Project number:	**288577**
Project acronym:	**urbanAPI**
Project title:	**Interactive Analysis, Simulation and Visualisation Tools for Urban Agile Policy**
Instrument: Call identifier: Activity code:	**STREP** **FP7-ICT-2011-7** **ICT-2011.5.6 ICT Solutions for governance and policy modelling**

Start date of Project:	**2011-09-01**
Duration:	**36 month**

Deliverable reference number and title (as in Annex 1):	
Due date of deliverable (as in Annex 1):	
Actual submission date: Revision:	***see "History" Table below***

Organisation name of lead contractor for this deliverable:
University of the West of England

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium	
CO	Confidential, only for members of the consortium (including the Commission Services)	X

Title:
D5.2 – User Evaluation Cycle 1
Author(s)/Organisation(s):
Zaheer Khan (UWE)/Kamran Soomro (UWE)
Working Group:
UWE, FH-IGD, AIT, AEW, City partners
References:
D5.1 – Assessment Methodology Other WP5 documents

Short Description:
This document presents the user evaluation cycle 1 results. It describes the evaluation process and presents the criteria based and scenario based evaluation results. In addition, evaluation issues, user response analysis and lessons learned are also presented.
Keywords:
Criteria Indicators and Metrics, Questionnaire, Testing scenarios

History:

VERSION	AUTHOR(S)	STATUS	COMMENT	DATE
001	ZAHEER KHAN (UWE)	NEW	FIRST TEMPLATE OF TOC	27-MAR-2013
002	ZAHEER KHAN (UWE)	RFC	REVISED TOCS BASED ON QUALITY REVIEW/FEEDBACK. AND ALSO, FILLED 'ABOUT THIS DOCUMENT' AND 'INTRODUCTION' SECTIONS	12-JULY-2013
003	ZAHEER KHAN (UWE)	RFC	COMPLETED CHAPTER 2	19-JULY-2013
004	ZAHEER KHAN (UWE)	RFC	REVISED CHAPTER 1 AND CHAPTER 2, ADDED NEW FIGURE 2, FIGURE 3 AND FIGURE 4.	16-17 NOV-2013
005	ZAHEER KHAN (UWE)	RFC	ADDED CHAPTER 3.1 AND 3.2 FOR EVALUATION RESULTS OF VITORIA-GASTEIZ AND BOLOGNA APPLICATIONS. ALSO ADDED CHAPTER 3.4 FOR RULE EDITOR EVALUATION	18-19 NOV-2013

006	**ZAHEER KHAN (UWE)**	**RFC**	**ADDED CHAPTER 4.1, CHAPTER 4.2, CHAPTER 4.3, CONCLUSIONS AND EXECUTIVE SUMMARY**	**20-22 NOV-2013**
007				

TABLE OF CONTENTS

Executive Summary

[This chapter briefly presents the overall evaluation summary – will be updated once all the city evaluation reports are ready – Nevertheless, each city evaluation report has separate executive summary provided]

Using the Criteria Indicators and Metrics (CIM2) methodology [2][3][4] a rigorous evaluation design process is followed that resulted in a detailed evaluation criteria and sub-criteria in compliant with ISO 25010 characteristics and their respective assessment indicators and questions. These evaluation questions are directly derived from the user requirements specification and hence are fully traceable back to user needs and goals. Finally, a user consultation activity is carried out where city partners indicated priorities for these questions as well as identified missing and unnecessary questions. Finally, a web based interface is developed to receive evaluation feedback from city stakeholders and to carry out an automated analysis of the question response.

The overall approach mostly followed for the evaluation was to provide training material (videos) two weeks in advance. Also, hands-on training sessions are conducted for each city so that the evaluators can understand specific functionality and learn how to use the applications and then respond to specific questions. Based on the experience it is wise to say that this training based evaluation approach was successful to the extent that evaluators got first-hand experience of using and getting better understanding of the specific functionalities these applications can support. It is understood that without training sessions it would have been difficult to fully understand the functionality and especially how to perform a specific operation. The results reflect positive evaluation responses for the aspects covered in the hands-on training sessions.

