FP7 – SEC- 2011-284725

SURVEILLE

Surveillance: Ethical issues, legal limitations, and efficiency

Collaborative Project

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 284725

SURVEILLE Deliverable 3.8
Report combining results of all effectiveness research

Due date of deliverable: 31.05.2014

Actual Submission date: 29.05.2014

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

Author(s): Michelle Cayford, Dr. Coen van Gulijk, Dr. Simone Sillem (TU Delft)

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)	

Executive summary

This document builds upon research conducted in Work Package 3 of the SURVEILLE project, summarizing earlier work. On the basis of work done, it reports on changes and discussions concerning the usability scoring applied in the assessment matrix of SURVEILLE deliverables D2.6 and D2.8. In that matrix, technical usability, fundamental (human) rights and ethical considerations are scored for surveillance technologies that are deployed in counter-terrorism and serious crime fighting.

This report is solely oriented on the (technical) usability scoring. The objective for this report was to review and update the usability scoring used in D2.6 and D2.8, based on the earlier reports in the technical work package in the SURVEILLE project (WP3). This deliverable is mostly a report of a reflection on earlier work and discussions within the SURVEILLE project and does not add new information as such; therefore, it does not add new literature references. For the scientific background, the authors refer to the earlier deliverables D3.1 to D3.6.

Reports D3.1 to D3.6 were briefly reviewed and their impact on the scoring was assessed. This assessment yielded the first scoring revision. This version was discussed at the April meeting of the SURVEILLE End User Panel of law enforcement officials, combined with e-mail discussions with other project partners to yield a second scoring revision to address the concerns of the End User Panel and other project partners. The resulting new version of the usability scoring table is given below.

Table 4: Usability scoring – second revision

Factor	Attribute	Sub-category	Sub-category yes/ no	Score
Effectiveness				0-3
	Delivery			0-1
	Context			0-1
	Sensitivity			0-1
Cost				0-3
	Initial cost			0-1
		Purchase price	y/n	
		Installation cost	y/n	
		Space requirement cost	y/n	
	Personnel requirements			0-1
		Number of personnel	y/n	
		Training required	y/n	
		External partners	y/n	
	Additional running costs			0-1
		Maintenance & sustainability	y/n	
		False-positive rate	y/n	
		Other (power, transport, etc.)	y/n	
Privacy-by-design				0-3
	Data collection			0-1
		Selective	y/n	
		Minimized	y/n	
		Overt or covert	y/n	
	Data access & use			0-1
		Who has access	y/n	
		Clear regulations	y/n	
		Protection against function creep	y/n	
	Data protection			0-1
		Encryption or otherwise access protected	y/n	
		Protected against manipulation	y/n	
		Secure against theft	y/n	
Proven technology				0-1

Each attribute scores 0, 0.5, or 1. If only one sub-category scores 'y,' the attribute scores 0. If two sub-categories score 'y,' the attribute scores 0.5. And if all three sub-categories score 'y,' the attribute scores 1.

Table of Contents

1. Introduction

In Deliverable D2.6 of SURVEILLE a matrix of surveillance technologies was developed, which scored technologies according to the categories of usability, ethics, and fundamental (or human) rights. This report discusses the usability scoring of this matrix, reporting on the findings of deliverables in Work Package 3 (WP3) – Perceptions and Effectiveness of Surveillance – and on input from the MERPOL End User Panel of law-enforcement officials (EUP). The subsequent changes and developments for the scoring are described, with an updated usability scoring system proposed.

Section 2 of this deliverable reviews the original framework for usability assessment as presented in D2.6. Section 3 summarizes WP3 deliverables, focusing on aspects that impact the usability scoring. This is followed by a first round of scoring revision in Section 4. This revised scoring was presented to the MERPOL End User Panel, and their feedback and the resulting second iteration of scoring revision is discussed in section 5. Lastly, we discuss the EUP's concern that the fact that any police decision to use surveillance technology is based first and foremost on justification, necessity, and proportionality, is not reflected in the usability scoring. We found that this concern was related to different understandings as to where in the overall multidimensional scoring and in what manner justification, necessity, and proportionality were assessed. The aim of the technological usability scoring assessment developed in SURVEILLE is not to determine if a law enforcement official had legal basis for using surveillance equipment, as this matter forms a distinct dimension of the legal (fundamental rights) assessments. There, fundamental (or human) rights provisions binding upon the EU and its Member States are applied both to assess any application of national or European law by those who operate surveillance technologies, and to evaluate the law itself as to its conformity with fundamental rights. The overall scoring matrix of SURVEILLE and the various assessments included in it are capable of producing results that can be used and further developed for reviewing policy and legislation, so that the regulation and the use of surveillance technology can have a stronger legal and justificatory basis.

2. Usability Scoring – D2.6

The development of a framework for a usability assessment began with D2.6. In this deliverable a semi-quantitative method was presented with numeric values, or scores, assigned to attributes of surveillance technologies relating to usability. The four factors used for scoring surveillance technology were: effectiveness, cost, privacy-by-design and excellence. Nominal scoring methods were chosen for these four factors, meaning that predefined attributes are either present or not present in a given technology.

In D2.6 the assignment of factors is as follows: 'Effectiveness' is scored from 0 to 3, depending on which three effectiveness attributes are present. 'Cost' and 'Privacy-by-design' are also scored from 0 to 3. 'Excellence' is a separate attribute, which scores either 0 or 1. The scores of the four factors are added together which yield an overall usability score ranging from 0 to 10.

