



D6.8

Human factors of MOBiNET

Work package: 6
Version number: 3.0
Dissemination level: PU
Date: 25/11/2016



7th RTD Framework Programme
Directorate General for Communications Networks, Content & Technology
Cooperative Systems for energy efficient and sustainable mobility (FP7-ICT-2011-6.7)
Contract Type: Collaborative project
Grant agreement no.: 318485

Version control

Version history			
Version	Date	Main author	Summary of changes
0.1	15/06/2015	Mandy Dotzauer	Table of Contents and main structure of the document.
0.2	20/11/2015	Mandy Dotzauer	First draft of the method section
0.3	02/12/2015	Mandy Dotzauer	First draft of result section
0.4	03/12/2015	Mandy Dotzauer	Finishing method section, first draft of the introduction
0.5	13/01/2016	Mandy Dotzauer	Including Release 3 results
0.6	15/01/2016	Mandy Dotzauer	Finishing introduction, first draft of discussion and conclusion, adding references and Annex 1, formatting the document
0.7	18/01/2016	Mohamed Mahmod	feedback
0.7	19/01/2016	Mandy Dotzauer	Incorporating feedback
0.75	26/1/2016	Olli Pihlajamaa	MOBiNET Introduction (subsection 1.1)
0.8	01/02/2016	Mandy Dotzauer	Revision of introduction (subsection 1.1)
1.0	04/02/2016	Mandy Dotzauer	Completion of executive summary, final formatting
2.0	24/02/2016	Mandy Dotzauer	Incorporating Feedback
2.1	25/10/2016	Mandy Dotzauer	Revision of the method section, incorporating information of Release 3.1; revision of the result section; changing analysis from parametric to nonparametric analysis
2.1	08/11/2016	Juho Kostianen, Guillem Bernatallada, Caroline Schießl	Peer-reviews
2.2	14/11/2016	Mandy Dotzauer	Incorporating feedback
3.0	25/11/2016	Mandy Dotzauer	Final revision after quality review
	Name		Date
Prepared	Mandy Dotzauer		25/10/2016
Reviewed	Juho Kostianen (VTT), Guillem Bernatallada (ACASA), Yanying Li (ERTICO)		17/11/2016

Authorised	Rasmus Lindholm	28/11/2016
Circulation		
Recipient	Date of submission	
European Commission	28/11/2016	
Project partners	28/11/2016	

Authors

Mandy Dotzauer (DLR), Olli Pihlajamaa (VTT)

Table of contents

Table of figures	6
Abbreviations and definitions	7
Executive summary	8
1. Introduction	9
1.1. MOBiNET platform.....	9
1.2. Objectives and approach	10
1.3. Scope and limitations.....	11
1.4. Structure of the document	12
2. Method	13
2.1. Participants	13
2.1.1. Assessment from a developer/service provider perspective	13
2.1.2. Assessment from an end user perspective	14
2.2. Questionnaires	14
2.2.1. System usability scale (SUS)	14
2.2.2. Acceptance scale	15
2.3. Study design and procedure	15
2.4. Analysis	15
3. Results	17
3.1. Service provider/developer evaluation.....	17
3.1.1. System usability scale	17
3.1.2. Acceptance scale	17
3.2. End user evaluation	18
3.2.1. System usability scale	18
3.2.2. Acceptance Scale.....	19
3.3. Changes in ratings over time	19
3.3.1. System usability scale	19
3.3.2. Acceptance scale	20
3.4. Hackathon	20
4. Discussion	22
4.1. Implications of usability ratings	22
4.2. Implications of acceptance ratings	22
4.3. Implications of the Hackathon survey	23

5. Conclusion/ Impact assessment	24
References	25
Annex 1	26

Table of figures

Figure 1: Overview of the main MOBiNET components.	10
Figure 2: Average scores of the System usability scale across all releases. Error bars present the standard deviation. Red line indicates the target score of 68. Left: Scores of developers/service providers. Right: Scores of end users.	20
Figure 3: Average scores of the acceptance ratings of all releases. Error bars present the standard deviation. Left: Acceptance scores of developers/service providers. Right: Acceptance scores of end users.	20

Abbreviations and definitions

Abbreviation	Definition
ANOVA	Analysis of variance
API	Application programming interface
B2B	Business-to-Business
B2C	Business-to-Customer
df	Degrees of freedom
IA	Impact assessment
IDE	Integrated development environment
ITS	Intelligent transport systems
M	Mean
ns	Not significant
SD	Standard deviation
SDK	Service development kit
SUS	System Usability Scale
p	Probability

Executive summary

In this document, results of the work in Task 6.2, Human factor assessment of MOBiNET, are presented. The main objective of this deliverable is to provide in-depth information on the rationale, the procedure, and the results of the human factor assessment. In the end, implications of the results are also given as well as limitations of the results.

The objective of document D6.8 is to show how acceptance and usability ratings are used to assess the impact of the MOBiNET platform with regard to human factors. In terms of human factors, quality of use and fit to context as well as the value of the service/application sought by the user are assessed. These aspects can be evaluated when assessing the acceptance and usability ratings of users. In turn, those ratings can be used to project the success of the MOBiNET platform with regard to human factors.

