UrbanData2Decide
Integrated Data Visualisation and Decision Making

D5.1 USABILITY TESTS AND FEEDBACK INTERVIEWS REPORT

PROJECT

Acronym:	**UrbanData2Decide**
Title:	Data Visualisation and Decision Making Solutions to Forecast and Manage Complex Urban Challenges
Coordinator:	SYNYO GmbH
Reference:	847511
Type:	Joint Programme Initiative
Programme:	Urban Europe
Start:	September 2014
Duration:	26 months
Website:	http://www.urbandata2decide.eu
E-Mail:	office@urbandata2decide.eu

Consortium:

SYNYO GmbH, Research & Development Department, Austria (SYNYO)

University of Oxford, Oxford Internet Institute, UK (OXFORD)

Malmö University, Department of Urban Studies, Sweden (MU)

Open Data Institute, Research Department, UK (ODI)

IT University of Copenhagen, Software Development Group, Denmark (ITU)

ZSI Centre for Social Innovation, Department of Knowledge and Technology, Austria (ZSI)

SYNYO · UNIVERSITY OF OXFORD · MALMÖ UNIVERSITY · ODI open data institute · IT University of Copenhagen · ZSI

DELIVERABLE

Number:	**D5.1**
Title:	**Usability tests and feedback interviews report**
Lead partner:	ZSI
Work package:	WP5: Demonstration and Validation
Date:	December 2016
Authors:	Dobner, Susanne ZSI Voigt, Christian ZSI
Contributors:	Contributions from all partners
Reviewers:	Jonathan Bright, OII

The UrbanData2Decide project is co-funded under the Joint Programming Initiative, 2nd call Urban Europe.

TABLE OF CONTENT

Content

LIST OF FIGURES

INTRODUCTION

"*True usability is invisible. If something is going well, you don't notice it*."

(Rubin & Chisnell 2008, 6)

Usually people don't notice good design as it is "*serving us without drawing attention to it*" (Norman 2013, xi) instead what we usually notice is if something is designed poorly. Then using a product or a tool can be an irritating or frustrating experience, which then draws attention to its design. Often in product or tool development the emphasis is put on a nice, fashionable design instead of its usability. The picture below (see figure 1) shows a well-known example of a product, a "teapot for masochists" which may not need a usability test to prove its impracticality.

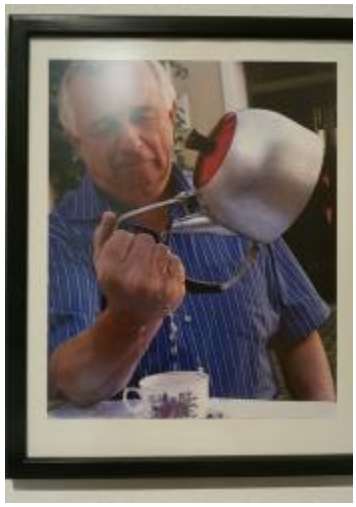


Figure 1 "Teapot for Masochists" (referring to the model by Jacques Carelman's)

Beginning in the 1960s, primarily in the US, the involvement of end-users in the design process has gained increasing attention. The engagement of users has primarily evolved in the context of social responsibility, urban development and citizen participation before it moved to the realms of product development (Sanoff 2006). More engineers, scientists and designers started to scrutinize predominant assumptions of modern design and product development enabling first discussions and attempts towards a socially responsive and responsible approach to design. The change of mindset is reflected in a refocus to the design process itself and increasing attention given to design approaches such as 'inclusive and accessible design' over the last decade (Cassim 2007). Since different tools to support urban decision-making are being developed within the 'UrbanDat2Decide' project for

different purposes and user groups, this report focuses on introducing various usability testing methods as well as results from usability testing within the project.

The report is structured as follows:

- The first chapter introduces usability, its history, definitions and goals. It further links the method of usability testing to the ‚UrbanData2Decide' project context and previous reports, especially D.3.3 and D3.4.1

- The second chapter describes five different usability testing methods, including the methods: cognitive walkthrough, think aloud and paper prototyping. For each method single steps as well as materials needed to successfully implement each method are provided.

- The third chapter discusses the results of the usability evaluations of different project tools with various test groups in each city, ranging from citizens to local authorities.

- The fourth chapter discusses the results from the usability tests of three tools developed within the 'UrbanData2Decide' project and draws comparisons of the applied methods and results.

- Finally, the fifth chapter concludes with main theoretical and practical findings from the usability tests.

1.1 Definitions

"*Usability is not a precise science consisting of formulas and black and white answers. […] Instead, usability testing can often be an imprecise, ambiguous enterprise, with varying and sometimes conflicting observations, not surprising for any venture that has human beings as its focus*."

(Rubin & Chisnell 2008, 50)

Defining usability proves to be a difficult endeavor as it, following the argument in the quote above, can be a contested concept as something may be usable for one end user group but not for another. However, following a more operational definition may capture essential features of a usable product or tool. Rubin and Chisnell (2008) name five features that make a product usable, including *usefulness, efficiency, effectiveness, satisfaction and accessibility*. Another explanation of what a

[1] For further information, please see the reports D3.3. „Interface Design" and D.3.4 „Integrated Framework Report" on: http://www.urbandata2decide.eu/media-centre/

usable tool or product implies may be the following: "*the user can do what he or she wants to do the way he or she expects to be able to do it, without hindrance, hesitation, or questions.*" (4)
Further, Norman (2013) describes two important characteristics of good design:

- **Discoverability**: what can I do with it? Where and how?
- **Understanding**: what do the different functions mean?

In regards to usability testing, there is a number of different formats in which usability tests can take place, some have developed and changed over time. Whereas in the beginnings of usability testing a rather clinical style was standard, a less formal style and a closer distance with the tester are commonly used now. **Usability testing** *employs techniques to collect empirical data while observing representative end users using the product to perform realistic tasks*" (Rubin & Chisnell 2008, 19). Usability testing is a research tool, with its roots in classical experimental methodology. There is a range of usability tests from more classical experiments with a large number of users to rather informal and small qualitative studies. Depending on the testing method, there are different underlying objectives, time and resource requirements.

Some basic elements of user testing are (see Koivuniemi 2013; Rubin & Chisnell 2008):

- Developing research questions or test objectives rather than hypotheses
- Use of a representative sample of end users which may or may not be randomly chosen
- Representation of the actual work environment
- Observation of end users who either use or review a representation of the product
- Controlled and sometimes extensive interviewing and probing of the participants by the test moderator
- Collection of quantitative and qualitative performance and preference measures
- Recommendation of improvements to the design of the product

1.2 What is Usability Evaluation Good For?

As briefly mentioned in the introduction usability testing and participatory design have evolved over the last 50 years. Before the turn towards a stronger focus on the design process itself and involvement of end users, a non-integrated approach to product development was predominant.

This implies that development groups have tended to work in silos rather than interdisciplinary groups including end users. Usability testing enables designers to better understand users' expectations, satisfaction or problems with single functionalities of a tool or a whole product. This, in turn informs future design decisions. Usability evaluations concern the "*human computer interaction and in particular, the user interface. In contrast usefulness is concerned with whether or not the system supports user activity*" (Buchanan & Salako 2009, 639)[2].

Usability testing aims to ensure that tools and products developed are

- useful to and valued by the target audience
- easy to learn
- help people be effective and efficient at what they want to do
- satisfying (and possibly even delightful) to use

Many scholars, including Rubin & Chisnell (2008, 28) argue that "*usability testing is most powerful and most effective when implemented as part of an iterative product development process. That is, a cycle of design, test and measure, and redesign throughout the product development lifecycle has the greatest probability of concluding with a usable product*".

1.3 The Object of Usability Evaluation

As already discussed in the previous chapter usability testing is (usually) not a one-time only method but happens throughout the development cycle of a tool or product. Usually the usability of a tool or a product is tested in the early development phase and then repeated at various stages in the development cycle until it is released. This ensures that the usability is tested at various stages and includes end-users continuously from the beginning.

Dabbs and colleagues (2013, 155) describe three main lifecycle phases of product development:

- needs analysis and requirements gathering
- prototyping, design and testing
- final testing and product launch

[2] Read more on the topic of usefulness and its application In our other report D.5.2 "Simulation and application scenarios report".

Depending on the phase and progress of tool development different objects, respectively materials are used for usability testing. Usually in the very beginning first hand-drawn sketches or paper mock-ups are created, whereas in the mid-to-final testing phase click dummies or early prototypes that resemble the final tool or product are used. Thus the object of usability testing can vary and evolves with an improved version after every testing phase.

What can be tested?

- **Sketches**

> "*Sketching is not only the archetypal activity of design, it has been thus for centuries*."
> (Buxton 2007, 111)

Sketching in the process of developing and designing artefacts or tools is not a new phenomenon. It has evolved since the late medieval period, since then design was more and more separated from the process of making. Sketches provide designers with a tool to explore and communicate their ideas with colleagues and end-users (105).

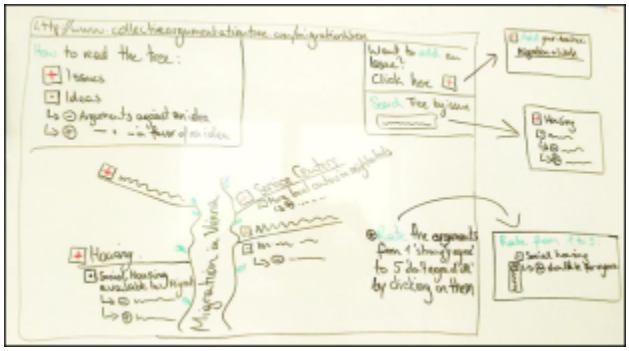


Figure 2 A sketch for an online discussion tool

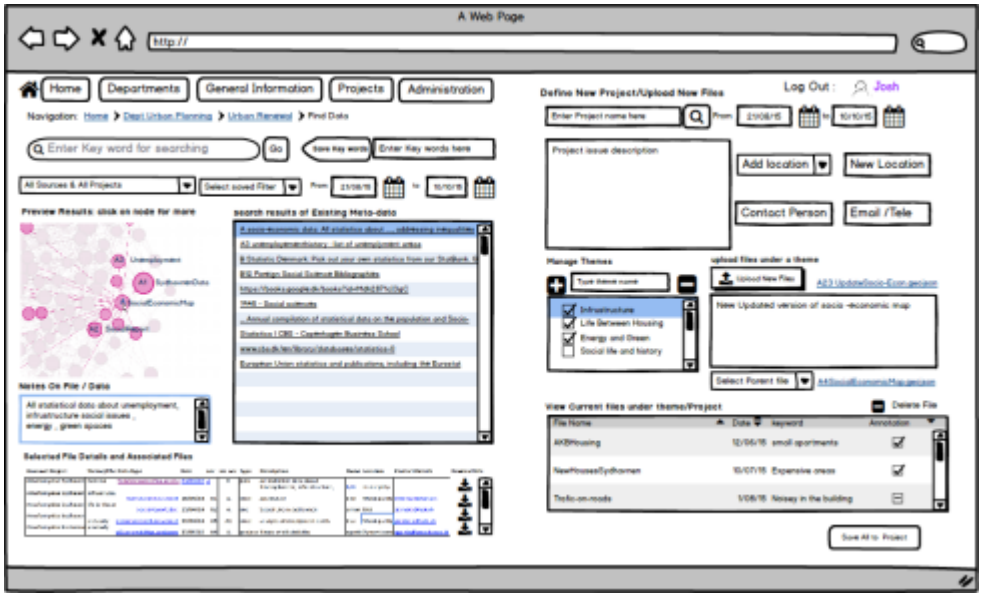
- **Wireframes**

Wireframes are usually two-dimensional illustrations of a page interface including rough illustrations of content and functions. They offer a fast creation of first prototypes in a quick and rather cost effective way. Generally there are two different wireframes: low-, and high-fidelity wireframes.

  - **Low-fidelity wireframes** can be created quite fast and encourage early communication between project teams and end-users early in the development process. Often they

contain simple images and boxes with mock content as fillers for not yet developed content and labels. A low-fidelity wireframe puts users focus on the general content and functionalities rather than its design, colors or fonts (Rettig 1994; Kankainen 2003).

- **High-fidelity wireframes** include a lot more detail in terms of content, information for single items or symbols than a low-fidelity wireframe (see picture 1 below). Thus their design matches the actual website more than a low-fidelity wireframe. Often paper-mock ups (see chapter 2.4) are used for both low-, and high fidelity wireframes. Though the latter, if more complex can also be designed as a click dummy.



Picture 1 Example of a high-fidelity wireframe

➢ **Click dummies**

A click dummy is an interactive prototype, in the sense that it allows users to click through the design or a tool. However without the actual full functionality, it is still a partly functioning demonstration of the tool. For a click dummy neither a full functionality of the tool nor real data is necessary. It should contain all pages a user would see and need while clicking through a website and fulfilling a specific task, for instance searching for specific information, uploading photos or signing up for a newsletter. Usability testing using a click dummy allows the elaboration of the visual and interaction design.

- **Early prototypes**

In their functionalities and design early prototypes are very close to the final tool or product. In comparison to click dummies early prototypes enable users to actually move between pages without a pre-given assemblage of screens (as is the case for click dummies). Early prototypes – following a participatory development approach - are usually a result of previous testing with paper sketches, mock-ups and/or click dummies.

2 APPLICABLE METHODS

2.1 The Process of Usability Testing

The following chapter outlines common steps towards finding a suitable testing method and steps guiding through a test regardless of the chosen method. Both steps raise general questions whose answers influence the choice of testing methods. Three questions below should encourage researchers to think about their motivations and objectives before deciding upon a testing method and further plans:

- Why are you testing?
- Who are you testing?
- What are you testing?

The chosen design is predicated by the test objectives – what is it that you want to learn about the product and user? The available resources, constraints and creativity in turn influence the chosen design.

Eight steps towards a successful completion of usability testing are listed and described below. The steps may be revisited and some researchers may follow an alternate order as some steps depend on choices, e.g. materials and time resources influence selection of participants (testers).

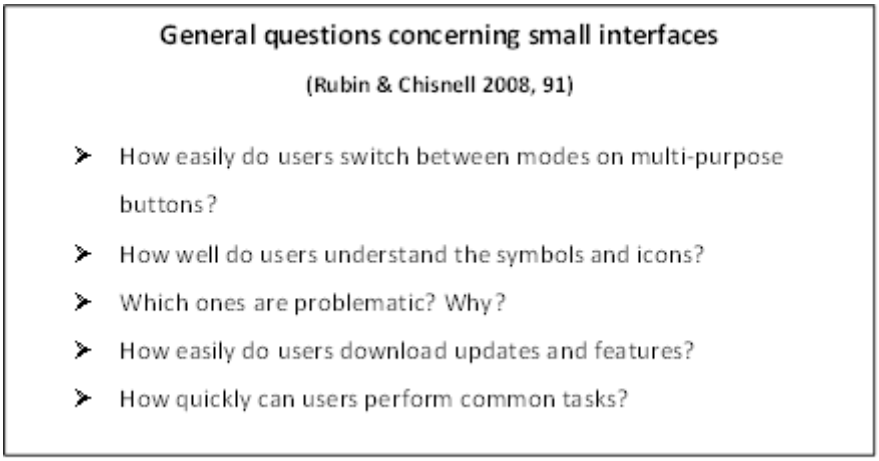
1) **Develop the Test Plan**

Choose an approach, plan research, and establish research questions / scenarios which should be tested.

- What is it that you want to find out? e.g. identify obstacles to completing a certain task using your tool, navigate through the application

- An example of a research question: how long do employees of Transport for London feel it takes them to find out about the number of passengers for the next arriving train at a specific station? Their perception of time spent and steps they had to take until they found out what they needed to know.

- How will you measure success? Completion of task e.g. define a benchmark time to complete a task, how long does it take the participants to accomplish a task – just one idea as time how long it takes may not be the most accurate measure, especially if using the "think aloud" method it will naturally take the participant longer; other ideas include counting errors or types of prompting



2) Select participants and moderator

- testing multiple product versions and/or multiple user groups; for instance testing a product with a lay person and with an experienced tester will most likely result in other findings

3) Set up a Testing Environment

- materials and machines required to perform tasks, including a laptop/PC, or print outs of the screens, printed wire frame drawings of the pages; these materials help to give developers a better idea of whether the whole systems is to be tested or parts of it

- preparation of test materials: for instance hand-drawn or computer based mock-up, a wireframe or a functioning tool. Prepare tasks or scenarios that the test person is supposed to complete; write down moderator guidelines

4) **Conduct and facilitate the Test**

5) **Debrief the Participant and Observers**
   - prepare an information sheet about the purpose of the testing and privacy issues including answers to questions such as is the testing anonymous? Or will information be passed on to other parties?

6) **Analyze Data and Observations**
   - compile and summarize data
   - analyze data

7) **Report Findings and Recommendation**
   - develop findings and recommendations in order to produce a final report

An example on how to structure a usability test sheet is provided in the Annex 1. The sheet is structured as follows: methodology, roles of facilitator, tasks and goals and reporting of results (critical errors – non-critical error; time on task; subjective evaluation).

(see Rubin & Chisnell, 2008, 65ff.85, 91; Unger & Chandler, 2009, 25)

2.2 Scenarios

Scenarios describe the stories and reasons why a user or target groups would visit a website or use a tool. Thus scenarios focus on detecting the specifics behind who, how and why a user visits a site and gives developers and designers a broader understanding of the context. Therefore it puts emphasis on the website or tool as a whole rather than on small or specific elements. Scenarios are very helpful for designing an interface as well as for usability testing. Scenarios for usability testing are

often formulated in specific tasks, e.g. on our emergency data map, you want to find out where all hospitals in Vienna are located. Hence the method of using scenarios can be very well combined with a think-aloud methodology.

