SEVENTH FRAMEWORK PROGRAMME

FP7-SEC-2011-284725

SURVEILLE

Surveillance: Ethical Issues, Legal Limitations, and Efficiency
Collaborative Project

SURVEILLE Deliverable 3.3: System effectiveness, efficiency and satisfaction assessment

Due date of deliverable: 30.09.2013
Actual submission date: 30.09.2013

Start date of project: 1.2.2012 Duration: 39 months

SURVEILLE Work Package number and lead: WP3 Dr. Coen van Gulijk (TU Delft)

Author(s): Ralf Eck (Fraunhofer)

Project co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission services)	
CO	Confidential, only members of the consortium (including the Commission Services)	

ABSTRACT

The goal of this work related to work package 3 of SURVEILLE was to assess system effectiveness by the evaluation of the aspect “user friendliness”. This evaluation should have in mind the technologies collected within the survey of surveillance systems conducted in work package 2. The idea is that the end-users that want to introduce a surveillance technology can look that technology up in the database of the survey and get to see an estimation of its usability. This user friendliness is assessed by the factor’s effectiveness, efficiency and user satisfaction. When analysing the aspects under which the survey of surveillance systems could be undertaken, it became clear that a direct evaluation of usability was impossible. This is because an evaluation of human factors has to be executed with a precise knowledge of work circumstances in the future use of the technology. But aspects such as working environment and level of training are unknown. This basis made it clear that a classic human factors test procedure is not applicable on the basis of the survey of surveillance systems.

Therefore Fraunhofer IOSB decided to take another course. To solve the problem in the way it was required by work package 3 a decision support tool AFUS – **A**pplication for **F**inding **U**sability **S**tandards – was developed. AFUS should support deciders of buying or developing surveillance technologies in regarding effectiveness, efficiency and user satisfaction. Based on the knowledge written down in human factors standards, Fraunhofer IOSB developed a decision tree to structure these standards. A decision tree is a possible technique to find those standards that fit best to the usability testing conditions defined by the end-users. The resulting lists of human factors standards vary by the path the user takes when going through the decision tree. The length of the lists with up to 13 documents is the next focus of this report. Within these documents the users get to many proposals of human factors evaluations. Even reading these many documents is time consuming.

To further differentiate the list of human factors standards a scoring of the clarity of the written knowledge and the applicability of its mentioned rules and examples was accomplished. All used standards were read and scored on these 2 aspects by Fraunhofer IOSB. A resulting measure of usefulness was calculated by a method from information retrieval for every standard in relation to the decisions in the decision tree. These values for usefulness allowed sorting the resulting list of human factors standards in a way the end-users get best candidates standards for their problem. Some of the standards are quite lengthy. To give the end-users further assistance for every standard a short description in two detail levels is offered as well as a proposal where to start reading inside of the standard in relation to the end-users path through the decision tree.

The decision support tool was built upon a client-server architecture. The goal was to minimize the demands on the client’s side. Having a central server additionally allows doing the necessary maintenance of the often revised standards and the software itself.

After finishing the prototype of AFUS an evaluation by the end-user panel via Internet was started. They were given two tasks and were asked to find the best human factors standard by AFUS. After fulfilling the task by using AFUS they were asked to fill in a questionnaire. This test by the end-users has not been completed by enough participants to come to consolidated findings so far. Such consolidated findings must be postponed until the SURVEILLE final report. First remarks of participants show the importance of having the standard documents themselves accessible by AFUS. Unfortunately this was not possible within this approach due to property rights. If the full text documents are added to AFUS licence fees have to be paid. That was not covered by the project budget. But if we decide to pursue this idea when an advisory board is introduced after the SURVEILLE project, AFUS can be further adapted to the findings of the end-user panel’s evaluation.

CONTENT

1. POSITION OF THIS REPORT IN SURVEILLE

This report is entitled 'Report on system effectiveness, efficiency and satisfaction assessment' is a deliverable in work package 3 of the SURVEILLE project. It contributes to the following objectives of work package 3:

- O3.1 To assess the benefits and costs of surveillance technology. (By 'benefits' we mean the delivery of improved security; by 'costs' economic costs, negative public perceptions, negative effects on behavior, and infringement of fundamental rights.) (project objective O2)
- O3.2 To produce proposals for improving the effectiveness of security surveillance, while taking fully into account perceptions, economic costs, legal limitations and ethical issues.
- With regard to O3.1 this report focuses on the user friendliness of surveillance technology which directly influences the efficiency. With regard to O3.2 it focuses on improving effectiveness by giving tools to improve user satisfaction with a given system.
- Note that other parts of these objectives are met in deliverables that are produced almost simultaneously with this report. They are:
- D3.2: European level study on perceptions, including an overview of effects and side-effects of surveillance and their perceived effectiveness.
- D3.3b: Report on system effectiveness, efficiency and satisfaction assessment
- D3.4: Report describing design of research methodology for assessing effectiveness.
- D3.5 Cost model for surveillance techniques.

These reports (and this work) focus on collecting prior knowledge in different fields (perception, financial cost, satisfaction, efficiency and effectiveness). The authors suggest reading those reports alongside this report because they demonstrate just how difficult technology assessment is.

Further developments for surveillance technology assessment requires design decisions for the research apparatus that have to be based on these reports and on a discussion with all SURVEILLE partners.

2. INTRODUCTION

The SURVEILLE project systematically analyses the impact of surveillance technology for the pursuit of organized and serious crime from various perspectives. In contrast to many approaches pursuing research from one or a few areas of expertise, the SURVEILLE project brings together social science, law, ethics, data security, human factor scientists and end users in multidisciplinary research. The SURVEILLE partners with a technical background are carrying out a survey of surveillance technologies. To do this a matrix for systematically classifying the various technologies and systems was developed [Guelke et al. 13]. To facilitate this matrix the involved project partners started collecting surveillance technologies for the survey [Gulijk et al. 12]. This report is one of a number of pieces of work in work package 3 on 'Perception and Effectiveness', specifically focussing on human factors. Human factors are one of the key factors in effective and efficient application of technology. The best sensors combined with high level analysis methods do not help if users cannot solve their tasks because they do not know how to use their tools. This problem is explicitly considered in SURVEILLE work package 3 "Perception and Effectiveness of Surveillance", and is specified in the Description of Work in the following tasks:

- Task 3.8 "Based on the survey of surveillance systems in WP2, assess system effectiveness relating also to aspects "data protection" and "user friendliness"…" and

- Task 3.8.1 "Based on results of T3.8 make proposals for improving system effectiveness and user satisfaction."

This report focuses on assessing user friendliness of surveillance systems. It is important to know the usability of a surveillance system before purchasing it. Modern surveillance systems have remarkable features to improve the user's abilities in increasing security. The number of features always raises the complexity of the system, which means that the importance of human factors of the system rises. If human machine interfaces and user experience are not adequate the users' goals are not reached.

In this report a description is provided of how usability generally can be assessed. Usability can be determined by the factors effectiveness, efficiency and user satisfaction. It will be shown that the basic conditions that are needed for assessing usability cannot be achieved in this project. This brought up an idea to develop a decision support system helping users to find procedures for their human factors evaluation by using generally accepted standards for human factors testing. This procedure was given to the SURVEILLE local authorities end-users for testing. An accompanying questionnaire is analysed and interpreted in the last section.

3. STATE OF THE ART

Our way into an information society is minted by technology. After a period, where information often came from databases, today information often comes from sensors or a combination of both, to produce higher quality and more up-to-date information. This process takes place in the development of surveillance technology too. The consequence is a rising complexity of systems caused by an increasing number of features to be handled by the users. This complexity is found again in the human machine interfaces. If this is not handled correctly, which is in the sense of this work by good usability; it results in longer times of use, in higher error rates, and increasing frustration of the user. This can be avoided by usability testing.

In the SURVEILLE project, work package 3 investigates in the topic of perception and effectiveness of surveillance. Because human factors have an impact on the effectiveness of system usage they have to be examined too. The task is to assess system effectiveness, based on the survey of surveillance systems in relation to the aspects of “data protection” and “user friendliness”. Data protection assessment will be reported in a separate deliverable D3.3B. How to assess effectiveness, efficiency and user satisfaction is described in this report.

First the focus is on assessment of usability in the sense of effectiveness, efficiency and user satisfaction. It is described how these factors are determined by usability testing. It will be shown, that a direct evaluation of the technologies collected in the survey of surveillance technologies is not possible. So the proposed way of solving the given task is to help the end users in testing the technologies they want to purchase themselves. These tests should be based on human factors standards that are subsequently explained. Due to the fact that not all human factors standards can be easily applied, a decision support system was developed. This system was given to the end users accompanied by a questionnaire. The examination of the results of the questionnaire is discussed. A presentation of further steps concludes this report.

3.1. Usability Testing of Interactive Products

[Sikorski 08] recapitulates that the quality of interactive systems is covered by ISO standards as follows:

- ISO 9126-1 specifies basic quality characteristics of a software product, among which functionality is critical for practical usefulness of the product.
- ISO 9241-10, 12-17 specifies ergonomic requirements for user interfaces and user-system dialog design.
- ISO 9241-11 provides guidance on usability resulting from productivity (effectiveness, efficiency, and satisfaction, as experienced by the user in specific context of use.

Figure 1 shows how design quality and quality of use are connected. The latter is the one in focus in this document. [Sikorski 08] further states that the majority of criteria used for user- and expert-based evaluations are covered by current ISO and software engineering standards.

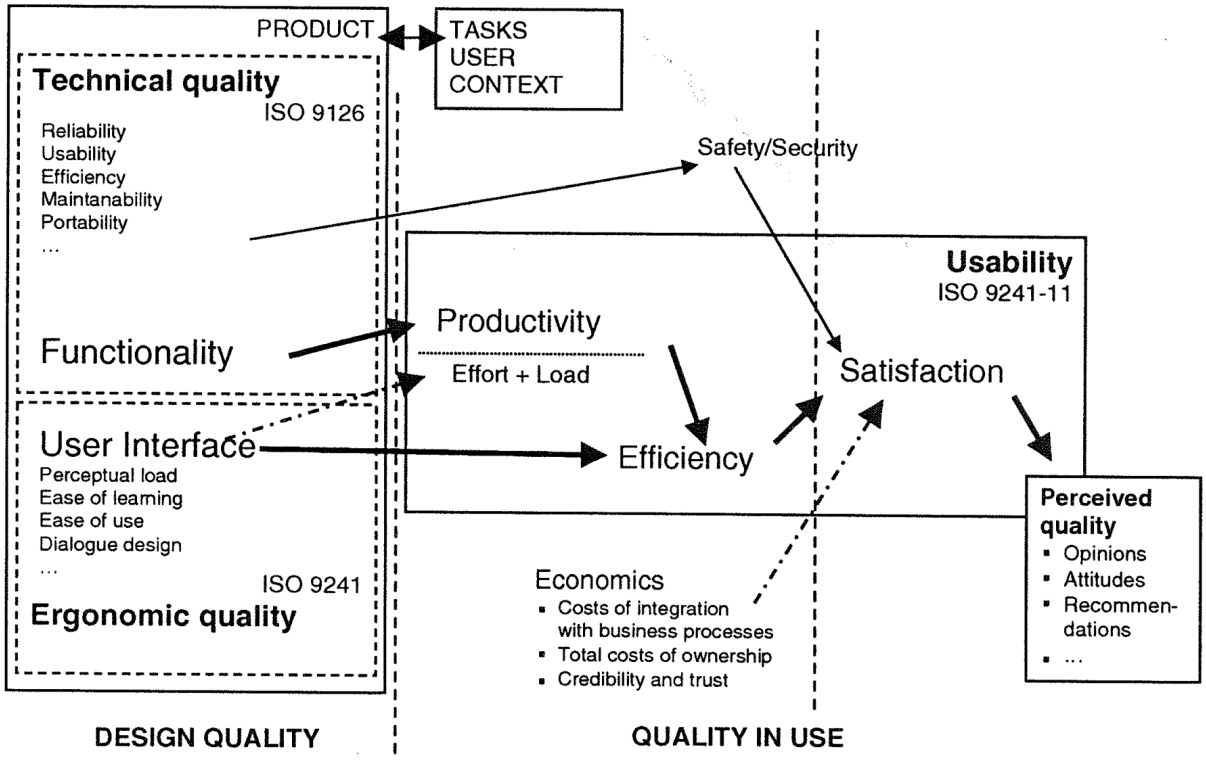

Figure 1: Quality components of interactive systems [Sikorski 08].

