact the results of the study. The exercise was conducted in a proctored environment, i.e. in computer labs and under supervision of a teaching fellow. The students were instructed to access the link to the online questionnaire, which was placed on the official e-learning website of the Faculty. They were instructed to close any other programs running on their computers. They were given enough time to complete the questionnaires finishing in approximately 30 minutes.

The questionnaire was divided in 5 parts. *Part 1* was designed to collect participants' demographic data and the information on whether users purchased online so far as well as an estimation of their own abilities in using web technologies. *Part 2* (Scenario) contained the link to the web page, (depending on a group the hyperlink led either to the design A or the design B) complemented with the instructions for executing four different tasks on the website. Participants received instructions to run the Mousotron software before starting each task and pause it after finishing it. After finishing each task, they also had to take the screenshot of Mousotron metrics and upload the picture to the foreseen place in part 2 of the survey. After performing each task, participants were required to evaluate the weight of the task as well as to evaluate time spent in comparison with their experience that they had in performing similar activities. *Part 3* contained the standard System Usability Scale questionnaire with 10 questions given to participants after briefly getting to know the system.

Part 4 contained Technology Acceptance Model Scale with nine Likert-type questions measuring: (1) perceived usefulness, (2) perceived ease of use as well as (3) intention for future usage. Three statements were allocated to each part of TAM. We used existing multi-item scales, adapted to suit the context of the study. Rating system was the same as in SUS, based on ratings 1 to 5 that correspond to the level of agreement with the two extremes: "Strongly agree" and "Strongly disagree". *Part 5* contained general questions regarding the overall satisfaction with the evaluated website.

Figure 1 illustrates the research procedure. Total number of the participants at the beginning was 161, but after allocating them to two different groups, some of the students did not access the survey as it was voluntary. Since the first task in the scenario was to register, to pick a random user name, and to enter the user name in the survey – we were able to match some survey data with the server-side data collected automatically. After excluding the uncompleted surveys (the reasons for leaving the survey remain unknown) and the ones that could not be matched with automatically collected data, the total number of participants in group A was 22 and in group B 36.

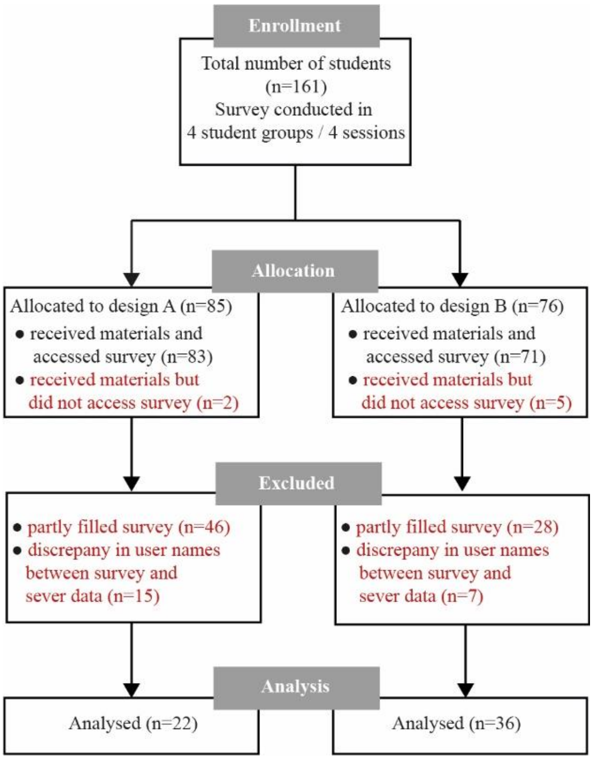

Figure 1. Research procedure with the number of participants (students)

