



DELIVERABLE 7.1

Unity – Report on Evaluation Criteria

Executive summary

Due date: 31st October 2015

Date of submission: 31st October 2015

Lead beneficiary: Centre of Excellence in Public Safety Management/Rotterdam
School of Management, Erasmus University Rotterdam

Project Title: Unity

Grant Agreement: 653729

Funding Scheme: Research and Innovation action – Safeguarding Secure Society

Duration Time: 36 months

Start date: 01/05/2015



Project funded by the European Commission within the H2020
Framework Programme

Introduction

This report provides the blueprint for technology evaluations in the Unity project. The development of the Unity platform is a step-wise procedure from concept to functioning prototype, and evaluations are an integral part of this process. The philosophy of Unity is based on participatory design, the premise of which is to include (potential) users of a technology as early as possible and optimally already throughout the design and development phase.

The objective of this report is to describe a high-level framework that will guide the planning and implementation of evaluations in the complex Unity setting. The focus is on the evaluation of the functioning as well as the effects of the Unity platform along consecutive stages of its development. Specifically, it outlines concepts, approaches to and operationalizations for (1) functionality and usability testing, (2) user acceptance and adoption and (3) short- and mid-term impacts.

Methodological approach

The suggested approach for evaluations in Unity is based on the following characteristics:

1. Consecutive piloting approach
2. Consideration of changes over time
3. Multi-stakeholder view
4. Multi-level approach
5. Triangulation of findings from various outcome levels

The above considerations translate into minimally two core aspects for our evaluation blueprint, namely a longitudinal setup of the evaluation and the consecutive logic of our pilots.

Timeline of the evaluation

To determine whether the technology implementation did lead to changes, we advocate for a baseline measurement to capture attitudes, behaviors, performance, etc. before the implementation. Subsequent measures after the platform has been deployed will then be compared to this baseline to determine whether changes have indeed taken place. This setup also allows us to test for causal effects. The suggested period for each pilot is approx. 9 months, with each measurement points three months apart.

The piloting approach

The design process from concept to functioning prototype is an integral part of the project and evaluations form important milestones for the improvement of the platform and/or its components. The whole design process will go through several iterations and the technology will increase in maturity over time. To test these modifications, consecutive pilots are planned. The pilots start with a first pilot in month 6 of the project for a proof-of-concept and end with a final pilot in month 31. Further, because of the multi-stakeholder approach (simultaneously considering various user groups such as community groups, intermediaries and police users) as well as the multi-national deployment (Unity is designed around six pilots in six different countries), transferability across contexts is an integral part of the design and development process.

We apply the consecutive piloting approach to maximize cross-cultural generalizability within UNITY. The fact that each pilot will take place in a different national context creates very good conditions to test for the localization versus generalizability of the platform. Yet again, this approach also clearly restricts the direct comparability of any impact assessments. This does not render impact assessments moot, but we are aware of the challenges for the generalizations of findings across (national) implementations contexts (cp. also discussions of cross-cultural data collection in D3.1).

Within Unity we integrate the development of the tool and the evaluation of our tool with the consecutive piloting approach. With this methodological setup we grasp step by step context-specific constraints and directly feed this information into the next development stage of the platform. In this scenario, the first pilot is primarily intended as proof of concept, i.e., no actual technology can be deployed to users. The second pilot is intended as demonstration of a first prototype, and so on until the platform obtains a certain stability. In consequence, we suggest to start with the full evaluation process only from the third pilot and concentrate on different aspects in the first two.

Cross-cultural aspects in the evaluation

The consecutive setup with pilots moving across countries introduces considerably variation in terms of national (political, economic, legal, social and technological) contexts as well as potential disparities in terms of norms, experiences, knowledge, etc. These national and contextual differences challenge the comparability of concepts, processes and results across pilot sites. The goal must be to establish “equivalent meaning of concepts and data in different social and cultural contexts” (Craig & Douglas, 2005). We advocate six basic principles of cross-cultural methodology (for a broader discussion see Jacobs & Bayerl, 2015). These are:

1. Sample equivalence
2. Data-collection equivalence
3. Measure equivalence
4. Conceptual equivalence
5. Functional equivalence
6. Translation equivalence

Target communities and groups

To allow for a better comparison of evaluation results across the different pilot countries, we will focus on the same type of communities in each pilot country. The equivalent groups across countries are:

- Young members of a minority group (above 18 years of age and below 26); which minority group is selected can be determined by each country itself;
- One intermediary, i.e., organization that supports police works in accessing or addressing this minority group;
- The local police force and subgroups within the police that deal with this minority group.