The international standard DIN EN ISO 9241-11 [EN ISO 9241-11] defines usability as: The effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments.

- **effectiveness:** the accuracy and completeness with which specified users can achieve specified goals in particular environments
- **efficiency:** the resources expended in relation to the accuracy and completeness of goals achieved
- **satisfaction:** the comfort and acceptability of the work system to its users and other people affected by its use

The specified users in SURVEILLE are the end-users in local authorities including police services and technology developers. They have the goals of improving the control of serious crime and terrorism and do this in their specific working environment. The relation of the aspects is shown in Figure 2.

To measure effectiveness, efficiency and satisfaction of users in the context of use when fulfilling tasks to reach the given goals, experiments are widely used. [Lazar et al. 10] summarizes that usability engineering includes expert-based testing, automated usability testing and user-based testing. The latter attracts the most attention and is considered the most effective. How such experiments with users can be conducted, may be found in [Cockton 2013]. The needed product features for a context of use can be found by experts with the aid of standards (evaluation by an expert).

To analyse the interaction process, modelling the interaction is a possibility. This is often done in research projects but is not yet found in standardization documents.

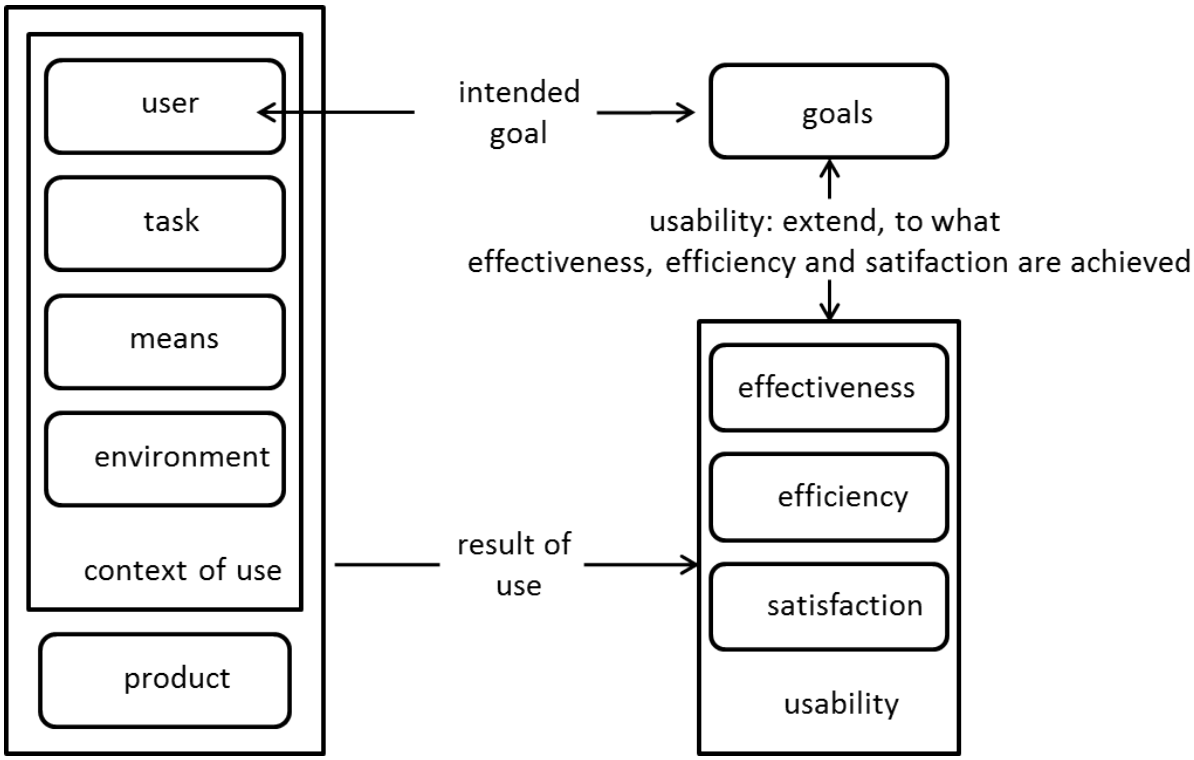

Figure 2: DIN EN ISO 9241-11 defines usability as shown [EN ISO 9241-11].

[Panic 11] carried into execution that the procedures to evaluate usability described in literature can be grouped into 2 categories. First there are procedures where experts make the evaluation of a system (Table 1). Secondly there are procedures that evaluate by having test subjects involved (Table 2). Similar tables can be found in [Nielson and Hackos 93] or [Timpe et al. 02].

Experts Analysis	**Description**
Cognitive Walkthrough	An Expert thinks through the steps, a user will have to make.
Heuristic Evaluation	Collection of general principles, whose abidance can be checked.
Using of Models	Models predict the usability of a system. Well-known GOMS-model.
Using of former work	Reuse of known evaluation test and comparison

Table 1: Analysis done by experts derived from [Panic 11].

Analysis with test subjects (users)	**Description**
Pure Observation	Observing directly or per video the behavior of users
Thinking aloud	User speaks their thoughts aloud while using the system.
Protocol analysis	Written or computer created protocols are analyzed
Physiologic indicators	Registering heart activity, Eytracking, skin resistance
Walkthrough	Discussing protocols or video recordings with the test subjects
Questionnaire	Questioning the users. Well-known NASA-TLX [Hart and Staveland 88]
Interview	Test subjects are interviewed

Table 2: Analysis using evaluations with test subjects derived from [Panic 11].

Before starting a human factors evaluation it has to be decided whether the evaluation will be formative (asking for concrete suggestions for improvement) or summative (general evaluation: good or bad). Summative methodologies can be used to compare products to each other or to compare them to criteria lists i.e. from standards [Nielson and Hackos 93], [Timpe et al. 02].

When doing an expert evaluation J. Nielson has shown that using his 'heuristic evaluation' good results are achievable without costly procedures (Nielson 92). A comparison of 6 studies using the heuristic evaluation showed that with 5 evaluators 75% of the usability problems of a human-machine interface were detected [Nielson 92]. [Nielson 94] later generally recommended 5 evaluators as a good mean and established a debriefing with the system developers, if it is a system development project. People from a product's development or sales department should be avoided. They know their product's strengths and weaknesses too well [Timpe et al. 02].

An expert's survey can be a good addition to an evaluation with users and a heuristic evaluation gives more hints on weakness of a product concerning costs, as an evaluation with users does. On the other side an experiment with users results in explicit hints as to what problems the users have with a system [Timpe et al. 02].

If an analysis with users is conducted it can be distinguished between inquiry and observation. Questionnaires and interviews, such as that in Table 2 belong to inquiry, the rest to observation.

The evaluation by observation results in a higher grade of details but is connected to a high effort (costs). Analyzing video records often needs 2 to 10 time longer than the actual recording time. Inquiries are usually much more time efficient [Timpe et al. 02], [Nielson and Hackos 93].

Because some usability problems are often found with one of the methods only, a combination of methods is usually best. This has to be decided on a case by case basis. Typically heuristic evaluation is used to get a first and rough impression of the quality of the usability of a human machine interface. If needed further evaluation with users may follow to find more problems in the interface. By this approach the various methods complement each other positively [Nielson and Hackos 93], [Jeffries et al. 91].

3.2. System Effectiveness and Usability

The idea in SURVEILLE was to have a procedure to evaluate surveillance systems' usability in the way ISO 9241-11 defines it above. It would be nice for decision makers who have to purchase a surveillance technology, to be able to use the survey of surveillance technologies of SURVEILLE [Gulijk et al. 12], open the corresponding fact sheet and to see a classification number for the human factors of the technology. The previous chapter pointed out to which degree the system and its context of use must be known when making an evaluation of human factors. But unfortunately it is not possible to define this context of use of a product completely and record it from the equipment fact sheet of the technology survey. Also it is not possible to define all the goals the user intends to reach by purchasing and using a surveillance technology and to derive all the single tasks the users have to fulfill to reach a goal. To evaluate the usability in the usability sense of DIN EN ISO 9241-11, a precise description of context of use with its four features as users, tasks, means and environment, and the goals as is depicted in Figure 2 is necessary.

So it has to be stated that the usability of a technology as depicted in an equipment fact sheet cannot be determined! And even if all the missing factors can be assessed, a usability

evaluation of all technologies in the survey of surveillance technologies of SURVEILLE would go far beyond the scope of SURVEILLE project frame.

So, if the end users cannot be helped in the intended way, what procedure may help them? The idea is that they cannot be provided with the human factors as effectiveness, efficiency and satisfaction for every surveillance system of the technological survey, they should be helped by an approach to come to a usability evaluation of their specific technology, with their specific use context, including their users (colleagues). If it is not possible to help them directly, it may be possible to enable them to help themselves. The idea is based on current quality of standards for human factors, as reflected in the above mentioned ISO 9241. These standards provide an established consensus of experts from industry, government and research for how-to handle human factors under certain circumstances. They are developed and publicized by international standardization bodies like ISO, CEN, ITU or ETSI. They frequently provide a knowledge base and procedures for evaluating human factors. Beside this the standards concerning human factors focus on a variety of human factor topics and are of diverging quality in applicable examples and methodologies. This means that it is of equivocal help to give just a collection of standards to the users and leave them to it. Therefore a decision support system was developed helping the end users through the standards. The system works on a decision tree by which the standards are sorted, so the users get a list of appropriate results. This resulting list of standards is finally sorted by scoring them on applicability and clearness.

