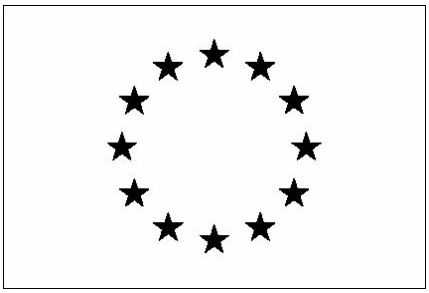

FP7 – SEC- 2011-284725

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 3.9
Final report of WP3

Due date of deliverable: 31.01.2015

Actual Submission date: 9.2.2015

SURVEILLE Work Package number and lead: WP3, Prof. Dr. Ir. P.H.A.J.M. van Gelder

(TU Delft)

Author(s): Michelle Cayford, Dr. Simone Sillem, Dr. Pei-Hui Lin, Dr. ir. Bert Kooij

(all 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

The purpose of this report is to provide criteria based on scientific literature for scoring surveillance technology according to the usability scoring developed in Work Package 3 (WP3). That is, according to what criteria or threshold does a practitioner decide if a technology scores a 1 or 0 for a given category within the 10-point scale developed in SURVEILLE. This usability scoring forms part of the scoring matrix developed in SURVEILLE and assesses the effectiveness of surveillance technology. The template for this assessment appears in Table 1 below.

This report serves to aid decision makers in determining what score to give surveillance technologies in their evaluations. Each category and sub-category of the usability scoring is covered, with discussion of what criteria is necessary for a piece of surveillance technology to be given a positive score. For some categories a threshold has been set. In other cases, it has proven to be difficult to prescribe a rule on how to decide on the scoring, and rather the opinion of experts should be employed to make a sound decision on the scoring. Where thresholds have been established they are meant to serve as a guide and can be altered over time and for different scenarios.

Certain concerns related to the scoring raised by the SURVEILLE law-enforcement End User Panel (EUP) are covered in the conclusion as well as the suggestion about continuous improvement of the methodology.

Table 1: Template for assessing the usability of surveillance technology

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 2.6 of SURVEILLE a matrix of surveillance technologies was developed, which scored technologies according to the categories of usability, ethics, and fundamental (or human) rights. In Deliverable 3.8 a summary of the research done in Work Package 3 (WP3) – Perceptions and Effectiveness of Surveillance – was given, and the usability scoring of the matrix was discussed, reporting on the findings of deliverables in WP3 and on input from the MERPOL End User Panel of law-enforcement officials (EUP). The subsequent changes and developments for the scoring were described, with an updated usability scoring system proposed.

This current report focuses on the usability scoring system and establishing criteria and thresholds for the different categories in the scoring. As such, it is not meant to be a summary of all the work done in WP3. Significant work on perceptions of surveillance technology was carried out in WP3 and this work is summarized in D3.8. It was ultimately decided not to include evaluation of perceptions in the usability scoring, as this would be better contained under the ethics portion of the scoring matrix.

The usability scoring system appears in Table 1 on pg. 3 of this report. This table can be used as a template to assess the usability of surveillance technology. The categories and sub-categories could be slightly modified or adjusted according to the user's needs. More importantly, the thresholds that determine which score the technology receives will vary widely depending on the technology and the situation in which it is used. This will be seen in the discussions in this report.

This report will evaluate all factors and attributes given in the usability scoring table as found in D3.8. Criteria are given for the choices between the different scoring possibilities. For some questions this choice is obvious given the nature of the factor, in other cases a threshold has to be set based on substantiation from the literature.

For example: Cost/initial cost – a threshold should be set to decide when a certain technology is "expensive" for the "purchase price cost." That is to say, it is expensive to buy. So, when is something expensive? 10,000 euros? 100,000 euros? There is a need to find a solid basis to set such a threshold. So the task is to somehow figure out a sensible rule for this threshold: a fixed number, an equation or perhaps something else. It is important that this rule has some basis in the literature or is based on some kind of investigation.

Section 2 of this deliverable reviews the factor, Effectiveness, Section 3 will look at Cost, Section 4 will discuss Privacy-by-design and finally, in Section 5, criteria for determining whether a technology is a Proven Technology will be given.

2 Effectiveness

The research undertaken for SURVEILLE revealed that a significant body of knowledge regarding the effectiveness of surveillance technology does not exist. While surveillance technology has developed rapidly over the past decade and its application continues to swiftly spread, there has not been a parallel development of evaluating the effectiveness of this technology. This is even the case with CCTV, which is perhaps the best-known technology when it comes to surveillance. CCTV use and experience is developed and widespread, but systematic evaluations of its effectiveness are still lacking. The various studies that have been carried out show a wide range of varying results regarding the success of deploying CCTV.[1]

Given the lack of study on effectiveness, it comes as no surprise that no definition of "effectiveness" even exists in the surveillance context. In deliverable 3.4 a definition was developed for the SURVEILLE project. This definition is 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.[2]

Due to the lack of literature and scientific study on the effectiveness of surveillance technology, the criteria for evaluating effectiveness were chosen based on expert opinion in discussion with the SURVEILLE End User Panel, and based on an evaluation of existing literature on CCTV. Since CCTV is the surveillance technology with the largest body of literature surrounding it, this literature was examined to determine what kinds of criteria were used in these studies. Academic consideration by engineers was also given as to what aspects of the technology would affect its effectiveness as defined in SURVEILLE deliverable 3.4. The three criteria included under Effectiveness are Delivery, Context, and Sensitivity.