Pilot countries can further decide on one additional community, which is of special relevance for their own national or local police force. All community groups should be eligible to participate on legal grounds, and their choice should add value to the tool development in that it increases diversity in the end-user groups.

Selection criteria and suggested sample sizes

We consciously refrain to name concrete numbers or selection criteria for participants at this point in time. Taking sample and data-collection equivalence into account the sampling technique, selection criteria and number of participant will be decided upon the experiences and insights gained through additional empirical data from WP2 and WP7 and in close engagement with the pilot countries.

Consideration of diversity aspects

We will consider explicitly the specific needs of different social groups. Gender will be considered in the various local, regional and national contexts, but also other relevant diversity strands (such as ethnicity, socio-economic status, disability, sexual orientation or religious views) may be included, where relevant.

Evaluation criteria

Functionality and usability testing

Functional testing is a process, in which software is tested to ensure that it conforms to all requirements as laid out in the design specifications. The most appropriate one(s) will be chosen based on the context of the testing once the platform has been developed to an appropriate stage. Some techniques that may be employed in this context are:

- *Smoke testing;*
- *Black box testing;*
- *White box testing;*
- *Unit testing;*
- *User acceptance testing.*

Non-functional testing addresses the quality of the component or system. Non-functional testing refers to a broad number of testing disciplines addressing several (non-functional) features. Among these:

- *Load testing;*
- *Stress testing;*
- *Performance testing;*
- *Recovery testing;*
- *Resilience testing;*
- *Security testing;*
- *Scalability testing.*

Integration of the platform and data into existing systems will also need to be assured, especially on the police side (*integration testing*). Suggested methods for usability testing are heuristics evaluations and A/B testing. Suggested methods for functionality testing include the creation of test plans together with standard, erroneous and destructive testing to ensure that the system is robust enough.

Tool acceptance and adoption

This part of the evaluation aims to understand 1) whether and to what extent users are willing to adopt the newly developed tool(s), and 2) the factors that may influence the willingness or actual usage of the tool(s) by individuals, groups or organizations.

Acceptance and adoption touch on a wide range of aspects to assess how user perceive a new technology. The most common concepts are shown in Table 1. Table 2 provides potential operationalizations and methods for these concepts as well as the user group(s) where each aspect should be assessed.

Table 1. Overview of main acceptance and adoption concepts (adapted from Jeyaraj et al., 2006)

Concept	Definition	Literature (selection)
Acceptance	The demonstrable willingness within a user group to employ IT for the tasks it is designed to support	Dillon & Morris, 1996
Resistance/rejection	Refusal of individuals to use the system or comply with ways of working prescribed by their organization	Kim & Kankanhalli, 2009
Adoption	Users' decisions to accept and integrate innovations into their work routines	Bhattacharjee et al, Perols, & Sanford, 2008
Continuance intentions	Intension to continue using the new system	Bhattacharjee et al, Perols, & Sanford, 2008
Satisfaction	Degree to which users are satisfied with the system	Au et al., 2008
Perceived system use	Amount of use of a new technology by a person or organization	Jeyaraj et al., 2006
Appropriation	Processes involved in the integration of new technologies into existing work practices	Dennis et al., 2001
Adaptation	Mutual influences of technology and users over time	Majchrzak et al., 2000

Table 2. Measurement of user acceptance an adoption (partly adapted from Jeyaraj et al., 2006)