The resulting software is implemented on a server reachable over the internet (Password protected) so the end users can reach it by using their internet browser. The system can be used in English and German. The system guides the users by questions on the basis of the decision tree. The result is a sorted list of standards convenient to the user's situation, ranked on the basis of usefulness. For every list entry a half page summary of the standard is given.

3.3. Standards

Why standards?

The expressions 'standardisation' and 'standard' have manifold definitions. To achieve a unification of terms the standard DIN EN 45020 with its definitions is helpful. [ISO 13] defines a standard as "a document that provides requirements, specifications, guidelines or characteristics that can be used consistently to ensure that materials, products, processes and services are fit for their purpose."

Standards are defined by national or international standardization bodies, examples are ISO, EN, ETSI, and ITU. Finally the quality rating of the standards is based on the quality of the institution that stands behind the standard. These institutions are national as well as international. The internationally accepted institutions often work in close collaboration. The participants of the development of a standard work tightly together and come to a consensus. This usually leads to good quality of guidelines and definitions. This is why the use of standards is proposed for the approach in this deliverable.

Sometimes standards are enforced by law. But even the voluntary following of standards guarantee a certain minimum quality. A well-known example is DIN EN ISO 9000ff, giving recommendations for quality management in enterprises. The seal of approval by a third party is used as proof of quality [Alex and Klein 08].

Although standards are not always mathematically clear, their use leads to better usability and is often a good basis for human factors evaluation [Sherehiy et al. 06]. This provides a motivation to solve the usability task in work package 3 of the SURVEILLE project by using

usability standards to find a good evaluation procedure to optimize effectiveness, efficiency and satisfaction of surveillance systems.

Which standards?

As pointed out the use of standards for the evaluation of system effectiveness and user satisfaction of surveillance systems has a certain quality rating. The next step is to decide which standards to take. Every standard has an application area where it is accepted. This area is technical on the one hand, which is human factors here, and geographical on the other. The geographical aspect comes from the sphere of influence of the standardization body. A human factors standard may be nationally approved, e.g. in Germany, but this does not mean it is necessarily accepted in other European countries, such as Great Britain. Due to the fact that SURVEILLE is a project co-funded by the European Commission and with the aim to provide a comprehensive survey of the types of surveillance technology deployed in Europe the standards used to evaluate human factors in SURVEILLE should be accepted in the whole European Union.

Whether a Standard is developed for a national, European or international level can be seen by its acronym. Fortunately the standardization bodies today do not work in isolation. To prevent chaotic normalization they increasingly work together. Table 3 shows examples of well-known standardization bodies. If one searches for a human factors standard and finds DIN EN ISO 9241 he or she finds guidelines for the securing of human factors of interactive systems. The acronyms show, that the standard is accepted in all 3 levels of Table 3.

Standardization body	DIN Deutsches Institut für Normung (German Institute for Standardization)	CEN Comité Europèen de Normalisation (European Committee for Standardization)	ISO International Organization for Standardization
Level	National	European	International
Acronym	DIN	EN	ISO

Table 3: Well-known standardization bodies.

DIN EN ISO 9241 shows also, that it was developed internationally under the rules of ISO and then it was accepted by the CEN. The standards of CEN have to be taken over by all national bodies in the EU, such asDIN, without changes [Mücke 05]. If DIN ISO <number> is found it means that ISO passed it and then DIN took it over as a national standard for Germany. Generally expressed this means that international standards can be transferred into a national standard optionally.

Beside standards, specifications are a comparable source also, if they are coming from a body that is as trustworthy as the above mentioned standardization bodies are. Specifications can be used earlier in a design and development process than standards and are often adapted faster to the state of the art. Some of them are therefore considered for the evaluation of human factors.

So this work will focus on international standards concerning human factors in the evaluation of system effectiveness and efficiency as well as user satisfaction. The focus is on ISO standards that are accepted nearly worldwide, supplemented by some international standards coming from ETSI and ITU. They cover the telecommunication area, which has a huge amount of devices and where they define human factors rules for. If one scans through the area of standards of the mentioned standardization bodies, more than 30 standards concerning human factors can be found. This is no real help for the end users in SURVEILLE, because they would have to read more than 1000 pages and find that not many of the mentioned standards meet their problems. The next chapters therefore describe how a decision

support tool can be developed that further helps the end users to find solutions inside the standards most fitting to their problem.

Body	Document type	Abbreviation	Example
CEN	Standard	DIN EN	DIN EN 614-1
ISO	Standard	DIN EN ISO	DIN EN ISO 9241-11
ISO	Specification	DIN ISO/TS	DIN ISO/TS 20282-2
ITU	Specification	ITU F.	ITU F.901
ETSI	Specification	ETSI Guide	ETSI EG 201472

Table 4: Types of documents used for the decision tree.

3.4. Software Tools

This chapter describes the process by which the software for the decision support tool was chosen. This tool allows it to find the right standard for usability evaluation described in chapter 3.3. The basis of the decision support is a decision tree that will be introduced in chapter 4.2. This means that the software has to be able to represent this. The software should help the end-users through the procedure by a simple question and answer procedure and thus support him in navigating through the decision tree.

The answering of questions is a typical scenario for learning environments. They present a dialogue in which the authors can enter their questions. Users can answer the question and get new questions depending on their previous answers. Fraunhofer IOSB developed such a learning environment under the name *Crayons* [IOSB 09].

Another possibility depicts an online questionnaire, as are used in psychology and market research. Most common here are simple 'single-choice-questions', which are questions where the user can choose one answer out of a group of given answers. They are an optimal solution representation of choosing a branch of the decision tree. A well-known open source software to develop online questionnaires is 'Lime-Survey' [Schmitz 12].

More flexible solutions present tools that allow building 'Rich Internet Applications'. These applications can be reached as web pages over the internet. By their user interface and short reaction time on user interaction they give the impression of a desktop application. A well-established tool is *jQuery* which has a JavaScript-library of subroutines as main features [Firtman 12]. Another example is Vaadin. The browser of the user runs a small client, a JavaScript program. All the rest is running on a powerful server. To use a tool like Vaadin a fast communication line is needed and technologies like Asynchronous JavaScript and XML (Ajax) as well as JavaScript Object Notation (JSON) are used [Frankel 11].A classical Java software can be used to automatically load the application on the user's computer on demand (Java Web Start).

The process to find the right tool for the decision process was:

1. Collection of possible software tools
2. Develop a short example (where possible)

3. Evaluate the tool by fulfilling a typical task

3.4.1. Evaluated software tools

3.4.1.1. Crayons

Crayons (Figure 3) is a Software Package developed by Fraunhofer IOSB, helping Authors to produce professional online courses. Based on the book metaphor, it allows users to produce and edit with no knowledge of programming. Information as text, Animations or images can be combined and linked. Crayons is used via web browsers like Microsoft Internet Explorer or Mozilla Firefox. The system is independently usable over all operation platforms worldwide. To exchange data the whole course can be exported by Sharable Content Object Reference [IOSB 09].

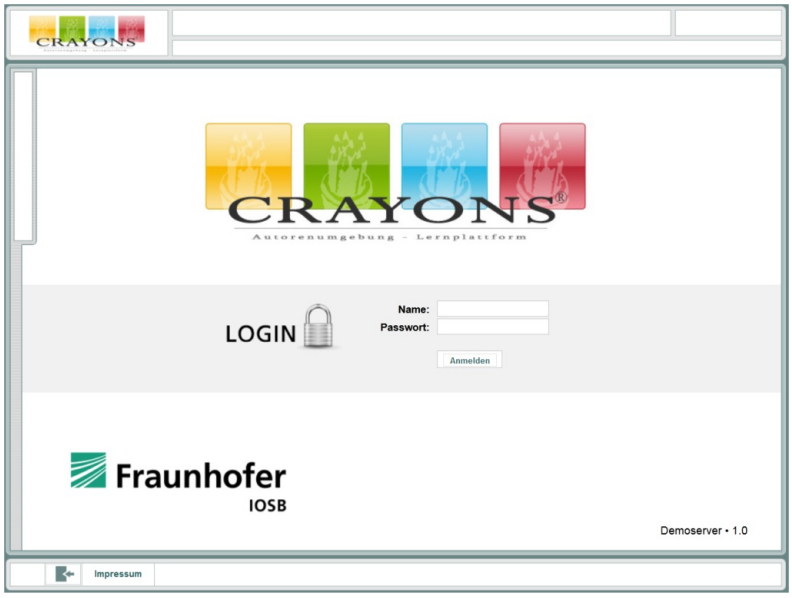

Figure 3: Crayons [http://crayons.iosb.fraunhofer.de/].

Its flexibility makes it adaptable to all needs but extensive changes in the software are needed to build a decision tree. A lot of functions specific for a learning environment could not be used in a decision support system (functional overload). So the conclusion is that it is possible to use crayons, but it is not efficient.

3.4.1.2. LimeSurvey

LimeSurvey is web based licence free open source software for online questionnaires. Used technologies are PHP script and for example MySQL [Schmitz 12]. After configuring the system it is possible to produce enquiries and administrate them (Figure 4). A lot of templates facilitate the building of complex questionnaires. Texts can be adapted to several languages. The templates can be edited directly by an HTML editor.

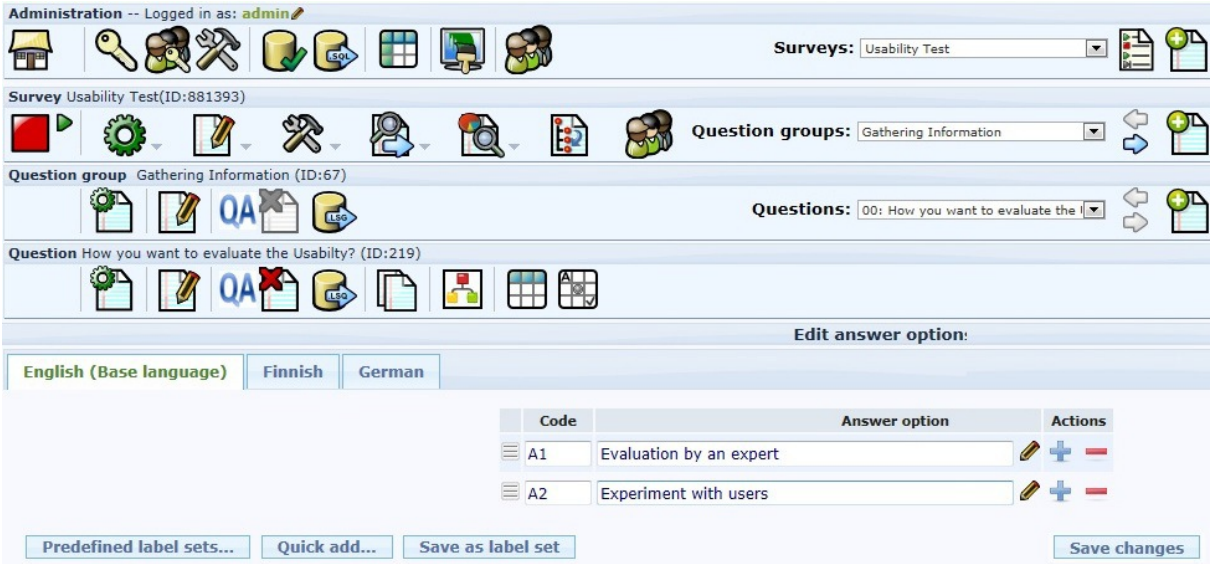

Figure 4: Administration in LimeSurvey.