3.2 Website and tasks for the A/B testing

Website we used for the study is from a Croatian start-up company selling natural cosmetics. The original website design is design A. Note that it was designed as a large number of other websites, without consulting the end-users. The website provides a portfolio of products and it has a functional webshop from where the users can buy available products. Additional functionality is the possibility register as a user which is useful if a user is purchasing the products often as they can earn additional discounts. The Figure 2 demonstrates the design A as a very clean design with minimalistic icons. Owners of the website would describe it as "trendy, plain and hipster". In order to make a fair comparison, the design B was built using the same text, the same pictures and the same functionalities, and the only thing that changed was the design. Design B was not plain or minimalistic; at the top of the website all the functionalities were listed. In general, the process of buying products required fewer clicks compared to the original design A. For the purpose of creating the alternative design, end-users were not consulted as well. As mentioned before, the scenarios contained four tasks in total and differed only with respect to the website design. In designing the websites and the tasks it was important to follow the well-established criteria for this kind of experimentation, i.e. the content in both scenarios had to be identical – except for the representational format used (website design); the representations used in both scenarios had to be equivalent in terms of the conveyed information (they had to be "informationally equivalent") and the required time to process these two scenarios had to be equivalent – without time limits.

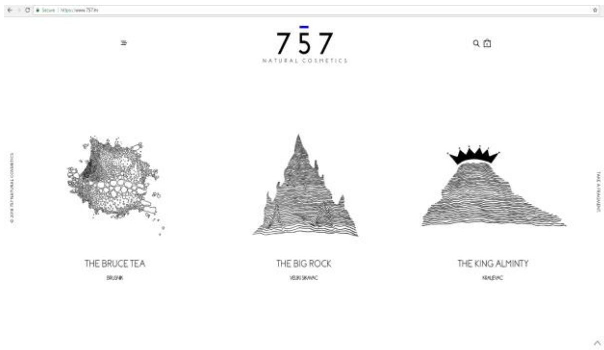

Figure 2. Design A of the website

Figure 3. Design B of the website

The set of four tasks which participants needed to perform are listed and described in Table 4.

Table 4. Task assigned to participants

Task name	Description
Registration	Find where you can register and complete the process. Please select a random username (e.g. NIVI89). For the purpose of completing the registration procedure, you do not need to enter a valid e-mail address. Please enter your username in the allocated field. When you finish the registration, close/leave the website.
Language change	Change the language of the website. When you change the language, close/leave the website.
Information	Find the information about the company and copy/paste it into the allocated field below.
Shopping cart	Find the soap "The Sea Heart – Galešnjak, 60g" and 'buy' it. Before selecting the payment type, take the screenshot of the shopping cart content, save and upload the picture.

3.2 Research participants

The participants of the study were the second-year undergraduate students of the University in Split, Faculty of Economics, Business and Tourism. General statistics of the sample population is presented in Table 5. The participants are from a relatively homogeneous group and mostly share a similar background in terms of education, economic situation etc. Based on the gender and age of the participants it is considered that the sample is representative when compared to the total population and between two groups. Majority of faculty's students are female (almost 80%) and with regards to age from 19 to 22 years old (around 97%).

Table 5. General statistics

		Group A (N=22)		Group B (N=36)		Total (N=58)	
		N	%	N	%	N	%
Gender	Female	19	86,4	27	75	46	79,3
	Male	3	13,6	9	25	12	20,7
Age	19-20	18	81,8	22	61,1	40	69,0
	21-22	3	13,6	13	36,1	16	27,6
	23 and more	1	4,5	4	11,1	2	3,5
Online buying	Yes	17	77,3	33	91,7	50	86,2
	No	5	22,7	3	8,3	8	13,8

4 Research results

As reported, the aim was to investigate whether there are significant differences between subjective (SUS and TAM) and objective data (server side data and mouse data) between groups A and B. The experiment in effect demonstrates the extent to which the objective data is sensitive to changes in design. The results are presented and interpreted hereinafter.

4.1 SUS and TAM results

Results of the SUS survey are presented in Table 6. SUS score does not represent a percentage; a mean score of 68 would represent a grade "C" and anything below a score of 51 is an "F" (putting a website in the bottom 15%) (Mclellan, Muddimer, & Peres, 2012). Following this interpretation, Group A (or design A) got an unacceptable usability score and design B got an acceptable usability score. The Mann-Whitney U test confirms that there is a statistically significant difference between the results of the two groups.