A few reasons why scenarios can be helpful in the design phase include for instance

> *"to remind a team, during both the design and validation phases, of the overarching goal(s) that users have when interacting with a solution. Scenarios help the team consider the design of the solution as a whole rather than getting caught up by specific pages, elements, or interactions. They note questions and goals and sometimes define the possibilities of how the user(s) can achieve them."*[3]

The three main questions to be discussed, before deciding upon a scenario are:

- Who is the user?
- Why does the user come to the site?
- What goals does the user have?

Three types of scenarios are described below:

- **goal-or task based scenarios:**
  - describe what the user wants to do, what kind of task they try to accomplish, where there is no further information on the single steps he/she may be taking to fulfill a task

- **elaborated scenarios:**
  - entail more information on the user (age, background etc.), enabling the developers to better develop suitable functionalities and content

[3] https://methods.18f.gov/decide/user-scenarios/

- **full scale task scenarios:**
  - describe each step a user is taking in order to fulfill a task in detail, full scale task scenarios are similar to use cases

2.3 Cognitive Walkthrough and Think Aloud

Overview

The cognitive walkthrough is a usability testing method that is usually applied early in the design process. Often once first prototypes and a clearer idea of who the target users are is established, a cognitive walkthrough can detect possible mismatches and flaws in the design at an early stage of tool development. Cognitive walkthroughs enable the analysis of a user interface by simulating a step-by-step user behavior for a predefined task.

First applied in 1990, the method was originally used to evaluate the usability of physical objects like postal kiosks or ATMs. Nowadays it is widely used also in areas such as software or app development. A cognitive walkthrough enables a researcher to "*evaluate each step necessary to perform a task, attempting to uncover design errors […] mismatches between users' and designers' conceptualization of a task, poor choices of wording for menu titles and button labels*." (Rubin & Chisnell 2008, 18) Olson et al., (1984) discuss the think-aloud method as "*one of the most effective ways to assess higher-level thinking processes*" (Charters 2003, 71). The cognitive walkthrough methodology focuses on one aspect of usability in particular: **ease of learning**. Whereas other usability methods, such as a heuristic evaluation (see chapter 2.3) focus preferably on system experts and people who have previously used the system, the cognitive walkthrough aims at new users. Based on the theory of learning, during a cognitive walkthrough participants are asked to perform certain tasks using the tool to be tested. It makes the **learnability of a system** for new users explicit. Common theory of learning by exploration suggest that with new software, website or other tools people without tool expertise, solve new problems and tasks using "*general problem-solving methods, general reading knowledge, and accumulated experience with computers*."[4]

Developers trying to imagine "people's thoughts and actions when they use an interface for the first time often perform cognitive walkthroughs" (Lewis & Rieman 1994, 46). Rubin and Chisnell (2008) advise to recruit real end-users. Each step of the interaction of the end-user or tester with the interface is thereby being evaluated asking questions about each of the steps. Thus a cognitive

[4] http://autocww.colorado.edu/~blackmon/Papers/CognitiveWalkEncycHCI2004.pdf

walkthrough allows to reveal "*implicit or explicit assumptions made by developers about users' knowledge*" (25) performing a specific task.

Combining Thinking Aloud and Cognitive Walkthrough

Thinking aloud is a usability method, which integrates well with the cognitive walkthrough because it allows a great insight into people's thought processes while performing a certain task. While testers perform certain tasks they are encouraged to say their thoughts out loud (**Usability Testing Essentials ,**2011, 19). The experiences of a user with the product become better understandable, if the tester doesn't see but also hears about the user's reactions, ideas or frustrations.

"When users think out loud, you don't have to guess what they're thinking. They tell you."

During a cognitive walkthrough, evaluators can ask participants to share their thoughts as well as take notes of their observations or in case the tester halts his or her thoughts or consciously or unconsciously not share all thoughts. Thus the thinking aloud method is not (only) an alternative to but also an addition to performing a cognitive walkthrough. While some suggest that if time is of relevance to the evaluation, e.g. if a certain task should be performed fast, then the think aloud method may not be optimal. It has been shown in some studies that it doesn't significantly slow down performance (Chisnell & Rubin 2008, 54). Some studies even suggest that thinking aloud improves the performance of testers, though other studies have shown different results.

In order to best capture the testers' comments and thoughts, Chisnell & Rubin suggest to either

● **type users' comments into a log**, you can export them into a spreadsheet, then sort the comments by keywords or by issues. If you created your log files in a spreadsheet or word processing document, you can sort within this document (…)

● or use **observer forms** or some other format that doesn't allow for quick sorting, you can organize the notes about user comments into positive and negative groupings (Chisnell & Rubin 2008, 54)

Also for the final report "*quoting users' comments to support your findings makes for a better understanding of the users ' experience*. These user quotes make for powerful statements in your report of the study findings." (258ff.)

Materials, Steps and Moderation

The following paragraphs describe the materials needed to perform a cognitive walkthrough, as well as the single steps towards a successful testing process. For an example of testing an online journey planner using the cognitive walkthrough method, please see Annex 2.

Materials

- representation of the user interface
  - depending on the available format a representation can be: printouts of the user interface, a click-dummy, a wireframe or an already usable web interface
- user profile
  - to know who potential end-users are and invite an appropriate test person
- task list
  - including all the tasks used during the walkthrough
- problem reporting sheet
  - to analyze and report results later

Single Steps

Steps towards a successful completion of a cognitive walkthrough include defining the end user and tasks to brainstorming of solutions to improve the tool/product from the very beginning. In the case of London end users and tasks could for instance include an employee of Transport for London checking the number of passengers in a specific train and station in the London use case.

1. **Define the users** of the product and conduct a context of use analysis

2. **Define the** appropriate **tasks** for the walkthrough

   - it is recommended to start with a simple task and move to more complex tasks later
   - A common theme in the research and case study literature is that only a few tasks can be examined in any cognitive walkthrough session. 1- 4 tasks in any given session depending on complexity are recommended
   - choose realistic tasks which include the core features of the tool

Finding a suitable task is not an easy endeavor as difficult or demanding tasks *"creating a high cognitive overload" interfere with verbalization, because other processes crowd verbal information out of working memory."* However, also tasks which are too simple may hinder the participant to describe *"these automatic or near-automatic happenings"* (Charter 2003, 73). Preferably, tasks should be broken down in smaller units.

3. **Choose an evaluator / or a group of evaluators (the once who guide through the test)**

4. Develop **principles for the walkthrough**, e.g.:

    - no discussions about ways to redesign the interface during the walkthrough
    - person who guides through the test does not defend the designs of the tool, e.g. does not explain why things were designed in a certain way

5. Conduct the actual walkthrough

- provide a representation of the interface to the evaluators.
    - *Walk through the action sequences for each task from the perspective of the "typical" users of the product.*

For each step in the sequence, see if you can tell a credible story based on the following questions (Wharton, Rieman, Lewis, & Polson 1994, 106):

- ✓ Will the user try to achieve the right effect?
- ✓ Will the user notice that the correct action is available?
- ✓ Will the user associate the correct action with the effect that the user is trying to achieve?
- ✓ If the correct action is performed, will the user see that progress is being made towards the solution of the task?

- capture findings during the walkthrough: fill out task sheet and take notes
    - including: success stories, failure stories, design suggestions, and problems that were not the direct output of the walkthrough, assumptions about users, comments about the tasks, and other information that may be useful in design. Use a standard form for this process.

Three types of Think-Aloud protocols – How to interact (or not) with participants?

Olmsted-Hawala et al. (2010) describe three types of think-Aloud protocols: the verbal protocol, speech-communication-based protocol, and the coaching protocol.

- **verbal protocol** developed by Ericsson and Simon is the most traditional protocol. The verbal protocol follows the Ericsson and Simon method stressing that the moderator does not probe words beyond "keep talking" (Olmsted-Hawala et al. 2010, 2384)

- **speech-communication-based protocol** firstly introduced by Boren and Ramey implies verbal feedback in form of "um-hum or un-hum" to keep participant talking, probing by test administrator in form of feedback tokens or questioning tone picking up on last word uttered by participant after 15 seconds of silence, e.g., Participant says "that was odd..." Test administrator says after pause "Odd?")

- **coaching protocol** (2384 ff.) entails a more active intervention, or coaching of the participant, e.g. more verbal feedback and probes where test administrator asks direct questions about different areas of Web site, such as areas where user is having difficulty/is pausing/or is describing area as confusing or frustrating; gives help or assists when participant is struggling

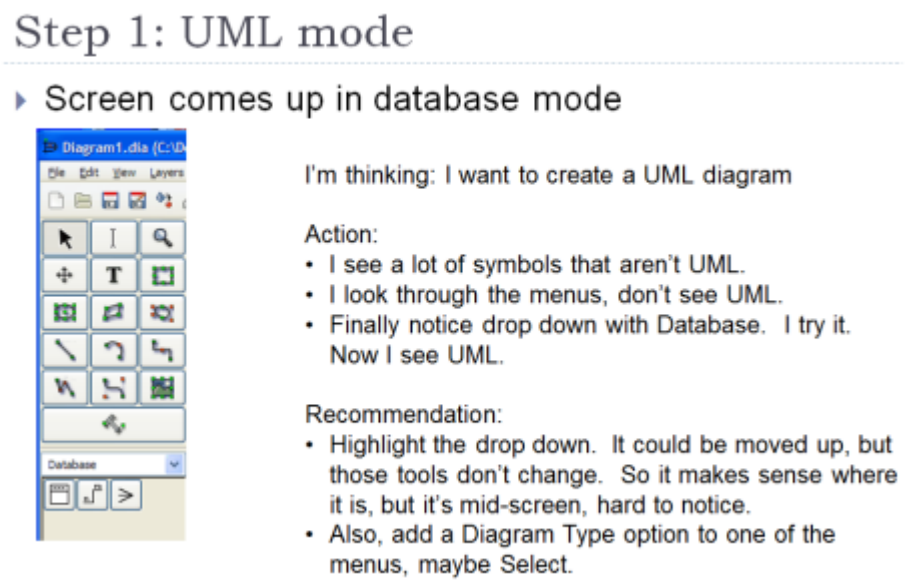
Out of the three protocols, the verbal and speech-communication-based protocols resemble a situation close to the real experience a user would have in a non-testing situation (given that users usually don't have external help). Hence if experiences close to real experience of users are wished to obtain, the coaching protocol is not the right technique (2388). In an ideal situation, participants in as testing setting should articulate

6. Discuss findings and experiences with colleagues after the walkthrough

7. Discuss, report and decide upon potential solutions to identified problems

- revisit the reporting sheet filled out during the walkthrough and summarize the result

The picture below shows parts of a Think Aloud Testing session, with the task to „create a UML diagram" and the verbalized thoughts and action by the tester.



Picture 2 Screenshot of a step in a Think-Aloud Testing Session.

Moderation

Some of the tips for successful moderation of cognitive walkthroughs listed by Dumas & Loring (2008, 15ff.) include the following:

- "**Let the participant speak!**" is the most common credo among think aloud specialists and researchers. It is thus important at the very beginning of a test session to make the participant aware that you, as a moderator will try to talk as little as possible. Then the participant is aware of the type of interaction intended during the test session.
- *Whenever possible, let participants work on tasks without interruption. This will make the experience seem more natural because users typically interact with a product without someone continually asking, "What is happening now?" or "Tell me what you see on this screen."*
- Decide ahead of time what you will do if one or more participants are not able to finish all tasks. For example, prioritize so that the last tasks are least important and can be omitted, or establish a time limit for each task.

- Prepare a data collection sheet or file (if typing notes) ahead of time, and have a good labeling and numbering scheme so that the notes don't get mixed up.
- Know well in advance the types of measures you will be collecting, and have a consistent way to record them. Know the product and the tasks well enough that you can anticipate problems participants might have, whether technical or usability related.

2.4 Heuristic Evaluation

A heuristic evaluation is a usability testing method widely used in human-computer interaction design developed by Jakob Nielsen. In a heuristic evaluation, usability experts evaluate a site's interface using a set of accepted evaluative principles. During the evaluation a group of usability experts "*inspect a user interface to find and rate the severity of usability problems using a set of usability principles or heuristics*" (Vukovac et al. 2010, 273).

> "*Heuristics in an evaluation are meant to guide evaluators to find and explain usability problems*" (Forsell 2014, 194)

The analysis of the filled out evaluative principles results in a list of potential usability issues[5] identified by the heuristic experts. This list of identified problems can be prioritized according to the severity of problems. In the usability literature between 3 to 5 experts are recommended for a thorough evaluation of a site's interface or a tool. Ratings based on the opinion of three evaluators are considered reliable (Forsell 2014, 186). Commonly, each evaluator works independently and discusses the findings afterwards.

An advantage of a heuristic evaluation, as Kantner and Rosenbaum (1997) point out is that rather quick and informal feedback is gathered (by having experts fill in a set of criteria). Additionally, it proves valuable when time and resources are rather shorts as skilled evaluators "*can produce high-quality results in a limited time*". Thus, the method is very much dependent on the skills and expertise of the evaluators as well as a thorough introduction to the tasks and heuristic principles by the facilitators. Furthermore, evaluators should feel encouraged to "*point out problems and give constructive critique, since obtaining such feedback is the aim of the evaluation*" (Forsell 2014, 194).

[5] http://www.usability.gov/how-to-and-tools/methods/heuristic-evaluation.html

Four main phases of usability heuristics testing are:

1. Pre-evaluation training
facilitators give evaluators the needed domain knowledge and information on the scenario
2. Evaluation
heuristic experts evaluate the tool/software and then aggregate results
3. Severity rating
rates the severity of each identified problem
4. Debriefing
discussing the findings with the other heuristic experts and/or faciliators

A heuristic evaluation can be conducted either individually or as a group evaluation where the interface is evaluated by the team altogether. A group evaluation requires more planning, as all evaluators have to come together and common issues of group discussions may occur (e.g. one evaluator dominating a discussion etc.); however the testing just needs to happen once.
Heuristic evaluation identifies usability problems as well as the severity of issues. In order to judge the severity of issues, testers can analyze and create the rating of identified problems according to the **frequency, impact and persistence** of an issue.

When going through the results of the evaluation, a possible problem rating scale discussed by Nielsen and Molich (1990) includes a scale of 4 severity stages:

1) Cosmetic problem – not necessary to fix, unless time allows
2) Minor usability problem – low priority, fixing it should be given low priority
3) Major usability problem – important to fix, should be given high priority
4) Usability catastrophe – imperative to fix the problem

The evaluators need to be equipped with a certain task to perform on the website or interface, e.g. in the London scenario this can include to find data of arriving trains at Waterloo station. Each task is then evaluated according to a pre-set list of usability heuristics. A set of example of the 10 most commonly known heuristics is provided below.

The **10 commonly known usability heuristics** introduced by Nielson (1994) are:

1. **Visibility of system status**
   The system should always keep users informed about what is going on, through appropriate feedback within reasonable time. I.e. “Where am I?” and “Where can I go next?” should always be clear for the user.

   - Do you know where to go next in the navigation?

2. **Match between system and the real world**
   The system should speak the user's language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

   - Do you understand the terms used on the website/the tool?

3. **User control and freedom**
   Users often choose system functions by mistake and will need a clearly marked "emergency exit" or “home” button to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

   - Do you know how to return to the main page?
   - Is the ‘home’ function easy to find?
   - Is the ‘home’ function available on every page?
   - Do you know how to get back to the last page or function?

4. **Consistency and standards**:
   Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

   - Do you understand the meaning of the icons?
   - Do you understand the meaning of the symbols?
   - Do you understand the meaning of the words/language used?

5. **Error prevention**:
   Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

   - Do you understand the error message?

6. **Recognition rather than recall**:
   Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another.

Instructions for use of the system should be visible or easily retrievable whenever appropriate.

- Is the structure of topics clear and logical for you?
- Is the structure of information clear and logical for you?
- Is the structure of actions you can choose clear and logical for you?

7. **Flexibility and efficiency of use**:
   Accelerators—unseen by the novice user—may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

   - Are shortcuts guiding you through the system available? (e.g. to not having to click through a user's manual every time)
     - ➔ If yes, do you find them useful?
     - ➔ If no, would you find it useful to have shortcuts while navigating the website?

8. **Aesthetic and minimalist design**:
   Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

   - Is the information provided on the website precise?
   - Is the information provided on the website too extensive? (i.e. information overload)

9. **Help users recognize, diagnose, and recover from errors**:
   Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

   - Do you understand the error message?
   - Do you understand how to solve the problem?

10. **Help and documentation**:
    Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

    - Do you find help or contextual explanation (i.e. to explain specific words or steps) where necessary?

In addition to Nielson's (1994) heuristics, Weinschenk & Barker (2000) compiled usability guidelines and heuristics from various sources and developed a set of twenty types. The types include Nielson's ideas but go beyond the ten common heuristics to provide a more holistic list.

1. User Control: heuristics that check whether the user has enough control of the interface.

2. Human Limitations: the design takes into account human limitations, cognitive and sensorial, to avoid overloading them.

3. Modal Integrity: the interface uses the most suitable modality for each task: auditory, visual, or motor/kinesthetic.

4. Accommodation: the design is adequate to fulfill the needs and behaviour of each targeted user group.

5. Linguistic Clarity: the language used to communicate is efficient and adequate to the audience.

6. Aesthetic Integrity: the design is visually attractive and tailored to appeal to the target population.

7. Simplicity: the design will not use unnecessary complexity.

8. Predictability: users will be able to form a mental model of how the system will behave in response to actions.

9. Interpretation: there are codified rules that try to guess the user intentions and anticipate the actions needed.

10. Accuracy: There are no errors, i.e. the result of user actions corresponds to their goals.

11. Technical Clarity: the concepts represented in the interface have the highest possible correspondence to the domain they are modeling.