The overall evaluation participation rate was very promising and different types of users – urban planners, policy makers, GIS experts, General public etc participated in the evaluation. That is all local stakeholders who attended training sessions responded to evaluation questions. However, the size of user sample set for individual cities is too small to generalise the results. After training sessions, evaluators were allowed to reuse application for a week and then respond to evaluation questions through online evaluation portal. This allowed evaluators to fully understand strengths and limitations of these applications in their own time and respond to evaluation questions based on educated knowledge of urbanAPI application scenarios. Based on the evaluation results a detailed analysis is performed and presented in city based evaluation reports. These analyses are provided from two perspectives: i) scenario based and ii) criteria based. In addition, evaluation results of rule editor are presented.

The overall assessment results are projected in the form of final mark values, which indicate whether or not the outcome is above acceptable threshold for application scenarios or criteria.

The 3DVR application is evaluated for Vitoria-Gasteiz, Bologna and Vienna and results indicate that ….

Similarly, PME application is evaluated for Vitoria-Gasteiz, Bologna and Vienna and results indicate that …

And, finally UGS application is evaluated for Ruse and results indicate that …

In addition to above applications, specific urbanAPI toolset is also evaluated e.g. Rule Editor. The evaluation result indicates that ….

It is found that many features were already implemented in the application but end users couldn't use them due to lack of detailed training and hence responded negatively. This suggests the need for better and detailed training needs during cycle 2. Also, some features couldn't be implemented for different cities due to lack of fine granularity of data available but similar features are implemented for other cities where data is of high quality. Evaluation responses and charts presented in city reports do not reflect such limitations which are beyond the control of application implementation capacity. Rather these limitations are explained separately in tables and covered in technical perspectives as learning outcomes.

Nevertheless, the overall purpose of the evaluation exercise is fulfilled, as it is able to identify the aspects, which require development focus during cycle 2. In addition, it is also understood that some of the limitations revealed are directly related to availability of quality and detailed data set e.g. high quality 3D city model or detailed GSM data to capture mobility patterns, or UGS socio-economic data etc. These aspects will be further explored during cycle 2 of the project development.

The evaluation results are also analysed from users level of expertise and roles point of view and the extent to which it affects the evaluation results. The results show that different types of evaluators (urban planner, policy maker, GIS experts, others) contributed to evaluation results more or less equally regardless of their level of expertise. However, role based user analysis indicate different types of users provide evaluation responded based on their domain knowledge.

An analysis of the response-behaviour of various user roles is also performed. The results show that different user roles are better suited for evaluating different aspects of the application. For example, evaluation results (and user perspectives) indicate that tertiary users are more suited to deal with basic functionality, usability and benefits-related evaluation questions while primary users are more appropriate to deal with detailed and complex functionality and usability related questions. Moreover, primary users are the best fit to validate application behaviour against user defined assessment criteria and system requirements for interim evaluation as secondary or tertiary users cannot relate different evaluation aspects/questions to original application requirements and available application features. Also, not all the application functionality is developed and suitable for secondary/tertiary users. For example in 3DVR application, general public or citizens are mainly concerned with the web based neighbourhood navigation, annotation or street refurbishment by placing selected objects and are not the right targeted audience for Admin tool, Rule editor or configuring the web based neighbourhood scenario. This is exactly why there is low participation rate in Bologna Rule Editor evaluation. And, despite having training material and getting hands on training, some users were unable to understand fully the available application features and respond to more functionality related questions for 3DVR and PME applications. This suggests that it will be useful to clearly identify that which components or features of the applications are for which type of users and training should be tailored according to different user groups.