To estimate the abilities and requirements of LimeSurvey an example questionnaire was developed. It was possible to represent the decision tree, and to have 3 languages. In lower part of Figure 4 one can see English is prepared as language. The language decision is followed by the associated questions. The result for the survey is seen in Figure 5.

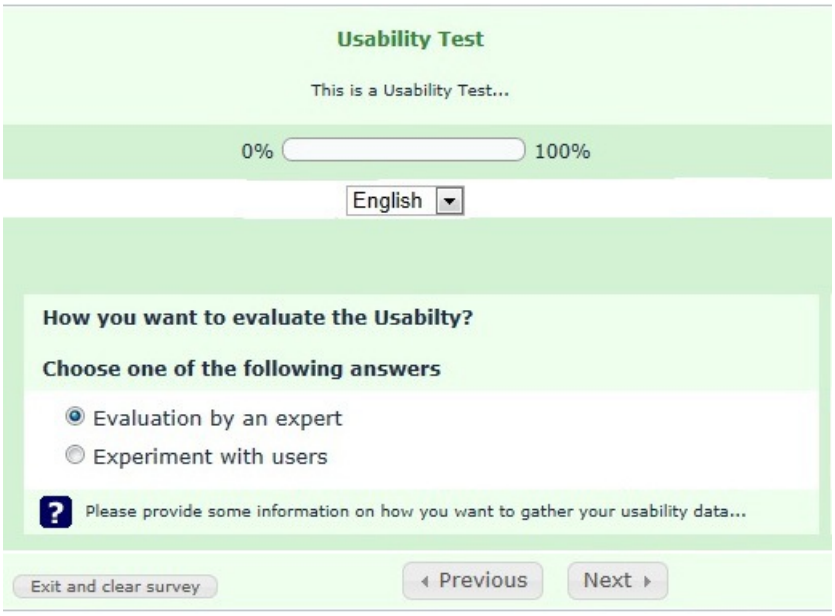

Figure 5: A question as a result of administration in Figure 4.

The first test showed that LimeSurvey is not an ideal tool for our decision support system. Too many modifications would have to be done. Indeed LimeSurvey allows Multilanguage questionnaires where texts can be adapted easily and the templates are flexible and modifications possible by HTML and JavaScript [LimeSurvey 13]. But a lot of question-templates are of type radio button only and have to be modified for use if explanations are wanted, as we do. Browser switching by web button was erroneous. Over all it can be stated, that for our project the cumbersome modifications and an inapplicable functional range lead to the rejection of this tool.

3.4.1.3. jQuery

jQuery basically is a JavaScript library used by many web sites. It has Document Object Level that enables the finding of elements in a document, as well as selecting and manipulating these elements. jQuery can be extended by plugins [jQuery 13]. With jQuery one easily can find components of web pages and manipulate them [Steyer 11] [Chaffer 09].

The script libraries may be hosted by a server under the users' control, or by other server services, e.g. Microsoft. If the library is compressed they are tiny – 32 kB. Developers need knowledge in JavaScript programming as deep as the programmed application is complex [Steyer 11]. Additional functionality can be reached by adding jQuery UI or jQuery Mobile. The library is very flexible and allows a lot of extensions to design human machine interfaces for desktop and mobiles. It is proper for the programming of a decision support system. The result of a test programming sample is shown in Figure 6.

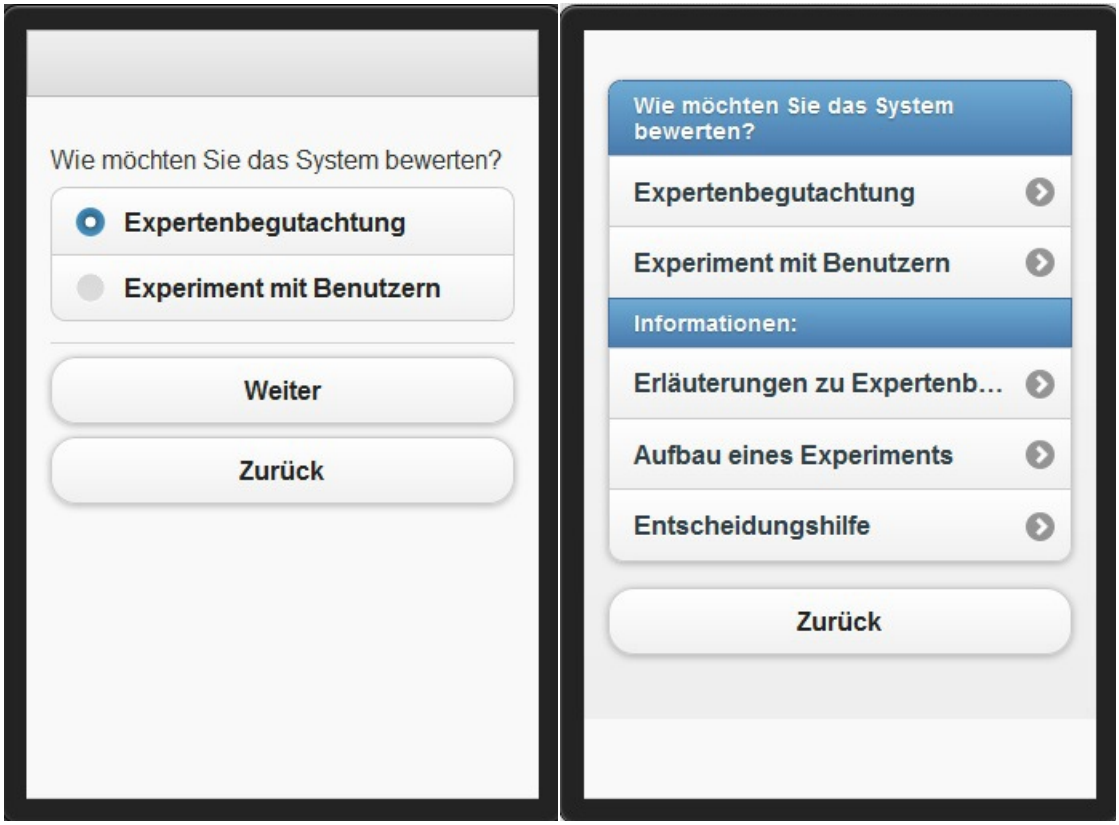

Figure 6: Test examples ‚radio button list', and 'list view' produced with jQuery.

3.4.1.4. Java

Java is a programming language and runtime environment as well. It was published first by Sun Microsystems. As a programming language it is object oriented. A compiler translates the program into bytecode, a machine code the Java Virtual Machine (JVM) can read. JVM can be hardware and software as well, usually the latter. JVM is an interpreter decoding and running the programmed code. JVM is optimized for faster running. Execution time was a problem in the beginning. JVM exists for all standard operation systems. [Flanagan 05] [Oracle 12] [Ullenboom04].

To produce an example for the decision support system the Java Standard Edition and Swing were used. Swing is a programming interface (API) and graphic library for programming graphical user interfaces [Oracle 11].

To build the decision tree, a database was taken into account but discarded because development time and effort were too high. A leaner solution employing an XML file was preferred because the number of data sets, that are human factors standards, was manageable by the XML file. The XML file contains the tree structure and all longer texts, tooltips texts in English and German as well as the standards to be proposed as the solution at the end of the decision process. An expansion for considering the relevance of every added standard was implemented too. The developed example is shown in Figure 7. It has classical menu options, easy language selection, navigation buttons to use the application, and radio button for entering the information for the decision support.

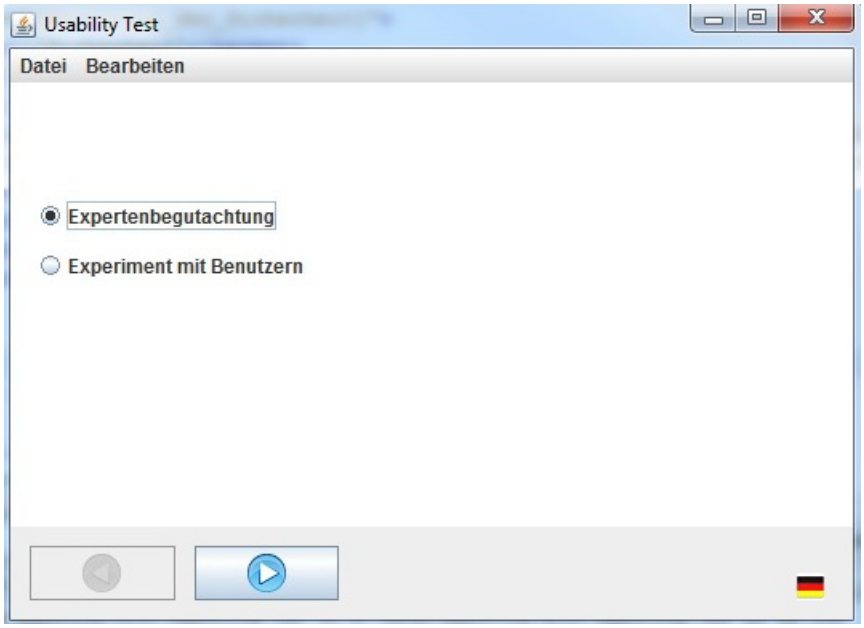

Figure 7: First example of a user interface implemented in Java Swing.

The application programmed in pure JAVA was designed to run locally without internet and as a Java Web Start application. This allows downloading the application from the web [Java 13]. This means that the needed files are downloaded on the user's computer by entering a web address and then run on the user's computer.

A web browser is not compellable, but a requirement is the local installation of a Java Standard Edition platform. If not present the user has to install it himself. Because the user group in SURVEILLE was not estimated to like installing additional software, Java Swing was rejected.

3.4.1.5. Vaadin

Vaadin is a Java framework for web applications to produce rich internet applications. Its architecture is designed in a way that the main part of the application is running on the server. Inside the web browser Ajax (Asynchronous JavaScript and XML) it is possible to send HTTP-requests to the server and change the content of the user's HTML-page. Vaadin includes many technologies such as Google Web Kit GWT, Ajax or JavaScript Object Notation JSON, which is a compact format for data exchange. On the client side JavaScript is running only, no plugins are needed. By using GWT a correct visualization of content is guaranteed in nearly all web browsers [Grönroos 11]. To have GWT running correctly JavaScript has to be activated on client site [Frankel 11].

The structure of Vaadin is presented in Figure 8. The client side engine is the only thing running in the user's browser as a JavaScript. All other components are running on the server's

side. The connection to the server is mandatory. Consequently there is no offline mode [Grönroos 11].