12. Flexibility: the design can be adjusted to the needs and behavior of each particular user.

13. Fulfillment: the user experience is adequate.

14. Cultural Propriety: user's cultural and social expectations are met.

15. Suitable Tempo: the pace at which users works with the system is adequate.

16. Consistency: different parts of the system have the same style, so that there are no different ways to represent the same information or behavior.

17. User Support: the design will support learning and provide the required assistance to usage.

18. Precision: the steps and results of a task will be what the user wants.

19. Forgiveness: the user will be able to recover to an adequate state after an error.

20. Responsiveness: the interface provides enough feedback information about the system status and the task completion.

2.5 Paper Prototyping

Overview

Paper prototyping is a usability testing method to obtain user feedback early in the development process. Typically, developers or designers create paper or software-based simulations of user interface elements in a static or dynamic way (Maguire & Bevan 2002). The method promotes fast iterative development of a product/tool as well as fostering close interactions between designers, developers and end-users. Synder (2003) describes paper prototyping as "*a subset of participatory design*" (14) emphasizing their similar approach towards end-user involvement and iterative development. First, a paper prototype of a developed tool is crafted and then secondly, the paper mock-up is given to a test user who is asked to complete a task in order to detect difficulties or pitfalls as the test user navigates through the prototype. Advantages of paper prototyping include the detection of usability problems quite early in the development process, the communication between designers and users is promoted as well as the minimal resources and materials required to build a first paper prototype[6].

The paper prototype can be a paper mock-up simply drawn by hand (see figure 6) or screens developed by using a tool, such as Balzamiq or (see figure 7). The prototype can also be a mixture of a screen shot and hand drawn additions.

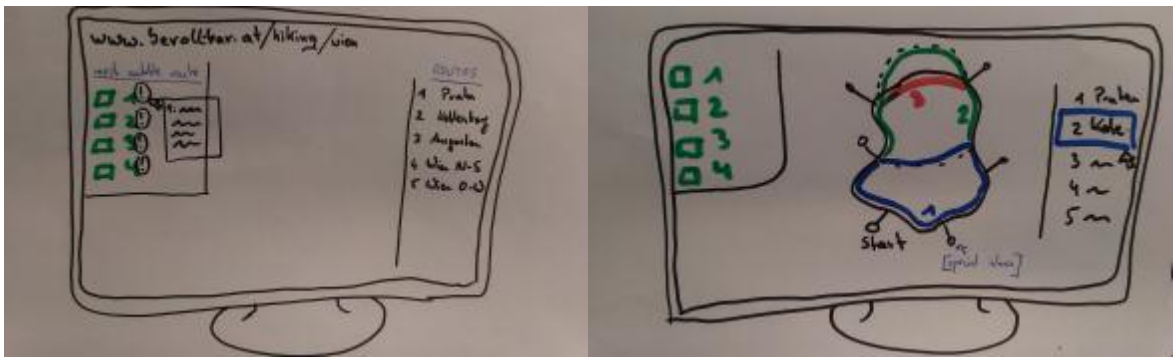


Figure 3 Hand-drawn paper mock-ups for a website, start screen (left picture) and second screen (right picture)

[6] see http://www.usabilitynet.org/tools/prototyping.htm

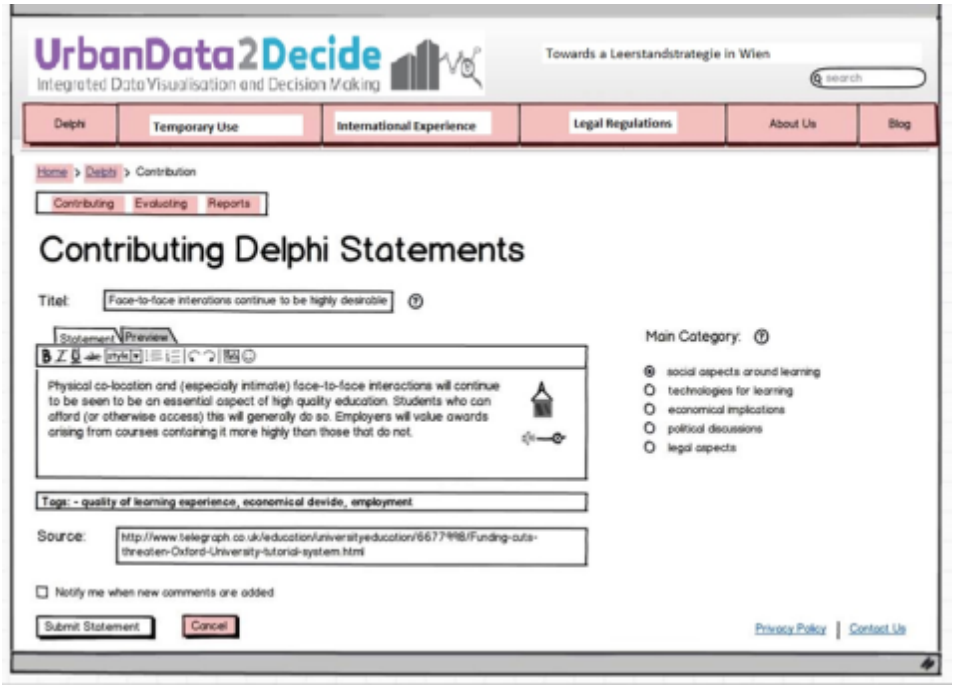


Figure 4 Paper mock-up created with Balsamiq tool; start screen for an online Delphi Study

Both pictures above show screen shots of the tools to be tested including the start screen, menus, dialog boxes etc. A paper prototype should show all the functions to enable users to fulfil specific tasks, e.g. upload a photograph on a website. During the testing phase users "click" by touching the prototype buttons or links and "type" by writing their data in the prototype's edit fields[7].

Some authors suggest differences between using computer- based and paper-based low-fidelity prototypes in regards to making users feel more comfortable to critize the prototype. In their comparative study Sefelin and colleagues (2003) concluded that "*paper- and computer-based low-fidelity prototypes lead to almost the same quantity and quality of critical user statements*" (Sefelin et al. 2003, 779) though users preferred computer prototypes, this does not significantly influence their criticism and feedback.

Possible findings from paper prototyping include the following topics

- ***Concepts and terminology***
  Do the target users understand the terms you've chosen? Are there key concepts they gloss over or misconstrue?

7 http://www.csee.umbc.edu/courses/undergraduate/345/spring12/mitchell/readings/paperPrototyping_Snyder.pdf

- ***Navigation/workflow***
  If there's a process or sequence of steps, does it match what users expect? Do they have to keep flipping back and forth between screens? Does the interface ask for inputs that users don't have, or don't want to enter?

- ***Content***
  Does the interface provide the right information for users to make decisions? Does it have extra information that they don't need, or that annoys them?

- ***Page layout***
  Although your scribbled screens may not be pretty, you'll still get a sense of whether users can find the information they need. Do you have the fields in the order that users expect? Is the amount of information overwhelming, not enough, or about right?

- ***Functionality***
  You may discover missing functionality that users need, or functionality you'd planned but users don't care about.

Due to the multitude of findings listed above typically both high-, and low-level issues and questions arise during paper prototyping. A high-level issue would be questioning the acceptance of the tool by the market or user group; whereas low-level issues concern feedback regarding specific functionalities, like single buttons.

Paper prototyping may not be the preferred usability testing method if the following issues should be detected: technical feasibility/capability; download time or other response time; scrolling, or colors and fonts.

Materials, Moderation Rules, Single Steps[8]

Four stages of paper prototyping can be required:

1. **Concept design**
   to explore different metaphors and design strategies, sketch out possible approaches in a brainstorming environment, evaluate the extent to which each approach meets the usability requirements and objectives agreed in the stakeholder meeting

2. **Interaction design**
   to organize the structure of screens or pages, document the sequence in which user tasks will make use of each set of post-it-notes and review how the screens/pages can be

[8] see http://www.usabilitynet.org/tools/prototyping.htm

organized to simplify user tasks

3. **Screen design**
   for initial design of each individual screen, ask the user to carry out a realistic task (based on the context of use and scenarios), as the user selects options on each screen, the developer explains what happens, and either points to the next screen or presents the next screen to the user (without giving any hints).

4. **Screen testing**
   to refine the screen layout, to test more detailed interaction, prepare pieces of paper with menus, scroll boxes, dialogue boxes, etc., and present these to the user; the user simulates pointing and clicking using a pencil, and simulates typing by writing on paper.

2.6 System Usability Questionnaire

In comparison to other usability methods discussed in the previous chapters, system usability questionnaires focus more on the subjective satisfaction of users with the interface (Vukocav et al. 2010, 273). Whereas, other methods such as the cognitive walkthrough puts emphasis on specific parts of an interface (by simulating step-by-step user behaviour) the questionnaire is filled out by participants themselves and enables a larger sample.

Some examples of what a system usability questionnaire can look similar are provided below

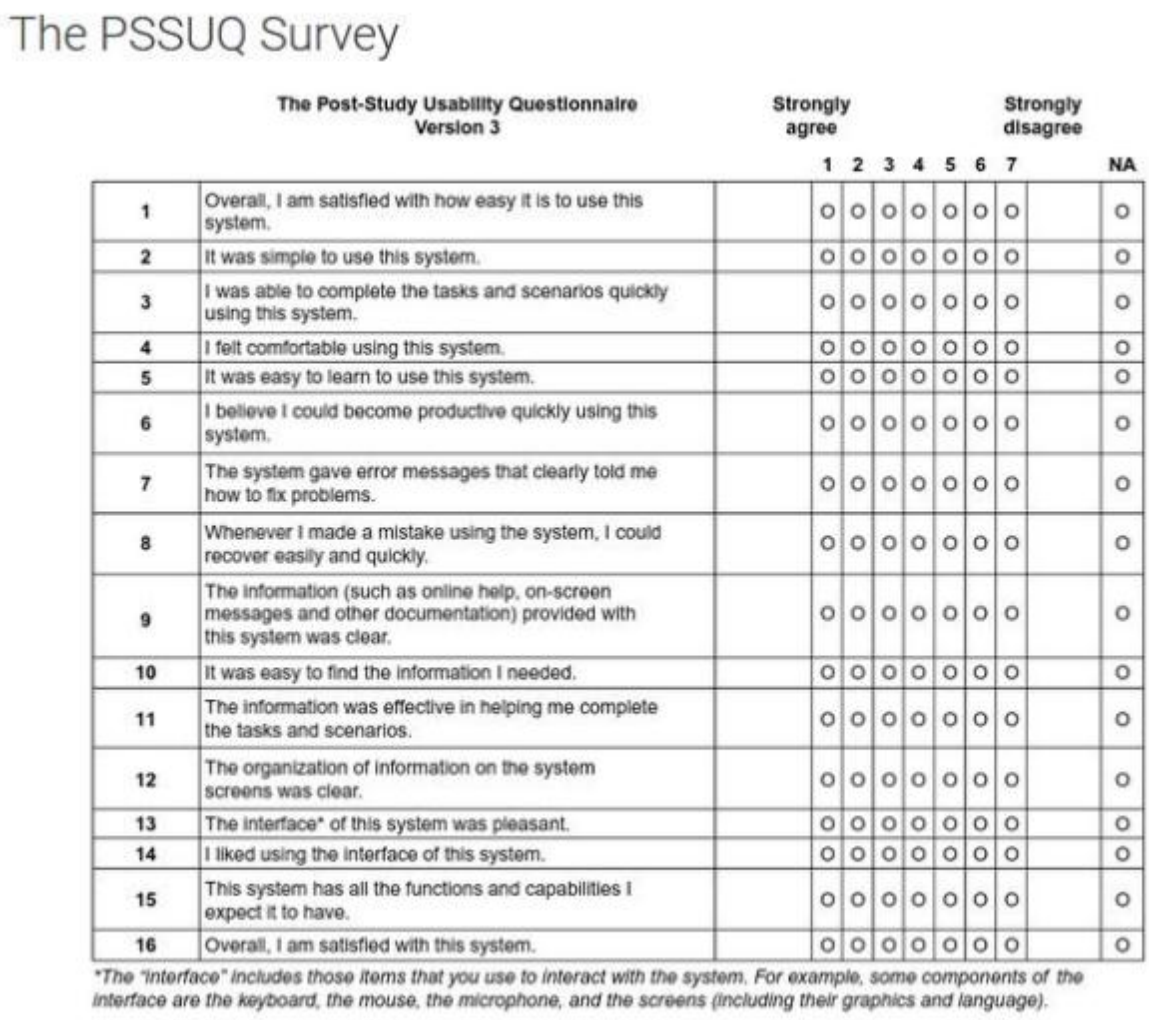
Figure 5 Example: System Usability Questionnaire[9]:

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Mean Rating	Percent Agree
Thought Website was easy to use			1	12		3.9	92%
Would use website frequently			2	6	5	4.2	85%
Found it difficult to keep track of where they were in website	3	6	3	1		2.1	8%
Thought most people would learn to use website quickly			5	8		3.6	62%
Can get information quickly		1	2	8	2	3.9	77%

[9] example template sheet retrieved from: http://www.usability.gov/how-to-and-tools/resources/templates/report-template-usability-test.html

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Mean Rating	Percent Agree
Homepage's content makes me want to explore site		1	5	2	5	3.9	54%
Site's content would keep me coming back			2	6	5	4.2	85%
Website is well organized			5	6	2	3.8	62%

Figure 6 Example: Post-Study Usability Questionnaire (PSSUQ)



The PSSUQ Survey

	The Post-Study Usability Questionnaire Version 3		Strongly agree 1	2	3	4	5	6	Strongly disagree 7		NA
1	Overall, I am satisfied with how easy it is to use this system.		O	O	O	O	O	O	O		O
2	It was simple to use this system.		O	O	O	O	O	O	O		O
3	I was able to complete the tasks and scenarios quickly using this system.		O	O	O	O	O	O	O		O
4	I felt comfortable using this system.		O	O	O	O	O	O	O		O
5	It was easy to learn to use this system.		O	O	O	O	O	O	O		O
6	I believe I could become productive quickly using this system.		O	O	O	O	O	O	O		O
7	The system gave error messages that clearly told me how to fix problems.		O	O	O	O	O	O	O		O
8	Whenever I made a mistake using the system, I could recover easily and quickly.		O	O	O	O	O	O	O		O
9	The information (such as online help, on-screen messages and other documentation) provided with this system was clear.		O	O	O	O	O	O	O		O
10	It was easy to find the information I needed.		O	O	O	O	O	O	O		O
11	The information was effective in helping me complete the tasks and scenarios.		O	O	O	O	O	O	O		O
12	The organization of information on the system screens was clear.		O	O	O	O	O	O	O		O
13	The interface* of this system was pleasant.		O	O	O	O	O	O	O		O
14	I liked using the interface of this system.		O	O	O	O	O	O	O		O
15	This system has all the functions and capabilities I expect it to have.		O	O	O	O	O	O	O		O
16	Overall, I am satisfied with this system.		O	O	O	O	O	O	O		O

*The "interface" includes those items that you use to interact with the system. For example, some components of the interface are the keyboard, the mouse, the microphone, and the screens (including their graphics and language).

Figure 7 Example: User Interaction Satisfaction[10]

1. **Tasks can be performed in a straightforward manner**:
Never Always
1 2 3 4 5 6 7
2. **Organization of information on the site**:
Confusing Very clear
1 2 3 4 5 6 7
3. **Use of terminology throughout the site**:
Inconsistent Consistent
1 2 3 4 5 6 7
4. **During the session, the test administer appeared to be**
Unfriendly Friendly
1 2 3 4 5 6 7
5. **Information displayed on the screens**:
Inadequate Adequate
1 2 3 4 5 6 7
6. **Census Bureau-specific terminology**:
Too frequent Appropriate
1 2 3 4 5 6 7
7. **Characters on the computer screen**:
Hard to read Easy to read
1 2 3 4 5 6 7
8. **Learning the site**:
Difficult Easy
1 2 3 4 5 6 7
9. **Experienced and inexperienced user's needs are taken into consideration**:
Never Always
1 2 3 4 5 6 7
10. **Finding what you were looking for**:
Difficult Easy
1 2 3 4 5 6 7
11. **During the session, the test administer acted in the following way**
Unhelpful Helpful
1 2 3 4 5 6 7
12. **Forward navigation:**
Impossible Easy
1 2 3 4 5 6 7
13. **Backwards navigation:**
Impossible Easy
1 2 3 4 5 6 7
14. **Overall reactions to the site:**
Terrible Wonderful
1 2 3 4 5 6 7
Frustrating Satisfying
1 2 3 4 5 6 7
Difficult Easy
1 2 3 4 5 6 7
15. **Please add any additional comments:**

[10] www.reginfo.gov/public/do/DownloadDocument?objectID=11471801

2.7 Participatory Design

"Participatory design is a process that involves developers, business representatives, and users working together to design a solution. It actively involves users in the design process to help ensure that the product designed meets their needs and is usable in the process."[11]

From its very beginnings, Participatory Design (PD) in organizations aimed at balancing power relations in order to ensure technology serves all stakeholders. Hence PD included the design of company policies as much as laws at the national level (Kensing & Blomberg, 1998). In a 'design for design' scenario, users need to find areas where their combined feedback can actually influence the future design. The methods of PD still apply, interviews, user diaries and visual artifacts (photos or drawings) bring people in and make their voices heard. However, what changes is the increased need for the systematic aggregation and analysis of user experiences to reflect the global nature of systems' user bases. In fact, design becomes a process of re-purposing existing designs and 'design-for-use' becomes 'design-in-use', that is, future use situations can be seen as design situations (Ehn, 2008).

One of the underlying motivations for participatory design (PD) is the need to align the design of ICT applications with the skills and conceptual models of the people who should not only use them but be effectively supported in their activities. When PD originated in Scandinavia in the 1970s, it went beyond productivity goals and explicitly aimed for better work-live balances. PD emphasized the need of users' perspectives, not only to safe time but also to make the software 'work' in people's overall life context (Kensing & Blomberg, 1998). Participatory design was to overcome the efficiency paradigm and include factors such as people's motivations, their carrier plans or their ideals of teamwork. These goals were new in times when workers had little influence on the technologies they should use afterwards. Paternalistic attitudes towards users of technology who were deemed unfit to understand the workings of technology characterized the introduction process of new technologies (Spinuzzi, 2005).