2.1 Delivery

In the usability scoring structure of SURVEILLE "Delivery" refers to whether or not the equipment yields a useful outcome when used correctly. When there is evidence of prior successes or success is reasonably achievable this attribute scores 1; when there is evidence of some success it scores 0.5; otherwise it scores 0.

[1] See, for example, Ditton and Short 1999, Gill et al. 2006, Phillips 1999, Armitage et al. 1999, Farrington et al. 2007.

[2] Note for the non-native English speaker: the words "effective" and "efficient" are different words with different meanings. They should not be used interchangeably. For this reason a definition of "efficient" was also developed for the SURVEILLE project in D3.4. As the scoring, however, focuses on effectiveness we have included only that definition here.

Determining if a technology yields a useful outcome is a logical and necessary measure of effectiveness evaluation. To be effective a technology must have success in obtaining the outcome for which it is deployed. Although this is not the sole measure for evaluating a technology or program, "[m]easurement is a fundamental aspect of effectiveness assessment."[3] A study evaluating the effectiveness of counter-terrorism strategies used measurable outcomes to determine success. These included outcomes such as the number of terrorist incidents, the decline in the number of terror groups, reduction of the public's fear of terrorism and increased ability to respond to events.[4]

In the area of CCTV much of the controversy surrounding its effectiveness hinges on this point of success. Numerous studies seek to establish if it accomplishes its goal. With regard to achieving the desired outcome of a reduction of crime, some studies show that CCTV has no effect (Ditton and Short 1999; Gill et al. 2006; Phillips 1999), while others find that it has a small to modest effect (Armitage et al. 1999; Farrington et al. 2007).[5] The point here is not to judge whether or not CCTV is effective, but to demonstrate the importance of this factor of delivery in evaluating effectiveness. It is the first measurement that comes to the professional or layperson's mind – does the given technology or program achieve the intended outcome?

As this is an intuitive and rather evident factor of measurement, it does not need further elaboration here.

2.2 Context

"Context" relates to the conditions of employment. Is the surveillance technology being used in the context for which it was designed and in which it performs well? Or are the conditions such that it cannot perform optimally – i.e. weather inhibits its performance; it is being deployed in a context for which it was not originally intended or which challenges its functionality (e.g. a sound recording bug on a public transport bus is a poor context as it is designed for recording conversation among a few people, not multiple conversations at once with significant background noise).

Context was strongly emphasized by members of the EUP as being important in evaluating surveillance technology. In instances of whether and what kind of surveillance equipment must be chosen at a certain point in an investigation, the deploying officer considers the context and determines whether a given piece of surveillance technology should be used. Many things are taken into account in

[3] Van Um, E. and D. Pisoiu, "Effective Counterterrorism: What Have We Learned so Far?" Economics of Security Working Paper 55, Berlin: Economics of Security, 2011.

[4] Lum, Cynthia, Leslie W. Kennedy, and Alison Sherley, "Are Counter-Terrorism Strategies Effective? The Results of the Campbell Systematic Review on Counter-Terrorism Evaluation Research," *Journal of Experimental Criminology* 2, no. 4 (January 10, 2007): 489–516.

[5] Caplan, Joel M., Leslie W. Kennedy, Gohar Petrossian. "Police-monitored CCTV cameras in Newark, NJ: A quasi-experimental test of crime deterrence." *Journal of Experimental Criminology* 7, no. 3 (September 2011): 255–74, p. 256.

making this decision including privacy impact, pressing social need, 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.

The importance of evaluating context can be seen particularly in the case of CCTV. CCTV has been used in a multitude of scenarios, making it the best example to use in considering "context" as a measurement of effectiveness. It is also arguably one of the most controversial technologies when it comes to measuring success. There have been numerous studies done by various parties yielding results from opposite ends of the spectrum. One article examining studies on CCTV effectiveness, showed that out of 11 studies, five found that CCTV had a negative impact on crime, five found no effect, and one found an uncertain effect.[6] This can be attributed, at least in part, to a lack of or vague statement of purpose, as well as to the context in which it is deployed. In certain contexts CCTV performs better than in others.

According to criminology course material at the University of Leicester, which draws on a number of CCTV studies, the contradictory results of CCTV studies are due to the varying circumstances in which the technology is used.[7] A study done by Tilley in 1993 on the use of CCTV in car parks establishes context as one of the necessary criteria for evaluating CCTV effectiveness. "Context" is described as the necessary conditions for a particular crime-prevention method to "fire its (potential) casual mechanism/s to produce an outcome-pattern."[8] The way in which CCTV is implemented will affect whether or not the desired outcome is attained.

Gill and Springs in their study on the effectiveness of CCTV identify nine contexts in which CCTV works best: 1) "small, enclosed areas" 2) greater camera density in a given area 3) against material crimes, particularly vehicle theft 4) for special initiatives 5) active police interest and response 6) in conjunction with other crime-reduction measures, 7) good lighting 8) adequate monitoring of cameras 9) permanent systems (in the context of measuring long-term crime reduction).[9]

As a whole, the body of research on CCTV has begun to identify contexts in which CCTV is most effective. Some of the necessary contexts are becoming apparent: clear lines of sight such as long, straight roads; simple target areas, particularly those with controlled entrance and exit; a crime prevention strategy with CCTV as a part of this strategy (as opposed to CCTV used in isolation). A study as old as 1992 found that "CCTV does not seem to be very useful in large complex and