This suggests that the overall limitations identified in the scenario based and criteria based evaluation results are not fully dependent on the evaluators level of expertise. Rather, these limitations can be attributed either to level of training provided (including training videos) to evaluators or required applications features were not found by evaluators. For interpretation purposes, explicit user perspective is included in city reports. Similarly, technical perspective is also provided that explains limitations beyond control of application implementation capabilities. These perspectives help to synthesise the evaluation results. In general, this is what exactly the CIM 2 based user evaluation cycle 1 was intended to do. And, consequently, evaluation results identified number of aspects in the urbanAPI applications which should be dealt with during development cycle 2.

About this document

[This chapter briefly describes what this document is about and who should read it]

This document briefly elaborates the process to prepare the user evaluation design and reports on the overall interim user evaluation cycle 1 results of different urbanAPI applications. This user evaluation follows the CIM2 approach as defined in the Deliverable D5.1 – Assessment methodology – accessible from http://www.urbanapi.eu/about/downloads.html. These interim user evaluation - cycle 1 - results reflect on the usability, functionality, relevance and benefits criteria and identify strengths and weaknesses and new requirements for urbanAPI applications. The intended audiences for this document are mainly urbanAPI project partners, especially technical partners – IT developers, local city stakeholders from different departments e.g. GIS experts, policy makers, surveyors, architects etc. and urbanAPI stakeholder board members. In addition, external stakeholders can benefit from this document to understand the overall evaluation process, issues and lessons learned. A more detailed list of stakeholders is identified in D2.1 – User requirements definition [1] and D5.1 – Assessment methodology [2].

1 Introduction and Objectives

[This chapter briefly introduces the overall concept of user evaluation using CIM methodology for urbanAPI applications. Also, what is expected during cycle 1 evaluation]

1.1 Background to User Evaluation

In urbanAPI project, user evaluation refers to validation of the end-to-end outcomes of the UrbanAPI applications against the stated and implied objectives and application scenario requirements of the project. The overall aim identified in the assessment methodology [2] is to validate functionality, usability and relevance of the applications and identify benefits and overall impact of the application. This user evaluation is planned for both development cycles of the project [5] and in this document cycle 1 evaluation planning and results are presented.

Overall the user evaluation will enable the evaluation stakeholders to contribute in verifying the following policy making and public participation objectives:

- indirect mechanism through usage data, such as number and type of interactions supported by the applications;
- direct mechanism through questionnaire for evaluation of qualitative and quantitative aspects of the implemented system by all city stakeholders, for example trust in the solutions;

- analysis of application specific results for different types of stakeholders and sharing of experiences and lessons learned.

In addition, the following ICT and policy modelling objectives will also be verified in both user and technical evaluation:

- usability study to analyse the impact of the planned 3D user interface elements and other GUIs designed for different applications;
- functional capacity of the applications based on the available technology and data;
- conduct task-based usability studies to show the effectiveness/efficiency of the rule-based approach and the corresponding user interfaces.

In general, the evaluation of each application consists of three main activities: i) *Preparation* – i.e. designing evaluation with specific criteria, indicators, test scenarios and questionnaire; ii) *User evaluation* – i.e. implementing or carrying out the actual evaluation; and iii) *Reporting* – i.e. documenting the evaluation outcomes.

1.2 Objectives of the Cycle 1 User Evaluation

The objective of the user evaluation cycle 1 is to perform an interim evaluation so that applications features can be improved during cycle 2. The cycle 1 evaluation aims:

i) to assess the extent to which urbanAPI applications have fulfilled user requirements, and
ii) to test current development status and indicate strengths and weaknesses to be handled in cycle 2 development stage.

This interim evaluation also helps to indicate technical challenges, such as data granularity and quality, affecting the usability and functionality related aspects of these applications. In addition, the overall evaluation approach will be examined to identify any gaps for improvement.

1.3 Report organisation

The remainder of this report presents the overall user evaluation process, planning and preparation, criteria and indicators in chapter 2. Chapter 3 presents application and city specific evaluation results followed by evaluation issues, user analysis and lessons learned in chapter 4. Finally, we conclude in chapter 5.