Developers/service providers and end users rated their acceptance and the usability of the platforms. The group of developers/service providers (except of the Hackathon participants) consisted of members of the MOBiNET consortium operating a pilot site. The group of end users also consisted of consortium members who took on the role of the end user. The number of participants varied across releases. Between 9 and 17 developers/service providers completed the questionnaires, while 13 hackathon participants filled in the questionnaires. MOBiNET consortium members taking on the role of the end user varied between 3 and 8 across all releases.

A standardized usability questionnaire (System usability scale, SUS) and acceptance questionnaire (Van der Laan Acceptance scale) were used to assess the quality of the MOBiNET platform from a human factors point of view and to inquire usefulness and satisfaction ratings. The scores are used to make predictions of the success of the MOBiNET platform. The results will also reveal shortcomings in usability, usefulness, and satisfaction. In order to ensure comparability of the results the same questionnaires were used across all releases and the hackathon as well as for all participants (i.e. developers/service providers, hackathon participants, and end users).

In terms of usability, changes/improvements from one release to the next do not suggest statistically significant improvement of the usability neither for service developers/providers nor end users. Nonetheless, visual inspection of changes suggests improvements of the platform's usability. During the last releases, the overall scores increased and moved closer to the target value. It appears that the usability of the platform increased making it easier and more pleasant to use the platform.

In terms of acceptance, results reveal higher ratings for the last two releases. It appears that the platforms function and functionality improved over time and therefore ratings increased. In the beginning, ratings were around zero indicating that participants neither rejected nor accepted the platform. The positive change in ratings might reflect the improvements of the last release.

1. Introduction

1.1. MOBiNET platform

In the MOBiNET project, the MOBiNET platform is realized as an e-marketplace and a technical platform enabling the interactions between suppliers, developers and users of mobility-related content and services. It offers a centralized directory for publishing and finding both Business-to-Business (B2B) and Business-to-Consumer (B2C) services as well as functionality enabling and supporting interoperability between data sources and offering services across Europe without the need for standardized hardware. In this research project also several example services are examined to evaluate the usefulness of the platform and its concepts.

As shown below in Figure 1 the MOBiNET platform is directed towards different user groups. End users interact with the platform and the services via the MOBiAGENT on their mobile device or by accessing the Service Directory through a web interface. Service providers rely on the functionality of the Dashboard to manage their services published to MOBiNET and service developers will mainly use the MOBiNET Service Development Kit (SDK) together with the Dashboard and APIs of the different components.

For end users, MOBiNET's Service Directory provides access to services from different providers and sources. Through the platform, applications can be found for different smartphone ecosystems, web services, services related to the user's personal vehicle, etc. from one source. In addition to a website, the service directory is accessible through a user interface on the end user device, the MOBiAGENT. The MOBiAGENT provides multiple communication channels (cellular, ITS-G5) to be used to receive ITS-related service information. The implementation done during the project runs on Android and can be installed through the common App Stores.

Dashboard contains a rich set of functionality for service providers and other business users. A central functionality is related to the *Service Directory* containing descriptions of all services published on MOBiNET. Services can be registered providing a service description and easily updated and deleted. It is also possible to search for other services and aggregate those.

SDK is a set of tools, APIs and documentation that enable service developers to easily and effectively utilize MOBiNET platform functionalities for their services. Both MOBiCENTRE components and MOBiAGENT can be accessed programmatically through REST-APIs. In MOBiNET, the software part of the SDK is provided for Eclipse integrated development environment (IDE). In addition, *Service description editor* is realized as an Eclipse plugin. A standalone windows application has been available since the hackathon/ R3.0 in October 2015).

Because MOBiNET does not only address service developers or providers or the end user, but designs the platform for all those different stakeholders, it is important to test the functionalities tailored to fit the needs of different stakeholder groups. Therefore, while developing the product, the different stakeholder groups need to assess it. Their feedback is needed to improve the platform and to make MOBiNET a successful mobility platform not only for service developers and service providers, but also end users.