Concept	Possible operationalization	Measurement level	Possible methods	Police	External
Intention to use	Self-reports on possible use in the future	Individual level	Self-reports through surveys or interviews	X	X
Acceptance	Self-reports on degree of willingness to use the technology	Individual level	Self-reports through surveys or interviews	X	X
Resistance	Self-reports on degree of unwillingness to use the technology	Individual level	Self-reports through surveys or interviews	X	X
Satisfaction	Self-reports on degree of satisfaction with the technology	Individual level	Self-reports through surveys or interviews	X	X
Perceived system use	Self-reports of system usage (e.g., in terms of frequency or intensity)	Individual and group level	Self-reports through surveys or interviews	X	X
Concept	Possible operationalization	Measurement level	Possible measurement	Police	External
Adoption	Binary variable: yes/no	Individual and group level	Objective information in tool logs or observation	X	X
Diffusion	Amount (percentage) of available features used by an individual, group or organization	Group and organizational level	Objective information in tool logs or observation	X	X
Rate of adoption	Percentage of adopters in a group and number of active users	Group and organizational level	Objective information in tool logs or observation	X	X
Actual system use	Amount of usage such as number of reports sent, frequency of reports	Individual, group, organizational level	Objective information in tool logs or observation	X	X
Length of adoption	Time elapsed since first adoption (e.g., in months or years)	Individual, group, organizational level	Archival documents or self-reports	X	X

To understand why a certain technology succeeds or fails, it is important to also consider possible antecedents to acceptance and adoption. Potential antecedents can be found in the following areas (cp. Jeyaraj et al., 2006; Venkatesh & Davis, 2000):

- Technology characteristics;
- Characteristics of the individual;
- Characteristics of the organization;
- Characteristics of the environment.

Tables 3 to 5 list factors that appear as the most consistent impact factors (cp. Jeyaraj et al., 2006). Individual level factors are shown in Table 3, organizational level factors on Table 4. Table 5 lists impact factors of the local or national environment with relevance for user and organizational acceptance and adoption decisions.

Table 3. Potential influence factors on user acceptance and adoption – individual level (partly adapted from Jeyaraj et al., 2006)

Concept	Definition	Literature (selection)
Perceived usefulness	Degree to which a person believes that using a particular system would enhance his or her job performance	Davis, 1989
Perceived ease-of-use	Degree to which a person believes that using a particular system would be free from effort	Davis, 1989
Results demonstrability	Tangibility of the results of using the innovation	Moore & Benbasat, 1991; Venkatesh & Davis, 2000
Job relevance	Individual's perception regarding the degree to which the target system is relevant to his or her job	Venkatesh & Davis, 2000
Task-technology fit	Degree to which a technology assists an individual in his or her portfolio of tasks	Goodhue & Thompson, 1995
Perceived output quality	Degree to which an individual believes that the system performs his or her job tasks well	Venkatesh & Davis, 2000
Subjective norms	Person's perception that most people who are important to him think he should or should not perform the behavior in question	Venkatesh & Davis, 2000
Image	Degree to which use of an innovation is perceived to enhance one's status in one's social system	Venkatesh & Davis, 2000
Experience	Extent of prior experience with comparable technologies	n/a
Voluntariness	Extent to which potential adopters perceive the adoption decision to be non-mandatory	Venkatesh & Davis, 2000
Demographic variables	Features of users, which might impact attitudes, emotions, behaviors, etc. (e.g., age, gender, nationality, etc.)	n/a

Table 4. Potential influence factors on user acceptance and adoption – organizational level (partly adapted from Jeyaraj et al., 2006)

Concept	Definition	Literature (selection)
Management support	Extent to which managers in the organization provide resources or guidance to support users	Rai & Howard, 1994
Organizational support	Extent to which the organization provides resources or guidance to support users	Thompson et al., 1991
Use of/burden on resources	Extent to which the tool reduces or increases burden on resources in the organization (e.g., personnel, finances, time)	n/a
Features of the organization	Characteristics of the organization that impact acceptance and adoption decisions of new technologies on the organizational level (e.g., size, structure, firm history, existing infrastructure)	Bajwa et al., 2005; Rai et al., 2009

Table 5. Potential influence factors on user acceptance and adoption in the environment

Concept	Definition	Literature (selection)
Urbanization	Degree of urbanization from rural to urban/central city	Damanpour & Schneider, 2006
Community wealth	Measured as income of the target population	Damanpour & Schneider, 2006
Unemployment rate	Percentage of unemployed in the population	Damanpour & Schneider, 2006
Education level	Educational attainment in the target population	Damanpour & Schneider, 2006

Impact evaluation

This part of the evaluation aims to understand the impact (or outcomes) of the citizen tool focusing on five areas:

1. Operational outcome criteria
2. Livability criteria
3. Relationships
4. Acceptance of CP
5. Police-internal attitudes and processes

Operational outcome criteria describe changes in police performance in terms of potential key performance indicators. An overview of relevant concepts, their possible operationalizations and methods for their measurement are provided in Table 6. It should be noted that the list in Table 6 is of course only a selection. This selection is largely based on the empirical data about indicators and measurements of community policing success in WP3.