2 User Evaluation – Planning and Preparation

[This chapter briefly elaborates user evaluation concepts including the evaluation process, planning and preparation done to perform overall assessment]

As the main objective here in cycle 1 is to perform end-to-end (i.e. black-box) evaluation [2], based on the requirements specifications [1] including application scenarios, use-cases and requirement definitions, which provide necessary input to design evaluation contents. In this chapter we briefly elaborate the overall evaluation process and present necessary criteria and indicators selected for the purpose of evaluating urbanAPI applications.

2.1 Evaluation Process

[The process we followed]

The overall process to conduct the urbanAPI applications evaluation is based on selected set of activities as depicted in Figure 1 below. The evaluation process is explained in detail in D5.1 – Assessment methodology [2]. With reference to user evaluation cycle 1, the following activities are performed: i) evaluation design, ii) design verification, iii) development of test cases, iv) evaluation implementation, v) results documentation.

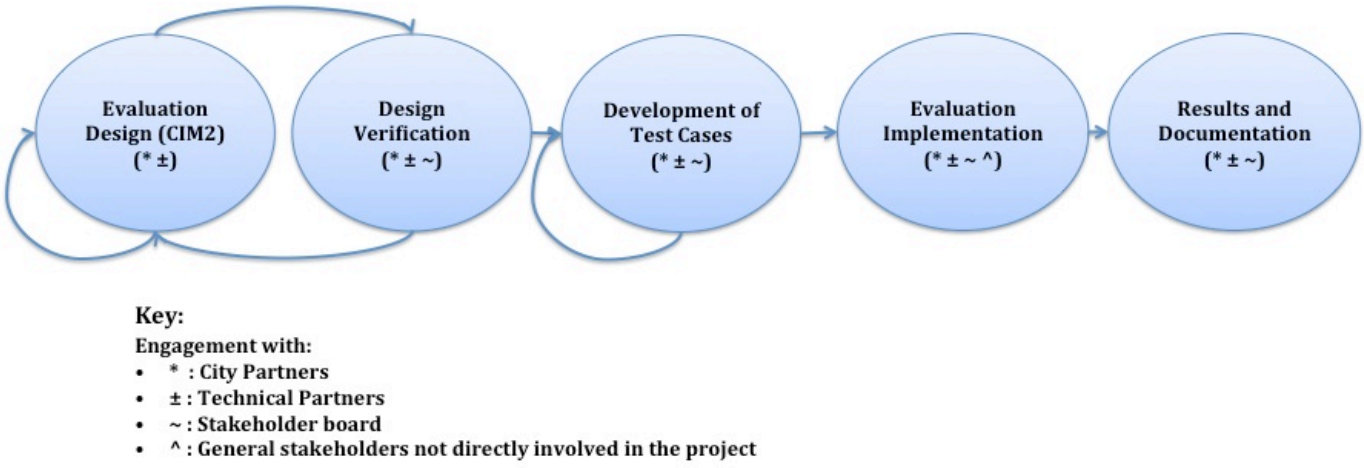


Figure 1: The evaluation process [2]

i) *Evaluation design:* This activity resulted in preparing a thorough and structured evaluation design that includes evaluation criteria, indicators and questions to be used for user evaluation. In this respect, ISO25010 characteristics compliant list of criteria identified and extended for relevance and benefits. Indicators are derived from the common user needs and requirements and mapped on to the already selected criteria. For each indicator a selected set of questions are derived directly from the user requirements. As a result, a

very comprehensive and detailed application and city specific evaluation design is prepared.

ii) *Design verification:* This activity resulted in consulting and getting feedback from the city stakeholders on the evaluation design. The idea is to improve the design by getting feedback from city stakeholders on criteria, indicators and questions. In addition, city partners identified the importance of each questions (i.e. weights) as well as more relevant local stakeholders to be involved in the evaluation exercises. Based on the feedback, questions are revised and more high priority questions are selected based on its importance to each application scenario and relevance to importance of original requirements. The aims of this activity were:

a) To verify that the evaluation design contents (i.e. criteria, questionnaire, etc) capture important assessment elements and are in line with stakeholder needs and requirements defined in Deliverable D2.1 – User requirements definition;

b) To highlight the importance (very low, low, medium, high, very high) of specific questions;

c) To indicate improvements in the indicators, questionnaire, response options, metrics etc, if possible;

d) Identification of additional criteria, indicators, questionnaire and metrics, if not covered in the existing ones;

e) to correctly identify evaluation stakeholders suitable to respond to specific questions e.g. urban planners or IT developers, etc, and

f) to get feedback on the online user evaluation.