Effectiveness

The three attributes of effectiveness described as either present or not present in the surveillance technology are delivery, simplicity, and sensitivity. 'Delivery' refers to whether a particular technology, correctly applied in a particular context, yields a useful outcome. Examples of useful outcomes are: detection of prohibited conduct, items or substances; sufficient facts for justifying pre-emptive actions or sufficient leads to continue an investigation. When there is evidence of prior successes or success is reasonably achievable this attribute scores a 1, otherwise it scores 0.

'Simplicity' relates to the structure and ease of operation that a surveillance technology provides. As a general rule, the simpler a surveillance technology, the more useful it is in crime control. When the structure required for a surveillance operation involves more than one external expert and/or stakeholders, it is considered to be a complex structure operation. When the surveillance technology has proven ease of use in prior cases or its ease of use is reasonably achievable this attribute scores a 1.

The sensitivity of the technology relates to the likelihood of error. A technology may deliver information that is open for multiple interpretations or provides vague data that enables a wrong conclusion. For instance, recording a telephone conversation could be done by phone tapping or by CCTV cameras. The phone tap is more sensitive when it comes to recording the exact conversation, making it less likely that false conclusions are drawn than with CCTV; by contrast, the CCTV footage is more sensitive when it comes to understanding whether someone was coerced into performing the phone call. When there is evidence of a high rate of errors in the interpretation, or errors could reasonably be expected, this attribute scores a 0. Otherwise it scores 1. When there is no evidence about the error rate and an assessment cannot be made it also scores 1.

Cost

The attributes for cost are purchase cost, personnel requirements, and additional resources. 'Purchase cost' is the money spent on buying the equipment and associated systems. Using the equipment list in D2.6, price ranges upwards of 50,000 euros are considered to be expensive. This scores a 0; if prices are lower, a 1 is scored.

'Personnel requirements' relates to the number of personnel involved in the use of the surveillance technology. Personnel in this attribute are restricted to the personnel of the organization that performs the surveillance task. This can be within a single police force, a single national coordination team or a dedicated technical surveillance team. Note that it is assumed that the personnel have received training and are experienced with the technology. When two or fewer persons are involved in handling the intelligence gathering process in an operation, it scores a 1, while if more personnel are involved the attribute scores 0.

The category of 'Additional resources' indicates whether external partners are required in the use of the surveillance technology. These could be commercial partners or vendors that operate the surveillance technology that must be contracted for their assistance. Regardless of the amount of money spent, it is a financial complication that drives cost up. So when a third party has to be contracted this attribute scores a 0, and if it does not, it scores 1.

Privacy-by-design

The associated attributes for the factor 'Privacy-by-design' are observation of persons, collateral intrusion, and hardware and software protection. 'Observation of persons' refers to what the surveillance technology actually observes – people or objects. When a surveillance technology only observes chemicals, objects or data, it scores a 1. When it records people or their behavior, or records their voices, it scores a 0.

The question related to 'Collateral intrusion' is whether a surveillance technology can perform targeted surveillance or whether it records a larger group of people where only one is relevant. When it is targeted to the individual or individuals under investigation this attribute scores a 1, otherwise it scores a 0.

'Hardware and software protection' indicates whether it is difficult, from a technological perspective, to insert privacy-by-design rules. This can be either in the design of the hardware or the software for the system. When it is difficult from a technological point of view to adhere to privacy-by-design principles, this attribute scores 0, otherwise it scores 1.

Excellence

The criterion for 'excellence' is whether a given technological system has proven its use beyond reasonable doubt. Explicit examples include iris-scans or DNA sampling for personal identification; their correctness and excellence have both

been proven scientifically and been successfully applied in crime fighting without doubt. Therefore, when a surveillance technology has proven its use beyond doubt, it scores a 1. Otherwise it scores 0.

In summary, the first framework for usability scoring, developed in D2.6, is given in Table 1:

Table 1: First framework for usability scoring (D2.6)

Factor	Attribute	Not present/ present	Score
Effectiveness			0-3
	Delivery	0/1	
	Simplicity	0/1	
	Sensitivity	0/1	
Cost			0-3
	Purchase cost	0/1	
	Personnel requirements	0/1	
	Additional resources	0/1	
Privacy-by-design			0-3
	Observation of persons	0/1	
	Collateral intrusion	0/1	
	Hardware & software protection	0/1	
Excellence			0-1

This scoring system was created as an initial starting point. It was expected that it would develop and change according to ongoing research in the SURVEILLE project and based on input from project members.

3. Review of Work Package 3 Deliverables

Work Package 3, Perceptions and Effectiveness of Surveillance, produced deliverables 3.1-3.6, each reporting on different aspects of effectiveness.

The deliverables in Work Package 3 are the following:

- D3.1: Report describing the design of the research apparatus for the European-level study of perceptions
- D3.2: Report of results of European-level study on perceptions, including an overview of effects and side effects of surveillance and their perceived effectiveness
- D3.3: Report on system effectiveness, efficiency and satisfaction assessment
- D3.3b: Report on system effectiveness, efficiency and satisfaction assessment
- D3.4: Report describing design of research methodology for assessing effectiveness of selected representative surveillance systems
- D3.5: Cost model for surveillance techniques
- D3.6: Report on methodology and criteria for incorporating perception issues in the design phase of new surveillance systems
- D3.8: Report combining results of all effectiveness research

Deliverables 3.2-3.5 are aimed at fulfilling SURVEILLE Objective 3.1: To assess the benefits and costs of surveillance technology. Deliverable 3.6 produces proposals for improving effectiveness in the future, thus fulfilling Objective 3.2: To produce proposals for improving the effectiveness of security surveillance, while taking fully into account perceptions, economic costs, legal limitations and ethical issues. This paper will also contribute to improving the effectiveness of security surveillance through proposals, in this case related to the scoring of surveillance technology.