Table 6. Overview of concepts to assess operational outcomes

Concept	Possible operationalization	Measurement level	Possible measurement	Police	External
Crime reduction	Number of crimes on a local or national level	Group level	Objective information in police data	X	
Increase in solved crimes	Number of solved crimes on a local or national level	Group level	Objective information in police data	X	
Faster solution of cases	Time taken for the solution of cases	Group level	Objective information in police data; Self-reports through surveys or interviews	X	X
Citizen satisfaction with police work	Degree of satisfaction with police work	Individual level	Self-reports through surveys or interviews		X
Perception of performance police	Subjective impression of police work in terms of , e.g., effectiveness, quality	Individual level	Self-reports through surveys or interviews	X	X
Reduction in wrong convictions	Number of wrong convictions	Group level	Self-reports through surveys or interviews; Objective information in police data	X	X
Availability	Extent to which police officers are available for the public, e.g., in terms of 1. Ease of making contact 2. Speed of reactions 3. Time spent for requests	Individual and group level	Self-reports through surveys or interviews; Objective information in police data	X	X
Visibility	Extent to which police officers are present in the public	Individual and group level	Self-reports through surveys or interviews; Objective information in police data	X	X

Crime reporting	1. Number of crimes reported from the target communities	Individual and group level	Objective information in police data	X
	2. Number of people reporting crimes			
	3. Quality of the information reported			

Citizen participation, and with this establishing and maintaining relationships between police, target groups and other stakeholders, is an important element of community policing (e.g., Myhill, 2012). Relevant criteria and their possible operationalizations are summarized in Table 7.

Table 7. Concepts to assess the relationship between citizens, police and other stakeholders

Concept	Possible operationalization	Measurement level	Possible methods	Police	External
Trust/confidence	Degree of trust/confidence in police as organization	Individual and group level	Self-reports through surveys or interviews	X	X
Legitimacy	Degree of perceived legitimacy of police as organization	Group and organizational level	Self-reports through surveys or interviews	X	X
Fairness	Degree of perceived fairness in treatment of target group members	Group and organizational level	Self-reports through surveys or interviews	X	X
Respect for police	<ol style="list-style-type: none"> 1. Degree of reported respect towards <ol style="list-style-type: none"> a. Members of target groups b. Police officers 2. Number of respectful and/or disrespectful acts 	Individual, group, organizational level	Self-reports through surveys or interviews; Objective information in police data	X	X
Perceived expertise	Degree of perceived expertise of police <ol style="list-style-type: none"> 1. In general 2. With respect to the target communities 	Group and organizational level	Self-reports through surveys or interviews	X	X
Perceived professionalism	Degree of perceived expertise of police <ol style="list-style-type: none"> 1. In general 2. Treatment of the target communities 	Group and organizational level	Self-reports through surveys or interviews	X	X
Image of the police	<ol style="list-style-type: none"> 1. Positive perception of police organization 2. Degree of positive media coverage, positive feedback/comments by the public on social media 	Organizational level	Self-reports through surveys or interviews; Objective data from media analysis	X	X
Closer cooperation	Extent of collaborative activities between police, citizens and other groups or organizations, e.g. in terms of: <ol style="list-style-type: none"> 1. Providing information 	Individual, group, organizational level	Self-reports through surveys or interviews; Objective information in	X	X