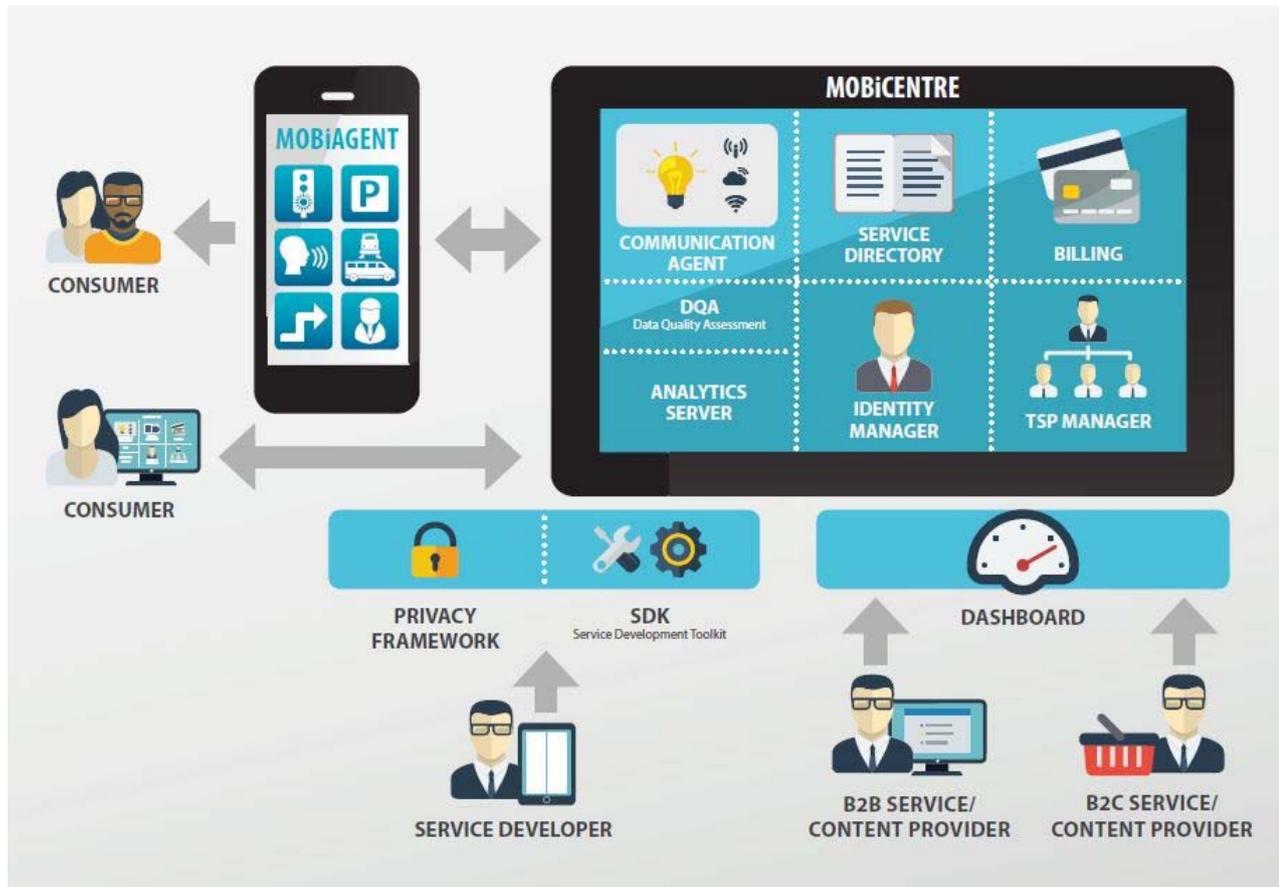


Figure 1: Overview of the main MOBiNET components.

1.2. Objectives and approach

From a human factors perspective, a user-centred as well as iterative design approach is indispensable not only to ensure the usability of a newly developed product/service, but also to heighten the probability of identifying design errors as well as technical deficits of the product. So that, in the end, the product will be accepted and used. While designing and developing, it is advisable to follow an iterative design approach meaning that from the early start, potential users are involved in assessing the product regarding design and functionality uncovering deficits. Results of an assessment need to be considered when further developing the product. After further developments and improvements, another user assessment needs to take place in order to examine changes and to ensure that fixes did not cause any new problems and further problems introduced by changes in design. This process continues until the product is finished and ready to be introduced into the market (Nielsen, 1993).

The goal of the human factors impact assessment is to use results obtained during validation to project the potential success of the MOBiNET platform. The human factors impact assessment is based on usability and acceptance ratings of the MOBiNET stakeholders. For analysis, usability and acceptance ratings were gathered during the validation process of the individual releases. Those ratings serve as indirect measures of impact because the usage of a product highly depends on the perceived usability of the individual as well as the acceptance of such a product. This concerns all stakeholders of MOBiNET. Services and application need to be easily published in MOBiNET. Data such as real traffic information need to be easily accessible, so successfully navigating through the Dashboard is crucial for developers and service providers. In addition, being able to find and download apps is crucial to win end users. The

applications also need to be designed to be user friendly to heighten the chance of being used. For each release of the MOBiNET platform (up to date: Releases 1, 2.1, 2.2, 3, and 3.1), acceptance and usability ratings were gathered by means of an online questionnaire (see Annex 1 for the survey) for project members using the platform as well as external hackathon participants at the 22nd ITS World Congress in Bordeaux.

Please note: The MOBiNET platform is a proof-of-concept and still in an early developmental stage. Therefore, the conclusions drawn here are neither final nor conclusive. They only reflect the potential success based on the current developmental stage.

1.3. Scope and limitations

In general, impact assessment is done after the introduction of, for example, a new law. Over a defined period of time, it is assessed whether or not passing a law has an effect (Blankenberg, 1995). In the case of MOBiNET, impact assessment should have been done after the platform has been fully developed, externally released, and used for a while. Because of the project-dependent time constraints, the general approach of impact assessment cannot be applied. Other methods had to be defined and used to project the potential of the platform based on an assessment from a human factors point of view.