Table 6. System Usability Scale (SUS) results

	Group A (N=22)	Group B (N=36)	Total (N=58)
Mean	**50,7955**	**79,9306**	68,8793
Median	51,2500	80,0000	72,5000
St. dev.	15,34	11,60	19,30
Mode	50,00	75,00*	50,00
Minimum	15,00	50,00	15,00
Maximum	77,50	97,50	97,50
Mean Rank	13,64	39,19	
		Mann-Whitney U	745,00
		Asymptotic Sig. (2-sided test)	0,000

In analysing TAM results, first internal consistency of the scales of the questionnaire was analysed for both groups to determine the internal consistency. Cronbach alpha coefficient results are presented in Table 7. According to the boundaries for the Cronbach alpha coefficients (0.90/excellent, 0.80/very good and 0.70/ satisfactory), all TAM variables (from 0.875 to 0.922) are considered acceptable for further analysis. Results in table 8 demonstrate that all TAM variables significantly differ between the two groups.

Table 7. Cronbach alpha coefficients

TAM factors	Cronbach's Alpha (N=58)
Ease of use	0,922
Usefulness	0,875
Intention to use	0,919

Table 8. Descriptive statistics for TAM

	Group	Mean	Median	Mode	Min	Max	Mean Rank	p
Ease of use	A	3,15	3,00	3,00	1,00	5,00	15,32	**0,000**
	B	4,60	5,00	5,00	2,00	5,00	38,17	
Usefulness	A	3,29	3,33	3,00	1,00	5,00	19,82	**0,001**
	B	4,26	4,50	5,00	2,67	5,00	35,42	
Intention to use	A	2,30	2,67	1,00	1,00	4,00	20,66	**0,002**
	B	3,38	3,17	3,00	1,33	5,00	39,90	

After performing each task, the part 2 of the questionnaire required from the participants to evaluate the weight of the task execution as well as to evaluate the time spent in comparison with the experience they had performing similar activities. The results of the Mann-Whitney U test for group A and group B showed statistical difference for each task (the highest p value among differences for each task was 0,039). All the answers from group A were ranked lower than the results from group B. The results regarding weight of the first task (registration) in group A showed mean rank 23,73 and for group B 33,03; as for the time spent – group A has mean rank of 21,23 and group B of 34,56. Second task in group A has mean rank 24,02 and group B 32,85, while for the time spent the group A has mean rank of 22,86 and group B of 33,56. For the third task, both the results for evaluating the weight (A=23,73 and B=33,03) and the time (A= 22,82 and B=33,58) also point that the results for group A have lower mean rank than for group B. Results for the fourth task are also consistent in that regard since mean rank for evaluating the weight of the task is 21,14 for group A and 34,61 for group B, as well as the results regarding time spent (A=21,11 and B=34,62).

4.2 Server side data results

Server side data was data collected automatically while participants interacted with the websites. Even though there was a huge amount of data, for the purpose of this research, several metrics were used:

- visit duration in seconds – is refers to the total amount of time spent on the website,
- number of searches – number of times when participants used the "search" option on the website,
- number of actions – action is every page participants' visit,
- number of extra actions – this metrics was calculated as the difference between the "number of actions" and the "number of actions spent for finishing a task"; therefore, it represents a number of extra actions which a participant performed, probably in order to get to know the website,
- extra time needed.

Results presented in Table 9 include only the metrics with statistically significant difference between the group A and group B. Statistical data analysis was also done at the level of each task, but there were no significant differences due to a small sample.

Table 9. Server side data results

Metrics	Group	Mean	Median	Mean Rank	U	p
Searches	A	0,50	0,00	33,86	300,00	**0,007**
	B	0,06	0,00	26,83		
Extra actions	A	7,54	7,00	35,73	259,00	**0,027**
	B	4,89	4,50	25,69		
Extra time	A	207,45	165,00	37,34	223,50	**0,006**
	B	93,14	86,50	24,71		

It should be noted that even though Google Analytics was set up, due to the controlled conditions of the survey, Google Analytics did not provide any other useful data (apart from number of visits, and time stamp of visits, all consistent with the number of participants and survey feedback).

4.3 Mouse metrics results

Mouse metrics were also collected automatically using the Mousotron tool. It enabled collecting several types of metrics: (1) keystroke, (2) left button, (3) right button, (4) double clicks, (5) mouse wheel, (6) speed, (7) seconds, (8) idle seconds and (9) centimetres. Since the participants needed to upload the results/collected statistics after each task, it was possible to process the data on the level of each task. For the first task (registration) the results of mouse metrics, which represent statically significant difference ($p < .05$), are presented in Table 10. For the second task (language change) only for the 'mouse wheel' movement there was statically significant difference ($p = .004$). For the third task, the collected mouse metrics differ significantly between groups for 'keystroke', 'mouse wheel' and 'centimetres'. Results are presented in Table 11. An examination of the findings of the mouse metrics referring to the last task (buying the product) revealed that average ranks for each metrics are not significantly different.