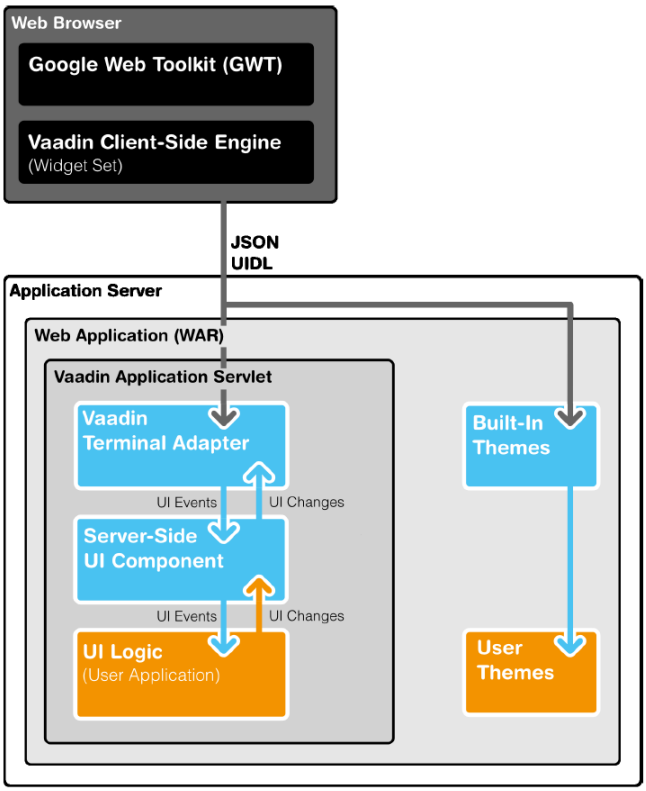

Figure 8: General Architecture of a Vaadin Framework. Modified from [Grönroos 11].

The look and feel of the reviewed examples is excellent. Beside radio button list one can use help texts as extended tooltips, and also images can be included Figure 9.

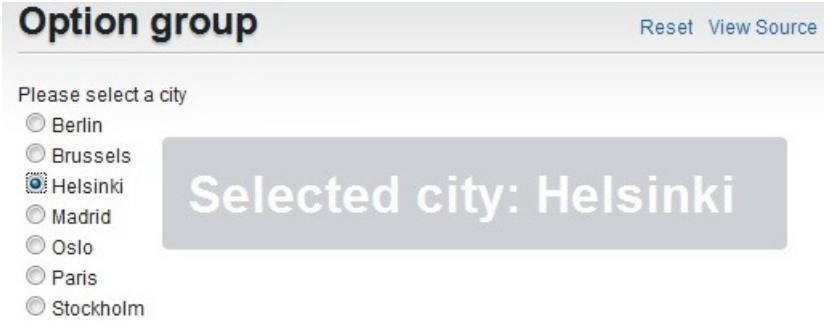

Figure 9: Example of an option group represented by a list of radio button [Grönroos 11].

The functionality of Vaadin is adequate to produce an application as our intended decision support tool. The development can be done in Java. Compared to pure Java on the basis of Swing it is easier accessible by the user and provides a very good visualization of the human-machine interface.

3.4.2. Evaluation of software tools

In the previous chapters software tools were presented. The purpose for analyzing them was to find an adequate tool for the implementation of a decision support system.

The tool Crayons (chapter 3.4.1.1) does not allow an efficient implementation fitting to SURVEILLE's conditions. Considerable changes in the software would be needed, which take too much time. It is possible to implement via Crayons but not feasible. Because of this no prototype was built with Crayons.

Analyzing the applicability of the other tools (LimeSurvey, jQuery, Java, Vaadin) we found that it is possible to build our decision support tool with all of them. To find out how efficient the tools are a prototype was built with all of them. Table 5 shows them together.

Product	Opportunities	Risks	Recommendation
LimeSurvey	• Fast and easy adjustment of the graphical interface • Access via web browser	• Cumbersome modifications • Unsuitable range of functions	Not applicable within the project.
jQuery	• Flexible at use • Various adjustments of the graphical interface possible • Access via web browser	• Extension and maintenance of JavaScript-projects	Not applicable within the project.
Java	• Experienced colleagues available • Access via web browser (Webstart) or local installation	• Executable on mobile devices • Webstart-technology is complicated to use	Not applicable within the project.
Vaadin	• Based on Java • Extensions • Minimum requirements for the client are low • Access via web browser	• Durability of product	Applicable within the project.

Table 5: The analyzed programming tools.

LimeSurvey allows easy adaption of the user interface. Multilingual implementation is therefore easy. The user can reach the application via internet. Complex installations are not needed. Risks may come by cumbersome modifications and inadequate functionality.

The user interface of jQuery can be designed to the user's needs by a lot of extensions. The JavaScript library is flexible to use and suitable for our goals. The risk is a reduced opportunity to make a structured software development as well as software maintenance on the basis of JavaScript.

Java is widely used in Fraunhofer IOSB. A produced application can be accessed locally and via Internet. Via Internet Java Web Start technology has to be used, which means a risk concerning the acceptance of later users by running third party software on their computers. A second risk concerning the acceptance by later users is that the application will not run on every system especially mobiles.

Using Vaadin applications can be built by the use of Java. The generated JavaScript Code is usable without further plugins in the user's browser. Specialties can be solved by extensions which can be embedded in Vaadin. A risk is posed by Vaadin's age. Even though it has been in use for 10 years so far (including its forerunner 'IT mill Toolkit') it is not clear how long Vaadin will be supported in its current version. It is unclear whether future versions will downwardly compatible.

The risks of LimeSurvey outbalance its potential benefits, so we do not recommend it here. We judge the possibility of rejection by future users of Java applications a relevant risk. Pure Java is not recommended either.

Vaadin and jQuery have a balanced proportion of risks and benefits. In the end we chose Vaadin 6.8, the good working knowledge of Vaadin at Fraunhofer IOSB and the possibilities this offered for assistance by colleagues, was deemed essential.

4. DECISION SUPPORT SYSTEM

Next follows a concept for a decision support tool for the selection of an appropriate human factors standard allowing for the evaluation of effectiveness, efficiency and user satisfaction for surveillance technology. Chapter 4.2 introduces a decision tree which categorizes the human factors standards and the 'leaves' of which provide a list of standardization documents, considered in this work.

The resulting list of standards suitable for human factor analysis is still long. To organize this list by an additional optimization process to facilitate searches by end users and reduce their amount of work is proposed in chapter 4.3. The optimization operates by scoring the candidate standards inside the list as well as giving additional data to users i.e. short descriptions of the standard.

Chapter 4.4 follows the proposal for the human-machine-interface of the decision support tool. The last chapter in this section presents the existing concept's implementation.

4.1. Concept

The decision support system AFUS **A**pplication for **F**inding **U**sability **S**tandards should support decision makers buying or developing security technologies in assessing effectiveness, efficiency and user satisfaction. The bases are human factor standards as a consensus of knowledge in human factor design and evaluation. They contend the knowledge that is needed to evaluate human factors of surveillance systems. The standards show, that there is quite a collection of theoretical and experimental methods that all have strengths and weaknesses concerning effort and results. For nearly all of them it is important to know how the future users are trained and skilled in their job and under which conditions they have to work. As pointed out before in chapter 3 this is fully unknown within the boundary conditions of the survey of surveillance technology in SURVEILLE [Gulijk et al. 12]. Neither the risk cycle nor the Bow-tie model [Gulijk et al. 12] give the needed data in the detail needed for usability testing. SURVEILLE's deliverable D2.6, where the matrix of surveillance technologies is described [Guelke et al. 13], goes into more detail but does not describe the usage situation and the surveillance staff in sufficient detail that is needed for human factors testing to get enumerable results of effectiveness, efficiency and user satisfaction.

This basis made it clear that a classic human factors test procedure is not applicable. The human factors tests will have to be made when the details of the intended surveillance technology are made clear, as the intended staff that will use it and working conditions under which the staff has to reach the goals require consideration. This will provide decision-makers, who typically come from SURVEILLE's user group, a basis on which they can decide what usability testing procedure is suitable for the parameters on which the decision support tool is proposed. It should allow choosing the appropriate human factors standard where the end-user can find information to assist in testing.

The next chapters describe the required components for the proposed decision support tool and an implementation of a first prototype.

4.2. Decision Tree

As pointed out in the previous chapters there are numerous human factors standards. They describe various techniques for human factors evaluation. A decision tree is a possible technique to find those standards that fit to the usability testing conditions. [Panic 11] from Fraunhofer IOSB developed such a decision tree for the evaluation of interactive systems where the leaves of the tree represent human factors standards. This tree was adapted to

the conditions in SURVEILLE. The standards considered in SURVEILLE have to be accepted in Europe because SURVEILLE focusses on the field of European surveillance. The intention was to facilitate the search for human factors standard describing a suitable evaluation of the surveillance system. On every point where the tree branches, the question arises which criteria of the investigated surveillance system has to be evaluated. After going through the tree the evaluator gets a corresponding list of standards. Figure 10 gives an example for a node in the decision tree.

```
- <node name="[Abfragetechnik]">
    <parent>[Experiment_mit_Benutzern]</parent>
    <label language="de">Befragung der Benutzer</label>
    <label language="en">User questioning</label>
    <description language="de">Die Befragung der Benutzer kann mithilfe von Fragebögen durchgeführt
      werden. Konzepte und Beispiele für die Gestaltung solcher Fragebögen können den Normen
      entnommen werden, die Ihnen in Folge angezeigt werden.</description>
    <description language="en">User questioning can be performed by means of questionnaires. Concepts and
      examples for the design of these questionnaires can be found in the standards provided
      shortly.</description>
    <image>null</image>
    <child>null</child>
    <result spage="44" page="40" chapter="D" details="1" applicable="3">[DIN EN ISO 9241-420]</result>
    <result spage="6" page="4" chapter="A.6.4" details="0" applicable="1">[ITU F.901]</result>
    <result spage="8" page="4" chapter="2.3" details="0" applicable="3">[ITU F.910]</result>
  </node>
```

Figure 10: A node on the decision tree (XML-code).

Due to the content of the initial standards, they can be separated into three categories. This represents the first level of the tree including the node usability. Usability is the branch we focus on in this report. [Panic 11] added an additional plane to usability and called the result core tree (Figure 11). To this core tree a set of branches is added to better structure the tree. On the leaves' plane the human factors standards are listed. The figures for usability "Expert rating" and "experiment with users" can be found in the appendix 9.2. They show in detail how 'expert rating' is broken down into model based methodologies and analytical methods. Model based forecasts of a system behaviour as GOMS [Card et al. 86] are not mentioned in human factor standards. So no standardization documents can be found in this branch.