Developing software or a tool is an iterative process that needs to go through a developmental cycle; therefore, it was decided to assess usability (quality of use) and acceptance (value sought by the user) for each release while doing the validation (for detailed information on the validation procedure, please see Deliverables D5.1.2 and D5.7.2). The reason behind this approach is that with each release, the platform's functioning (does it work) and functionality (what does it do) should be further developed and improved. These further developments and improvements should be reflected in the usability and acceptance scores. Changes in those scores might serve as an indicator of the success of the platform. In addition to collecting data from MOBiNET consortium members, an independent assessment of the platform is also necessary in order to collect unbiased data. As a first step, participants of Bordeaux's Hackathon (<http://itsworldcongress.com/programme/competitions>) filled in the questionnaire assessing the usability of the platform. (Please note: Validation results as well as providing further input to improve the MOBiNET platform and/or components of the platform are not subjects of this deliverables. For more information on this, please see Deliverable D5.7.2)

Except for the Hackathon, all ratings were completed by MOBiNET consortium members whose views may be biased. Data collected for assessing the human factors impact need to be considered with care. The sample size per release was very small. If the number of participants is very small, extreme responses gain more weight and skewed results in one direction. The more participants the more accurate the picture will be. It might be that validators responded more positive than a validator not associated with MOBiNET would do. Besides the fact that developers and service providers assessed how well a B2B service can be published, the same persons took on the role of the end user assessing end user functionalities. Here again, real or potential end user should have been involved in the validation process. Therefore, results presented here only serve as the first indication. It should be noted that MOBiNET serves as a proof-of-concept of an e-marketplace platform that provides a number of functionalities enabling service development and delivery, not intended to be commercially ready at the end of the project. Therefore, the usability assessment only reflects the potential of the MOBiNET platform at the momentary developmental stage. When further developments occur, the platform needs to be reassessed and results analysed again.

1.4. Structure of the document

The first part of the document provided an overview on what impact assessment is and why a human factors perspective cannot be neglected when assessing the potential of a new system. The following section (**Error! Reference source not found.**) describes the method of data collection. The method includes information on participants, study design, procedure as well as test material used and analysis performed. In **Error! Reference source not found.**, results of usability and acceptance ratings are presented. In **Error! Reference source not found.**, those results are interpreted and discussed in [Conclusion/](#) Impact assessment. This chapter will also include a section of shortcomings as well as suggestions to gather more valid data needed for more valid impact assessment in the future.

2. Method

2.1. Participants

Developers/service providers and end users rated their acceptance and the usability of the platforms. The group of developers/service providers (except of the Hackathon participants) consisted of members of the MOBiNET consortium operating a pilot site. The group of end users also consisted of consortium members who took on the role of the end user. Detailed information on these two groups is given below.

2.1.1. Assessment from a developer/service provider perspective

Table 1 shows the number of persons participating in the survey. As can be seen, the number varies across releases. The table also summarizes the overall experience with information technology, knowledge about the MOBiNET concept as well as experience with the MOBiNET tools (Note: Participants of the Hackathon did not answer the question about their experience with information technology). Participants had to rate their experience and knowledge on a 5-point Likert scale, where 1 = very low, 2 = low, 3 = moderate, 4 = high, 5 = very high. Results show that participants involved in the validation process are very familiar with information technology and the concept of MOBiNET. Their experience with MOBiNET tools increased from one release to the next. On the other hand, participants of the Hackathon were neither very familiar with the concept of MOBiNET nor the tools of MOBiNET.

Table 1: Overview of participant information (developer/service provider perspective)

	Release 1	Release 2.1	Release 2.2	Release 3	Release 3.1	Hackathon
Number of participants (N)	11	19	8	13	9	13
Mean age (SD)	37.1 (10.3)	40.6 (9.7)	37.4 (9.3)	42 (10.6)	38.1 (9.3)	29.6 (6.0)
Mean ratings of experience with information technology (SD)	4.45 (.68)	4.42 (.61)	4.38 (.52)	4.23 (.83)	4.56 (.53)	
Mean ratings of knowledge about the MOBiNET concept (SD)	3.91 (.7)	3.74 (1.1)	3.75 (1.2)	4.08 (.86)	4 (.71)	2.23 (1.4)
Mean ratings of experience with	2.55 (1.1)	2.74 (.93)	3.0 (1.1)	3.38 (.96)	3.56 (.73)	2.38 1.5)

MOBiNET tools (SD)						
--------------------	--	--	--	--	--	--

2.1.2. Assessment from an end user perspective

The table below provides an overview of the number of participants, the average age and their experience and knowledge filling in the questionnaires from an end user perspective for each of the releases. In general, participants indicated that they are very familiar with information technology. They also knew the MOBiNET concept and had some experience with MOBiNET tools. Please note: MOBiNET-related personnel filled in the questionnaires from an end user perspective.

Table 2: Overview of participant information (end user perspective)

	Release 1	Release 2.1	Release 2.2	Release 3	Release 3.1
Number of participants	3	5	8	8	8
Mean age (SD)	36.7 (14.2)	39.6 (1.5)	33.4 (6.9)	47.25 (10.4)	42.75 (9.8)
Mean ratings of experience with information technology	4.67 (.58)	4.6 (.55)	4.5 (.54)	4.38 (.74)	4.88 (.35)
Mean ratings of knowledge about the MOBiNET concept	4.0 (1.0)	3.8 (1.3)	4.4 (.74)	4.13 (.99)	3.88 (.64)
Mean ratings of experience with MOBiNET tools	2.67 (.58)	4.4 (.74)	3.3 (1.0)	3.5 (1.3)	3.63 (.71)

2.2. Questionnaires

A standardized usability questionnaire and acceptance questionnaire were used to assess the quality of the MOBiNET platform from a human factors point of view and to inquire usefulness and satisfaction ratings. The scores are used to make predictions of the success of the MOBiNET platform. The results will also reveal shortcomings in usability, usefulness, and satisfaction. In order to ensure comparability of the results the same questionnaires were used across all releases and the hackathon as well as for all participants (i.e. developers/service providers, hackathon participants, and end users).