[6] University of Leicester, Criminology, Technologies of Control, 8.4 Surveillance and CCTV, http://www.le.ac.uk/oerresources/criminology/msc/unit8/page_10.htm.
[7] Ibid.
[8] Tilley, Nick. *Understanding Car Parks, Crime and CCTV: Evaluation Lessons from Safer Cities*. Crime Prevention Unit Series Paper No. 42. Home Office Police Department, London, 1993, p.3.
[9] Gill, Martin and Angela Spriggs. *Assessing the Impact of CCTV*. Home Office Research Study No. 292. Home Office Development and Statistics Directorate, London, 2005, pp.117-119.

crowded environments to deal with surreptitious behaviour such as pick pocketing or shoplifting."[10]

As can be seen from the case of CCTV, establishing the exact context for each kind of surveillance technology takes time and a great deal of study. CCTV in the UK has been studied for over two decades. Although it may be one of the more problematic kinds of technology to evaluate, CCTV nonetheless demonstrates that establishing precisely in which context a technology is most effective can be a lengthy process and require rigorous study. CCTV has attracted a lot of attention and consequently numerous studies. Other technologies evaluated in SURVEILLE (such as those covered in D2.1 or in the scenarios in D2.6, D2.8, and D2.9) have not attracted the same level of attention, and subsequently similar studies evaluating effectiveness are scarce to non-existent. The evaluations of the "context" category of surveillance technologies in SURVEILLE deliverables have been based on what the technology is designed for and a subsequent logical conclusion as to the contexts in which it performs best.

2.3 Sensitivity

"Sensitivity" relates to the likelihood of error – information is open to interpretation or vague data enables wrong conclusions.

Several studies on CCTV demonstrate the importance of this factor. Some surveillance technologies' collection of data, such as CCTV, is more subject to human action and interpretation than others. A "fundamental part of the CCTV system is a reliable image evaluation by a human observer, whose effectiveness (i.e. incident detection rate) is influenced by many variables."[11] In a study investigating factors influencing CCTV operators three variables were investigated: frame rate (or frames per second), brightness, and the number of monitors the operator had to observe. The results showed that with lower frame rate and decreased brightness the detection rate fell. The decreased detection rate for brightness, however, was minimal, while that for lower frame rate was more significant. And as would be expected, detection rate fell when operators were monitoring four screens as opposed to one or two.[12]

As another study by Näsholm et al. states, this type of error – misinterpretation or vague data – is often not focused on.[13] The study speaks specifically with regard to CCTV, but it can be applied to other technologies, as considerations of error typically focus on error related to the technology itself, such as image

[10] Webb, Barry and Gloria Laycock. *Reducing Crime on the London Underground: An Evaluation of Three Pilot Projects*. Crime Prevention Unit Paper No. 30. Home Office, London, 1992, p.23.
[11] Van Voorthuijsen, G., H. van Hoof, M. Kilma, K. Roubik, M. Bernas, P. Pata. "CCTV Effectiveness Study," 39th Annual IEEE International Carnahan Conference on Security Technology, 2005, p. 105.
[12] Ibid.
[13] Näsholm, Erika, Sarah Rohlfing, James D. Sauer. "Pirate Stealth or Inattentional Blindness? The Effects of Target Relevance and Sustained Attention on Security Monitoring for Experienced and Naïve Operators," *PLoS One*, 9:1, Jan. 21, 2014.

quality. This can also result in error as the unclear image can lead to wrong conclusions. The error, however, is related to the human conclusion or interpretation of the data collected, not to the technology itself making an error. Näsholm et al.'s study related to human factors in CCTV systems, focused on operators' inability to detect unexpected stimuli. It found that 66% of 171 operators failed to detect the unexpected stimuli. (Based on the criteria of the study, this figure includes those who detected the stimuli but failed to correctly identify it.) The point here is not so much the study's findings but the indication that evaluation of this kind of factor of sensibility is necessary in measuring the effectiveness of surveillance technology. "[T]here has been little academic interest in CCTV monitoring… A scientific approach to security surveillance research is required, and such an approach must consider human factors in addition to technical capacity. Knowledge gained from such research may influence the design of monitoring systems, and the training given to operators."[14]

Research also shows that even in optimal conditions, correctly matching and identifying an unfamiliar face from CCTV footage is a task prone to error. (The success rate for correctly identifying known faces is much higher.)[15]

This factor of sensibility combines the technical elements such as image quality which could lead to poor recognition or misinterpretation, as well as human factors, such as the volume of data being delivered being too much for an operator to process and act upon. The example of CCTV demonstrates the importance of including this factor in an evaluation of surveillance technology effectiveness.

3 Cost

The research undertaken for SURVEILLE revealed that surveillance costs are very difficult to estimate in advance. To study the actual costs of a surveillance technology one must consider not only its initial cost, which covers the costs of materials and technological devices as well as the necessary infrastructure, but also the cost of personnel, maintenance of the technology, and other costs such as power and insurance.

The literature review performed for D3.5 revealed that there is no general guideline for the cost modelling of security devices. The review produced three topics that contributed to the cost study of D3.5. One of those topics – selecting baggage-screening technologies – will be used in this current report to establish cost thresholds. Researchers like Virta et al. (2003) have developed a cost model that captures the cost of deploying, maintaining, and operating a single baggage-

[14] Ibid, p.7.
[15] Hillstrom, Anne P., Lorraine Hope, Claire Nee. "Applying psychological science to the CCTV review process: A review of cognitive and ergonomic literature." Home Office Scientific Development Branch, March 21, 2008.

screening device over a one-year period.[16] Later Jacobson et al. (2006) enhanced the cost model to take into account the cost and benefit of deploying multiple devices at the same time.[17] In 2007 Feng determined the optimal specifications for continuous security responses provided by various screening technologies in terms of risk and cost-effectiveness assessment.[18]

Over the years an enormous amount of money has been spent on surveillance technologies. Butler and Poole show that the implementation of the 100 per cent reliable baggage-screening mandate resulted in investments of many billions of dollars for additional screening devices.[19] To be cost-effective, the benefits of a security measure must outweigh the costs of the measure. The following paragraphs will discuss how thresholds can be set, which indicate the point at which a security device becomes expensive.