	<ol style="list-style-type: none"> 2. Providing assistance 3. Reporting (potential) crimes and disorder 4. Become a police volunteer [where applicable depending on country context] 5. Become a police officer/apply for a police job 		police and external data		
Awareness of security as common good	Degree to which individual or groups perceive the creation of security as a common tasks	Individual and group level	Self-reports through surveys or interviews	X	X
Willingness to work for the police	<ol style="list-style-type: none"> 1. Number of volunteers from the target communities [where applicable depending on country context] 2. Number of applicants for police jobs [where applicable depending on country context] 	Individual and group level	Objective information in police data		X
Complaints about police work	Number of official complaints against police officers	Individual and group level	Objective information in police and/or governmental data [depending on country context]	X	

The success of community policing efforts is tied to the willingness of citizens to accept community policing as a concept and to engage with police officers in their area. Therefore, acceptance of community policing will be assessed with respect to general acceptance as well as the willingness of the public to directly engage in community policing. Table 8 provides possible operationalizations for both aspects.

Table 8. Concepts to assess acceptance of community policing

Concept	Possible operationalization	Measurement level	Possible methods	Police	External
Acceptance of community policing	Degree to which activities for CP are accepted and supported by citizens/external stakeholders, e.g., with respect to: <ol style="list-style-type: none"> 1. Visibility/presence of officers in public space 2. Emergency assistance and aid 3. Information gathering for crime prevention (in person, through IT) 4. Reaction to help requests 	Individual and group level	Self-reports through surveys or interviews	X	X

	5. Engaging with CP officers on the street				
Willingness to help police	Willingness in the target community to:				
	1. Share information with police				
	2. Voice concerns about safety				
	3. Report (potential) crimes and disorder	Individual and group level	Self-reports through surveys or interviews	X	X
	4. Become a police volunteer [where applicable depending on country context]				
5. Become a police officer/apply for a police job					

Livability describes the extent to which citizens consider their environment as a space that affords sufficient quality of life across various areas (e.g., work, family, free time, etc.). Community policing can support livability, for instance, by increasing citizens’ feeling of safety or by improving social cohesion by fostering contacts between diverse groups. Table 9 lists possible ways to assess livability in communities.

Table 9. Overview of concepts to assess livability

Concept	Possible operationalization	Measurement level	Possible methods	Police	External
Feeling of safety	Extent to which citizens perceive themselves, their groups or their environment to be safe	Individual and group level	Self-reports through surveys or interviews		X
Effects on society	Responsibility for and ownership of problems, crime and disorder	Individual, group, community level	Self-reports through surveys or interviews	X	X
Improvements in image	Better image of a group, city, area, etc. internally, by other citizens or by police	Group and community level	Self-reports through surveys or interviews; Media analysis	X	X
Social cohesion	Community cohesion and trust within the community	Group and community level	Self-reports through surveys or interviews		X

The next aspect addresses effects within the police and for individual police officers in terms of internal work processes and officers’ attitudes. Table 10 lists relevant concepts identified in D3.1.

Table 10. Overview of concepts to assess police internal attitudes and processes

Concept	Possible operationalization	Measurement level	Possible measurement	Police	External
Improved intelligence gathering	Amount of information/intelligence; Quality of information	Organizational level	Objective information in police; Self-reports through surveys or interviews	X	
Efficiency	Less resources needed in terms of personnel, time, budget, etc.	Individual and organizational level	Objective information in police; Self-reports through surveys or interviews	X	X
Knowledge about communities	Degree of knowledge about target communities in terms of e.g., amount, validity, quality	Individual and organizational level	Self-reports through surveys or interviews	X	X
Knowledge of problem causes	Better problem solving; improved assessment of CP effectiveness	Individual and group level	Self-reports through surveys or interviews	X	
CP as integral part of the police	Degree to which community policing is integrated in organizational structures, practices, trainings, etc.	Organizational level	Objective information such as structures and practices; Self-reports through surveys or interviews	X	
Localized approach	Degree of localization of community policing structures and practices	Organizational level	Objective information such as structures and practices; Self-reports through surveys or interviews	X	X
Professionalism of officers	Status and respect of CP within the police	Individual level	Self-reports through surveys or interviews	X	X
Less right violations, fewer victims of police violence	Frequency of official violations and complaints, severity of violations and complaints	Organizational level	Objective information in police data; Self-reports through surveys or interviews	X	X

Unintended consequences

Technologies are implemented with a “general intent” in mind “with regard to values and goals underlying a given set of structural [technology] features” (DeSanctis & Poole, 1994, p. 126). Users, however, often deviate from this general intent, choosing to ignore features designers or their organization deem important, modifying them according to their own preferences or using them in unintended ways.