Recognizing the crucial importance of users' practice-based knowledge, PD put forward the argument that involving workers early on, in fact taking workers' interests as starting points, was good design practice if changes were to be sustainable (Kensing & Blomberg, 1998).

[11] http://www.usabilitybok.org/participatory-design

Picture 3 Picture from a participatory design workshop[12]

Eventually PD is also evaluated against the degree of empowerment people get from a process that takes on their goals and values. A process that is highly iterative, integrating feedback on design changes and their impacts (Spinuzzi, 2005).

"It is important to note that, while the users are a valuable source of information and ideas in Participatory Design, they are not allowed to make end decisions and are never empowered with the tools that the experts use. Participatory Design sessions are simple exercises in which we give our users the tools to create and design mockups of software or products they would love to use in the "perfect world" scenario while also asking them to explain why they built their perfect software or a product in that particular way."[13]

3 USABILITY EVALUATIONS

For the usability evaluation two methods were chosen, namely heuristic evaluation (see 2.3) and cognitive walkthrough (see 2.2).

The usability of three tools: ODI Train Usage Data Tool, ODI Open Data TFL tool and the SYNYO Emergency Data map, developed within the 'UrbanData2Decide' project have been tested throughout September and October 2016. The following two chapters (3.1 and 3.2) briefly discuss the two usability methods focusing on the materials and the process of the tool testing. The other chapters (3.2, 3.3 and 3.4) analyze the results of the usability tests according to each tool. Firstly, the tasks and scenarios given to the testers are described, and secondly, the findings are analyzed and recommendations for tool developers given.

[12] http://www.participatory-design.com/

[13] http://www.uxpassion.com/blog/participatory-design-what-makes-it-great/

3.1 Methodologies Used: Heuristic Evaluation and Cognitive Walkthrough

3.1.1 Heuristic Evaluation

In a heuristic evaluation, usability experts evaluate a site's interface using a set of accepted evaluative principles. During the evaluation a group of usability experts "*inspect a user interface to find and rate the severity of usability problems using a set of usability principles or heuristics*" (Vukovaet al. 2010, 273). Between three to five experts are recommended for a thorough evaluation of a site's interface or a tool. Ratings based on the opinion of three evaluators are considered reliable (Forsell 2014, 186). Commonly, each evaluator works independently and discusses the findings afterwards. The evaluators need to be equipped with a certain task to perform on the website or interface. Each task is then evaluated according to a pre-set list of usability heuristics. For this evaluation the ten commonly known heuristics by Nielsen and Molich (1994) were used – but slightly adapted. Given that a heuristic evaluation is highly dependent on the expertise of the evaluators we opted to slightly adapt the heuristics. Whereas the ten heuristics are originally described with a short paragraph we opted for questions that the participants had to rate according to a severity scale (see Figure 12 below)

1) Cosmetic problem – not necessary to fix, unless time allows
2) Minor usability problem – low priority, fixing it should be given low priority
3) Major usability problem – important to fix, should be given high priority

Figure 8 Scale of severity stages

In addition to the questions, there was also space on the working sheet for other comments and questions not covered by the ten given heuristics.

For instance the heuristic '**Visibility of system status'** was translated into the following two questions:

a) Do you know where to go next in the navigation?

b) Is it clear if the content rendering of a page is completed?

Process and Materials: All heuristic evaluation sessions took place online, participants joined a Skype call. Participants received an email beforehand; including two documents (see Annex 6.3 and 6.4)

1. **a hand-out** with information of methodology, access information for the tool to be tested, steps and time plan of the evaluation

2. **a working sheet** where participants filled in their results and later sent the Word Document electronically to the facilitators afterwards

Participant selection (for the heuristic evaluations and cognitive walkthroughs): The evaluators were chosen according to their expertise and relation to the tool, i.e. people who were involved or part of the organization, which developed a tool, were not chosen as evaluators. Between three to six evaluators took part in each of the three usability sessions.

The online sessions lasted about one and a half hours. All sessions started with a short introduction round where participants shared their background and their previous experiences with usability testing of applications. The testing itself included two steps:

- *Heuristic Evaluation training* (app. 20minutes): facilitators gave evaluators the needed domain knowledge and information on the tasks and scenarios on some PowerPoint slides using a shared screen.

- *Evaluation* (app. 1hour): after the short training, the evaluators had about an hour to execute the tasks at hand and fill out the working sheet. Meanwhile the facilitators were on standby on Skype in case questions or problems occurred during the evaluation phase.

3.1.2. Cognitive Walkthrough

Cognitive walkthroughs enable the analysis of a user interface by simulating step-by-step user behavior for a predefined task. Firstly applied in 1990, the method was originally applied to evaluate the usability of physical objects like postal kiosks or ATMs. Nowadays it is widely used also in areas such as software or app development. The cognitive walkthrough methodology focuses on one aspect: **ease of learning**. Based on the theory of learning, during a cognitive walkthrough participants are asked to perform certain tasks using the tool to be tested. It makes the **learnability of a system** for new users explicit.

While testers perform certain tasks they are encouraged to say their thoughts out loud (**Usability Testing Essentials** 2011, 19). The experiences of a user with the product is often better understandable, if the tester doesn't only see but also hears about the user's reactions, ideas or frustrations.

Process and materials: All three tools (ODI Train Usage Data tool, ODI Open Data TFL and SYNYO Emergency Map) were tested with face-to-face cognitive walkthrough sessions in Vienna. Altogether three participants took part in testing all of the tools.

1. After a short introduction to the procedure and steps of the testing session, an important principle of a cognitive walkthrough, namely "**let the participant speak!**" was stressed again. Thus the participant is aware that the moderator will try to talk as little as possible. Within one session all three tools were tested with short breaks after each tool.

2. Each session lasted between an hour to an hour and a half. In regards to documentation of the walkthrough sessions, the moderator had problem-reporting sheets that beside problems also included the reporting of success stories, design suggestions, and other problems that were not the direct output of the walkthrough.

3.2 Train Crowding Data Tool/ London (developed by ODI)

Tool Description

The Train Crowding Data tool visualizes crowding data for train carriages due to arrive at a particular Victoria Line station (north- and southbound trains) at a given date and time, together with simulated data for subsequent trains. As applicable in Figure 12. the visualization shows crowding data for the next four incoming trains at a selected station.

Figure 9 screenshot of train data tool

The heatmap functionality (see Figure 13.) simulates, visualises, and demonstrates the occupancy of the Victoria line network at a given date and time. The visualization is based on both actual data and simulations based on it.

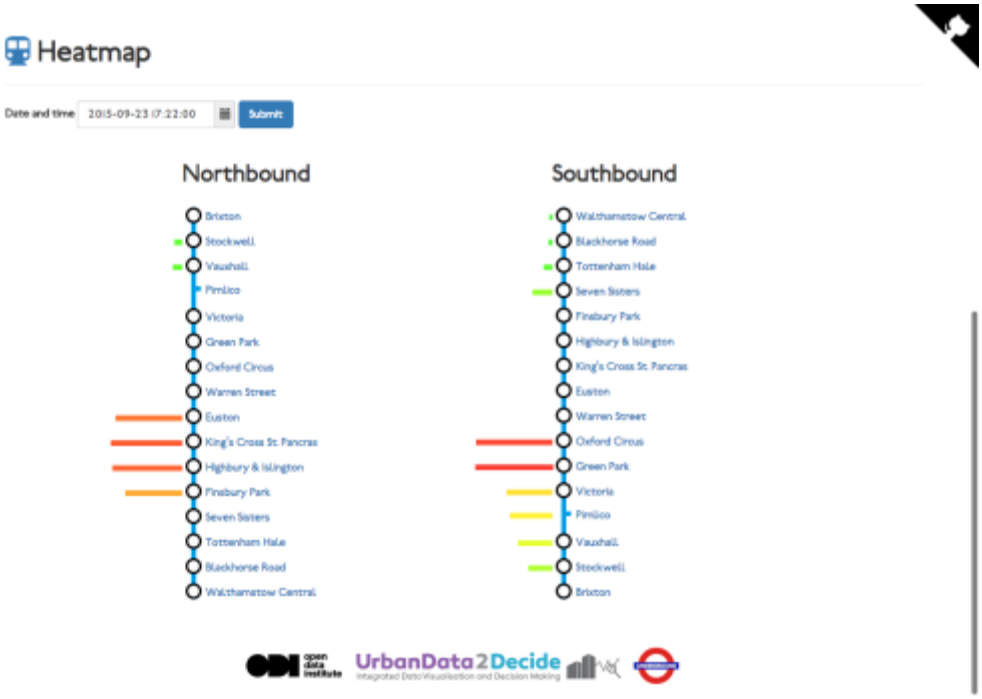


Figure 10 screenshot of the heatmap function

Heuristic Evaluation

1) Tasks

As described earlier (see chapter 3.1.1) the evaluators were given a handout which also included a brief contextualization of the tool and the given tasks.

> ***Context:*** *The site is targeting developers and public transport operators who want to make use of open train data. The site contains crowd data of trains operated by the London Underground.*

For the train crowding data visualization tool the following tasks were formulated:

- **Task 1**: You want to get on a train at *King's Cross St. Pancras* station. Go to "Crowding Data". Check the information for the incoming trains at *King's Cross St. Pancras* station to see which train is the least crowded.
- **Task 2:** Go to the function "Heatmap" to see which of the station going Southbound had the trains with the least number of passengers arriving on 1st January, 2PM (14:00).
- **Task 3:** You are getting on the train at Oxford Circus station regularly. You are interested to see the data for yesterday. Go to "Heatmap". Oh no sorry, you are wrong. Go back to the main page and go to 'Crowding Data' again. Then look for the data for arriving trains at Oxford Circus station on the 25th January 2016, around 10AM (10:00).

2) Analysis of Results

- **Visibility of system status, e.g.**
  - Do you know where to go next in the navigation?
  - Is it clear if the content rendering of a page is completed?

The navigation of the tool didn't pose a problem to the evaluators. However, whereas the rendering sign of a page was clear the loading time was very long "*you have got to have a lot of patience*" (It took more than one minute and sometimes up to 5 minutes for a page to finish loading).

- **Match between system and the real world?**
  - Do you understand the terms used on the website/the tool? Labels, Headings, Explanations etc.
  - Do you understand the meaning of the icons?

- **User control and freedom**
  - Do you know how to return to the main page / 'home' function ?
  - Is the 'home' function available on every page?

- **Consistency and standards**:
  - Do symbols and labels repeat?

- Are existing standards for symbols / metrics used? (Home = House; Help = Question mark)

Two of the evaluators suggested that the home button (see Figure 14.) should be a house not a train (as it is the case currently) – this was regarded as a minor and a major usability issue and was mentioned several times also regarding other heuristics than 'match between system and real world', e.g. 'consistency of symbols'. A house as the symbol for the home function is a broadly used symbol and users search for this symbol intuitively when visiting a website. Generally, other symbols and labels used in the tool didn't pose a problem to the evaluators.

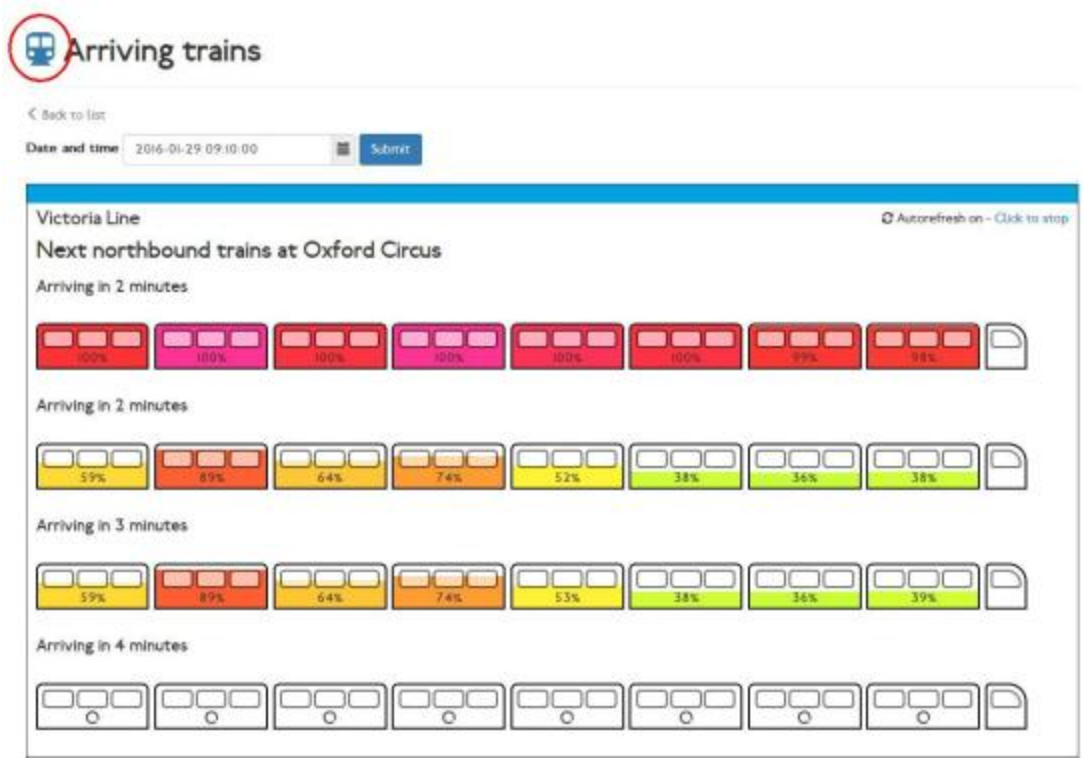


Figure 11 screenshot showing 'home button' (train symbol at the top left corner)

Two evaluators positively pointed out that the same font and colour line was used as the official London tube uses. Furthermore the length and colour of bars provided in the Heatmap function was regarded as easy to understand and to interpret.

- **Error prevention:**
  - Are there sufficient error messages?
  - Do you understand messages trying to prevent you from entering invalid data?

The fact that there are no error messages provided was considered as a major usability problem, because despite relatively long (page) loading times and no changes in the percentages of the train wagons, no message was provided; instead only the loading icon continued to be shown.

Similarly, when certain times were selected (e.g. data for 2AM on Victoria line), no valid data was shown, but still no error message appeared. A suggestion by an evaluator was to better and more clearly indicate that 'there is no valid data' or 'users can't select a certain hour or time' e.g. time

when the tube doesn't operate. A similar comment from another evaluator pointed to the fact that it is currently not possible to see data for the future, however the user is still able to choose a time and date in the future.

- **Recognition rather than recall**:
  - Is the structure of topics clear and logical for you?
  - Is the structure of information clear and logical for you?
  - Is the structure of actions you can choose clear and logical for you?

The structure of topics, information and actions was clear to all evaluators and didn't cause problems. The structure of actions, e.g. select a subway station and change date and time seemed logical for the evaluators; thus a step-by-step guide (as applicable within other websites) doesn't seem necessary for this tool.

- **Aesthetic and minimalist design**:
  - Is the information provided on the website precise (e.g. correct and specific)?
  - Is the information provided on the website too extensive or too sparse?

Three evaluators rated the aesthetic and design as a major usability problem, mainly because the information provided on the website was considered too sparse. Evaluators suggested that a short info paragraph would help to better understand what the page is all about. A comment from an evaluator:

"*Generally, the information provided seems incomplete or incorrect: relevant information is not shown because loading time takes too long; although auto fresh is on, and arrival timings of the next train kept updating the relevant data (percentile point of crowdedness) stayed the same; regarding heatmap function: suggestion to not only have the coloured bars but more information, e.g. medium green = 30-39% crowdedness*."

Help users recognize, diagnose, and recover from errors:

- Do you understand why an action was erroneous?
- Do you understand how to solve the problem?

Similarly to issues raised for 'error prevention' the fact that there are no error messages was considered as a major usability problem. One evaluator explained that

"*I did not understand why an action was erroneous, and how to solve it when loading for Task 1 and 3 took too much time, and when percentile point of crowdedness was not updated at all. I was able to see relevant data for the 1st and 2nd train in Task 1 and 3 at my default times (= 2016-01-29 15:00). However, when I tried different times (e.g., 2016-01-30 09:17:00, and 10:17:00), there was no data shown at all*."

Help and documentation:

- Do you find help or contextual explanation (i.e. to explain specific words or steps) where necessary?

Although the main page of the tool contains short information on each tool, only one evaluator found this function. The evaluator rated 'Help and Documentation' as a cosmetic usability problem and thought the "Show documentation" was helpful.

One evaluator commented that in order

"to understand the context of these datasets and visualisation (e.g., only victoria line is available). I started the tasks before reading 'documentation', and thought information is not useful for developers because other lines such as Bakerloo and Central line for Oxford Circus station, and Northern, Piccadilly, and Metropolitan line for King's Cross St. Pancras ones were not available at all."