3.1 D3.1 – Report describing the design of the research apparatus for the European-level study of perceptions

Deliverable 3.1 gives an overview of European projects on surveillance perception that have been completed or are ongoing. A description of each project and how it relates and can contribute to SURVEILLE is given.

Moreover, this deliverable gives an overview of the effects and side effects of using surveillance:

1. Surveillance technologies being perceived as a threat themselves – they may interfere with various aspects of peoples' lives and may infringe on personal rights (e.g., privacy and freedom of movement)
2. Chilling effect – on free speech, free association, etc.
3. Security dilemma – may have the effect of over-sensitizing people to the perception of threats and thereby making them feel unsafe
4. Self-surveillance and normalization – people obey not because they are monitored but because of their fear that they could be watched

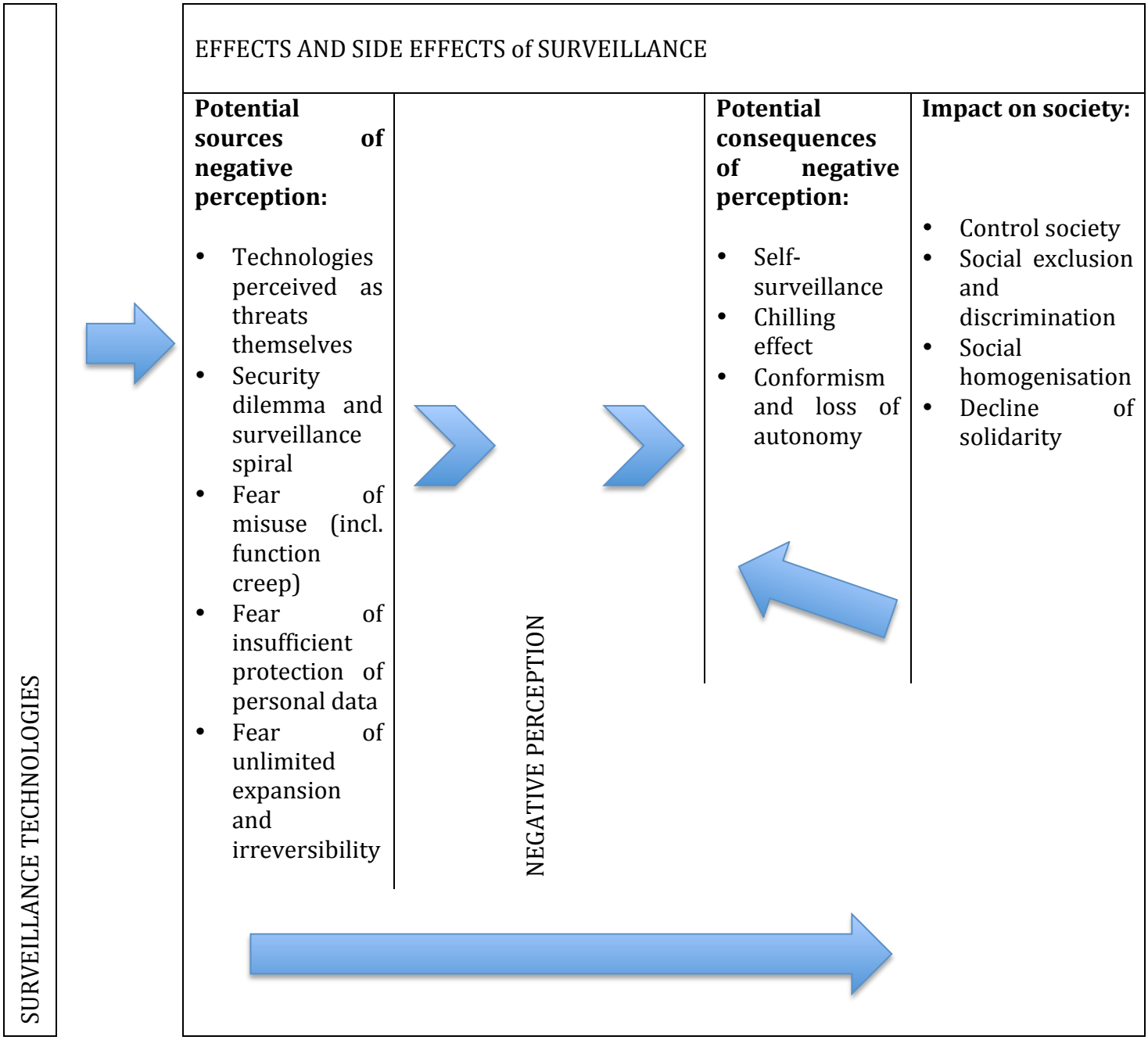
5. Function creep – a technology is designed for one purpose or context, but in fact is used for another (may imply infringement of privacy)

3.2 D3.2 – Report of review of European-level studies on perceptions of surveillance: negative perception, effects, side effects and perceived effectiveness

Deliverable 3.2 goes deeper into the literature on the perception of surveillance. More effects and side effects are found and they are categorized into sources of negative perception, possible consequences and their impact on society. However, based on a literature study, big differences between the occurrence of these effects and side effects are found depending on culture, goal of the use of the surveillance technology and other factors. A negative perception of surveillance is very context-dependent. Further, the relationship between perception and effectiveness is complex, with no cause-consequence link between them.