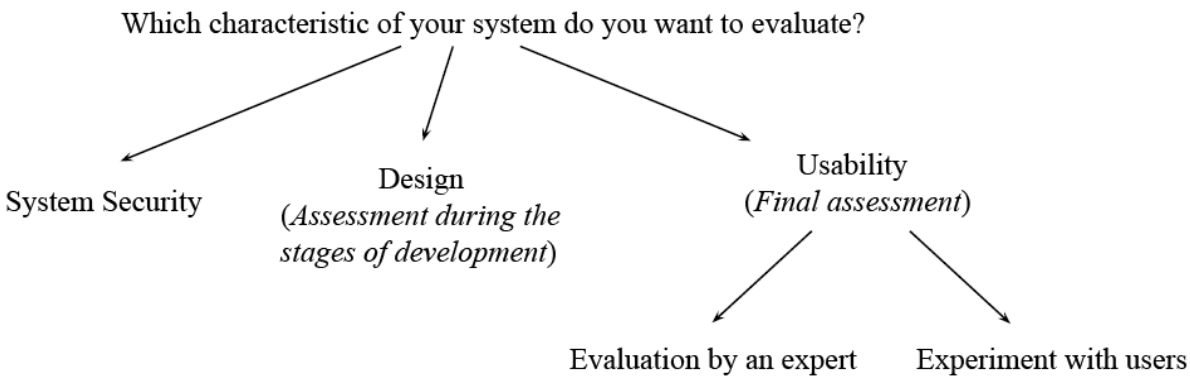

Figure 11: Core part of the decision tree [Panic 11].

4.3. Results differentiation

In the leaves of the decision tree described in chapter 4.2 predefined lists of standards can be found. A typical result list of standards is shown in Figure 12. These resulting lists vary by the path the user takes when going through the decision tree. The length of the lists with up to 13 documents is the next focus of this report. Within these documents the users get to many proposals of human factors evaluations. But even reading these many documents is time consuming. The list in Figure 12 is sorted alphabetically but doesn't present an intuitive search strategy for the users. The following proposal should optimize this.

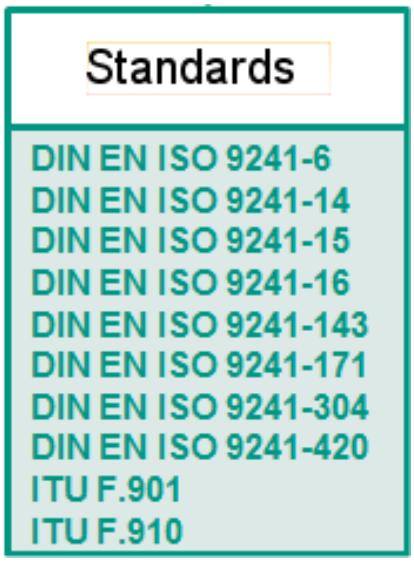

Figure 12: Resulting list of standards after passing the decision tree (not optimized).

The resulting list of standards should support the user in finding a good evaluation procedure fitting to his boundary conditions. The list should be sorted in the order of usefulness. This usefulness has to be determined by the features of the standard combined with the features of the user. Besides this order of usefulness, further important information of the standard should be presented to the users. This includes a digest and a hint where to start reading in the document; some of the standards exceed 100 pages.

Metadata

To have a clear database of standards and specifications concerning human factors some metadata of the documents have to be stored too. These are generally

- Identifier e.g. 'DIN EN ISO 9241-11'
- Year of enactment combined with the body e.g. 'EN'
- URL to the full standard document (for future use)

Some metadata is dependent on the context. Every standard fits varyingly to a user's evaluation intention. In context 1 a standard A may be very useful, but it might be not even applicable for another context 2. So the following metadata are dependent on the path by which the user goes through the decision tree.

- Context (positions in the decision tree)
- Chapter (good starting point for reading the standard document)
- Physical page for paper version, electronic page for electronic reading
- Applicability
- Clarity

The applicability and the clarity of a standard are dependent on the context. Both features may be crucial for the decision of the appropriate standard. They build the basis for the optimization of the result list.

In the future the decision support tool should have a link leading to the text of the full standard directly. This was not possible so far because most of the documents are subject to intellectual property rights. Having a link inside of the tool that opens the full document text to the public is an infringement on these rights. But if the proposed decision support tool should go public, the intellectual property rights can be considered, which means that licenses have to be paid.

Evaluation of applicability and clarity

The next step is finding a scoring for standards concerning its applicability and its clarity. As mentioned this has to be done within the resulting list of standards, which was found by a particular path through the decision tree and building the context. The score was set up in whole numbers. For each feature the number should express how strong the feature is. The quantity chosen is $\{0,1,3\}$. The higher the number the stronger the feature is. The numbers taken are subjective and nonlinear to get a stronger differentiation of the resulting score. To assess the characteristics 'little, normal, strong' all used human factors standards had to be analyzed. This analysis had to be context sensitive because the list of results the standard was in depends on the decisions made through the decision tree. If a standard hits more than 1 list of results it could have different grades because it might fit differently to the contexts the decision tree built.

	Grade	**Characteristic**
	0	Little
Clarity	1	Normal
	3	Strong
	0	Little
Applicability	1	Normal
	3	Strong

Table 6: Scoring of clarity and applicability.

The clarity in the context of an expert's review is characterized as 'little' when it contains no remarks, examples or other clarifying material. In such a case the clarity gets the score '0'. The clarity is 'normal' when the standard has remarks giving the reader important indications, and thus the standard gets the score '1'. Does a standard for further clarification contain examples; the clarity is strong and gets the score '3'. The examples may be valid for very detailed images too. If an example is very marginal it is downgraded. Contrary to this is an explanation that is rich in depth and detail and is evaluated as an example.

The applicability is expressed by the scores $\{0,1,3\}$. These scores depend on the path through the decision tree the end users take. The features needed for a specific score are listed in appendix 9.1. The tables there show which features are taken for the decision "test procedures" or for general assessment criteria.

To take a subjective measure is a first attempt to get a fast result for differentiating the list of standards. Expert's knowledge about the standards content is the basis for projecting how future users can profit from the human factors standard under the given context in every possible branch. To raise the quality in the scoring of Table 6, the numbers in the table have to be adapted due to further knowledge by practicing the subjective rating of the standards

by more people. This could clarify whether the figures are of value and how much they scatter within different people.

Evaluation of usefulness

In the previous chapter a possible evaluation of applicability and clarity was presented. The features have been scored by the values of the quantity $\{0,1,3\}$. By this nonlinear factor of which one of three values may be ascribed we assume a better differentiation of the resulting standards in the decision tree. From these single values a total score, representing the usefulness of a human factors standard for a user, has to be calculated. As shown before a standard document has to be understood by an evaluator. More important than the text itself are examples. The feature clarity should represent this. The second feature of the total score is the applicability, the feature that shows how a human factors document can be used in a specific context. Both features clarity and applicability result in the usefulness of the standards in this context.

The possibility of calculating a total score comes from information retrieval and weights the single features in dependence of areas [Harman 92]. Often a binary score represents whether a feature is present ('1') or not ('0'). The weight factor here sets the portion of the total score. We think that clarity is more important for the user panel than applicability. (Clarity gets a value of '2;' Applicability gets a value of '1'). We differentiate applicability from clarity on the basis of the results of a study by [Souza and Bevan 90]. [Souza and Bevan 90] found out that when building user interfaces by using human factors standards, that the clarity of the standard is more important for the quality than the applicability. If a text is presented clearly, and is understandable by non-human factors experts it is assumed that the text can compensate a reduced applicability. A user e.g. can derive a checklist himself from a list of requirements in the standard. This is easier for non-human factors experts if the standard is formulated well. Vice versa we assume that an existing checklist in a standard not structured and formulated clearly, helps the non-human factors expert much less, because the non-expert will not comprehend a text of insufficient clarity well. The results of the studies of the chapter suggest taking a high weight factor for non-human factors experts. Therefore we assigned clarity a weight factor of 2.

Thus, we come to the following formula for calculating the usefulness of one standard:

$$usefulness = \sum_{i=1}^{2} g_i * s_i$$

where g_i is the weight factor for each feature, thus for applicability it is $g_1 = 1$ and for clarity it is $g_2 = 2$. In our case we do not look for the existence of a feature by a binary value pair but differentiate the rating of applicability and clarity. s_i denotes the single scores with values in {0,1,3}. If for example a standard document has applicability with a value of 3 and clarity value 0 than usefulness can be calculated as

usefulness = 1*3 + 2*0 = 3

4.4. User Interface

Figure 1 gives an impression of the user interface of AFUS. The user is going through questions and can give his answers by clicking the radio buttons. The system prototype is realized in German and English. The initial language is set by the given browser language. If the browser language is not German, then English is set. This can be changed manually afterwards. For each question help text is given, explaining the wider sense of e.g. the question itself and the alternatives in the answer. In the lower part of the user interface the general

navigation button allow the user to go back and on as well as restarting. This way the user has full control of the software at every decision level.

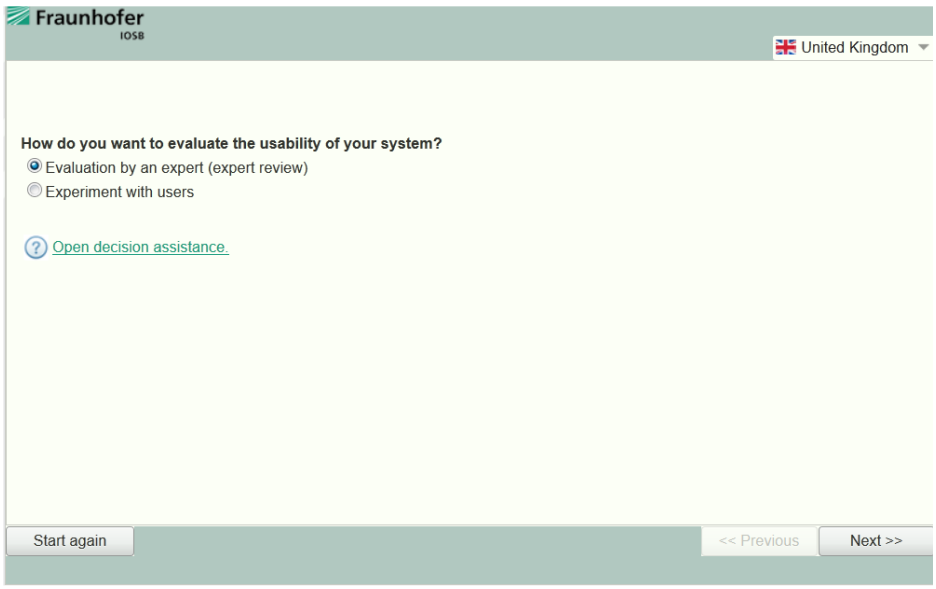

Figure 13: User interface of AFUS.

After going through the dialogue step by step, the resulting list of standards is shown (Figure 14). The list is organized from the best fitting standard at the top to the less useful standards at the bottom. Here 4 standards fit excellently, receiving the highest possible value 9. The smiley gives an additional graphical impression. If the user moves the mouse cursor over the brief description of the standard, a hint where to start reading is given. This helps the user to concentrate on the important part of the standard. Clicking on the standards code gives further information on the standard.