The usability scoring in SURVEILLE divides Cost into three categories – Initial cost, Personnel requirements, and Additional running costs. Each of these three categories is further subdivided into three sub-categories. The sub-categories are scored either "yes" or "no," with "yes" being a positive score that the technology in question meets the threshold cost requirement. If none or only one of the sub-categories scores yes then the category scores 0. If 2 of the sub-categories score yes then the category scores 0.5. If all 3 of the sub-categories score yes then the category scores 1.

3.1 Initial cost

For the estimation of the initial costs of a security technology, the purchase price, the installation cost and the space requirement costs are required.

Although our scoring system separates the initial cost into three sub-categories, in the literature these two costs are combined. In the cost models developed by Virta et al, Jacobson et al. and Feng, the purchase price of the technological device as well as the installation costs have been combined in one cost parameter. Virta et al. consider the purchase price and installation cost together, while the other authors only mention purchase price. Therefore, in this report these three parameters are examined together.

In the cost model developed by Virta et al. the purchase price and installation cost of a baggage screening security device is estimated to be $1,000,000.[20] Some

[16] Virta, J.L., Jacobson, S.H., Kobza, J.E., "Analysing the cost of screening selectee and non-selectee baggage," *Risk Analysis*, Vol. 23, No. 5, 2003.
[17] Jacobson, S.H., Karnani, T., Kobza, J.E. and Ritchie, L., "A cost-benefit analysis of alternative device configurations for aviation-checked baggage security screening," *Risk Analysis*, Vol. 26, Issue 2, 2006, pp. 297–310.
[18] Feng, Q., "On determining specifications and selections of alternative technologies for airport checked-baggage security screening," *Risk Analysis*, Volume 27, Issue 5, 2007, pp. 1299–1310.
[19] Bulter, V. and Poole, R.W., "Rethinking checked-baggage screening," Policy Study 297. Reason Public Policy Institute, 2002.
[20] Virta et al., pp. 898, 901.

of the parameters in their cost model deal with values that are security sensitive, and therefore, not available to the public. Thus, not all the values are real-world data. This is always a difficulty when dealing with security devices such as surveillance equipment. Using data that is not confirmed to be real figures as a basis for our threshold is therefore a realistic exercise. These thresholds, of course, serve as guidelines and can be adjusted according to the situation and needs of the user.

Feng only considers purchase price and not installation cost in his cost model of four different types of baggage-screening technologies. These figures appear in Table 1 under the parameter C_D. Jacobson et al. create a cost model examining three kinds of explosive trace detection machines, explosive detection systems, backscatter, and dual-energy x-ray machines, which includes the purchase price of each device (see C_F in Table 2). If we average the purchase costs presented in the cost models of these three reports we arrive at a purchase price just over $500,000. Therefore we can set $500,000 as the threshold figure. Initial costs that are at or below this amount score 1; initial costs from $500,000 to $750,000 score 0.5, and those above $750,000 score 0.

Table 2: Cost Parameters for Four Baggage-Screening Devices
Source: Feng, 2007.

Device Parameters	EDS	BX	DX	MVT
C_D	$800,000	$333,333	$500,000	$1,000,000
C_M	$125,000	$41,667	$62,500	$80,000
C_I	$0.19	$0.09	$0.03	$0.12
T_1	10 years	10 years	10 years	10 years
T_2	10 years	10 years	10 years	10 years
#bags/hour	180	250	1500	1500
N_C	394200	547500	3285000	3285000
K	7	5	1	1

Table 3: Showing the Purchase Cost of Six Baggage-Screening Devices
Source: Jacobson et al., 2006.

Parameter	Expected Value (EDS)	Expected Value (ETD-TC)	Expected Value (ETD-TO)	Expected Value (ETD-AG)	Expected Value (XRAY)	Expected Value (BACK)
P_{FA}	0.125	0.1	0.05	0.02	0.2	0.05
P_{FC}	0.05	0.2	0.15	0.1	0.05	0.05
C_F	$1,000,000	$45,000	$45,000	$45,000	$500,000	$333,333
C_o	$125,000	$14,000	$14,000	$14,000	$62,500	$41,667
C_I	$0.19	$0.30	$0.83	$1.29	$0.03	$0.09
N_1	10 years	5 years	5 years	5 years	10 years	10 years
N_2	10 years	5 years	5 years	5 years	10 years	10 years
S_C	10,000,000	10,000,000	10,000,000	10,000,000	10,000,000	10,000,000
S_{CAP}	270,000	164,160	60,480	38,880	1,728,000	540,000
Bags/hour	125	60	28	12	800	250

3.2 Personnel requirements

In the literature discussed above, explicit personnel cost is not dealt with in the cost models. Instead it is included in operational or maintenance costs. Virta et

al. include the cost of operating the baggage screening system per checked bag inspected. This cost is $0.525.[21] Feng also estimates personnel cost based on the operation cost of the screening cost per bag.[22] This is a difficult figure to work with since it includes costs other than personnel cost and it is per inspected bag. Jacobson et al. include personnel cost in the purchase price.[23] In a study by Butler and Poole, the annual cost of one Transportation Security Administration (TSA) operator is put at $45,000.[24] Because of the difficulty of working with a price per bag figure, the threshold of personnel costs will be set at $45,000 (per employee/ per year). Costs of $45,000 and below score 1, from $45,000 to $75,000 scores 0.5 and above $75,000 scores 0. The fact that this is a very initial estimated threshold and is the cost of only one employee should be taken into account, and the threshold figure consequently modified when applied to a specific case.