The 12 effects and side effects are summarized and categorized in Table 2:

Table 2: Effects and side effects of surveillance (D3.2)



The aspects discussed in this deliverable are part of the broadly understood societal 'cost' of surveillance technologies. As is concluded from this literature, what the actual cost is depends on many local factors. This means that, in order to rate this parameter, the knowledge of local experts has to be used.

3.3 D3.3 – Report on system effectiveness, efficiency and satisfaction assessment

This report assesses system effectiveness by evaluating user friendliness. It was found that the basic conditions to assess user friendliness cannot be met in this project. Therefore an alternative solution was found – develop a system to help users find procedures (via standards) to evaluate human factors, i.e. user friendliness. A decision support system, Application for Finding Usability Standards (AFUS), is developed in this report. This tool proposes an adapted set of human factors standards by having the user answer a manageable set of questions. Based on the user's responses a list of standards is given, with a rating of how useful they are for the given context. A critical question is how non-human factors' experts can handle a decision tree through responding appropriately to the questions in AFUS. This procedure has been initially tested by SURVEILLE end-users, but needs further testing.

3.4 D3.3b – Report on system effectiveness, efficiency and satisfaction assessment; Data protection

This report was added to complement the above discussed, originally planned D3.3. It develops a method to score Privacy Impact (Privacy Impact Assessment or PIA), measuring the usability parameters 'Observation of persons,' 'Collateral intrusion,' and 'Hardware and software protection.' Earlier assessment methods had been developed for specific implementations but not for technology concepts. Moreover, the assessment process was too time and resource demanding for a broad variety of technologies. A new rating scale is developed in this report with nine questions on data collection, data access and use, and data protection. The sum of scores gives a total score between 0-22. A high score means high attention to privacy. An advantage of this method is that the rating of a technology can be done in hours instead of months; a disadvantage is that the results obtained are less detailed. It nonetheless adds more detail to the usability factors in D2.6 for 'Privacy-by-design' and therefore is potentially an improvement to the scoring.

3.5 D3.4 – Design of a research methodology for assessing the effectiveness of selected surveillance systems in delivering improved security

D3.4 begins by defining "effectiveness" and "efficiency" as these terms pertain to surveillance technology. There are no clear definitions of these words as they are used in the security realm in relation to surveillance technology. Therefore they are defined for use in SURVEILLE as follows:

Effective surveillance technology has the technical capacity to deliver the intended security goals, and when employed for a defined goal within the necessary context (good location, trained operators, a larger security system, etc.), achieves the intended outcome.

Efficient surveillance technology delivers the intended security goals with low use of resources in terms of cost, time and/or physical and mental efforts.

There is also a lack in open-source literature on frameworks or methods for assessing the effectiveness and efficiency of surveillance technology. Several existing models from other domains were examined in this report to determine which could form a basis for examining surveillance technologies. The models examined were: the value analysis process by Roland; ISO 31.000: risk management; FEMA's point-scoring method for assessing terrorist threats to buildings; the CDC model for ex-ante evaluation of surveillance systems for infectious diseases; a RAND quantitative operations analysis for evaluating RPAs. An evaluation was done based on the MERPOL crime investigation scenario used in D2.6 to determine which method could be used in the SURVEILLE project. The CDC as well as the RAND model demand an enormous effort in both time and money and do not seem to be usable for the project. The ISO and Roland model on the other hand, seem to be quite general and qualitative and thus, do not give enough detail. The FEMA model is a point-scoring method, which seems to be very usable for the SURVEILLE project. Moreover, the method can be adjusted fairly easily for this purpose. The FEMA method was chosen as the best suited for further development in the SURVEILLE project.

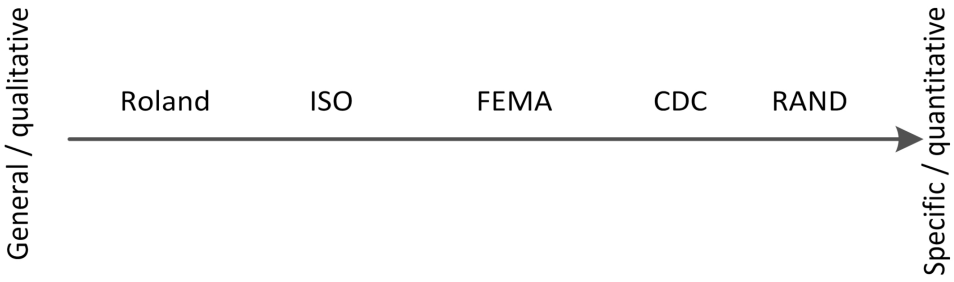


Figure 1: Scale of existing models for assessing effectiveness (D3.4)

3.6 D3.5 – Cost model

This report discusses the basic formulation of cost-benefit analysis for security devices. This is mostly a mathematical exercise where the basic formulas yield a relatively straightforward expression for cost-benefit analysis of security devices. However, as the factors in the model are more closely examined, this relatively simple model rapidly becomes mathematically challenging, involving questions of uncertainties (e.g. probability of attack, amount of risk reduced) and modeling direct and indirect costs (e.g. loss of lives after an attack, fear or social tensions in relation to terrorist activity, infringement of fundamental rights to privacy). In this regard, the report yields insights into the cost analysis of security devices; however, it also shows that complex mathematical modeling is required for any sensible cost estimate. It was found that this would require a substantial research effort, beyond the scope of the methods in the SURVEILLE project.