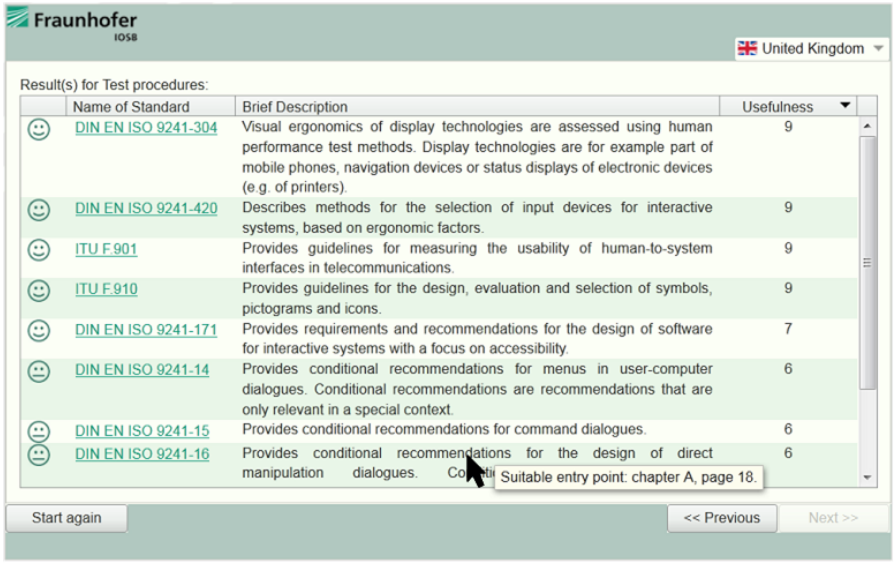

Figure 14: Resulting list of weighted standards of AFUS.

Unfortunately it was not possible to implement the standard document itself into the software, due to intellectual property rights. Fees have to be paid for the text if the text is given publicly or to SURVEILLE partners. This was not foreseen in the proposal. We therefore leave this out for the purposes of the project, but in the event of an advisory panel including an ergo-

nomic counsel is set up, it makes sense to improve the AFUS tool e.g. by adding the actual document of the standard. The software of AFUS allows this without much programming effort.

4.5. Implementation

As mentioned above we decided to use a client-server architecture for the AFUS software. This allows minimum influence on the user's computer, the client side. Even tablet-PCs or smart phones can be used on the client side. The architecture allows central software maintenance and a central maintenance of the standards database as well.

AFUS is a special kind of decision support system. Because of its structure, the system can be classified as an Expert System (ES) [Marakas 03]. Some simplifications have been made, to assure AFUS's compact proportions. For example we use an XML-file as a replacement for a typical relational database because we do not have so many standards that a database is really needed. During the design stage, the system was filled with domain-specific expert knowledge. This knowledge is not a set of automatically retrieved data; in fact the data, which are all human factors standards mentioned in the resulting lists, was collected and interpreted by experts of Fraunhofer IOSB before adding it to the system.

```
- <resultItem name="[DIN EN ISO 9241-307]">
    <label language="de">DIN EN ISO 9241-307</label>
    <label language="en">DIN EN ISO 9241-307</label>
    <version>2009</version>
    <source>http://google.com/?q=DIN EN ISO 9241-307</source>
    <description language="de">Enthält verschiedene Prüfverfahren für optische Anzeigen.</description>
    <description language="en">Provides different test methods for display technologies.</description>
  </resultItem>
```

Figure 15: One standard within a resulting list (XML-code).

To deliver the data the singleton "EvaluationSession" was used, where the language, position of the file and the administration of the node can be managed. Further a "HierarchyObject" has to be defined from which the "HierarchyNode" and the "HierarchyResultItem" can inherit. The system was built by using Vaadin 6.8 (chapter 3.4). The panel of the user interface was built in a flexible form, so that future findings e.g. by the user group test can be considered easily. All clickable parts of the user interface have tool tips, see Figure 14.

5. EVALUATION OF THE DECISION SUPPORT SYSTEM

A very important part of an expert system is the user interface, or in other words: "… the success or failure of an expert system can often be attributed to the quality and functionality of its interface…" [Marakas 03]. As a consequence, the evaluation of the AFUS interface by future end-users is of high importance.

The first attempts of evaluation were carried out in an informal way, by asking the end-users for their comments. Later on, a user group test, including a tailored questionnaire, was carried out.

The procedure for the user group test was as follows: First, the participants had to walk through two different scenarios for which they had to find the most appropriate usability standard document. They were also asked which kind of usability evaluation they would perform in the given scenario. Afterwards a questionnaire was used to assess the users' opinion about the AFUS application.

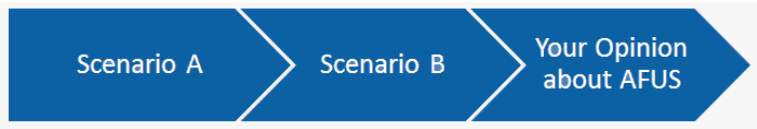

Figure 16: Sequence of the test procedure.

The questionnaire used in the user group test consists of two parts:

Part one contains the questionnaire ISONORM 9241-110-S, which evaluates software upon the international standard ISO 9241-110. This includes the seven dialog principles "Suitability for the task", "Self-Descriptiveness", "Conformity with user expectations", "Suitability for Learning", "Controllability", "Error Tolerance" and "Suitability for Individualization" covered by 21 questions. ISONORM 9241-110-S is based on the former version ISONORM 9241-10 proposed by J. Prümper [Prümper 99]. The standard document ISO 9241-110 only focuses on dialogue principles related to the ergonomic design. It does not consider any other aspect of design.

Part two covers the design quality and the attractiveness of the AFUS interface. It is a shortened version of the "User Experience Questionnaire" (UEQ) developed by [Laugwitz et al. 08]. UEQ allows the assessment of the user experience of an interactive product. Users can express their feelings, attitudes and impressions that arise from using the product by answering a 26 item questionnaire. The original UEQ covers the aspects of attractiveness, design quality (stimulation, novelty) and use quality (efficiency, perspicuity, dependability). Since only the aspects of attractiveness and design quality of the AFUS interface are to be assessed, a smaller set of 14 items is sufficient.

To evaluate the gathered data, a minimum number of participants are required. Until now, the recorded number of responses is too low to make any statistical calculations. Nonetheless, there are some tendencies:

Part one of the questionnaire (ISONORM 9241/110-S) delivers excellent values for six out of seven dialog principles (Table 7)Table 7: Preliminary results from user group test with ISONORM 9241/110-S. It seems as the principle of "Self-descriptiveness" is not fulfilled in all aspects. It appears that the users strongly agree that AFUS "provides *on request* context-sensitive explanations, which are helpful", but don't think the explanations are offered automatically. Additionally it seems that AFUS offers insufficient information about which entries are valid and necessary. Consequently, the overall score of the dialog principle "self-descriptiveness" is low although the users stated the application is self-declaring. The other

six dialog principles are rated high, resulting in an excellent overall score for the AFUS application.

Suitability for the task	high values
Self-Descriptiveness	low values
Conformity with user expectations	high values
Suitability for Learning	high values
Controllability	high values
Error Tolerance	high values
Suitability for Individualization	high values

Table 7: Preliminary results from user group test with ISONORM 9241/110-S.

Part two seems to show that users like the appearance of AFUS. The application reached high values in attractiveness and design quality (Table 8).

Attractiveness	high values
Stimulation	high values
Novelty	high values

Table 8: Preliminary results from user group test with UEQ (shortened version).

Within the framework of the questionnaire, some user comments are gathered. For example, there was a request to provide the whole standard document text in the result list of AFUS (because of licences issues, the links to the full scale documents are not accessible). It was stated that the application is easy to handle and self-declaring.

To sum up, it can be said, that AFUS reached high values in the current available user responses. The number of participants is too low to make any statistical calculations, but the first results seem to be promising. Furthermore an active support of the end-users is indispensable to gain more information about the variety and extent of future improvements. AFUS is one of the main agenda items at the SURVEILLE End-User Panel meeting being held on 24 October 2013 in Mannheim where we expect to get more input for an evaluation.

6. CONCLUSION

The approach to help the end-users in the SURVEILLE project to find an applicable human factors standard for an evaluation of effectiveness, efficiency and user satisfaction by a decision support tool is practicable. The users can find an applicable standard quickly by using the developed tool AFUS. The standard proposed by AFUS is derived by the decisions of the user. These decisions represent the user's edge conditions. By answering a manageable set of questions the user gets an adapted set of human factors standards that are ranked by the clarity of the text document and the applicability of mentioned rules and examples.

The result we achieved with this work package can be optimized in several directions. We did not succeed in offering the users the full text document because of property rights. But the users want to have them to evaluate the quality of AFUS results.

The decision tree is a first projection. It has to be advanced by further research. Standards are constantly maintained by national or international standardization bodies that lead to permanent change. These changes have to be maintained in the database of AFUS. The differentiation of the resulting list of standards the user gets after going through the decision tree has to be validated by user tests.

When developing AFUS the goal was to end up with one test procedure, because we thought efficiency forces end-users into a lean frame of time and effort. First talks gave the hint that in big projects there is room for more testing. This situation was not borne in mind and has to be discussed at the end-user meeting in October 2013. A main question to be answered there is how non human factors experts can handle a decision tree as proposed here, in giving the proper answers to AFUS.

Most important is to verify the encouraging results by resilient user tests.

7. REFERENCES

[Alex and Klein 08]: D. Alex, M. Klein. Einführung in die DIN Normen. Viehweg and Teubler Publisher, Berlin. 1999

[Card et al. 86]: S. Card, P. Moran, A. Newell. The Psychology of Human-Computer Interaction. Lawrence Erlbaum Associates, Hillsdale. 1986.