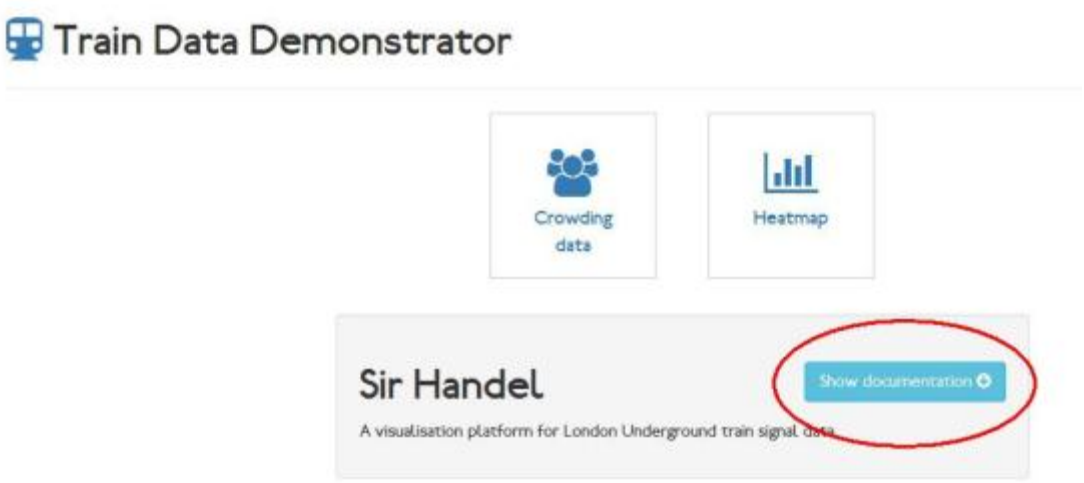


Figure 12 screenshot of 'Show documentation' function (encircled in red)

For other evaluators the apparently non-existing documentation and information was rated as a major usability problem as there "*could be more explanation on website, intro to the sites and its goals would be helpful*." Thus, the "Show documentation" function could be made more clearly visible and possibly renamed in order make it more prominent on the main page.

Cognitive Walkthrough

1) Tasks:

- Task 1: Go to "Crowding Data". Check the information available for the Northbound trains at *King's Cross St. Pancras station* to see which train is the least crowded. (Note: for now please ignore date and time).
- Task 2: You are getting on a southbound train at *Oxford Circus station* regularly. You are interested to see the data for this station on the 25th January 2016 10AM. Go back to the main page and go to 'Crowding Data' again.

- Task 3: Go to the function "Heatmap" to see which station on the southbound line has the train with the least number of passengers arriving on 1st January, 2PM (14:00).

2) Analysis of Results:

Figure 13 screenshot of the date and time entry function (encircled in red)

Similarly to results of the heuristic evaluation, the oftentimes very long loading times were negatively mentioned. One tester thought that the train symbol used as the home button is good and intuitive. The automatic "autorefresh" function, as one evaluator suggests "*can be confusing if current data changes all the time, thus it can be made more explicit to de/or activate this function*."

In regards to the Heatmap looking at lines indicating the crowdedness for trains main issues were:

One issue was the meaning of the line next to the trains which was unclear to many participants (see Figure 18.)

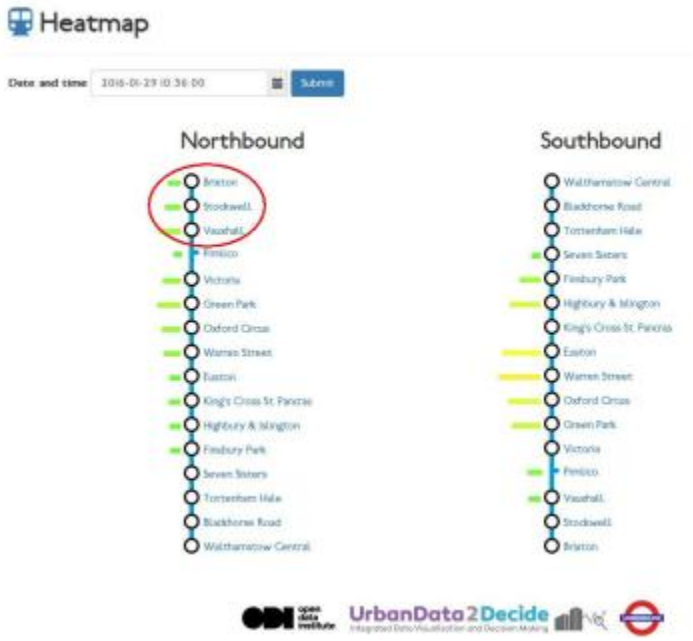


Figure 14 screenshot of heatmap showing train stations (encircled in red)

It turned out that the current visualizations are not enough to detect the 'least number of passengers' or differences between trains (see stations circled in red above), thus numbers or percentages could be added next to the stations.

3) Recommendations and Conclusions

The results of both usability tests – the heuristic evaluations and cognitive walkthroughs – were analysed in the previous paragraphs. The most pertinent and commonly raised problems and suggestions for improvements, including single functionalities that should be renamed, or put somewhere else, etc. are listed below.

- The current home symbol (a train) should be changed to a house symbol
- The loading times are very long (sometimes more than 5 minutes)
- A clearly visible (de-)activate function for autorefresh function should be available
- An introduction to the purpose of the page should be added or the 'show documentation' function needs to be more clearly visible and possibly renamed

\+ **Positive feedback:**

- Same font and colour used as real TFL trains
- No need for shortcuts or guidelines for website

3.3 Open Data TFL / London (developed by ODI)

Tool Description

The Open Data TFL tool visualizes open Transport for London (TFL) train data. The site contains a multitude of different data (e.g. passenger load, car temperature etc.) that the user can select according to a time period and also compare different data sources. The site is mainly targeting developers who want to make use of open train data.

The visualization below (see Figure 14) shows data for the occupancy of a specific passenger car during a certain time period. This allows the signal data to be viewed and evaluated as a time series graph. There is also an API that exposes signal data for a given time period in JSON format.

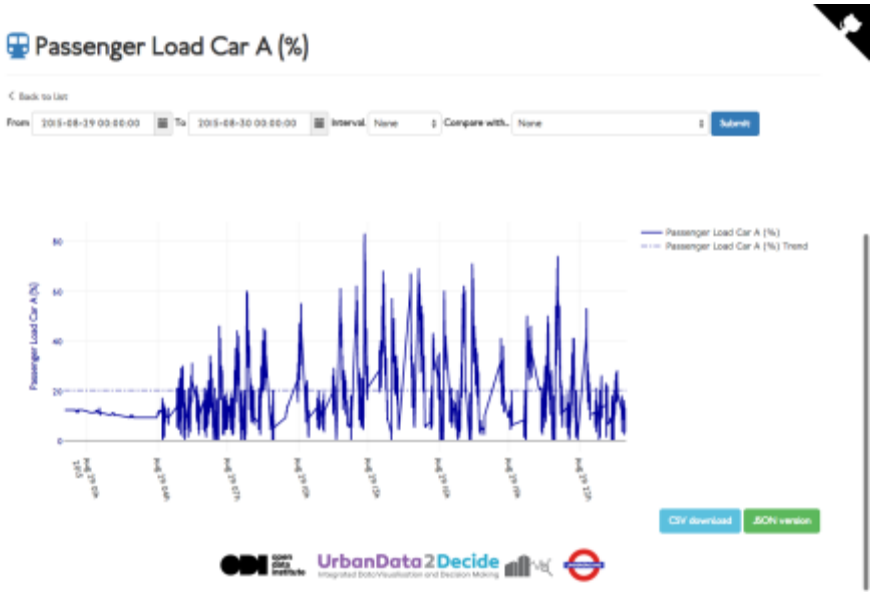


Figure 15 screenshot of signal graph tool

Heuristic Evaluation

1) Task

- Task 1: Let's concentrate on ' Customers and Enthusiasts Dashboard'. Gain an overview of this category. Imagine you want to write an App about 'Passenger Experience' using the London Underground. Now, check whether you can find any useful open data. (10 - 15 min)

2) Analysis of Results

- **Visibility of system status, user control and freedom** and c**onsistency and standards**:
    - Do you know where to go next in the navigation?
    - Do you know how to return to the main page?
    - Is the 'home' function easy to find?
    - Is the 'home' function available on every page?
    - Do you know how to get back to the last page?

For almost all evaluators it was unclear what home is, i.e. how to return to the main page (see Figure 19.). It was not clear how to return to the main page other than by using the arrow. One participant suggested that it would be better to have a home symbol rather than train symbol for the 'return to main page' function.

Figure 16 screenshot of 'home button' and 'Back to list' symbol on top left corner

Another evaluator suggested referred to changing the name '*back to list*' to '*back to overview*'. Additionally the lack of navigation guidance was rated once as a major usability problem and once as a usability catastrophe. The content rendering, i.e. loading signal of a page was clear to all evaluators.

- **Match between system and the real world?**
    - Do you understand the terms used on the website/the tool?

For most evaluators the meaning of most terms used on the website was not clear, the common ratings were major and minor usability problem. Especially for non-trained users terms like ATO or HSBC are not clear. Evaluators who rated this heuristic as "no problem" still suggested adding explanations when a user places the cursor on terms like '*ATP worst case forward location Offset*'.

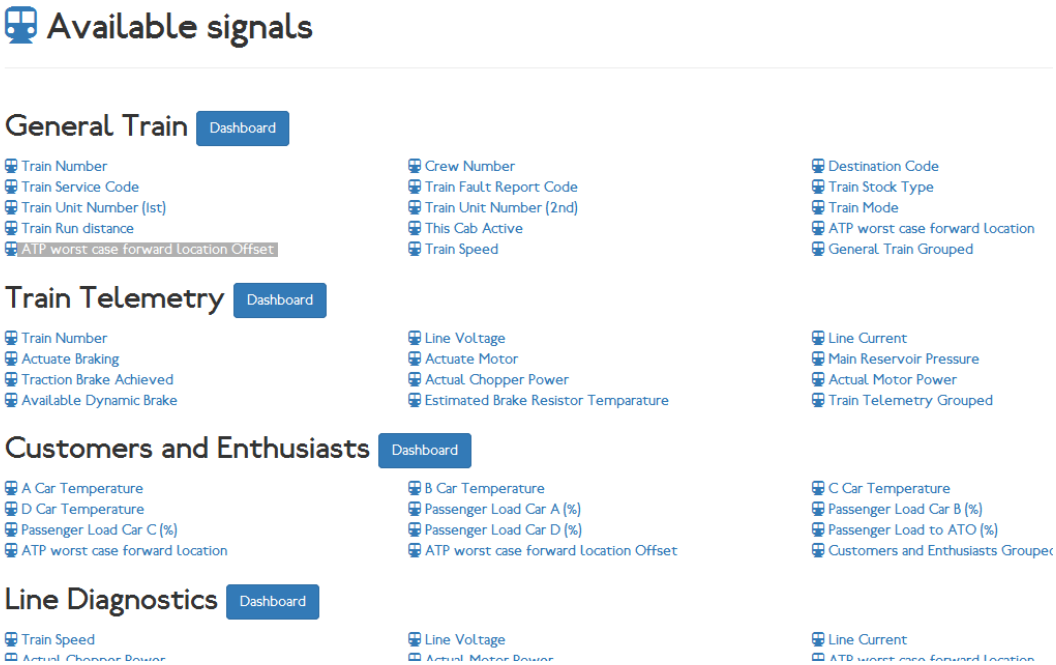


Figure 17 screenshot of main page of the tool showing list of data types

- Do you understand the meaning of the icons?

Whereas the icons shown on right corner of each visualization were clear to most evaluators, some icons as commented by participants don't have a direct functionality, e.g. '*produced with plotly*'. Also, some icons like '*show closest data on hover*' are not intuitively clear, and when clicking on the icon there is no action. Thus the available symbols should be explained.

Error prevention and help users recognize, diagnose, and recover from errors:

- Do you understand messages trying to prevent you from entering invalid data?
- Do you understand the error message?
- Do you understand how to solve the problem?

The fact that there was no error messages produced was confusing to most evaluators. One evaluator rated the non-existing error prevention as a minor usability problem as there were no messages shown when data was missing, i.e. a data source file shown on the list contained no data (see figure 21).

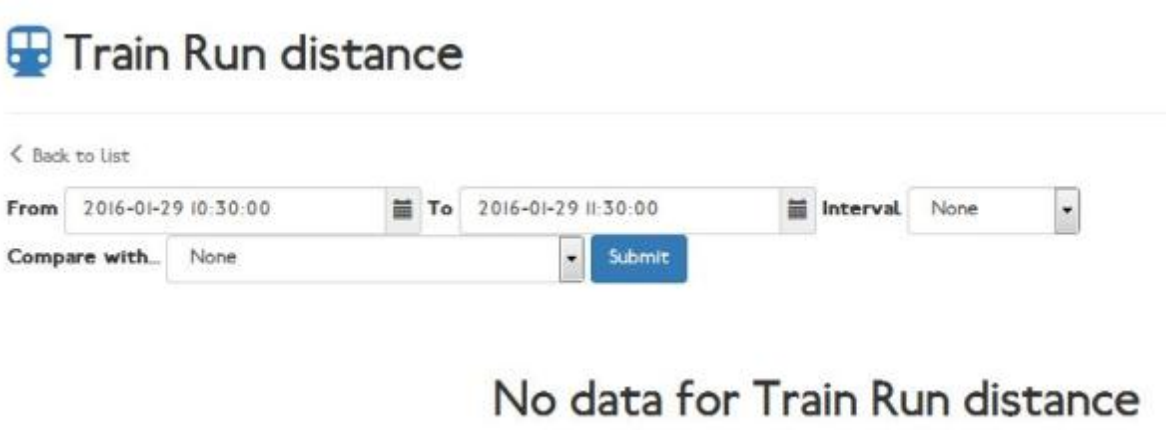


Figure 18 screenshot of empty data category 'Train Run distance'

Another evaluator reported that when changing the time intervals manually the display doesn't correspond, and although this is a clear error no error message was received (rated as major usability problem). When trying to download a csv files, no error message was given, which was rated by a participant as a minor usability problem.

- **Recognition rather than recall**:
    - Is the structure of topics clear and logical for you?

For three evaluators the structure and arrangement of data was not entirely clear (rated as a minor usability problem). Suggestions to improve the structure included adding subcategories to some topics and headlines which would help to understand, e.g. what is telemetry? Additionally, no data context is given to explain where the data is from etc. The fact that there is no purpose description of the page was rated as a major usability problem by two evaluators.

- **Flexibility and efficiency of use**:
    - Are shortcuts guiding you through the system available? (e.g. to not having to click through a user's manual every time)

      -If yes, do you find them useful?

      -If no, would you find it useful to have shortcuts while navigating the website?

For this heuristic no issues were reported, only one evaluator mentioned that there aren't a lot of shortcuts or guidance provided. But no suggestion was made to necessarily add a manual or shortcuts to the website.

- **Aesthetic and minimalist design**:
    - Is the information provided on the website precise?

Two evaluators commented that there could be more information on what the website contains, i.e. a purpose description is missing (rated as a minor usability problem). Another participant mentioned that

"generally, the information provided on the website is precise and specific. However, if the unit of temperature is provided, information would look more specific. After all, customers have a preferred or acceptable range of temperature. The specific temperature would be more useful for customers than the relative temperature difference between cars." (rated as cosmetic usability problem)

- Is the information provided on the website too extensive? (i.e. information overload)

Generally, the information provided on the website was rated as too scarce by some evaluators (rated as major usability problem).

One evaluator commented that

"Although there is an indicator of each line on the right side of the plot, it would be great if there are the full names of each line on the plot directly even when those names are a bit long (e.g., 'passenger load car A' instead of 'passenger …') About the csv file downloading, it was impossible to download them (error occurred) on a few pages. Plus, it will be greatly helpful to provide preview of 1-2 pages of those csv files to users to help them understand how those files look like. (cf. data.gov.uk sites provides preview of most datasets (excel, csv formats) online." (rated as cosmetic usability problem)

Help and documentation:

Other comments regarding the usability of the London Signal Tool included, to "*check with the ISO 9241 Parts 11, 110 and 151 on implementation guidelines especially with respect to navigation and data display - will drastically improve the solution outputs, i.e.:*

- *a sitemap for clear navigation structure (avoid dead ends like with potential "home" button)*
- *appropriate scaling of data diagrams in the dashboard (currently almost illegible)."*

Cognitive Walkthrough

1) Tasks

- Task 1: Let's concentrate on 'Customer and Enthusiasts Dashboard'. Gain an overview of this category on the 25th of January.
- Task 2: Imagine you want to write an App about 'Passenger Experience' using the London Underground. Check the data for car temperature. Now, check the data for car temperature on the 29th of January and explore the different intervals.

2) Analysis of Results

Similarly to results from the heuristic evaluation, it became evident that the legend should be described (more), as especially for novice users abbreviations like ATO or ATP are not clear. In addition, one participant suggested to add more categories instead of having so much data in "Ungrouped" on the start page (see Figure 21).

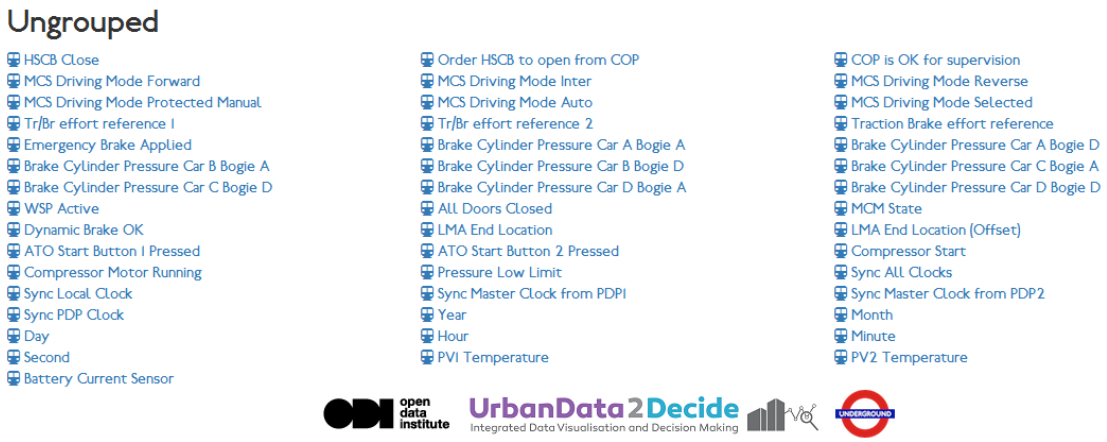


Figure 19 screenshot of data type in 'Ungrouped' category

In addition, someone mentioned that there is too much data shown and in relation very little visualization, she suggested that instead of so much data to "compare with" it might be more practical and functional to have less data but more clear visualizations, as the drop down list now is too long – and here again, the abbreviations are not clear. When evaluators tried to enter the given date (25th January) all had difficulties and needed several tries to figure out that despite the time being e.g. 10:30, all 6 digits had to be filled out (giving 10:30:00).