The deliverable nevertheless provides sensible inputs for the usability scoring developed in D2.6. It recommends that the Cost factor in the scoring be further developed such that the categories of 'Cost' – purchase cost, personnel cost, additional resource – are refined by more detailed elements in the calculations of costs. Further, that the attribute of 'Additional resource,' referring to whether external personnel are required for an operation, be changed to 'Additional running costs,' which comprises of a number of relevant parts that all need to be considered for scoring. These changes render a setting of thresholds for cost scoring that is more accurate.

3.7 D3.6 – Report on methodology and criteria for incorporating perception issues in the design phase of new surveillance systems

In this last deliverable a methodology is proposed to incorporate perception issues into the design phase of new surveillance technologies. The methodology corresponds to the two broad groups of negative perceptions and perceived effectiveness. As to negative perceptions, the methodology envisages three levels of intervention: Minimum Harm by Design (MHbD), Transparency by Design (TbD) and Accountability by Design (AbD). The first level aims to minimize the negative impact of technologies on individuals and societies, the second to make the way surveillance functions and its improvements transparent to the public and to the people affected by surveillance, and the third to make accountable the misuse of technologies and enable the sanctioning of those responsible.

The second part of the methodology, pertaining to perceived effectiveness, rests on the idea that interventions should address the background conditions affecting perceived effectiveness and not merely take measures to give the impression that security is being increased or something is being done. It envisions two levels of interventions:

- higher effectiveness
- Transparency by Design

The first level aims at improving effectiveness compatibly with legal, ethical and societal restraints. The second aims at making success rates and improvements in effectiveness transparent to the public and to people affected by surveillance. For both negative perceptions and perceived effectiveness, measures at the institutional, societal and legal levels are also required in order to make design interventions fruitful.

The MHbD proposed here is very similar to Privacy-by-design (PbD). The deliverable argues that the PbD approach is unsatisfactory because it is both too broad and too narrow. On the one hand, it targets the design phase of new surveillance technologies and the whole life cycle of complex surveillance systems, and on the other hand, it focuses only on information privacy, not addressing other fundamental rights.

As this deliverable states, the research concerning MHbD, TbD, and AbD is initial and pioneering. Further research is necessary, including technical research that examines how to design technologies according to the methodology proposed.

3.8 Conclusion

The review of the above deliverables yielded the first iteration of the usability scoring revision. This review was not a straightforward action. The principal reason being that deliverables 3.1 to 3.6 show that it is far from simple to derive a semi-quantitative point scoring system for the complicated discussions that are associated with surveillance technologies. A clear example is D3.5 on cost modeling: literally dozens of individual factors contribute to the overall cost for the use of any given surveillance technology and they are not easily reflected by the 3 point score in the usability scoring system. Some suggestions for modification of the usability score are mentioned in that report. These suggestions were assessed and discussed which led to two changes in the scoring system. We believe that this discussion greatly contributes to the deeper understanding of the problem but capturing a complex cost model in a few proximity indicators remains a difficult task.

Deliverable 3.2 on perceptions of technologies did not end up in affecting the usability scoring. This decision was reached after discussions with the project consortium as a whole. It was discussed that this aspect is not intrinsic to the technology but belongs in the ethical discourse, affecting the ethics scoring method and is therefore omitted from the usability scoring. This will mean that the notion of 'cost' must be addressed with caution: the technological usability score will still reflect mainly or exclusively financial costs, while broader issues of societal cost will need to be incorporated into the ethics assessments.

Deliverables 3.3 and 3.4 demonstrated that there is no such thing as a clear definition of the effectiveness and efficiency of surveillance technologies. Some ways in which to develop instruments for this were described in D3.4 but are not developed as part of this project. In the end, definitions were postulated and they have proven useful for the development of the usability score. Also, no clear answer could be found for the user-friendliness of a technology (D3.3); an intricate network of ISO standards defines what user-friendliness is in a broad variety of contexts. In the usability scoring it is reduced to the simplicity of a system.

Deliverable 3.3b is important for the usability score; it describes the complications of privacy-by-design rules. There was a considerable debate between project members whether this is a purely technical property. Though it is not purely a technical property it is included and assessed in the usability score because the technical components and software programming of the technical components determine whether a technology can be made intrinsically robust for privacy-by-design. That assumption excludes misuse by personnel working with the technologies and the data that it yields.

4. Scoring Revision – 1st iteration

Based on WP3 deliverables, the usability scoring was revised to incorporate knowledge and factors that influence the usability scoring. The most important change was that the scoring is more detailed to address the complicated discussions in deliverables 3.1 to 3.6.