We suggest two ways of capturing unintended side-effects: firstly, analyze behavioral traces from tool use; secondly, interviews and focus groups with selected users in the community and in the police

forces. The first approach obtains concrete usage data to pinpoint unexpected patterns and usages, the second approach can identify users' motivations, concerns and interpretations of the tool, their rationales for (not) using it and how the tools changed attitudes, knowledge, behaviors, etc. of themselves or their community group. Optimally interviews would therefore target individuals, which continue using the tools (high or low usage) and individuals, which discontinued the use. For each target group (communities, external stakeholders and police) we suggest to conduct ten to twelve interviews, half of them with continued/heavy users, half of them with discontinued/non-users.

References

- Au, N., Ngai, E., & Cheng, T. (2008). Extending the understanding of end user information systems satisfaction formation: An equitable needs fulfillment model approach. *MIS Quarterly*, 32 (1), 43-66.
- Bajwa, D.S., Lewis, L.F., Pervan, G., & Lai, V.S. (2005). The adoption and use of collaboration information technologies: International comparisons. *Journal of Information Technology*, 20, 130-140.
- Bhattacharjee, A., Perols, J., & Sanford, C. (2008). Information technology continuance: A theoretic extension and empirical test. *Journal of Computer Information Systems*, 49 (1), 17-26.
- Craig, C. S., & Douglas, S. P. (2005). *International Marketing Research*. John Wiley & Sons, Chichester.
- Damanpour, F. & Schneider, M. (2006). Phases of the adoption of innovation in organizations: Effects of environment, organization and top managers. *British Journal of Management*, 17, 215-236.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13 (3), 319-339.
- Dennis, A.R., Wixom, B.H., & Vandenberg, R.J. (2001). Understanding fit and appropriation effects in group support systems via meta-analysis. *MIS Quarterly*, 25 (2), 167-193.
- DeSanctis, G., & Poole, M. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science*, 5 (2), 121-147.
- Dillon, A., & Morris, M. G. (1996). User acceptance of information technology: Theories and models. *Annual Review of Information Science and Technology*, 31, 3-32.
- Goodhue, D.L. & Thompson, R.L. (1995). Task-technology fit and individual performance. *MIS Quarterly*, 19 (2), 213-236.
- Jacobs, G. & Bayerl, P.S. (2015). Accounting for cultural influences in big data analytics. In B. Akhgar, G.B. Saathoff, H.R. Arabnia, R. Hill, A. Staniforth & P.S. Bayerl (Eds.), *Application of Big Data for National Security* (pp. 250-261). Amsterdam: Elsevier.
- Jeyaraj, A., Rottman, J. W., & Lacity, M. C. (2006). A review of the predictors, linkages, and biases in IT innovation adoption research. *Journal of Information Technology*, 21(1), 1-23.
- Kim, H.-W. & Kankanhalli, A. (2009). Investigating user resistance to information systems implementation: A status quo bias perspective. *MIS Quarterly*, 33 (3), 567-582.
- Majchrzak, A., Rice, R. E., Malhotra, A., & King, N. (2000). Technology adaptation: The case of a computer-supported inter-organizational virtual team. *MIS Quarterly*, 24 (4), 569-600.
- Moore, G.C. & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research*, 2, 192-222.
- Myhill, A. (2012). *Community engagement in policing. Lessons from the literature*. NPIA (National Policing Improvement Agency).
- Rai, A. & Howard, G.S. (1994). Propagating CASE usage for software development: An empirical investigation of key organizational correlates. *OMEGA*, 22 (2), 133-147.
- Rai, A., Maruping, L.M., & Venkatesh, V. (2009). Offshore information systems project success: The role of social embeddedness and cultural characteristics. *MIS Quarterly*, 33 (3), 617-641.

- Thompson, R. L., Higgins, C. A., & Howell, J. M. (1991). Personal computing: Toward a conceptual model of utilization. *MIS Quarterly*, *15* (1), 124-143.
- Venkatesh, V. & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, *46* (2), 186-204.