2.2.1. System usability scale (SUS)

The System usability scale (SUS) was used to assess the overall usability of the MOBiNET platform. The SUS is a reliable tool for measuring usability meaning how easy and pleasant the product is to use. The overall score serves as an indication of how well the overall usability is. The scale contains ten statements, which are rated on a 5-point Likert scale. Participants indicate their degree of agreement

with each statement (e.g. I think that I would like to use this system frequently, please see Annex 1 for the complete questionnaire.) ranging from *strongly agree* to *strongly disagree*. It is an easy to use and fast to administer industrial standard (Brooke, 1996).

Scores are converted and can range from 0 to 100. Validation of the questionnaire and extensive analyses of scores of over 500 studies revealed that an average score of 68 can be considered as satisfactory usability. Anything below a score of 68 is considered below average indicating that improvements are needed. A score greater than 68 is considered as above average. Test data has been normalized and scores translated into US grades. A score of 68 translates into the grade C (Brooke, 1966).

2.2.2. Acceptance scale

The Van-der-Laan-acceptance scale assesses acceptance on two dimensions: usefulness and satisfaction. The scale consists of nine-Likert items (5-point scale) such as *useful-useless* or *nice-annoying*. Corresponding scores range from -2 to +2 indicating either rejection or acceptance of a product/system (please see Annex 1 for the complete questionnaire). Therefore, zero indicates neither rejection nor acceptance. Positive or negative deviance from zero serves as an indicator of how well the system is accepted. Scores for the subscales can develop in opposite directions. It is possible that a system is considered useful indicated by positive scores. But participants might not be satisfied with it as indicated by a negative score. Therefore, a system might be practical but using it might not be satisfying (Van der Laan, et al., 1997).

2.3. Study design and procedure

Questionnaires used in order to assess the impact from the human factors point of view were integrated in the validation questionnaires. Validation questionnaires were administered after each of the platform releases. Therefore, a within-subject design with repeated measures was used to assess the human factors impact. At the end of each validation process, validators were asked to rate their overall impression of the MOBiNET platform. The focus laid on assessing the usability of the platform as well as their acceptance of the platform. The results of these two questionnaires are used for the human factors impact assessment (i.e. projecting the potential of MOBiNET from a human factors perspective). The repeated measure design also offers the opportunity to identify changes over time (e.g. increases in ratings) indicating improvements of the platform

Validation was done from two points of views: developers/service providers and end users. In addition, during the Hackathon in Bordeaux, participants also filled in the questionnaire assessing the usability and their acceptance of the platform. Therefore, conclusion can be drawn from three perspectives:

- (1) MOBiNET project members taken on the perspective of a developer/service provider,
- (2) MOBiNET project members taken on the role of an end user,
- (3) MOBiNET-independent developers using the platform to integrate services

2.4. Analysis

Not all questionnaires were filled in completely and not all pilot sites completed all release validations (different services at the pilot sites), therefore the number of participants varies across releases.

Because the number of participants is low and varies across releases, analyses of the results were done using nonparametric test procedures. Nonparametric statistics make no assumptions about the probability distributions of variables; therefore, they are robust measures. In addition, non-parametric statistics are applied when the sample size is very low.

Results are analysed according to the different perspectives. For each group of interest usability ratings and acceptance ratings were analysed for each release. Wilcoxon signed rank test were used in order to test results against a hypothetical median. In order to assess the usability ratings, results were tested against a hypothetical median of 68. Of relevance is whether the average rank differs significantly from the test value as well as the direction of the score (i.e. being below or above the test value). Assessing acceptance, a hypothetical median of zero (indicating neither rejection nor acceptance) was used. It was also tested whether the ranks of the subscales, usefulness and satisfaction, differed significantly from the test value and the direction of the difference (i.e. positive or negative deviance from zero).

The repeated measure design only applies to the assessment of MOBiNET project members. For their scores, it was also analysed whether rating of usability and acceptance changed over time. Over time, the platform was developed further (e.g. added functionalities as well as changes to interfaces); therefore, these changes might be reflected in later scores. Friedman's two-way analysis of variance was used to assess changes in ratings. Unlike ANOVAs, the Friedman test does not test for differences between groups; therefore, Mann-Whitney-U tests were used to assess differences between groups.

3. Results

3.1. Service provider/developer evaluation

3.1.1. System usability scale

Results of the usability ratings are shown in the table below (Table 3); information includes the number of participants per release, mean ratings (on a scale from 0 to 100 and the corresponding standard deviation (SD). In addition, statistical information (i.e. median, z-score, and p-value) is also provided. As presented in the table, on average, mean scores of SUS are below the target value of 68 for all releases. Analysing the data using the Wilcoxon signed rank test reveals significant difference between the hypothetical and actual median for releases 1, 2.1, and 3. Results of release 2.2 indicate a trend towards being significant meaning a negative deviation from the hypothetical median. Contrary, the results of release 3.1 show no significant differences between medians: medians are similar.