[Chaffer 09]: J. Chaffer. Learning jQuery 1.3 Packt Pub Ltd. Birmingham. 2009

[Cockton 2013]: G. Cockton. 15. Usability Evaluation. In „The Encyclopedia of Human-Computer Interaction", 2nd Ed. Soegaard et al. (eds.) Interaction Design Foundation. Aarhus, Denmark 2013

[EN ISO 9241-11]: DIN EN ISO 9241-11: Anforderungen an die Gebrauchstauglichkeit – Leitsätze. (Guidance on usability). 1998

[Firtman 12]: M. Firtman. jQuery Mobile: Up and running. O'Reilly Media, Sebastopol, 272 pp. 2012

[Flanagan 05]: D. Flanagan. Java in a Nutshell. O'Reilly Media, Incorporated, Sebastopol. 5th edition. 1256 pp. 2005

[Frankel 11]: N. Frankel. Learning Vaadin. Packt Pub Ltd., Birmingham UK, 412 pp. 2011

[Grönroos 11]: M. Grönroos: Book of Vaadin. Vaadin Ltd. Turku. or https://vaadin.com/book

[Guelke et al. 13]: J. Guelke, T. Sorell, K. Hadjimatheou, M. Scheinin, J. Andrew, J. Lavapuro, T. Ojanen, M. Garzia, P. Vermeulen, M. Vermeulen, B. McNeill, C. van Gulijk, S. Sillem, P.-H. Lin, B. Kooij. SURVEILLE Deliverable 2.6: Matrix of Surveillance Technologies. Report of EU Seventh Framework Programme FP7-SEC-2011-284725. 192 pages. 2013

[Gulijk et al. 12]: C. van Gulijk, H. Vagts, S. Höhn, O. Yaroyvi. SURVEILLE Deliverable 2.1: Survey of surveillance technologies, including their specific identification for further work. Report of EU Seventh Framework Programme FP7-SEC-2011-284725. 68 pages. August 2012

[Harman 92]: D. Harman. Ranking Algorithms. In Information Retrieval: Data structures and Algorithms, W. Frakes, R. Beaza-Yates (eds.). Prentice-Hall. 1992

[Hart and Staveland 88]: S. Hart. L. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Human mental workload. P. Hancock and N. Meshkati (eds.) Amsterdam: North Holland. Pp. 139-183. 1988

[IOSB 09]: Fraunhofer IOSB. Computer-assisted learning – Fun and usefulness combined. 2009. http://www.fraunhofer.de/en/press/research-news/2009/02/Presseinformation250209.html assessed August, 6th 2013

[ISO 13]: Definition 'What is a standard?' http://www.iso.org/iso/home/standards.html, assessed August 06th, 2013

[Java 13]: Java. What is Java Web Start and how is it launched? http://java.com/en/download/faq/java_webstart.xml assessed August 8th, 2013

[Jeffries et al. 91]. Robin Jeffries, James R Miller, Cathleen Wharton, and Kathy Uyeda: User interface evaluation in the real world: a comparison of four techniques. In Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology, pp. 119-124. ACM, 1991

[Jquery 13]: jQuery. Jquery homepage. 2013. http://jquery.com/ assessed August 7th, 2013

[Koivunen and May 13]: M.-J. Koivunen, M. May. Exploring Usability Enhancement in W3C Process. Online Source http://www.w3.org/2002/Talks/0104-usabilityprocess/slide3-0.html accessed July 29th 2013

[Laugwitz et al. 08]: B. Laugwitz, T. Held, M. Schrepp. Construction and evaluation of a user experience questionnaire. In HCI und usability for education and work. A. Holzinger (ed.) Springer Berlin, Heidelberg. Pp. 63-76. 2008

[Lazar et al. 10]: J. Lazar, J.H. Feng, H. Hochheiser. Research Methods in Human-Computer Interaction. John Wiley and Sons, Publisher. Chichester UK. 426 p. 2010

[LimeSurvey 13] LimeSurvey: Features of LimeSurvey, 2013. http://www.limesurvey.org/en/about-limesurvey/features, assessed August 06th, 2013

[Marakas 13]: G. Marakas. Decision support systems in the 21st century. Prentice Hall New Jersey. 2013

[Mücke 05]: E. Mücke. Kennzeichen von DIN-Normen und der korrespondierenden europäischen und internationalen Normen. Beuth Publisher, Berlin. 90 pp. 2005

[Nielson 92]: J. Nielson. Finding usability problems through heuristic evaluation. In proceedings of the SIGCHI conference on Human factors in computing systems, pp. 373-380. ACM, 1992

[Nielson 94]: J. Nielson. Heuristic Evaluation. In “Usability inspection methods”. J. Nielson, L. Mack (eds.). John Wiley and Sons. New York. Pp. 25-62. 1994.

[Nielson and Hackos 93]: J. Nielson, J.T. Hackos: Usability engineering. Academic Press. San Diego. 1993

[Oracle 11]: Oracle. Swing Enhancements in the JavaTM Standard Edition. 2011 http://docs.oracle.com/javase/6/docs/technotes/guides/swing/6.0/index.html 6th edition assessed August 8th, 2013

[Oracle 12]: Oracle. Differences between Java EE and Java SE. 2012. http://docs.oracle.com/javaee/6/firstcup/doc/gkhoy.html assed August 8th, 2013

[Panic 11]: D.L. Panic. Bewertung interaktiver Systeme. Seminar Paper of Karlsruher Intitut of Technology and Fraunhofer IOSB, Karlsruhe. 50 p. 2011

[Prümper 99]: J. Prümper. Test IT: ISONORM 9241/10. In Human-Computer Interaction – Communication, Cooperation and Application Design, H.J. Bullinger, J. Ziegler (eds.). Lawrence Erlbaum Associates, Mahwah New Jersey. Pp. 1028-1032. 1999

[Schmitz 12]: C. Schmitz. LimeSurvey Manual, 2012. Online Ressource http://manual.limesurvey.org/wiki/LimeSurvey_Manual, assessed August 06th, 2013

[Sherehiy et al. 06]: B. Sherehiy, D. Rodrick, W. Karwowski. An Overview of International Standardization Efforts in Human Factors and Ergonomics. In Handbook on Stand-

ards and Guidelines in Ergonomics and Human Factors. W. Karowski (ed.). Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey and London. Pp. 3-46. 2006

[Shneiderman and Plaisant 05]: B. Shneiderman, C. Plaisant. Designing the user interface. PEARSON, Addison Wesley Publishing Company, Boston et al. 653 p. 2005

[Sikorski 08]: M. Sikorski. HCI and the Economics of User Experience. In E. L.-C. Law, E. T. Hvannberg, G. Cockton (eds.). Maturing Usability – Quality of Software, Interaction and Value. Springer Publishing Company, London. Pp. 318-343. 2008

[Souza and Bevan 90]: F. de Souza, N. Bevan. The Use of Guidelines in Menu Interface Design: Evaluation of a Draft Standard. In Proceedings of the IFIP TC13 Third International Conference on Human-Computer Interaction, 1990

[Steyer 11]: R. Steyer. jQuery: das JavaScript-Framework für interaktives Design. Pearson Education, Munich, p. 51, 2011

[Timpe et al. 02]: K.P. Timpe, J. Jürgensohn, H. Kolrep. Mensch-Maschine-Systemtechnik: Konzepte, Modellierung, Gestaltung. Vol. 2. Symposion Publishing GmbH, Düsseldorf, Germany. 2002

[Ullenboom 04]: C. Ullenboom: Java ist auch eine Insel. Galileo Press, Bonn. 2004

8. INDEX OF ABBREVIATIONS

Abbreviati-on	Meaning	Remarks
AFUS	Application for Finding Usability Standards	
Ajax	Asynchron JavaScript and XML)	
CEN	Comité Européen de Normalisation	
ETSI	European Telecommunications Standards Institute	
GOMS	Goals, Operators, Methods, Selection Rules	A model based evaluation method by Card, Moran and Newell
GWT	Google Web Kit	
HW	Hardware	
ISO	International Organization for Standardization	
ITU	International Telecommunication Union	
JSON	JavaScript Object Notation	
JVM	Java Virtual Machine	Can be HW or SW
NASA-TLX	NASA Task load index	Assess work load by 5 7-point scales
PHP	Hypertext Preprocessor	Former: **P**ersonal **H**ome **P**age Tools
SW	Software	
UEQ	User Experience Questionnaire	

9. APPENDIX

9.1. Scoring Tables

	Score	Present features
Clarity	3 1 0	Examples Annotations No annotations, examples etc.
Applicability (Test procedures)	3 1 0	Checklist, test report Requirements Guidelines
Applicability (General assessment criteria)	3 1 0	General guidelines Guidelines for product groups Product specific guidelines

Table 9: Scores for clarity and applicability in the context of “evaluation by experts”.

	Score	Present features
Clarity	3 1 0	Examples Annotations No annotations, examples etc.
Applicability (Test procedures)	3 1 0	Observation procedure Requirements for observation General principles
Applicability (General assessment criteria)	3 1 0	General guidelines Guidelines for product groups Product specific guidelines

Table 10: Scores for clarity and applicability in the context of “user observation”.

	Score	Present features
Clarity	3 1 0	Examples Annotations No annotations, examples etc.
Applicability	3 1 0	Questionnaire Preparation principles General principles

Table 11: Scores for clarity and applicability in the context of “user questioning”.

9.2. Decision Trees

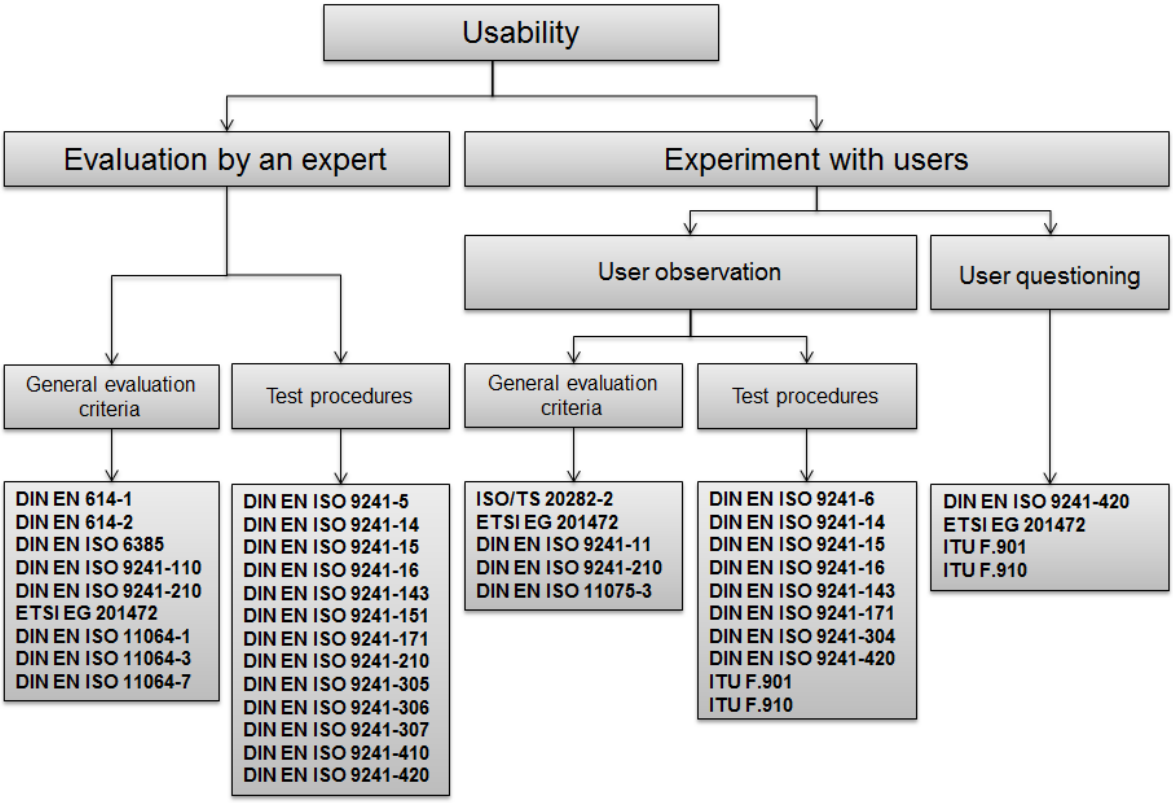

Figure 17: Decision Tree for Usability Testing in SURVEILLE.

Figure 1: Decision Tree [Panic 11]