3.3 Additional running costs

The additional running costs category includes maintenance and sustainability, false-positive rate, and other costs. These values are highly user dependent and device dependent. For example, some devices may have a very high maintenance cost and other devices may have low maintenance cost. Therefore, it is difficult to provide universal criteria for all the devices. Here the values are mainly from the literature and only recommended to serve as a starting point for users to consider their own requirements, but it is not intended to provide universal criteria for all the users/devices.

The information on the criteria for scoring on additional running costs is quite limited. In the following paragraphs, each aspect of the additional running costs is discussed.

3.3.1 Maintenance & sustainability (e.g. renewing license, rental cost, and necessary manpower)

Maintenance costs include scheduled maintenance cost and unscheduled maintenance cost. Scheduled maintenance is a preventive form of maintenance conducted at pre-set intervals to inspect and ensure that the device works properly. Unscheduled maintenance is needed after any failure event or the discovery of any unexpected technical wear or anomaly. The estimated amount of maintenance cost per device per year can be very different. Legg & Powell suggest that maintenance cost should include expenditures for services provided to maintain device software and hardware. This cost can incur as a result of the work performed by internal staff or by contractor or vendor.[25] In our project, we

[21] Ibid.
[22] Feng, p.1302.
[23] Jacobson et al., p.299.
[24] Bulter and Poole, p.6.
[25] Legg, D. and Powell, J. "Measuring inputs: guidance for evaluators, Research, Development and Statistics Directorate," Home Office, UK, 2000.

also need to estimate the amount of renewal of software license to pay to software vendors per year per unit.

M. Stewart and J. Mueller[26] focus their paper on Advanced Imaging Technologies (AIT). "The TSA [Transportation Security Administration] will use AITs as a primary screening measure, and to this end plans to procure and deploy 1,800 AITs by 2014 to reach full operating capacity. The costs are considerable. The DHS FY2011 budget request for 500 new AITs includes $214.7 million for their purchase and installation ($430,000 each), $218.9 million for 5,355 new Transportation Security Officers (TSOs) and screen managers to operate the AITs at the checkpoints, and $95.7 million for 255 positions to fund the support and airport management costs associated with the 5,355 new TSOs and screener managers (DHS 2010b). In addition, this equipment will require maintenance, support and upgrading…." [27] This paper states that the annualised cost of purchasing, installing, staffing, operating, supporting, upgrading, and maintaining the first 1,000 units is about $650 million per year. However, the authors do not separate the maintenance cost from the total cost.

Virta et al. analysed the cost of screening selectee and non-selectee baggage in airports. They identify the expected annual maintenance costs (operational) for a baggage screening security device, including annual lease expenses, to be around $125,000 US dollars; this cost does not include purchase price and installation cost and is independent of the volume of checked bags screened.[28]

Feng discusses the alternative technologies for checked baggage security screening. The maintenance cost for four devices for checked baggage security screening in airports is estimated (see C_M in Table 3).

Table 4: Showing the Estimated Maintenance Cost for 4 Baggage Screening Devices
Source: Feng, 2007.

Device Parameters	EDS	BX	DX	MVT
C_D	$800,000	$333,333	$500,000	$1,000,000
C_M	$125,000	$41,667	$62,500	$80,000
C_I	$0.19	$0.09	$0.03	$0.12
T_1	10 years	10 years	10 years	10 years
T_2	10 years	10 years	10 years	10 years
#bags/hour	180	250	1500	1500
N_C	394200	547500	3285000	3285000
K	7	5	1	1

There is no general information on the maintenance cost for surveillance devices. The only information the authors can find is from Virta et al. and Feng. If the

[26] Stewart, M.G. and Mueller, J. "Cost-benefit analyisis of advanced imaging technology full body scanners for airline passenger security screening," *Journal of Homeland Security and Emergency Management*, 8:1, 2011.
[27] Ibid, p.3-4.
[28] Virta et al.

expected values of the maintenance cost are averaged, the threshold can be set to $86,000. If the maintenance cost per year is below this amount it would be considered a positive score and the sub-category of "maintenance cost" would thus score "yes." If it is at or above this amount it would be a negative score of "no."

3.3.2 False-positive rate

A false positive is when a system identifies a hazard when it is not. This leads to additional steps, which slow operation down, cause inconvenience to customers, cost money and reduce people's confidence in the security system.

Virta et al. analysed the cost of screening selectee and non-selectee baggage in airports. The authors identify that the expected values of false-positive rate is 0.125, and the cost of falsely indicating threats is $50 US dollars.

Feng identifies the expected cost of a false alarm or of wrongly indicating a threat to be around $9.16 US dollars in alternative technologies for checked baggage security screening. The author also gives the false alarm probabilities for different devices (see "α" in Table 2).

Table 5: False Alarm Probabilities for Four Devices
Source: Feng, 2007.

	EDS	BX	DX	MVT
α^*	0.0151	0.0790	0.2186	0.0305
β^*	0.0038	0.0100	0.0100	0.0074
u^*	0.4507	0.3981	0.3162	0.4410
EC_1^*	$0.9892	$1.2574	$2.3762	$0.6897

In Butler and Poole's table below, the false-positive rate of an EDS machine is 30%, which is quite different from the number stated by Feng.[29]

Table 6: Showing False-Positive Rate
Source: Butler and Poole, 2002.

[29] Butler and Poole, p.4.