Table 10. Mouse metrics for the first task

Task Registration	group	Mean	Median	Mode	Min	Max	Mean Rank	U	p
Keystroke	A	148,75	145,50	20,00	20,00	257,00	34,75	235,00	0,033
	B	115,97	114,00	98,00	0,00	244,00	25,03		
Left button	A	66,45	64,00	65,00	15,00	168,00	38,60	158,00	0,001
	B	40,19	39,00	23,00	0,00	99,00	22,89		
Mouse wheel	A	41,20	31,50	0,00	0,00	120,00	34,32	243,50	0,045
	B	24,08	13,00	0,00	0,00	133,00	25,26		
Seconds	A	416,15	410,00	51,00	51,00	667,00	39,10	148,00	0,00
	B	250,47	237,50	8,00	8,00	640,00	22,61		
Idle Seconds	A	62,90	58,00	0,00	0,00	155,00	35,12	227,50	0,022
	B	26,53	22,00	0,00	0,00	122,00	24,82		
Centimeters	A	2159,09	1945,40	2038,94	761,48	4694,00	38,63	140,00	0,000
	B	1260,25	1088,75	40,00	40,00	3900,00	22,39		

Table 11. Mouse metrics for the third task

Information	group	Mean	Median	Mode	Min	Max	Mean Rank	U	p
Keystroke	A	26,90	4,00	0,00	0,00	301,00	35,05	251,00	0,036
	B	16,78	3,00	0,00	0,00	202,00	25,47		
Mouse wheel	A	139,86	92,00	0,00	0,00	527,00	40,00	147,00	0,000
	B	27,64	14,50	12,00	0,00	156,00	22,58		
Centimeters	A	1952,19	767,00	212,48	212,48	9186,36	35,05	251,00	0,036
	B	967,94	496,00	496,00	239,00	6888,50	25,47		

5 Discussion

The general idea of this paper was to examine the potential of data driven approach for (re)designing the digital services. Although there are a number of limitations to this research, especially the number of participants, restricted conditions and not using all the available objective data – the research reveals some interesting results.

Our main question was whether objective data are sensible to changes in design, and the general answer is positive. There was significant statistical difference in the subjective and objective data results. The results are consistent with theory implications that, if something takes more actions or requires more clicking and scrolling, the subjective opinion of users is worse. Even though the two provided designs kept the same functionalities, the results demonstrate the difference in both objective and subjective data.

Both **subjective methods** (SUS and TAM) show that users prefer design B to design A. The results for the overall website, they were not done at the level of each task. Results demonstrate that users' evaluation of the weights of performing each task as well as the time spent on each task, statistically differs for each task in favour of design B.

As mentioned, one of the restrictions of this research is the number of participants, consequently after analysing the results of **server side data** for each task of design A and design B – it turns out that there was no statistically significant difference in server side data logs. However, by analysing the results for four tasks altogether, for example, total number of "extra time", the results were statistically different between design A and design B. Server side data was consistent with above-mentioned results. After analysing the data from Mousotron software, the results also point to statistically different results at the level of each task (task 1, task 2 and task 3). Results of **mouse metrics** were therefore also consistent with the above results.

In the subsequent phases, the framework we mention throughout the paper would be further developed and validated through pilot studies such as this one.

6 Conclusion

To conclude, even though without actually asking the end users how they feel about a digital service, or whether there is something that could be improved – by using and following the data-driven approach enough data could be obtained to use it for (re)designing the digital services. This could be very practical when there is huge amount of users with different backgrounds and in the conditions where it could be hard to collect user specifications.

This approach should not be used in isolation, it can be helpful to provide additional information so the development teams can predict possible issues without waiting for formal users' feedback, as mentioned in proposed methodological framework for user-oriented data-driven information systems modelling that is under development by the authors of the paper.

Acknowledgment

This work has been supported by the Croatian Science Foundation under the project UIP-2017-05-7625: User-oriented process (re)design and information systems modelling – a case of smart city services.