Table 3: Statistical results of the Wilcoxon tests of the usability ratings of developers/service providers. ‘’ indicate statistically significant deviance from the hypothetical median.**

	N	Mean	SD	Median	Z	p
Release 1	11	49.09	19.63	45	-2.5	.01*
Release 2.1	17	45.14	17.39	47.5	-3.5	< .001*
Release 2.2	8	43.75	22.12	36,25	-1,8	.07
Release 3	13	51.92	15	55	-2.9	.003*
Release 3.1	9	60.28	16.1	62,5	-1,4	ns

3.1.2. Acceptance scale

Acceptance ratings are divided into usefulness and satisfaction. The table below (Table 4) shows the results of the usefulness and satisfaction ratings of participants taking on the perspective of a developer/service provider. For all releases, the number of participants, mean ratings and SD as well as statistical results, are presented. Again, Wilcoxon signed rank tests were used for analysis. Actual medians were tested against the hypothetical median of zero (i.e. indicating neither acceptance nor rejection).

Analyses of the usefulness ratings reveal no significant differences for releases 1, 2.1, and 2.2. Ratings of releases 3 and 3.1 differed significantly from the hypothetical median. Actual medians were higher than the hypothetical medians. The results of the satisfaction ratings are similar to the usefulness ratings. No significant differences were found for releases 1, 2.1, 2.2, and 3. Results of release 3.1 reveal a positive trend because the difference between the hypothetical median and the actual median is close to being statistically significant ($p= 0.07$).

Table 4: Statistical results of Wilcoxon tests of the acceptance ratings of developers/service providers. ‘’ indicate statistically significant deviance from the hypothetical median.**

	Subscale	N	Mean	SD	Median	Z	p
Release 1	Usefulness	11	.36	.89	.2	1.07	ns
	Satisfaction	11	.07	.87	.25	.31	ns
Release 2.1	Usefulness	17	.04	.64	.2	.4	ns
	Satisfaction	17	.09	.95	.5	.26	ns
Release 2.2	Usefulness	8	.07	.66	-.2	-.17	ns
	Satisfaction	8	-.03	.94	.12	.08	ns
Release 3	Usefulness	13	.41	.58	.4	2.1	.04*
	Satisfaction	13	.25	.67	.5	1.1	ns
Release 3.1	Usefulness	9	.62	.27	.6	2.5	.01*
	Satisfaction	9	.61	.78	.75	1.8	.07

3.2. End user evaluation

3.2.1. System usability scale

The same analysis as in 3.1.1 was performed in order to analyse the usability ratings of participants taking on the perspective of an end user. Similar to the results of the developer/service provider perspective, mean SUS scores are below the target value of 68 (see Figure 2). Results of release 1 indicate no significant differences between actual and hypothetical median, of releases 2.1 and 3.1 indicate a negative trend, while results of releases 2.2 and 3 show significant differences.

Table 5: Statistical results of the Wilcoxon tests of the usability ratings of end users. ‘’ indicate statistically significant deviance from the hypothetical median.**

	N	Mean	SD	Median	Z	p
Release 1	3	50	9.01	52,5	-1,6	ns
Release 2.1	4	56.8	10.8	58,7	-1.8	.07
Release 2.2	8	51.87	10.83	53,7	-2,5	.01*
Release 3	8	52.8	5.89	53,7	-2,5	.01*
Release 3.1	8	60	9.1	58,7	-1,8	.07

3.2.2. Acceptance Scale

Table 6 summarizes the results of the acceptance ratings of participants taking on the role of an end user. Analysing usefulness and satisfaction ratings, Wilcoxon signed rank tests did not reveal statistically significant differences between medians for releases 1 to 3. Significant differences between the hypothetical median and actual medians were observed for usefulness and satisfaction ratings of release 3.1. The actual medians differed positively from the hypothetical median.

Table 6: Statistical results of the Wilcoxon tests of the acceptance ratings of end users. “*” indicate statistically significant deviance from the hypothetical median.

	Subscale	N	Mean	SD	Median	Z	p
Release 1	Usefulness	3	.07	.81	.2	.27	ns
	Satisfaction	3	-.33	.8	0	-.44	ns
Release 2.1	Usefulness	4	.25	.37	.1	1.3	ns
	Satisfaction	3	.33	.57	0	1	ns
Release 2.2	Usefulness	8	.25	.64	.3	1.2	ns
	Satisfaction	8	.15	.63	.38	.68	ns
Release 3	Usefulness	8	-.21	.44	-.2	-.95	ns
	Satisfaction	8	-.25	.82	.38	-.68	ns
Release 3.1	Usefulness	8	.45	.29	.5	2.4	.017*
	Satisfaction	8	.41	.46	.5	2	.04*

3.3. Changes in ratings over time

3.3.1. System usability scale

Further analysis was done in order to examine changes in ratings over time. Figure 2 depicts the average SUS scores (i.e. mean scores) of developers/services (right) and end users (left) across all releases. Friedman’s ANOVAs did not reveal any significant changes over time for ratings of developers ($Z = 2.8$, $p = ns$) or end users ($Z = 2.59$, $p = ns$). Differences between groups were not revealed.