Table 1: Comparison of Baggage Inspection Alternatives					
Type	Bags /hour*	False Positive Rate	False Negative Rate	Initial Cost/ Unit	Unit Operating Cost/Year
Hand Search	12-30	n.a.	n.a.	$0	$45K
Dogs	400	n.a.	n.a.	$20K	$50K
Trace (Closed)	76	n.a.	30-50%	$450K	$90K
Trace (Open)	24-30	n.a.	15%	$45K	$90K
Trace (Non-directed)	15-20	n.a.	15%	$45K	$90K
Automated X-ray**	1,200-1,500	n.a.	n.a.	$250-400K	$90K
EDS Machine	150-200	30%	n.a.	$1,000K	$510K

*not including time to "clear" false positives

**not certified for use in the United States, though approved in Europe

n.a. = no generally accepted figure available.

As mentioned above, there is no general information on the false-positive rate for general surveillance devices. The only information available is from airport security screening. If we average the expected values of the different devices mentioned in the literature, the criteria can be set at 0.13. So if the false-positive rate is below 13%, the technology scores "yes." If at or above it scores "no."

3.3.3 Other Running costs (power, office furniture, transport, insurance, stationery items, etc.)

Butler and Poole discuss the different methods to screen checked baggage. They list the unit operating cost per year for different inspection alternatives. However, the unit operating cost includes staff cost and maintenance cost.

Virta et al. identify the expected cost of operating a baggage screening security device, per checked bag inspected, as equal to $0.525 US dollars. Currently we can set running cost threshold per scan to $0.525. If later more information is available on average running cost, this value can be changed. If the running cost per scan is $0.525 or less the technology scores "yes." If the cost is above this figure it scores "no."

3.4 Future Considerations for Cost

Members of the EUP suggested that this scoring template should include "savings obtained as a result of the use of the technology" in the Cost category.[30] That is, through the proper and lawful deployment of a particular piece of surveillance technology, the length of an investigation could potentially be significantly shortened, contributing to savings in hours and monetary cost. Savings could also be extended to opportunity costs, meaning the prevention of further criminal activity by the shutting down of a criminal organization or the like. These savings would be achieved not only by law enforcement agencies, but also by health services, local authorities, etc. In other words, by not using the given

[30] Minutes, SURVEILLE End User Panel Meeting, Mannheim, Germany, 24 October 2013.

technology in the investigation, costs would be incurred. EUP members felt that this savings figure was "relatively easy to quantify when considering a specific offence/ investigation."[31]

A way was not found to incorporate this "savings cost" into the current usability scoring template. This template, however, is meant to serve as a guide and to be adjusted according to particular users needs and to individual uses of surveillance technology. Practitioners are encouraged to find a way to incorporate savings into the cost scoring/ considerations. This is the first production of a surveillance technology usability scoring, and every scoring system should be improved upon.

4 Privacy-by-design

Like the Cost category, Privacy-by-design (PbD) is composed of three attributes, each of which is further divided into three sub-categories. The sub-categories are scored like those in the Cost category (see Section 3). The three attributes of PbD are Data collection, Data access and use, and Data protection.

4.1 Data collection

Data collection is divided into three sub-categories: is the data collection 1) selective 2) minimized 3) overt or covert.

The most important aspect is the selectivity of the surveillance measure.[32] In an ideal case only the subject or subjects should be under surveillance. Data minimization concerns only the collected data of the subjects in the focus of the video surveillance not the selectivity. While it is not always feasible to collect data in a transparent way – i.e. an overt covert observation is by definition not possible – transparency of surveillance measures has a big impact on data protection.

The following paragraphs will give an elaboration of the way in which the scores for the three sub-categories can be given and what issues are taken into account when deciding upon this.

4.1.1 Selective

The scoring for this category is fairly straightforward. If only the person of interest is being recorded a "yes" is scored. If persons other than the target are being recorded a "no" is scored. Unfortunately this method of scoring does not allow for some selectivity, such as if a couple non-targets are also recorded versus every possible person being recorded. A way to compensate for this could

[31] Ibid.

[32] In SURVEILLE deliverable 3.3b the importance of this aspect was conveyed by a corresponding weight, but it was ultimately judged to be to cumbersome to incorporate weights into the usability scoring template.

be sought in the future. The amount to which a technology has to be selective can be up for discussion.

4.1.2 Minimized

The objective of data minimisation as is given by the Commission Nationale de l'Informatique et des Libertés (CNIL) is "to reduce the severity of risks by limiting the amount of personal data to what is strictly necessary to achieve a defined purpose."[33] This is very similar to the definition that is used by the European Data Protection Supervisor (EDPS), which states that: "a data controller should limit the collection of personal information to what is directly relevant and necessary to accomplish a specified purpose."[34] This definition will be used throughout this deliverable. When only necessary data is collected, the score of "yes" is given. When also other, personal or irrelevant information is collected this yields a score of "no."

According to the CNIL, good practices that can ensure minimisation of data collection are:

- Ensuring that the personal data that is collected is sufficient, relevant and not excessive with regard to the purpose for which it is gathered.
- Ensuring that the personal data does not reveal any information having to do with race, political or religious views, information about health, etc.
- Ensuring that the data are not related to criminal convictions or security measures, although in the case of this research project that may prove to be difficult.
- Ensuring that no (or as little as possible) other personal data are collected.
- Ensuring that the transmission of electronic documents containing personal data is restricted to the individuals who need them in connection with their work.
- Ensuring that personal data that are no longer useful are deleted.