References

Anderson, C. (2015). *Creating a Data-Drivem Organization*. Sebastopol, Ca, SAD: O'Reilly.

Andrica, S., & Candea, G. (2011). WaRR: A tool for high-fidelity web application record and replay. *Proceedings of the International Conference on Dependable Systems and Networks*, 403–410. https://doi.org/10.1109/DSN.2011.5958253

Au, F. T. W., Baker, S., Warren, I., & Dobbie, G. (2008). Automated Usability Testing Framework. *Proceedings of the Ninth Conference on Australasian User Interface - Volume 76*, (January), 55–64. Preuzeto od http://dl.acm.org/citation.cfm?id=1378337.1378349

Bakaev, M., Bakaev, M., & Mamysheva, T. (2016). Current Trends in Automating Usability Evaluation of Websites Can You Manage What You Can ' t Measure ? Current Trends in Automating Usability Evaluation of Websites. *Ieee*, (August), 510–514.

Bao, Y., Wu, H., & Liu, X. (2017). From Prediction to Action: Improving User Experience with Data-Driven Resource Allocation. *IEEE Journal on Selected Areas in Communications*, *35*(5), 1062–1075. https://doi.org/10.1109/JSAC.2017.2680918

Blacksunsoftware. (2018). Blacksunsoftware. Preuzeto od http://www.blacksunsoftware.com/

Brooke, J. (1996). *SUS - A quick and dirty usability scale*. London, UK: Taylor and Francis.

Brooke, J. (2013). SUS : A Retrospective. *Journal of Usability Studies*, *8*(2), 29–40. https://doi.org/10.1074/jbc.R115.675280

Davis, F. D. (1989). Perceived Usefulness , Perceived Ease of Use , and User Acceptance of. *MIS Quarterly*, *13*(3), 319–340.

Frantz, T. L. (2018). Blockmap: an interactive visualization tool for big-data networks. *Computational and Mathematical Organization Theory*, *24*(2), 149–168. https://doi.org/10.1007/s10588-017-9252-6

Garcia, J. E. . c, & Paiva, A. C. R. . c c. (2016). An automated approach for requirements specification maintenance. *Advances in Intelligent Systems and Computing*, *444*, 827–833. https://doi.org/10.1007/978-3-319-31232-3_78

Garrido, A., Firmenich, S., Grigera, J., & Rossi, G. (2017).

Data-driven usability refactoring: Tools and challenges. *SoftwareMining 2017 - Proceedings of the 2017 6th IEEE/ACM International Workshop on Software Mining, co-located with ASE 2017*, (October).https://doi.org/10.1109/SOFTWAREMINING.2017.8100854

Geogy, M., & Dharani, A. (2016). A Scrutiny of the Software Requirement Engineering Process. *Procedia Technology*, *25*, 405–410. https://doi.org/10.1016/j.protcy.2016.08.125

Gordillo, A., Barra, E., Aguirre, S., & Quemada, J. (2014). The usefulness of usability and user experience evaluation methods on an e-Learning platform development from a developer's perspective: A case study. *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*, 1–8. https://doi.org/10.1109/FIE.2014.7044340

Grigera, J., Garrido, A., Rivero, J. M., & Rossi, G. (2017). Automatic detection of usability smells in web applications. *International Journal of Human Computer Studies*, *97*(October 2017), 129–148. https://doi.org/10.1016/j.ijhcs.2016.09.009

Halvorsrud, R., Kvale, K., & Følstad, A. (2016). Improving service quality through customer journey analysis. *Journal of Service Theory and Practice*, *26*(6), 840–867. https://doi.org/10.1108/JSTP-05-2015-0111

Harrati, N., Bouchrika, I., Tari, A., & Ladjailia, A. (2015). Automating the evaluation of usability remotely for web applications via a model-based approach. *NTIC 2015 - 2015 1st International Conference on New Technologies of Information and Communication, Proceeding*. https://doi.org/10.1109/NTIC.2015.7368757

Harrati, N., Bouchrika, I., Tari, A., & Ladjailia, A. (2016). Exploring user satisfaction for e-learning systems via usage-based metrics and system usability scale analysis. *Computers in Human Behavior*, *61*, 463–471. https://doi.org/10.1016/j.chb.2016.03.051