4.1 A More Detailed Scoring

4.1.a Effectiveness

For the factor 'Effectiveness,' the attributes remained the same, with three sub-categories added under each attribute. The categories appeared as follows:

Effectiveness

1. Delivery
    - yields a useful outcome
    - conditions of deployment
    - transparency
2. Simplicity
    - user friendliness
    - structure
    - data extraction
3. Sensitivity
    - technical error
    - interpretation
    - accountability

These categories were added based on deliverables 3.3, 3.6, and our consideration of detailed elements that could be important in evaluating the effectiveness of surveillance technology. 'User friendliness' was addressed in D3.3 and was added here as an element contributing to the simplicity of a technology. 'Transparency' and 'accountability' were both discussed in D3.6 and were considered to contribute to the delivery and sensitivity of a technology, respectively. 'Conditions of deployment' or context, was judged by us to be an important consideration in judging the effectiveness of surveillance equipment. It reflects how difficult is it to ensure that the right conditions or usage of a piece of technology is possible – e.g. random placement of CCTV is always a problem, but covert placement of a single camera in a room is not. The factor of context also formed part of the definition of 'effectiveness' developed for SURVEILLE, and thus seems a necessary element to include.

Under 'Simplicity' we added the sub-category of 'Data extraction,' which refers to how much time or data has to be processed to get a positive result. In other words, how difficult is it to extract the right data? For example, there is one minute of relevant footage in 1000 hours of CCTV data. For 'Sensitivity' it was decided to divide "the likelihood of error" from the original definition in D2.6

into two categories – 'Technical error' and 'Interpretation,' referring to whether the data is open to misinterpretation.

4.1.b Cost

For 'Cost' some of the attribute names were changed to better reflect the sub-categories, but the basic element is the same. Again, three sub-categories were added to each attribute. These sub-categories reflect the detailed elements as found in D3.5 to contribute to and better reflect the costs involved in purchasing and deploying surveillance equipment. They are the following:

Cost

1. Initial cost
    - purchase price
    - installation cost
    - space requirement cost
2. Personnel requirements
    - number of personnel
    - training required
    - external partners
3. Additional running costs
    - maintenance and sustainability
    - false-positive rate
    - other (power, transport, insurance, office supplies)

4.1.c Privacy-by-design

The names of the Privacy-by-design attributes were also changed, but again what the category represents was not. The nine questions related to data collection, data access and use, and data protection, outlined in D3.3b were easily transferred into a total of nine sub-categories, as they appear below:

Privacy-by-design

1. Data collection
    - selective
    - minimized
    - overt or covert
2. Data access and use
    - who has access
    - clear regulations
    - protection against function creep
3. Data protection
    - encryption or otherwise access protected
    - protected against manipulation
    - secure against theft

4.1.d Proven technology

Lastly, we changed the name of the final element from 'Excellence' to 'Proven technology.' The definition of this parameter does not change – technology proving itself beyond a reasonable doubt.

4.2 From Nominal to Ordinal Scoring

The second significant revision proposed from the D2.6 scoring was to change from nominal to ordinal scoring. Nominal scoring responds to the question, "Does the technology have the given attribute – yes or no?" If it has the attribute it scores 1, if not it scores 0. Ordinal scoring responds to the same question with a range of scores. That is, the technology is rated as having the given attribute according to no/ low/ medium/ high scores. For our purposes, the corresponding scores given were 0, 1/3, 2/3, and 1. This allows us to stay within the total score of 1 for each attribute. Each sub-category scores either 0 or 1/3. This means each attribute with its three sub-categories scores 0, 1/3, 2/3, or 1. An example of the scoring for the 'Cost' factor of a technology could be the following:

Cost

1. Initial cost – 2/3 score out of possible 1 point
    - purchase price – 1/3
    - installation cost – 1/3
    - space requirement cost – 0
2. Personnel requirements – 1/3 score out of possible 1 point
    - number of personnel – 1/3
    - training required – 0
    - external partners – 0
3. Additional running costs – 2/3 score out of possible 1 point
    - maintenance and sustainability – 0
    - false-positive rate – 1/3
    - other (power, transport, insurance, office supplies) – 1/3

Total scoring for Cost = 1 and 2/3 points

4.3 Conclusions on 1st iteration of revised scoring

The proposed revisions were made to reflect more detailed elements of usability based on findings in WP3 deliverables. The result was a more complex, but also more comprehensive scoring that includes not only a more detailed accounting of the technical workings of the technology, but the context of deployment and some elements of public perception (transparency, accountability).

The complete revision of usability scoring appears as in Table 3.

Table 3: Usability scoring – first revision

Factor	Attribute	Sub-category	Sub-category score	Score
Effectiveness				0-3
	Delivery			0-1
		Yields a useful outcome	0-1/3	
		Conditions of deployment	0-1/3	
		Transparency	0-1/3	
	Simplicity			0-1
		User friendliness	0-1/3	
		Structure	0-1/3	
		Data extraction	0-1/3	
	Sensitivity			0-1
		Technical error	0-1/3	
		Interpretation	0-1/3	
		Accountability	0-1/3	
Cost				0-3
	Initial cost			0-1
		Purchase price	0-1/3	
		Installation cost	0-1/3	
		Space requirement cost	0-1/3	
	Personnel requirements			0-1
		Number of personnel	0-1/3	
		Training required	0-1/3	
		External partners	0-1/3	
	Additional running costs			0-1
		Maintenance & sustainability	0-1/3	
		False-positive rate	0-1/3	
		Other (power, transport, etc.)	0-1/3	
Privacy-by-design				0-3
	Data collection			0-1
		Selective	0-1/3	
		Minimized	0-1/3	
		Overt or covert	0-1/3	
	Data access & use			0-1
		Who has access	0-1/3	
		Clear regulations	0-1/3	
		Protection against function creep	0-1/3	
	Data protection			0-1
		Encryption or otherwise access protected	0-1/3	
		Protected against manipulation	0-1/3	
		Secure against theft	0-1/3	
Proven technology				0-1

5. Scoring Revision – 2nd iteration

5.1 Reason for a second revision

The changes in the first iteration were presented to the MERPOL End User Panel at their April 2014 meeting. Their feedback and the resulting, second iteration of revised scoring is discussed in this paragraph.