4.1.3 Overt or covert

This issue is specific to surveillance technologies and therefore differs from issues only related to privacy. By definition, some surveillance technologies cannot be used overtly and thus have a privacy-infringing nature. This does not mean however, that the infringement of privacy should not be limited.

According to the CNIL, good practices when selecting measures to treat this risk are shown below. Many of these may not be applicable in the case of a covert surveillance technology:

- Ensuring that the processing is not covered by an exception and is not subject to the specific conditions set forth.

[33] Commission Nationale de l'Informatique et des Libertés. Measures for the Privacy Risk Treatment, Translation of edition of June 2012.

[34] European Data Protection Supervisor. Data Protection Glossary, https://secure.edps.europa.eu/EDPSWEB/edps/EDPS/Dataprotection/Glossary/pid/74.

- Ensuring that the means that will be used to inform the subjects for which data will be collected are known.
- Ensuring that the notification to the target audience is complete, clear and appropriate, explaining the nature of the personal data and the practical means chosen to collect it.
- Ensuring that a notification to the targeted audience is provided before the data is actually collected.
- Ensuring that the data cannot be collected without providing this information.
- Whenever possible, provide a notification when the target audience has been notified.

4.2 Data access and use

The sub-categories of Data access and use are who has access, are there clear regulations, and is there protection against function creep.

4.2.1 Who has access

To determine the accessibility and use of data while maintaining the required privacy, one has to examine in detail the persons that have access to private data and in what way measures have been taken to avoid leakage of the data to the outside world. For example, in the case of military or medical data, the people that are handling the data should be responsible (swearing of an oath) with regards to data being leaked to the outside world. One has to determine the respective responsibilities based on risks connected to the personal data. According to CNIL, it is recommended to use the "RACI matrix"; that is, determine who is responsible for carrying out each action (R=Responsible), who is accountable (A=Accountable), who is consulted (C=Consulted) and who is kept informed (I=Informed). Having these kind of measures in place would result in a "yes" score.

4.2.2 Clear regulations

The CNIL report recommends clear regulations in order to secure privacy protection. A clear and transparent privacy protection policy should be managed according to the following most important recommendations:

- Set out important aspects relating to data protection within a documentary base making up the data protection policy in a form suited to each type of content.
- Distribute the data protection policy to those in charge of enforcing it.
- Have acceptable exceptions to the regulations approved and amend the data protection policy accordingly.
- Establish a multi-annual action plan and monitor this.

When in the managing of the privacy protection policy these main principal recommendations are followed with regard to data protection regulations, this would compare to a usability score of "yes."

4.2.3 Protection against function creep

To protect against function creep, the CNIL report recommends taking measures such as:

- Deleting unnecessary accounts (e.g. guest accounts).
- Authorizing user functionalities only according to what is necessary. In other words, only giving advanced functionalities to those who need it.
- Deleting unnecessary services.
- Disabling autorun when inserting a removable device.

When these kinds of measures are taken, a "yes" would be scored.

4.3 Data protection

The assessment questions concerned with the topic of data protection are encryption, protection against theft and protection against manipulation. As was mentioned in D3.3b, this category, while extremely important for data protection assessment cannot be rated for a technology in general but has to be evaluated for every concrete implementation of a surveillance technology. Therefore, a number of issues that have to be discussed will be mentioned in the subsequent paragraphs. Examples of assessments have been given in the Appendix of D3.3b.

4.3.1 Encryption or otherwise access protected

When storing data some important questions about data protection have to be answered. As data from surveillance systems are critical to privacy, the data should be encrypted or otherwise protected against unauthorized access.

Good practices on data encryption are given in the CNIL report. This report gives both an objective for encryption (to make personal data unintelligible to anyone without access authorization) and a list of good practices to make sure the encryption is done in a correct manner. This report is, however, not specific to surveillance, so some practices may not be applicable.

Good practices as mentioned in the CNIL report:

- Determine what data should be encrypted and in what form.
- Choose the type of encryption that will be used to encrypt the data.
- Making sure the encryption algorithms used are chosen from public algorithms that are known to be adequate.
- Making sure measures are available to ensure the availability, integrity and confidentiality of the information necessary to recover possible lost secrets.

4.3.2 Protected against manipulation

Measures should be in place to protect the data against internal attackers, i.e. malicious employees, and external attackers, i.e., targeted attack against material that could become evidence.

The objective is "to be warned in the event of an unwanted modification or disappearance of personal data."[35] Good practices as given by the CNIL are:

- Identify and indicate what data has to be monitored for integrity.
- Making sure a method is chosen to monitor the integrity of the data.
- Determine when the function is to be applied and when the monitoring should be performed based on implementation of the business process.

When measures are taken to protect against external and internal manipulation, a score of "yes" can be given.

4.3.3 Secure against theft

When deploying mobile technology, like an audio bug or a hidden camera, that stores the data internally, this storage should be protected against attackers who might try to extract the data. When it is possible to extract the information from the device this will render a score of "no."

5 Proven Technology

The criterion for "excellence" is whether a given technological system has proven its use beyond a 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. It is very difficult to set fixed criteria for this factor. Moreover, when surveillance technologies can be deployed for different objectives, their effectiveness can differ for each goal, rendering it possible for one technology to obtain different scores depending on the goal for which it is used. Expert opinion should play an important role in determining what technologies have proven to be effective in their use as a surveillance technology.

6 Conclusion

In this report, effort was made to support decision makers in determining what scores to give different surveillance technologies when deciding what technology to use. For some factors a threshold has been set. In other cases, it has proven to be difficult to prescribe a rule on how to decide on the scoring, and rather the opinion of experts should be employed to make a sound decision on the scoring. Where thresholds have been established they are meant to serve as a guide and can be altered over time and for different scenarios.