Further, the function to change the interval (e.g. 1 second to 1 day) didn't change anything in the current visualization, however no error message was provided. Whereas a more "experienced" or "technically inclined" user zoomed in and out using the mouse function, another participant had difficulties as the 'zoom in/out' function on the right bar atop the visualization in the task bar didn't work.

Similar, to what has already been discussed for the results of the heuristic evaluation; evaluators mentioned that the used terms (i.e. data types) could be explained more as it is not clear what they mean. Additionally, one evaluator repeatedly 'said out loud' that the visualization is not easily readable, e.g. in regards to the car temperature the following questions were raised:

- What do I actually see? Too small and unclear and what do the number means?
- Which scale is this? – 3360 – numbers, percentages?
- Is it Celsius or Fahrenheit OR train speed trend: average speed?
- How exactly was this measured?

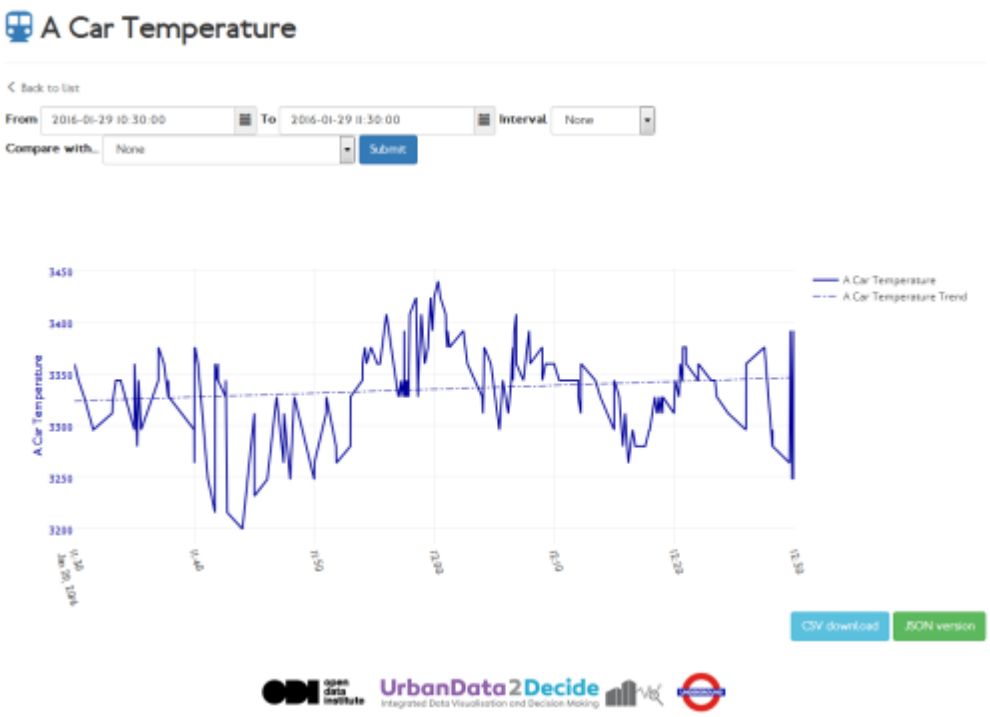


Figure 20 screenshot of visualization of A Car Temperature data

Further no description of x/y axis was found, and the submit button right next to the 'compare with' function "*has a misleading position*" as one participant mentioned.

Figure 21 screenshot of 'Submit' button

Recommendations and Conclusions

The results of both usability tests – the heuristic evaluations and cognitive walkthroughs – were analysed in the previous paragraphs. The most pertinent and commonly raised problems and suggestions for improvements, including single functionalities that should be renamed, or put somewhere else, etc. are listed below.

- A description of the visualizations should be added: what exactly does the user see?
- There is a strong need for providing error messages, especially if a data file is empty
- A short description of the meaning of the terms and abbreviations used should be added
- Some functions don't work, e.g. the top bar of visualizations
- A brief introduction to the purpose of the tool was suggested by almost all participants

3.4 Emergency Map in Vienna (developed by SYNYO)

Tool Description

The emergency map visualizes open emergency data in Austria, ranging from geo locations of fire departments to defibrillators. Users can select different types of emergency data in specific cities or rural areas. The site is targeting municipalities or NGOs who are working on urban disaster or safety management.

Heuristic Evaluation

1) Tasks

- **Task 1**: Explore the available data and name a few data sources you would find useful for an app that could support citizens in case of a national emergency.
- **Task 2**: Find a map of the fountains in Graz, Austria.
- **Task 3**: You are interested in whether emergency data also exists for Malmö, Sweden.

2) Analysis of Results

- **Visibility of system status, user control and freedom and consistency and standards**

- Do you know where to go next in the navigation?
- Do you know how to return to the main page / 'home' function ?
- Do symbols and labels repeat?
- Are existing standards for symbols / metrics used? (Home = House; Help = Question mark)

For all evaluators the home symbol (currently the 'UrbanData2Decide' Logo wasn't clear (rated twice as major usability problem and twice as minor usability problem) (see Figure X). Some thought that there is no home button at all, whereas others found the logo but suggested to use the common house symbol instead of the project's logo. The fact that these existing standards (e.g. house for home and question mark for help) symbols are not used was rated as a major usability problem twice.

Further, one evaluator mentioned that he used the 'UrbanData2Decide' Logo but got redirected to his home country instead of the current location of the map being Austria. This might have happened due to the IP address.

"Generally, navigation appears to be self-explanatory. However, there are some glitches and bugs: E.g. Even though having accessed the tool with a focus for Austria, clicking the home button brings me" to my home location. Is this due to my IP address?

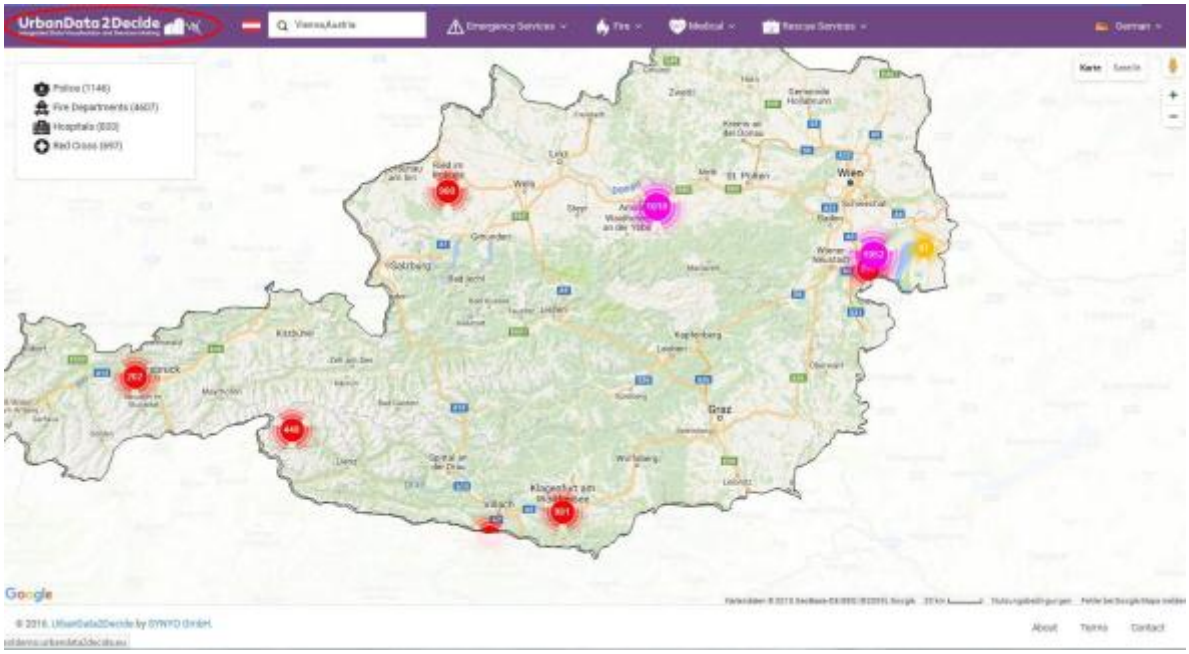


Figure 22 screenshot of tool main page and home button ('UrbanData2Decide' logo on top left corner)

Another issue raised by most evaluators was the deletion of selected data; it wasn't clear how to delete already selected data and whether this function even exists (rated as twice as major usability problem). All other symbols, e.g. emergency service symbols, on the top header were clear as well as the general navigation of the page, as other than zooming in and out of a map there are no pages to navigate to.

- Is it clear if the content rendering of a page is completed?

The content rendering was relatively clear to evaluators, although there is no explicit rendering sign, like e.g. a moving circle (rated twice as a minor usability problem, once as major usability problem as "*the rendering sign should be made clear*"). One evaluator suggested to better avoid the white shading over country borders and that the borders should match the national borders.

Figure 23 screenshot of country borders next to Salzburg, Austria

- **Match between system and the real world?**
  - Do you understand the terms used on the website/the tool? Labels Headings Explanations ...

Whereas most evaluators had no problems with the terms used, two evaluators commented that the language function didn't work and the naming of some emergency services were not clear. Someone didn't know what fire fountains were, and further suggested to combine rescue and emergency services categories.

"Broadly, used language and terms are clearly expressed and understandable. However, switching languages doesn't seem to have any effect on the terms displayed (i.e. all terms remain English anyway). Additionally, the categories to select data could be seen as overlapping (e.g. what other emergency services than police are available?). Additionally, I'd also say "emergency services" is a very broad category, encompassing firefighting, medical, etc. Hence, may re-label this as public safety and crime to create more distinctive categories? Regarding the sub-category "fire fountain", maybe relabel this as "water fountain" (a fire fountain, as far as I know, is a literally a lava fountain, I think…)"

  - Do you understand the meaning of the icons?

The meaning of the icons were clear to almost all evaluators, one suggested to increase the size of the symbols to make them better visible and another suggested to add information on the symbols when scrolling over them; also the relation between color of symbols and icons was not clear.

- **Error prevention and help users recognize, diagnose, and recover from errors:**

    - Are there sufficient error messages?
    - Do you understand messages trying to prevent you from entering invalid data?

The fact that there is no error message was rated as a cosmetic, a minor and a usability catastrophe among evaluators. Commonly all evaluators stressed the need for error messages, as now a user can only make assumptions as to why a function doesn't work, why data for e.g. Sweden is nonexistent or why some features are not available in some areas in Austria.

- **Recognition rather than recall**:
    - Is the structure of topics clear and logical for you and is the structure of information clear and logical for you?

The structure of topics was clear to most evaluators (rated as no problem twice), one feedback regarded overlaps of some filter categories (minor usability problem). Two evaluators mentioned that not all icons which are selected show on the legend on the left hand side and the logic for this isn't clear (rated as major usability problem twice). Additionally, the fact that in order to see data in detail a user has to zoom in a lot was mentioned negatively. For instance, if hospitals and police stations are selected a user gets some bubbles with numbers referring to both items (see Figure X).

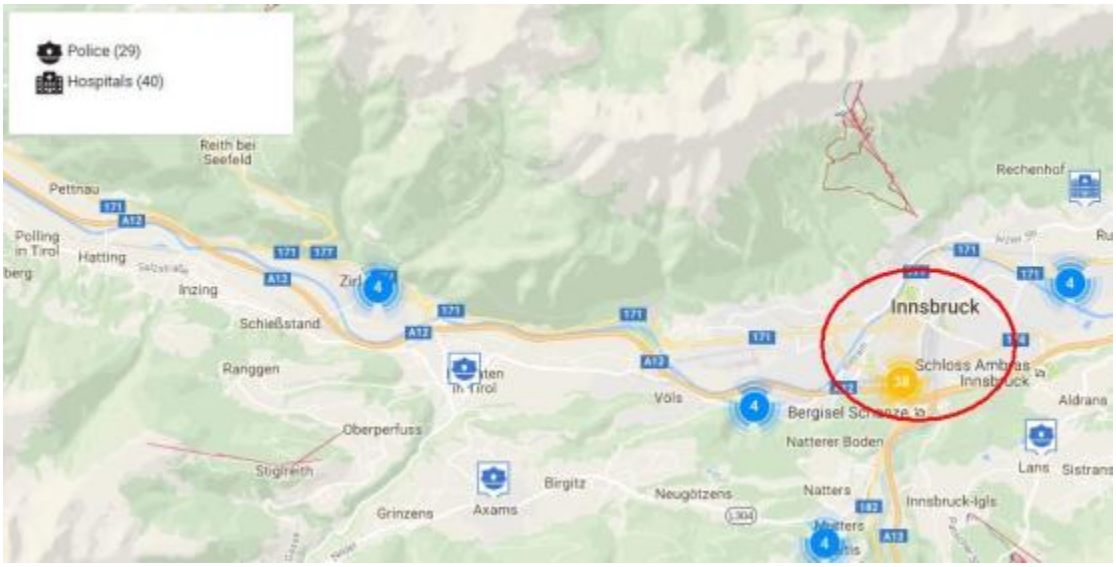


Figure 24 screenshot of bubbles showing selected emergency data in and around Innsbruck, Austria

An evaluator mentioned that the

> *"structure of information is generally ok, but summary information could be represented better (in the box on the top left corner of the map). The problem is that summary stats change based on the entire map section the user looks at. Hence, it is not always clear what the numbers relate to (e.g. when hovering over Germany one might also catch a small section of Austria on the map, which means that numbers are no longer limited to Germany)."*

- Is the structure of actions you can choose clear and logical for you?

The main issues (raised by two evaluators and rated as a major usability problem) concerned the deletion of selected data and the fact that the legend is not clickable. One evaluator expected to be able to download shown data but couldn't find the function to do so.

- **Flexibility and efficiency of use**

Are shortcuts guiding you through the system available? (e.g. to not having to click through a user's manual every time)

- If yes, do you find them useful?

- If no, would you find it useful to have shortcuts while navigating the website?

Some evaluators thought shortcuts would be helpful, yet not essential, whereas other mentioned that no shortcuts are necessary due to the nature of the site.

- **Aesthetic and minimalist design and help and documentation:**
  - Is the information provided on the website precise (e.g. correct and specific)?
  - Is the information provided on the website too extensive or too sparse?
  - Do you find help or contextual explanation (i.e. to explain specific words or steps) where necessary?

Two evaluators feedback was that the purpose of the map is not clear and that it would need additional information on the site, e.g. a description of the overall purpose of the site (rated twice as major usability problem). Another evaluator mentioned that the information provided is very limited and that the core function seems to be missing (rated as usability catastrophe). The lack of pop ups or sets of tabular information was mentioned negatively, and suggestions were made to add an introduction page (rated twice as major usability problem). Thus some contextual explanations and more guidance would help users to navigate and understand the site better.

Other comments included the following five points:

1. Country borders of Google Maps and as drawn by the shape files used by the tool do not match.

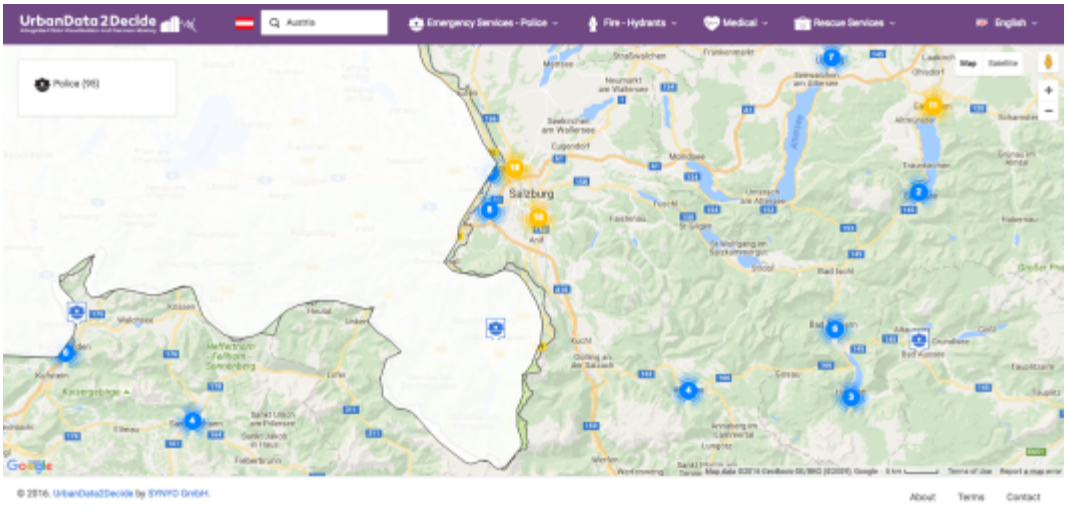


Figure 25 screenshot of country borders of Austria and Germany

2. The size and positioning of the purple page header changes depending on the word length of the filters selected.

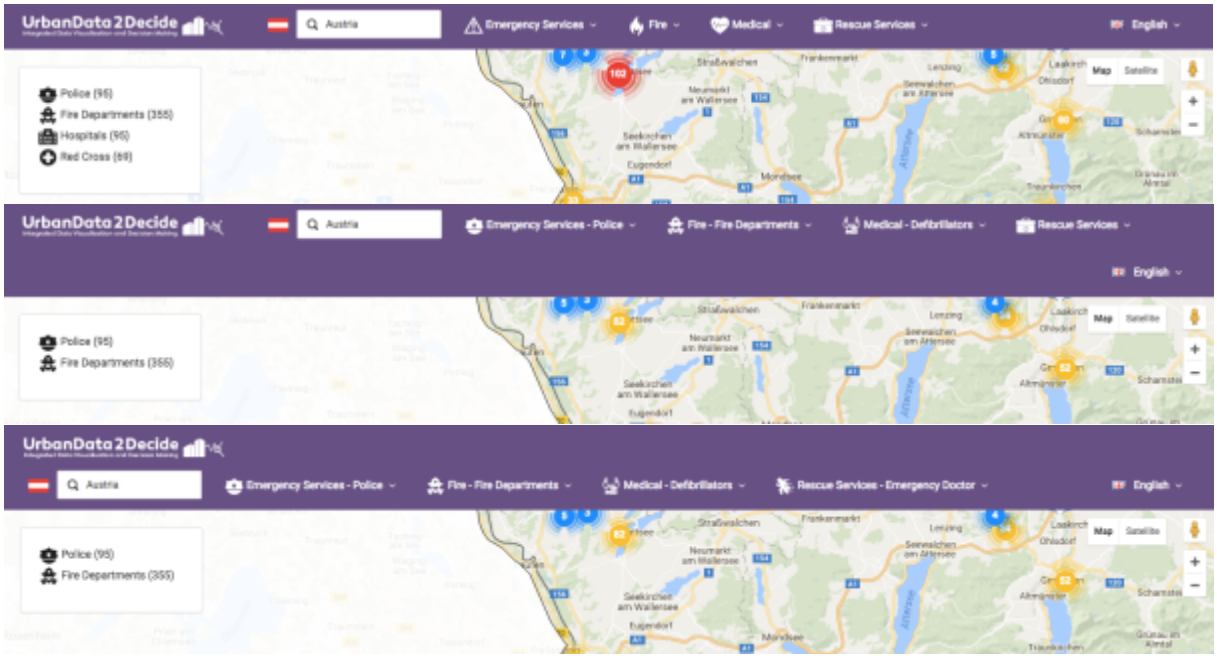


Figure 26 screenshot of page header and its changing size

3. *"The positions of "summary bubbles" are inconsistent and as one zooms into the map, they change positions. Additionally, even though largely relating to larger cities (e.g. Graz), the*

summary bubbles are not always plotted on top of the respective cities but somewhere besides them."