The EUP members expressed some points of concern related to the current approach of scoring as a whole. They consider the current approach to be potentially problematic in that it attempts to score "a number of highly complex issues that are unique to: a) the technology under consideration b) the circumstances of the deployment c) bespoke authorization d) decision making process." [1] EUP members believe that if these issues are not taken into consideration, the scoring process risks being flawed and open to misinterpretation.

More specifically related to the usability scoring, EUP members discussed the importance of measuring high yields. The benefits gained through using a piece of surveillance technology require more emphasis. For example, if deploying one specific piece of surveillance equipment removed the problem of a violent gang member, the benefits of that should be reflected and given emphasis in the scoring. Using this technology was more successful and cheaper than deploying a whole surveillance team. And the cost benefits extend into society as a whole – the local health services are saved the cost of treating his victims, and human lives are potentially being saved.

Another concern was that the scoring should take further account of the human involvement in deploying surveillance technology and the decision-making process surrounding it. That is, the decision to deploy surveillance technology can involve several persons, such as the investigating officer, technical support, the authorizing officer, and an operational security advisor. And when considering the deployment of surveillance technology in a given context, the authorizing officer takes many things into account, including the privacy impact, pressing social need, the minimization of collateral intrusion (this is separate from any minimization the technology does itself), product management, the deployment plan, and the appropriate decision-making model. In the case of the UK there is the National Decision Model that police officers should follow in making policing decisions. The third stage of this model – Powers and Policy – involves "considering what powers, policies and legislation might be applicable in this particular situation."[2]

[1] Minutes and Actions from EUP Meeting, April 23, 2014.
[2] "The National Decision Model," Association of Chief Police Officers, http://www.acpo.police.uk/documents/president/201201PBANDM.pdf, Accessed 23 April 2014.

It also became clear through discussion with the EUP that in policing considerations the first and foremost criteria are necessity, proportionality, and justification. Only when these are met are other factors, such as cost considered.

To address some of these concerns and in response to these discussions, additional changes were made to the scoring. Other issues raised, that are not dealt with in the usability scoring, such as necessity, proportionality, and justification, are discussed in section 5.3.

5.2 Changes in the second revision

Regarding the importance of context and measuring benefits: these factors were already treated under the Effectiveness category, however, they were only attributed 1/3 point each. To weigh these matters more heavily we returned to the original approach of three attributes under 'Effectiveness' with no sub-categories. Thus, each attribute receives 1 full point. 'Delivery' returns to referring solely to a technology yielding a useful outcome and having 1 full point. The second attribute of 'Sensitivity,' retains its original reference, referring to the likelihood of error, and also returns to 1 full point with no sub-categories. The attribute of 'Simplicity' was changed to 'Context.' For End User Panel members, the importance of the simplicity of a piece of surveillance technology appeared to be much less than that of the context of deployment. 'Context' refers to the conditions of deployment. Thus the EUP's concern regarding high yields is dealt with by the Delivery attribute. Their concern that the uniqueness of each and every deployment should be expressed in the scoring is addressed by the Context attribute. It was found that it was not possible to incorporate "human involvement" and "decision-making" into the current scoring system. We, therefore, address these matters separately in section 6 of this paper.

'Effectiveness' in the usability scoring now appears as follows:

1. Delivery – 1 point
    - whether a technology in a particular context, applied in a correct way, yields a useful outcome

2. Context – 1 point
    - relates to the conditions of deployment

3. Sensitivity – 1 point
    - relates to the likelihood of error; information is open to interpretation or vague data enables wrong conclusions

The attributes and sub-categories of 'Cost' and 'Privacy-by-design' as presented in chapter 5 remain the same.

Due to the sub-category changes with 'Effectiveness' and to keep the scoring uniform across all three categories, we revised the ordinal point division from 0, 1/3, 2/3, 1 to 0, 0.5, 1. Thus, each attribute of 'Effectiveness' – Delivery, Context, Sensitivity – scores either 0, 0.5, or 1. The sub-categories of 'Cost' and 'Privacy-

by-design' are scored as follows: if none of the sub-categories or only one sub-category is met, the technology scores zero for that attribute. If two of the sub-categories are met, it scores 0.5 and if all three are met it scores 1. Thus, if a technology scores positively for only one sub-category under each of the three attributes, it still scores 0. This would look like the following:

Cost

1. Initial cost
    - purchase price – yes
    - installation cost – no
    - space requirement cost – no
2. Personnel requirements
    - number of personnel – yes
    - training required – no
    - external partners – no
3. Additional running costs
    - maintenance and sustainability – no
    - false-positive rate – no
    - other (power, transport, insurance, office supplies) – yes

'Initial cost' scores positively for 'Purchase price,' 'Personnel requirements' for 'Number of personnel,' and 'Additional running costs' for 'Other.' However, the final score for 'Cost' would be 0. Meeting one sub-category for each attribute is still a weak score and not enough to merit the technology scoring a 0.5. This gives firstly and most importantly more weight to the attributes of 'Effectiveness,' which are the most crucial in deploying a piece of equipment. Secondly, this gives more importance to a technology scoring positively in more than one sub-category of an attribute. Thus, a technology scoring negatively in all sub-categories save two under 'Human resources' would score 0.5 for cost. This seems logical as surveillance equipment that is stronger in one attribute merits a higher score than technology that is weak in all three attributes.