Havice, J. (2017). How to Create Customer Personas with Actual, Real Life Data. Preuzeto od http://bit.ly/2FrhRdx

Huang, Z., & Benyoucef, M. (2014). Usability and credibility of e-government websites. *Government Information Quarterly*, *31*(4), 584–595. https://doi.org/10.1016/j.giq.2014.07.002

Ilbahar, E., & Cebi, S. (2017). Classification of design parameters for E-commerce websites: A novel fuzzy Kano approach. *Telematics and Informatics*. https://doi.org/10.1016/j.tele.2017.09.004

Inversini, A., Cantoni, L., & Bolchini, D. (2011). Design, User Experience, and Usability. Theory, Methods, Tools and Practice, *6770*(May 2014). https://doi.org/10.1007/978-3-642-21708-1

ISO. (1998). ISO 9241-11:1998. Preuzeto od https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-1:v1:en

King, R., Churchill, E. F., & Tan, C. (2017). *Designing with Data*. O'Reilly.

Kujala, S., Kauppinen, M., & Rekola, S. (2001). Bridging the Gap between User Needs and User Requirements. *Advances in Human-Computer Interaction I (Proceedings of the Panhellenic Conference with International Participation in Human-Computer Interaction PC-HCI 2001)*, (February), 45–50.

Lee, Y. Y., Smith, A., Calvert, L., & Snajdr, E. (2016). Innovative Data-Driven Methods to Improve Digital User Experience. *Qualitative & Quantitative Methods in Libraries*, *5*(2), 461. Preuzeto od http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,cookie,url,cpid,uid&custid=s8863137&db=edb&AN=120969018&site=eds-live&scope=site&authtype=ip,uid

Lemon, K. N., & Verhoef, P. C. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing*, *80*(6), 69–96. https://doi.org/10.1509/jm.15.0420

Maalej, W., Nayebi, M., Johann, T., & Ruhe, G. (2015). Towards Data - Driven Requirements Engineering, *33*, 1–6. https://doi.org/10.1109/MS.2015.153

Marangunić, N., & Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal Access in the Information Society*, *14*(1), 81–95. https://doi.org/10.1007/s10209-014-0348-1

Mclellan, S., Muddimer, A., & Peres, S. C. (2012). The Effect of Experience on System Usability Scale Ratings. *Journal of Usability Studies*, *7*(2), 56–67.

Mijač, T., Jadrić, M., & Ćukušić, M. (2018). The Potential and Issues in Data-Driven Development of Web Personas. U *MIPRO 2018* (str. 1427–1432).

Oertel, K., & Hein, O. (2003). Identification of Web usability problems and interaction patterns with the RealEYES-iAnalyzer. *Interactive Systems Design, Specification, and Verification 10th International Workshop, DSV IS 2003 Revised Papers Lecture Notes in Comput Sci Vol 2844*, *2844*, 77–91. Preuzeto s https://www.scopus.com/inward/record.uri?eid=2-s2.0-0142184064&partnerID=40&md5=62bbbc0cdfe9cbbd7cbd4f86da6a4cde

Rodden, K., Hutchinson, H., & Fu, X. (2010). Measuring the User Experience on a Large Scale : User-Centered Metrics for Web Applications. *SIGCHI Conference on Human Factors in Computing Systems*, 2395–2398. https://doi.org/10.1145/1753326.1753687

Rodriguez, M. G. (2002). Automatic data-gathering agents for remote navigability testing. *IEEE Software*, *19*(6), 78–85. https://doi.org/10.1109/MS.2002.1049396

Sengers, P., Boehner, K., Mateas, M., & Gay, G. (2008). The disenchantment of affect. *Personal and Ubiquitous Computing*, *12*(5), 347–358. https://doi.org/10.1007/s00779-007-0161-4

Spiess, J., Joens, Y. T., Dragnea, R., & Spencer, P. (2014). Using Big Data to Improve Customer Experience and Business Performance. *Bell Labs Technical Journal*, *18*(4), 3–17.

Tomitsch, M. (2018). *Making cities smarter*. Berlin: Jovis.

Venkatesh, V., Chan, F. K. Y., & Thong, J. Y. L. (2012). Designing e-government services: Key service attributes and citizens' preference structures. *Journal of Operations Management*, *30*(1–2), 116–133. https://doi.org/10.1016/j.jom.2011.10.001