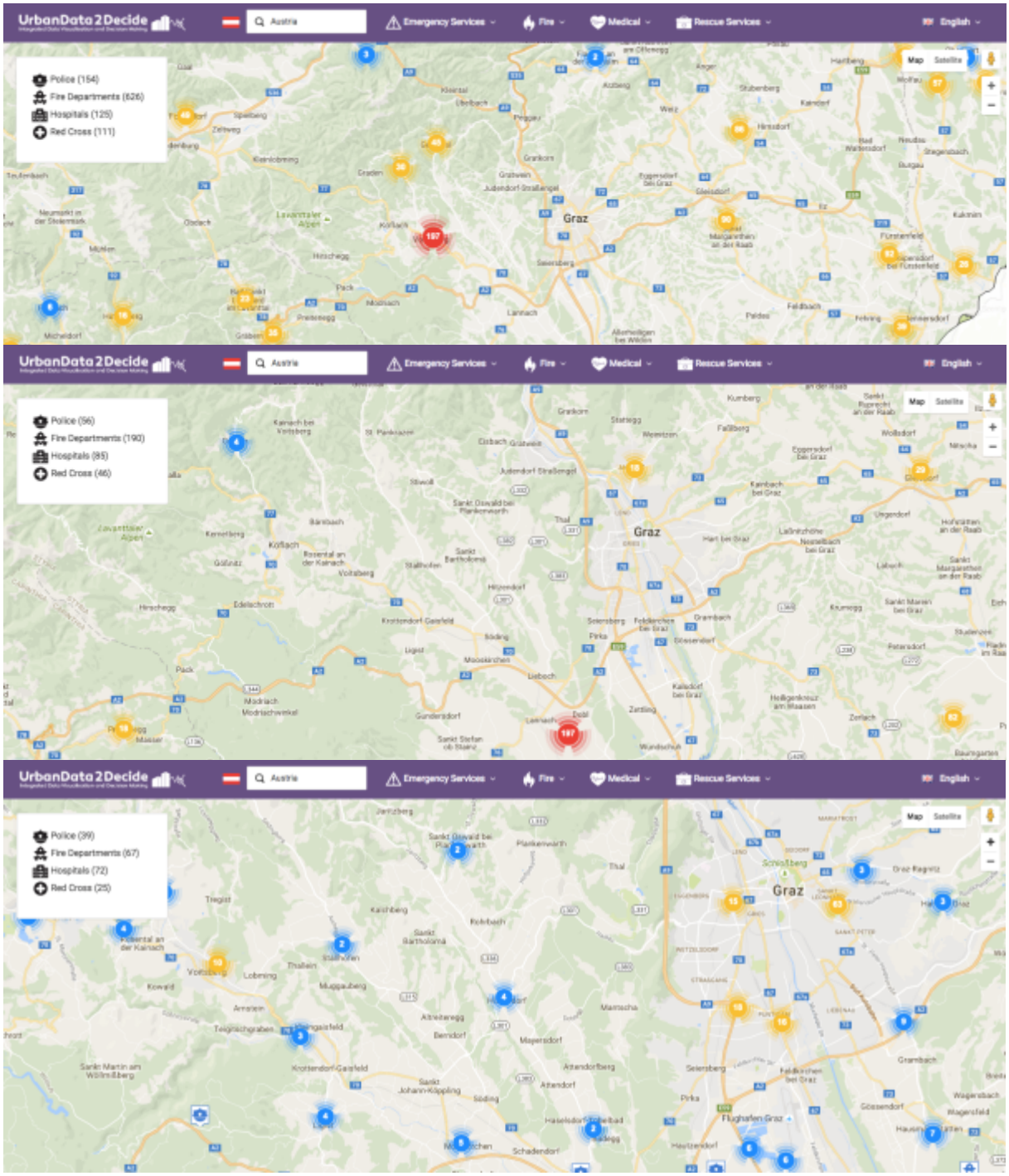


Figure 27 screenshot of changing locations of bubbles

4. Numbers in the summary box change depending on the map section, which means that they constantly change as one move through the map. Additionally, once the tool is loaded with more data, this would mean that it will be hard to see summary statistics for just one country.

5. Does the selection of a specific browser make a difference? (The evaluator used explorer)

Cognitive Walkthrough

1) Tasks

- Task 1: Find a map of all hospitals in Vienna, Austria. Once you have found the map, return to the start screen.
- Task 2: Find a map of the fountains in Graz, Austria.
- Task 3: You are interested whether emergency data also exists for Malmö, Sweden. Check whether you find any emergency data in the city of Malmö.

2) Analysis of Results

Similar to what has been reported during heuristic evaluation sessions the fact there are no clear boundaries when a city is selected, and the need to zoom in a lot was confusing for users who look for city-specific data. Further, most evaluators thought that the legend on the left side clickable which is not possible. Also, error messages should be provided as one tester mentioned, especially if there is no data for some emergency types. Currently there is no information available; hence if there are no bubbles the user can assume that there is no data. An 'uncheck' function respectively the possibility to delete selected data was mentioned by all testers. Additionally, the search field on the top right was confusing to testers as the place in search field stays the same even when going back to "home" – it might be better to have the city name in the search field light gray or something to make explicit that the city is currently not actively selected. In line with what has been reported during the heuristic evaluations, testers mentioned that the selection of language doesn't change anything, and a brief description of single components and purpose of the page would help the user's orientation.

3) Recommendations and Conclusions

The results of both usability tests – the heuristic evaluations and cognitive walkthroughs – were analysed in the previous paragraphs. The most pertinent and commonly raised problems and suggestions for improvements, including single functionalities that should be renamed, or put somewhere else, etc. are listed below.

- A function enabling to delete already selected data should be added
- The country borders need to be adapted as the shape files and information from Google Maps is not entirely matching
- The Home button should be changed from the project's logo to a house symbol
- A brief information about the purpose of the tool should be added
- Need for error message, especially if data is not available (e.g. in other countries or places in Austria)
- The select language function is not working

4 CONCLUSIONS

This report discussed multiple usability methods, ranging from scenarios, cognitive walkthrough's combined with a think aloud approach, to heuristic evaluation, paper prototyping and usability questionnaires. These methods are used depending on the available objects, more specifically their formats, ranging from paper prototypes to already functioning and clickable websites. Additionally, also time and personnel resources are of essence when deciding for a usability method. Thus, the application of a method needs to be well planned and thought through as the available type (**what** can be tested?), as well as **who** should be the testers, **when** and **where** does the testing take place, needs to be clarified. Within the 'UrbanData2Decide' project several tools have been developed, and the usability of three project tools has been tested. The tools that were tested are the ODI tools '*Train Crowding Data*' and '*Open Data TFL'* and the SYNYO '*Emergency Data Map'*.

The selection of usability methods – cognitive walkthrough combined with think aloud and heuristic evaluation – proved successful given the tools status and geographic location of our selected

usability testers (locations were in Austria, Sweden and the UK). Both usability tests brought similar issues and user problems to the fore. They further added valuable information that we wouldn't have received by just using one of the methods. Interestingly regardless of the tool, comparable feedback concerning recommendations was feedback by participants. Thus it became evident that certain features are oftentimes expected when visiting a website or using a tool. Some examples include the expectations of users to have a home symbol for the 'return to main page' function. Additionally, short information on the aim or context of a website, as well as the importance to have error message if a function doesn't work to diminish uncertainties by users are crucial to a good design of a tool or a website. Concluding, the usability testing of the three '*UrbanData2Decide*' project tools resulted in concrete recommendations for the developers as well as the designers of these tools in order to further improve them and strive for the invisibility of usability or in other words:

"*True usability is invisible. If something is going well, you don't notice it*."

(Rubin & Chisnell 2008, 6)

5 REFERENCES

Barnum, C. (2011) Usability Testing Essentials. Ready, Set … Test! Elsevier; Morgan Kaufmann.

Buchanan, S., Salako, A. (2009) Evaluating the usability and usefulness of a digital library. Library Review, 58(9), p. 638 – 651.

Buxton, B. (2007) Sketching User Experiences. Getting the design right and the right design. Elsevier; Morgan Kaufmann.

Cassim, J. D., H. (2007) Empowering Designers and Users: Case Studies from the DBA Inclusive Design Challenge. Design for Inclusivity: A Practiacal Guide to Accessible, Innovative and User-Centred Design. J. C. Roger Coleman, Huan Dong and Julia Cassim. Hampshire, England, Gower Publishing Limited.

Charters, E. (2003) The Use of Think-aloud Methods in Qualitative Research. An Introduction to Think-aloud Methods. Brock Education, 12(2), p. 68-82.

Dabbs DeVito A., Beach, S., Courtney, K., Matthews J. (2013) Evaluating Quality of Life Technologies. In: Schulz, R. (Ed.) Quality of Life Technology Handbook, CRC Press; p. 147 – 162.

Ddamba, J., Dittrich, Y. (2016) Data for Participation and Participation as Data: Supporting Incremental Participatory Decision-Making in Urban Planning. Forthcoming.

Dumas, J., Loring, B. (2008) Moderating Usability Tests. Principles & Practices for Interacting. Elsevier. Morgan Kaufmann Publishers.

Ehn, P. (2008) Participation in design things. Proceedings of the Tenth Anniversary Conference on Participatory Design 2008, 92–101.

Forsell, C. (2014) Heuristic Evaluation: A Guide for Use in Information Visualisation. In: Information Visualisation: Techniques, Usability and Evaluation. Banissi, E., Marchese, F., Forsell C., Johansson, J. (Eds). p. 177-199.

Koivnuniemi, A. (2013) Developing usability method for agile software development – Case study on RISE for Traffica. M.Sc.Thesis, School of information sciences, University of Tampere.

Kankainen, A. (2003) UCPCD: User-Centrered Product Concept Design. 2003 conference on Designing for user experiences.

Kantner, L., Rosenbaum, S. (1997) Usability Studies of WWW Sites: Heuristic Evaluation vs. Laboratory Testing. SIGDOC 97 Proceedings, Salt Lake City, US.

Kensing, F., Blomberg, J. (1998) Participatory design: Issues and concerns. *Computer Supported Cooperative Work (CSCW)*. 7, 3–4 (1998), 167–185.

Olmsted-Hawala, E. L., Hawala, S., Murphy, E. D., Ashenfelter, K. T. (2010) Think-aloud protocols A Comparison of Three Think-Aloud Protocols for use in Testing Data-Dissemination Web Sites for Usability. Proceedings of CHI 2010, Atlanta, GA. p. 2381 – 2390.

Olson, G., Duffy, S., Mack, R. (1984) Thinking-out-loud as a method for studying real-time comprehension processes. In: Kieras, D. E. & Just, M. A. (Eds.). New methods in reading comprehension research, p. 253-286.

Lews, C., Rieman, J. (1994) Task-Centered User Interface Design. A Practical Introduction. Clayton Lewis and John Rieman, California.

Maguiere, M., Bevan, N. (2002) User requirements analyis: A review of supporitn methods. IFIP 17th World Computer Congress, Montreal, Canada, Kluwer Academic Publishers.

Nielsen, J. (1994). Usability Engineering. San Diego: Academic Press. p. 115–148.

Nielsen, J., Molich, R. (1990) Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing System. ACM, p. 249-256.

Norman, D. (2013) The Design Of Everyday Things. Basic Book, New York.

Rettig, M. (1994) Prototyping for Tiny Fingers. Communications of the ACM 37(4), p. 21-27.

Rubin, J., Chisnell, D. (2008) Handbook of Usability Testing. How to Plan, Design, Conduct Effective Tests. Wiley Publishing, Inc. Indiana.

Sanoff, H. (2006) Origins of community design. Progressive Planning 166, p. 14-17.

Sefelin, R., Tscheligi, M., Giller, V. (2003) Paper Prototyping – What is it good for? A Comparison of Paper- and Computer-based Low-fidelity Prototyping. Conference proceedings. CHI 2003: New Horizons. Computer-Human Interaction conference, Florida.

Snyder (2003). Paper Prototyping: The Fast and Easy Way to Design and Refine User Interfaces. Morgan Kaufmann.

Spinuzzi, C. The methodology of participatory design. *Technical Communication*. 52, 2 (2005), 163–174.

Unger, R., Chandler, C. (2009) A Project Guide To UX Design. For User Experience Designers In The Field Or In The Making. Peachpit Press.

Vukocav, P., Kirnic, V., Klicek, B. (2010) A Comparison of Usability Evaluation Methods For E-Learning Systems. DAAAM International Scientific Book chapter 27 (9) p. 271-288.

Wharton, C., Rieman, J., Lewis C., Polson, P. (1994) The cognitive walkthrough method: a practitioner's guide. In: Nielson, J. (Ed.) Usability inspection methods, John Wiley & Sons, New York. p. 105 – 140.

Harvard reference style: http://libweb.anglia.ac.uk/referencing/files/Harvard_referencing_2015.pdf

6 ANNEX

6.1 Annex 1: Example for a Usability Test Plan

An example for a usability test plan is provided below, derived from the usability.gov website. Link: http://www.usability.gov/how-to-and-tools/resources/templates/usability-test-plan-template.html

Usability Test Plan

Table of Contents

Document Overview

This document describes a test plan for conducting a usability test during the development of [web site name or application name]. The goals of usability testing include establishing a baseline of user performance, establishing and validating user performance measures, and identifying potential design concerns to be addressed in order to improve the efficiency, productivity, and end-user satisfaction [add or delete goals].

The usability test objectives are:

- To determine design inconsistencies and usability problem areas within the user interface and content areas. Potential sources of error may include:
  - Navigation errors – failure to locate functions, excessive keystrokes to complete a function, failure to follow recommended screen flow.
  - Presentation errors – failure to locate and properly act upon desired information in screens, selection errors due to labeling ambiguities.
  - Control usage problems – improper toolbar or entry field usage.
- Exercise the application or web site under controlled test conditions with representative users. Data will be used to access whether usability goals regarding an effective, efficient, and well-received user interface have been achieved.
- Establish baseline user performance and user-satisfaction levels of the user interface for future usability evaluations.

[Add a paragraph that summarizes the user groups that the application or Website will be deployed/launched to, the user groups that will participate in the usability test and the number of participants from each user group that are expected to participate. Indicate whether the testing will occur in a usability lab or remotely and the expected date range for testing.]

Executive Summary

[Summarize specific details of the usability test for the given application or Web site; describe specific functions to be evaluated. Summarize the usability goals.]

Upon review of this usability test plan, including the draft task scenarios and usability goals for [web site name or application name], documented acceptance of the plan is expected.

Methodology

[Describe briefly the number of participants, the setting of the usability test sessions, the tools used to facilitate the participant's interaction with the application (ex., browser), and the measures to be collected, such as demographic information, satisfaction assessment, and suggestions for improvement.]

Participants

[Thoroughly describe the number of participants expected, how they will be recruited, characteristics of their eligibility, and expected skills/knowledge.]

The participants' responsibilities will be to attempt to complete a set of representative task scenarios presented to them in as efficient and timely a manner as possible, and to provide feedback regarding the usability and acceptability of the user interface. The participants will be directed to provide honest opinions regarding the usability of the application, and to participate in post-session subjective questionnaires and debriefing.

[Describe how the team will select test participants to meet stated requirements. Explain if participants will have certain skills and/or background requirements, if they will be familiar with the evaluation tasks, or have experience with performing certain tasks.]

Training

[Describe any training provided as an overview of the Web application or Web site.] The participants will receive and overview of the usability test procedure, equipment and software. [Describe any parts of the test environment or testing situation that may be nonfunctional.]