After this second round of revision the usability scoring now appears as in Table 4.

5.3 Items for future discussion

The discussion with EUP members made clear that the central and primary matter in deploying surveillance technology in a criminal investigation is establishing necessity, proportionality, and justification. EUP members would like this and the decision-making process surrounding the deployment of equipment reflected in the usability scoring. Several factors, however, prevent us from doing so. Firstly, these matters are not quantifiable in the context of usability. Secondly, the usability scoring focuses on the technology and not on the laws and processes surrounding the use of the technology. This leads to the final point, which is that necessity, proportionality, and justification are addressed in the (legal) fundamental rights assessments in Work Package 2 (Survey of Surveillance Technologies) and Work Package 4 (Law and Ethics of Surveillance Technologies). It should be made clear that there are in fact two different levels of addressing necessity, proportionality, and justification. One is

on the abstract level of legislation and policy. The other is on the level of concrete application by law enforcement. When the EUP members speak of these matters in the context of deploying surveillance technology they are referring to the concrete application level. The fundamental rights assessment in SURVEILLE, however, is addressing both the concrete level and the abstract level. That is, the assessment is reviewing not only whether the practitioner is deploying the surveillance technology according to the (national) law, but rather, even where there was a legal basis in national law, was the fundamental rights interference in fact necessary and proportionate according to an external fundamental rights assessment, based on European and international standards. On the abstract level, these assessments address how the law itself measures up against a fundamental rights assessment.

The assessment and scoring of surveillance technologies in SURVEILLE is intended to inform decision-makers and influence and shape future law regarding surveillance technology. While who will ultimately use the surveillance technology scores and scoring system is as yet an open question, it is envisioned that the scores could potentially have a broad application. Their use could extend across multiple levels, from policy-makers at national and European levels to technology developers to law enforcement officials. This will be a matter for further discussion at a later date in the SURVEILLE project.

Another matter for future joint discussion is the consequences of combining the usability, ethics, and fundamental rights assessments. According to the scores a technology has received in these categories, what are the resulting requirements detailing the justifications for its use.

Table 4: Usability scoring – second revision

Factor	Attribute	Sub-category	Sub-category yes/ no	Score
Effectiveness				0-3
	Delivery			0-1
	Context			0-1
	Sensitivity			0-1
Cost				0-3
	Initial cost			0-1
		Purchase price	y/n	
		Installation cost	y/n	
		Space requirement cost	y/n	
	Personnel requirements			0-1
		Number of personnel	y/n	
		Training required	y/n	
		External partners	y/n	
	Additional running costs			0-1
		Maintenance & sustainability	y/n	
		False-positive rate	y/n	
		Other (power, transport, etc.)	y/n	
Privacy-by-design				0-3
	Data collection			0-1
		Selective	y/n	
		Minimized	y/n	
		Overt or covert	y/n	
	Data access & use			0-1
		Who has access	y/n	
		Clear regulations	y/n	
		Protection against function creep	y/n	
	Data protection			0-1
		Encryption or otherwise access protected	y/n	
		Protected against manipulation	y/n	
		Secure against theft	y/n	
Proven technology				0-1

Each attribute scores 0, 0.5, or 1. If only one sub-category scores 'y,' the attribute scores 0. If two sub-categories score 'y,' the attribute scores 0.5. And if all three sub-categories score 'y,' the attribute scores 1.

6. Conclusion

Following the initial development of a framework for usability scoring in D2.6, findings in WP3 deliverables and discussion with the EUP led to further refinement of this scoring. It was found that for the factors of 'Cost' and 'Privacy-by-design' more granularity was helpful. Breaking each attribute into three sub-categories provides for a more detailed analysis of surveillance technology in these areas. For the factor of 'Effectiveness' the weight of the attributes is more important than the detail. Certain attributes related to 'Effectiveness' – 'Delivery' and 'Context' – are crucial to deployment considerations and central to the kind of surveillance technology employed. These elements are considered first and foremost, with other matters, such as cost, being secondary. Therefore more weight is given to these central attributes by not assigning sub-categories to the 'Effectiveness' attributes.

To better handle the sub-categories that were added, the scoring was changed from a nominal to ordinal scale. Each attribute can now score 0, 0.5, or 1. 'Effectiveness', 'Cost', and 'Privacy-by-design' factors still score between 0-3 points, and 'Proven technology' (formerly 'Excellence') still scores between 0-1 points. The total possible usability score remains 10.

The EUP's concern that the scoring as a whole does not account for the required evaluation by law enforcement of the necessity, proportionality, and justification of deploying surveillance equipment in a given scenario, was found to lie in different understandings of how and where these three matters are considered in the combined SURVEILLE scoring matrix. They are evaluated in the fundamental rights assessment, but the evaluation is not only regarding whether or not law enforcement had a legal basis for the technology's deployment. Rather, it is an assessment from a fundamental rights point of view of whether the use of the technology and any subsequent interference of a fundamental right (such as privacy), was in fact necessary and proportionate. As such, it may require a critique of the national law itself, against European and international standards.