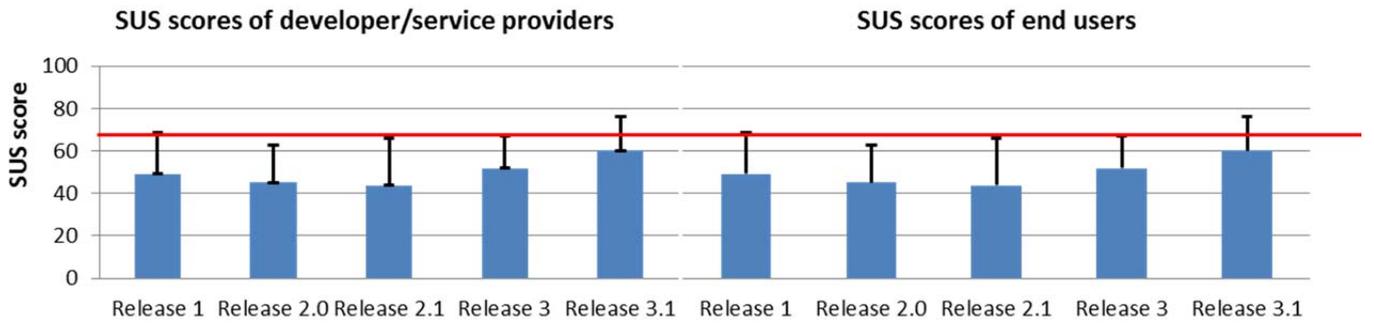


Figure 2: Average scores of the System usability scale across all releases. Error bars present the standard deviation. Red line indicates the target score of 68. Left: Scores of developers/service providers. Right: Scores of end users.

3.3.2. Acceptance scale

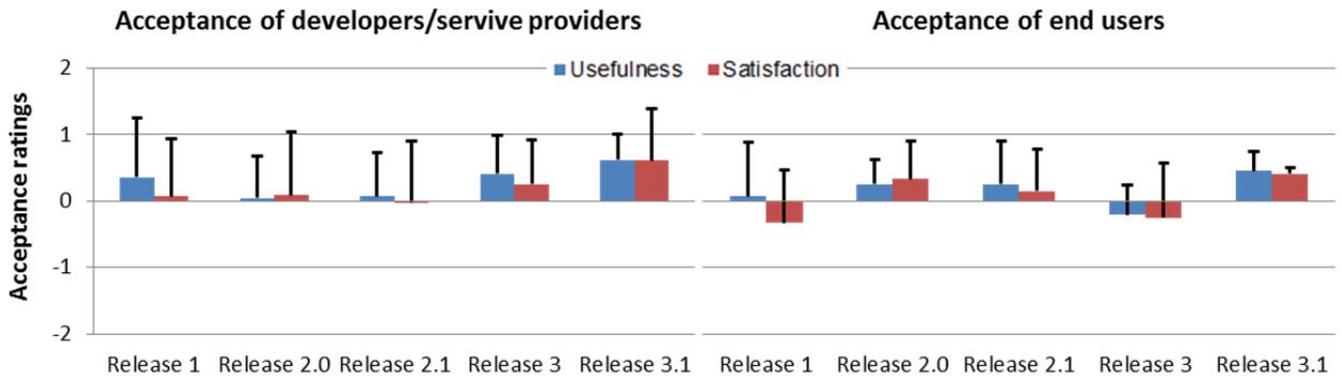


Figure 3: Average scores of the acceptance ratings of all releases. Error bars present the standard deviation. Left: Acceptance scores of developers/service providers. Right: Acceptance scores of end users.

In Figure 3, average ratings of usefulness and satisfaction are depicted across all releases for developers/service providers and end users. Friedman’s ANOVAs were used to examine changes in ratings over time. No significant difference could be found in the medians of ratings of usefulness (developers/providers: $Z = 5.13$, $p = ns$; end users: $Z = 5.64$, $p = ns$) and satisfaction (developers/providers: $Z = .98$, $p = ns$; end users: $Z = 4$, $p = ns$) over time for the groups. In addition, Mann Whitney U tests also did not yield significant differences in ratings between the groups.

3.4. Hackathon

Usability and acceptance ratings were analysed utilizing Wilcoxon signed rank test. Results of the SUS ratings reveal a significant negative difference between the hypothetical median of 68 and the actual median of 45. Usefulness differed positively from the hypothetical median of zero, while no difference was observed for the satisfaction ratings.

Table 7: Statistical results of Wilcoxon signed rank tests of the usability and acceptance ratings of participants of the Hackathon. '*' indicate statistically significant deviance from the hypothetical median.

		N	Mean	SD	Median	Z	p
SUS		13	45.38	11.67	45	-3.18	.001*
Acceptance	Usefulness	13	.29	.39	.25	2.16	.03*
	Satisfaction	13	.21	.44	.33	8.26	ns

4. Discussion

The goal of the human factors impact assessment is to draw conclusions with regard to the success of the MOBiNET platform meaning whether the platform will be used to implement and provide services as well as whether users see the benefit of the it and access services offered. Usability and acceptance ratings served as indirect measures. The assessment is based on the current status quo of the platform. Therefore, the conclusions drawn are not conclusive and only reflect results associated with the current developmental status of the platform. In order to finally conclude on the success of the platform based on human factors aspects, the platform needs to be finalized (made into a commercial product with selected functionalities and focus based on a business plan) and a final assessment made with actual users.