Procedure

[Usability Lab Testing]

Participants will take part in the usability test at [put the name of the testing lab here] in [location here]. A [type of computer] with the Web site/Web application and supporting software will be used in a typical office environment. The participant's interaction with the Web site/Web application will be monitored by the facilitator seated in the same office. Note takers and data logger(s) will monitor the sessions in observation room, connected by video camera feed [describe if lab has one-way mirror or video feed]. The test sessions will be videotaped.

[If the facilitator is seated in a control – describe the environment and the equipment and how communication is supported.]

The facilitator will brief the participants on the Web site/Web application and instruct the participant that they are evaluating the application, rather than the facilitator evaluating the participant. Participants will sign an informed consent that acknowledges: the participation is voluntary, that participation can cease at any time, and that the session will be videotaped but their privacy of identification will be safeguarded. The facilitator will ask the participant if they have any questions.

Participants will complete a pretest demographic and background information questionnaire. The facilitator will explain that the amount of time taken to complete the test task will be measured and that exploratory behavior outside the task flow should not occur until after task completion. At the start of each task, the participant will read aloud the task description from the printed copy and begin the task. Time-on-task measurement begins when the participant starts the task.

The facilitator will instruct the participant to 'think aloud' so that a verbal record exists of their interaction with the Web site/Web application. The facilitator will observe and enter user behavior, user comments, and system actions in the data logging application [describe how these metrics will be recorded if a data logging application is not used.]

After each task, the participant will complete the post-task questionnaire and elaborate on the task session with the facilitator. After all task scenarios are attempted, the participant will complete the post-test satisfaction questionnaire.

[For Remote Testing]

Participants will take part in the usability test via remote screen-sharing technology. The participant will be seated at their workstation in their work environment. Verbal communication will be supported via telephone.

The facilitator will brief the participant and instruct that he or she is evaluating the Web site/Web application, rather than the facilitator evaluating the participant. Participants will complete a pretest

demographic and background information questionnaire. Sessions will begin when all participant questions are answered by the facilitator. The facilitator will inform the participant that time-on-task will be measured and that exploratory behavior outside the task flow should not occur until after task completion.

The facilitator will instruct the participant to read aloud the task description from the printed copy and begin the task. Time-on-task measure will begin. The facilitator will encourage the participants to 'think aloud' and that a verbal record will exist of the task-system interaction. The facilitator will observe and enter user behavior and comments, and system interaction in a data logging application.

After each task, the participant will complete the post-task questionnaire and elaborate on the task session. After all tasks have been attempted, the participant will complete a post-test satisfaction questionnaire.

Roles

The roles involved in a usability test are as follows. An individual may play multiple roles and tests may not require all roles.

Trainer

Provide training overview prior to usability testing

Facilitator

Provides overview of study to participants

Defines usability and purpose of usability testing to participants

Assists in conduct of participant and observer debriefing sessions

Responds to participant's requests for assistanc

Data Logger

Records participant's actions and comments

Test Observers

Silent observer

Assists the data logger in identifying problems, concerns, coding bugs, and procedural errors

Serve as note takers.

Test Participants

Provides overview of study to participants

Defines usability and purpose of usability testing to participants

Assists in conduct of participant and observer debriefing sessions

Responds to participant's requests for assistance

Ethics

All persons involved with the usability test are required to adhere to the following ethical guidelines:

The performance of any test participant must not be individually attributable. Individual participant's name should not be used in reference outside the testing session.

A description of the participant's performance should not be reported to his or her manager.

Usability Tasks

[The usability tasks were derived from test scenarios developed from use cases and/or with the assistance of a subject-matter expert. Due to the range and extent of functionality provided in the application or Web site, and the short time for which each participant will be available, the tasks are the most common and relatively complex of available functions. The tasks are identical for all participants of a given user role in the study.]

[Describe the application's test setup up such as special development environments or test databases; concurrent development activities that may impact the test application's availability or performance; and impact to real data or workflows outside the testing situation.]

The task descriptions below are required to be reviewed by the application owner, business-process owner, development owner, and/or deployment manager to ensure that the content, format, and presentation are representative of real use and substantially evaluate the total application. Their acceptance is to be documented prior to usability test.

[Describe the scenarios and groups of participants whom will attempt to complete tasks and documented in sufficient detail to warrant customer sign-off. Describe how typical and encompassing these scenarios are in the overall scope of tasks that the application or Web site will support.]

Usability Metrics

Usability metrics refers to user performance measured against specific performance goals necessary to satisfy usability requirements. Scenario completion success rates, adherence to dialog scripts, error rates, and subjective evaluations will be used. Time-to-completion of scenarios will also be collected. [include or delete any metrics not used in the planned test]

Scenario Completion

Each scenario will require, or request, that the participant obtains or inputs specific data that would be used in course of a typical task. The scenario is completed when the participant indicates the scenario's goal has been obtained (whether successfully or unsuccessfully) or the participant requests and receives sufficient guidance as to warrant scoring the scenario as a critical error.

Critical Errors

Critical errors are deviations at completion from the targets of the scenario. Obtaining or otherwise reporting of the wrong data value due to participant workflow is a critical error. Participants may or may not be aware that the task goal is incorrect or incomplete.

Independent completion of the scenario is a universal goal; help obtained from the other usability test roles is cause to score the scenario a critical error. Critical errors can also be assigned when the participant initiates (or attempts to initiate) and action that will result in the goal state becoming unobtainable. In general, critical errors are unresolved errors during the process of completing the task or errors that produce an incorrect outcome.

Non-critical Errors

Non-critical errors are errors that are recovered from by the participant or, if not detected, do not result in processing problems or unexpected results. Although non-critical errors can be undetected by the participant, when they are detected they are generally frustrating to the participant.

These errors may be procedural, in which the participant does not complete a scenario in the most optimal means (e.g., excessive steps and keystrokes). These errors may also be errors of confusion (ex., initially selecting the wrong function, using a user-interface control incorrectly such as attempting to edit an un-editable field).

Noncritical errors can always be recovered from during the process of completing the scenario. Exploratory behavior, such as opening the wrong menu while searching for a function, [will, will not (edit Procedure)] be coded as a non-critical error.

Subjective Evaluations

Subjective evaluations regarding ease of use and satisfaction will be collected via questionnaires, and during debriefing at the conclusion of the session. The questionnaires will utilize free-form responses and rating scales.

Scenario Completion Time (time on task)

The time to complete each scenario, not including subjective evaluation durations, will be recorded.

Usability Goals

The next section describes the usability goals for [web site name or application name].

Completion Rate

Completion rate is the percentage of test participants who successfully complete the task without critical errors. A critical error is defined as an error that results in an incorrect or incomplete outcome. In other words, the completion rate represents the percentage of participants who, when they are finished with the specified task, have an "output" that is correct. Note: If a participant requires assistance in order to achieve a correct output then the task will be scored as a critical error and the overall completion rate for the task will be affected.

A completion rate of [100%/enter completion rate] is the goal for each task in this usability test.

Error-free rate

Error-free rate is the percentage of test participants who complete the task without any errors (critical or non-critical errors). A non-critical error is an error that would not have an impact on the final output of the task but would result in the task being completed less efficiently.

An error-free rate of [80%/enter error-free rate] is the goal for each task in this usability test.

Time on Task (TOT)

The time to complete a scenario is referred to as "time on task". It is measured from the time the person begins the scenario to the time he/she signals completion.

Subjective Measures

Subjective opinions about specific tasks, time to perform each task, features, and functionality will be surveyed. At the end of the test, participants will rate their satisfaction with the overall system. Combined with the interview/debriefing session, these data are used to assess attitudes of the participants.

Problem Severity

To prioritize recommendations, a method of problem severity classification will be used in the analysis of the data collected during evaluation activities. The approach treats problem severity as a combination of two factors - the impact of the problem and the frequency of users experiencing the problem during the evaluati

Impact

Impact is the ranking of the consequences of the problem by defining the level of impact that the problem has on successful task completion. There are three levels of impact:

High - prevents the user from completing the task (critical error)

Moderate - causes user difficulty but the task can be completed (non-critical error)

Low - minor problems that do not significantly affect the task completion (non-critical error)

Frequency

Frequency is the percentage of participants who experience the problem when working on a task.

High: 30% or more of the participants experience the problem

Moderate: 11% - 29% of participants experience the problem

Low: 10% or fewer of the participants experience the problem

[For studies with less than ten participants in a group, the percentages may to be adjusted. For example, for a study with 8 participants the low frequency should be 12.5% (1/8 = .1250]

Problem Severity Classification

The identified severity for each problem implies a general reward for resolving it, and a general risk for not addressing it, in the current release.

Severity 1 - High impact problems that often prevent a user from correctly completing a task. They occur in varying frequency and are characteristic of calls to the Help Desk. Reward for resolution is typically exhibited in fewer Help Desk calls and reduced redevelopment costs.

Severity 2 - Moderate to high frequency problems with moderate to low impact are typical of erroneous actions that the participant recognizes needs to be undone. Reward for resolution is typically exhibited in reduced time on task and decreased training costs.

Severity 3 - Either moderate problems with low frequency or low problems with moderate frequency; these are minor annoyance problems faced by a number of participants. Reward for resolution is typically exhibited in reduced time on task and increased data integrity.

Severity 4 - Low impact problems faced by few participants; there is low risk to not resolving these problems. Reward for resolution is typically exhibited in increased user satisfaction.

Reporting Results

The Usability Test Report will be provided at the conclusion of the usability test. It will consist of a report and/or a presentation of the results; evaluate the usability metrics against the pre-approved goals, subjective evaluations, and specific usability problems and recommendations for resolution. The recommendations will be categorically sized by development to aid in implementation strategy. The report is anticipated to be delivered to the Project UCD Contact by [enter date].

6.2 Annex 2: Example for a Cognitive Walthrough

An example for applying the cognitive walkthrough methodology for testing the usability of a journey planner website. After defining a task for the test user, three steps to complete the task are analysed and 'success' and 'failure' stories are listed. For the entire usability test, please see: http://elisa.dyndns-web.com/teaching/hiit/cognitive.pdf

Cognitive Walkthrough

Design and User Evaluation of Augmented-Reality Interfaces

Jorma Nieminen

Cognitive Walkthrough - Example

Define user an choose a sample task

Heikki is a regular third year computer science student from Helsinki. He has just had a lunch at the Unicafe in **Chemicum** building in **Kumpula** and now wants to go to the central library of the university to study for tomorrow's classes. Heikki knows that library is **near the railway station**.

Cognitive Walkthrough - Example

Analysis: Step 1

Will the user be trying to achieve the right effect?
Success Story 1: Defining the users current location is part of the original task.

Will the user notice that the correct action is available?
Success Story 1: The input field is clearly visible *(e.g. Not behind some menu)*.

Will the user associate the correct action with the desired effect?
Failure Story 1: User may try the timetable search on the right column (likelihood only 10% because the reading direction is from left to right and top to bottom)

If the correct action is performed, will the user see that progress is being made?
Success Story 1: User sees the written text appear in the input field.

Cognitive Walkthrough - Example

Analysis: Step 2

Will the user be trying to achieve the right effect?
Success Story 1: Entering the destionation is part of the user's original goal to get to the library

Will the user notice that the correct action is available?
Success Story 1: The input field is clearly visible.

Will the user associate the correct action with the desired effect?
Success Story 1: Input field is near the label "To" and there are no competing alternatives visible.

If the correct action is performed, will the user see that progress is being made?
Success Story 1: User sees the written text appear in the input field.

Cognitive Walkthrough - Example

Analysis: Step 3

Will the user be trying to achieve the right effect?
Success Story 1: User wants to see the alternative journey options because this is a part of his original goal.

Will the user notice that the correct action is available?
Success Story 1: Search button is clearly visible on the screen.

Will the user associate the correct action with the desired effect?
Failure Story 1: User clicks the bicycle button because it's on different colour (likelihood 2%).
Failure Story 2: User clicks the search button related to timetable search on the right column (likelihood 0,0001%).

If the correct action is performed, will the user see that progress is being made?
Success Story 1: The available routes appear on the screen after clicking the button.

Cognitive Walkthrough - Example

Record problems, reasons and assumptions

Will the user be trying to achieve the right effect?

Failure Story 1: User's goal is to get to the library. He may assume that the system know he's location automatically (likelihood 0,5%).

Will the user associate the correct action with the desired effect?

Failure Story 1: User may try the timetable search on the right column (likelihood only 10% because the reading direction is from left to right and top to bottom).

Will the user associate the correct action with the desired effect?

Failure Story 1: User clicks the bicycle button because it's on different colour (likelihood 2%).

Failure Story 2: User clicks the search button related to timetable search on the right column (likelihood 0,0001%).

6.3 Annex 3: Usability Testing Handout Example

Heuristic Evaluation

In a heuristic evaluation, usability experts evaluate a site's interface using a set of accepted evaluative principles. During the evaluation a group of usability experts "*inspect a user interface to find and rate the severity of usability problems using a set of usability principles or heuristics*" (Vukovac et al. 2010, 273). Between three to five experts are recommended for a thorough evaluation of site's interface or a tool. Ratings based on the opinion of three evaluators are considered reliable (Forsell 2014, 186). Commonly, each evaluator works independently and discusses the findings afterwards. The evaluators need to be equipped with a certain task to perform on the website or interface. Each task is then evaluated according to a pre-set list of usability heuristics.

Setup

Join Skype and access tool

Tool http://goingunderground.herokuapp.com/signals
Login tfl
Password UrbanData

Steps

1. Heuristic evaluation training (app. 15 minutes – group session)

facilitators give evaluators the needed domain knowledge and information on the scenario

2. Evaluation (approx. 1 hour – individual session)

* we are on standby in Skype to ensure you can proceed as planned

Structure

a) access tool
b) have your printed materials ready
   i) handout (including 10 heuristics criteria)
   ii) printed working sheet
c) execute the task and capture briefly (5-10 sentences) the results (which data have you chosen and why)
d) now, go through each of the 10 heuristics and write down your observations for each question in the working sheet. Don't forget the severity rating! Should be about 20 observations in total.
e) add any observations that didn't fit within the heuristics categories at the end of the working sheet
f) please send us the completed worksheet after the session, preferably within 2 days (typed not scanned).
g) You have completed the mission ;)

In a face-to-face session we would also have these steps:

- Aggregating results and consolidation of individual severity ratings
- Discussing selected usability issues as a group

When going through the results of the evaluation, a possible problem rating scale discussed by Nielsen and Molich (1990) includes a scale of 4 severity stages:

> **1) Cosmetic problem** – not necessary to fix, unless time allows
> **2) Minor usability problem** – low priority, fixing it should be given low priority
> **3) Major usability problem** – important to fix, should be given high priority
> **4) Usability catastrophe** – imperative to fix the problem

Context: *The site is targeting developers who want to make use of open train data. We introduce the interface (can you recognize the overarching classification of data)*

Task*: Let's concentrate on ' Customers and Enthusiasts Dashboard'. Gain an overview of this categorie. Imagine you want to write an App about 'Passenger Experience' using the London Underground. Now, check whether you can find any useful open data. (10 - 15 min)*

The **10 commonly known usability heuristics** (Nielson, 1994) translated into questions:

2) **Visibility of system status, e.g.**
   - Do you know where to go next in the navigation?
   - Is it clear if the content rendering of a page is completed?

3) **Match between system and the real world?**
   - Do you understand the terms used on the website/the tool? Labels Headings Explanations ...
   - Do you understand the meaning of the icons?

4) **User control and freedom**
   - Do you know how to return to the main page / 'home' function ?
   - Is the 'home' function available on every page?

5) **Consistency and standards**:
   - Do symbols and labels repeat?
   - Are existing standards for symbols / metrics used? (Home = House; Help = Question mark)

6) **Error prevention**:
   - Are there sufficient error messages?
   - Do you understand messages trying to prevent you from entering invalid data?

7) **Recognition rather than recall**:
   - Is the structure of topics clear and logical for you?
   - Is the structure of information clear and logical for you?
   - Is the structure of actions you can choose clear and logical for you?

8) **Flexibility and efficiency of use**: *** not always applicable ***
   - Are shortcuts guiding you through the system available? (e.g. to not having to click through a user's manual every time)
     i) If yes, do you find them useful?
     ii) If no, would you find it useful to have shortcuts while navigating the website?

9) **Aesthetic and minimalist design**:
   - Is the information provided on the website precise (e.g. correct and specific)?
   - Is the information provided on the website too extensive or too sparse?

10) **Help users recognize, diagnose, and recover from errors**:
   - Do you understand why an action was erroneous?
   - Do you understand how to solve the problem?

11) **Help and documentation**:
   - Do you find help or contextual explanation (i.e. to explain specific words or steps) where necessary?

Bibliography

Forsell, C. (2014) Heuristic Evaluation: A Guide for Use in Information Visualisation. In: Information Visualisation: Techniques, Usability and Evaluation. Banissi, E., Marchese, F., Forsell C., Johansson, J. (Eds). p. 177-199.

Nielsen, J. (1994) Usability Engineering. San Diego: Academic Press. p. 115–148.

Nielsen, J., Molich, R. (1990) Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing System. ACM, p. 249-256.

Vukocav, P., Kirnic, V., Klicek, B. (2010) A Comparison of Usability Evaluation Methods For E-Learning Systems. DAAAM International Scientific Book chapter 27 (9) p. 271-288.

6.4 Annex 4 Example of Evaluation Sheet

Reference to Heuristic	**Observation + Interpretation (reasons for severity rating); Suggestions for improvement are welcome**	**No Problem**	**Severity *Cosmetic problem***	**Severity *Minor usability problem***	**Severity *Major usability problem***	**Severity *Usability catastrophe***
Example: 5.a	Not all folders contain data, but there is no information provided beforehand. Suggestion: one could colour the topics who contain data differently				x	