The template for evaluating the effectiveness of surveillance technology as outlined here is the first time such a scoring system has been developed. As such,

[35] Commission Nationale de l'Informatique et des Libertés.

it should and will develop and improve over time. There have already proven to be differing views even within the SURVEILLE project itself on the appropriateness of this scoring process. The project has decided to adopt the scoring as it appears in this document. However, the law-enforcement end-user-panel (EUP) disagreed with the elements chosen for, or rather omitted from, the usability scoring. The EUP expressed the view that the issues of justification, proportionality and necessity are a critical and fundamental part of any consideration to deploy surveillance technology; and the scoring as a whole does not account for the required initial evaluation by law enforcement of the necessity, proportionality, and justification prior to of deploying surveillance equipment in a given scenario. Without the inclusion of these factors any scoring of the equipment would, according to the EUP, be futile.

According to the Coordinator of SURVEILLE, these comments suffer from a misunderstanding, related to the fact that the EUP was reviewing WP3 work on the effectiveness of surveillance, rather than SURVEILLE work as a whole. In the multidimensional and multidisciplinary framework of SURVEILLE, issues about necessity, proportionality and justification are considered in the fundamental rights intrusion score and in the final phase of merging the results from the three parallel assessments, namely the usability assessment reflected in the score discussed in this paper, together with the separate ethics and fundamental rights assessments. The EUP is definitely right in that issues of necessity, proportionality and justification must be included in decision-making about the use of a surveillance technology. In the SURVEILLE methodology they are included elsewhere in the overall matrix than in the usability score. The reader is referred to deliverables D2.6, D2.8, D2.9 and D4.10.

As stated above, this methodology is the first of its kind developed for evaluating surveillance technology. And as is the case with any methodology, it will be continuously improved over time.

References

Blejcharova, N., Cheu, R. L., and Bina, L. "Framework for selecting screening technologies for checked baggage inspection system at airport," *Journal of Transportation Security*, 2012.

Bulter, V. and Poole, R.W. "Rethinking checked-baggage screening," Policy Study 297. Reason Public Policy Institute, 2002.

Caplan, Joel M., Leslie W. Kennedy, Gohar Petrossian. "Police-monitored CCTV cameras in Newark, NJ: A quasi-experimental test of crime deterrence." *Journal of Experimental Criminology* 7, no. 3: 255–74, September 2011.

Commission Nationale de l'Informatique et des Libertés. Measures for the Privacy Risk Treatment, June 2012.

European Data Protection Supervisor. Data Protection Glossary, https://secure.edps.europa.eu/EDPSWEB/edps/EDPS/Dataprotection/Glossary/pid/74.

Gill, Martin and Angela Spriggs. "Assessing the Impact of CCTV," Home Office Research Study No. 292. Home Office Development and Statistics Directorate, London, 2005.

Hillstrom, Anne P., Lorraine Hope, Claire Nee. "Applying psychological science to the CCTV review process: A review of cognitive and ergonomic literature." Home Office Scientific Development Branch, March 21, 2008.

Hornick, J., Paetsch, J., and Bertrand, L. "A manual on conducting economic analysis of crime prevention programs," Canadian Research Institute for Law and the Family, 2000.

Jacobson, S.H., Karnani, T., Kobza, J.E. and Ritchie, L. "A cost-benefit analysis of alternative device configurations for aviation-checked baggage security screening," *Risk Analysis*, vol. 26, issue 2, 2006.

Legg, D. and James, P. "Measuring inputs: guidance for evaluators," Research, Development and Statistics Directorate, Home Office, UK, 2000.

Lum, Cynthia, Leslie W. Kennedy, and Alison Sherley, "Are Counter-Terrorism Strategies Effective? The Results of the Campbell Systematic Review on Counter-Terrorism Evaluation Research," *Journal of Experimental Criminology* 2, no. 4: 489–516, January 10, 2007.

Näsholm, Erika, Sarah Rohlfing, James D. Sauer. "Pirate Stealth or Inattentional Blindness? The Effects of Target Relevance and Sustained Attention on Security Monitoring for Experienced and Naïve Operators," *PLoS One*, 9:1, Jan. 21, 2014.

Stewart, Martin G. and John J "Cost-benefit analysis of advanced imaging technology full body scanners for airline passenger security screening," *Journal of Homeland Security and Emergency Management*, 8:1, 2011.

Tilley, Nick. "Understanding Car Parks, Crime and CCTV: Evaluation Lessons from Safer Cities." Crime Prevention Unit Series Paper No. 42. Home Office Police Department, London, 1993.

University of Leicester, Criminology, Technologies of Control, 8.4 Surveillance and CCTV, http://www.le.ac.uk/oerresources/criminology/msc/unit8/page_10.htm.

Virta, J.L., Jacobson, S.H., Kobza, J.E., "Analysing the cost of screening selectee and non-selectee baggage," *Risk Analysis*, vol. 23, no. 5, 2003.

Van Um, E. and D. Pisoiu, "Effective Counterterrorism: What Have We Learned so Far?" Economics of Security Working Paper 55, Berlin: Economics of Security, 2011.

Van Voorthuijsen, G., H. van Hoof, M. Kilma, K. Roubik, M. Bernas, P. Pata. "CCTV Effectiveness Study," 39th Annual IEEE International Carnahan Conference on Security Technology, 2005.

Webb, Barry and Gloria Laycock. "Reducing Crime on the London Underground: An Evaluation of Three Pilot Projects." Crime Prevention Unit Paper No. 30. Home Office, London, 1992.