4.1. Implications of usability ratings

The overall usability score indicates how easy and pleasant it is to use the MOBiNET platform. During the first releases, usability scores were significantly lower than the average score of 68 indicating that the MOBiNET platform was not easily and pleasantly used. This is true for service developers/providers and end users, respectively. Considering that all participants were familiar with MOBiNET meaning participants were project-related staff, it indicates that the platform had major usability flaws that needed to be tackled. Those flaws may range from difficult or time-consuming to use up to a number of functionalities not available at the time of testing. Changes/improvements from one release to the next also do not suggest statistically significant improvement of the usability neither for service developers/providers nor end users.

Nonetheless, visual inspection of changes suggests improvements of the platform's usability. During the last releases, the overall scores increased and moved closer to the target value. It appears that the usability of the platform increased making it easier and more pleasant to use the platform.

4.2. Implications of acceptance ratings

The subscales, usefulness and satisfaction, were used to assess acceptance of the platform. Overall, results for service developers/providers and end users were similar. Analysis of changes over time did not reveal any significant effects. According to the results, subjective ratings associated with the developmental state of the platform did not increase significantly from one release to the next release. Results may indicate that MOBiNET platform did not improve with regard to human factors aspects from one release to the next. It might also be the case that changes in design did not meet users' expectations. Changes might have not been to a degree noticeable for end users as MOBiNET platform is still a prototypical development. In addition, not all functions and functionalities were improved from one release to the next, new ones were also added.

Nonetheless, testing against the hypothetical median revealed higher ratings for the last two releases. It appears that the platforms function and functionality improved over time and therefore ratings increased. In the beginning, ratings were around zero indicating that participants neither rejected nor accepted the platform. The positive change in ratings might reflect the improvements of the last release.

4.3. Implications of the Hackathon survey

Participants of the Hackathon in Bordeaux were MOBiNET-independent persons who accepted the challenge to implement either a service or modify an existing service. Overall, participants were not very familiar with the concept of MOBiNET and have not worked with the platform before; therefore, they can be considered novice users. The Hackathon took place in June 2016 meaning that the developmental stage of the platform was that of Release 2.1. Compared with acceptance and usability ratings of MOBiNET-affiliated partners at that particular stage of development, results did not differ. Usability scores were significantly below the average score of 68 also indicating major usability flaws. Even though, the MOBiNET platform is still under development and a score of 68 might be difficult to accomplish at this stage, results indicate major usability flaws. In addition, acceptance ratings were right around the test value of zero indicating neither acceptance nor rejection. Participants were more or less indifferent.

5. Conclusion/ Impact assessment

Up until Release 3.1, participants were not yet fully convinced of the MOBiNET platform at its current stage as indicated by the results. It is important to keep in mind that the MOBiNET is a research project focusing on proof-of-concept. In addition, results need to be considered with caution because the sample size in each release was very low. When the number of participants is low, outliers (extreme responder) have more weight on the average result. In the future, the sample size needs to be significantly larger in order to have more valid results and be able to draw more exclusive conclusions. Because most of the participants were MOBiNET-affiliated this might have also introduced some bias. The next evaluation should be done with MOBiNET-independent participants.

As of Release 3.1, MOBiNET platform needs to improve usability before reaching the final development stage as reflected in usability ratings. In addition, acceptance of the platform is rather low. Participants were rather indifferent. Using these results to project the impact of the platform from a human factors perspective, it appears that ratings improve and therefore, the success of the platform improves. Nonetheless, more effort needs to be put in the development of the platform to make the platform as successful as intended in the beginning of the project. Evaluation of the final release will lead to more exclusive conclusions as platform development will reach a more finalized stage.

References

Nielsen, J. (1993). *Usability Engineering*. Academic Press, London

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189, 194.

Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, 5(1), 1-10.

Blankenberg, F. (1995)]. *Methods of Impact Assessment Research Programme, Resource Pack and Discussion*. The Hague: Oxfam UK/I and Novib.

Annex 1

System Usability Scale

Instructions:

You explored and experienced the MOBiNET platform. What is your overall impression?

Below you find pairs of contracting statements. Please indicate your agreement or disagreement with the statements.

	Strongly disagree				Strongly agree
	1	2	3	4	5
I think that I would like to use this system frequently.	<input type="radio"/>				
I found the system unnecessarily complex.	<input type="radio"/>				
I thought the system was easy to use.	<input type="radio"/>				
I think that I would need the support of a technical person to be able to use this system.	<input type="radio"/>				
I found the various functions in this system were well integrated.	<input type="radio"/>				
I thought there was too much inconsistency in this system.	<input type="radio"/>				
I would imagine that most people would learn to use this system very quickly.	<input type="radio"/>				
I found the system very cumbersome to use.	<input type="radio"/>				
I felt confident using the system.	<input type="radio"/>				
I needed to learn a lot of things before I could get going with this system.	<input type="radio"/>				

Acceptance scale

Instructions:

Please indicate how you find the MOBiNET platform (please tick a box on every line).

Useful	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Useless
Pleasant	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Unpleasant
Bad	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Good
Nice	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Annoying
Effective	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Superfluous
Irritating	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Likeable
Assisting	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Worthless
Undesirable	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Desirable
Raising alertness	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Sleep-inducing