The feedback is collected in specific templates and as a consequence, evaluation design is improved by adding/removing indicators/questions, specifying importance/weights of each question, indicating potential end users who can participate in evaluation and respond to specific questions, etc. Using these inputs and based on the relevance of each question to essential requirements (D2.1 – User Requirements Definition), high priority questions are selected for user evaluation cycle 1.

iii) *Test case development:* For this activity scenario descriptions defined for each city and applications are used. These scenario descriptions are defined in [1] and easy to understand as well as more specific to user requirements that helps assessing the extent to which urbanAPI applications meet specified or implied user needs and goals.

iv) *Evaluation implementation/exercises:* For each city and application evaluation questionnaire are uploaded on a web-based tool (Figure 3) in order to receive responses from specific city stakeholders on city-based applications. In this regard, a preliminary evaluation exercise for the urbanAPI tools and applications (3DVR and PME), more specifically rule editor with representatives of Vitoria-Gasteiz, Bologna and ASDE (Ruse representative) was performed during training workshop in Darmstadt, Germany on 2-4 July 2013. The idea was to get city partners familiar with online evaluation portal and suggest any improvements. This workshop report is covered in deliverable D3.6 – Beta Applications 1-3. Moreover a report on preliminary usability evaluation for Rule Editor was also prepared with several recommendations. As urbanAPI applications are expected to be web-based, this allows city stakeholders to perform remote evaluation and respond to evaluation questions. Nevertheless, the following evaluation exercises were performed in cities: i)

Vitoria-Gasteiz – from 23rd – 24th September 2013; ii) Bologna –18th October 2013; iii) Ruse – 21st November onwards. For Vienna a full remote training and evaluation exercise was experimented in order to test whether or not remote evaluation can be successfully performed. Please note that for cycle – 2 evaluation, it is planned that remote evaluation exercises will be performed.

v) *Result documentation:* The results are documented and analysis is presented in different reports for each city. These results and analysis include scenario based analysis, criteria based analysis, user response/behaviour analysis. In addition, qualitative feedback based on user and technical perspectives are also included and issues related to training, evaluation and user analysis are discussed. In summary, these reports provide a useful insights about the overall development status, issues and users perception about different application features, which could be used as requirements for the development cycle 2. Here in this document a summary of these results is presented.

All the above activities are shown in the following Figure 2.

Figure 2: The evaluation design activities and potential outcomes

2.2 Evaluation Criteria and Indicators

[Main criteria and indicators]

In [2] it is identified that the main purpose of the evaluation is based on the following criteria:

- **Relevance** – The relevance of the UrbanAPI tools and applications in achieving the ***project goals*** for the stakeholders and for illustrating the goals and potentials of ICT enabled ***participatory urban planning and governance, policy making process, information flow between different levels of governance,*** etc; More specific sub-criteria for relevance can be:

  i) ***Relevance for targeted stakeholders*** *– potential to be relevant for specific stakeholders, e.g. policy makers, urban and transport planners, estate-agents, citizens, etc*
  ii) ***Relevance for real-life use*** *– potential to be relevant in the real-life-use e.g. supporting local planning processes, bottom-up democracy, public engagement, inspiration for behavioural change, etc.*
  iii) ***Relevance for INSPIRE*** *– potential to be relevant with European directives for data harmonisation like INSPIRE*
  iv) ***Relevance to urban planning*** *– suitability of urbanAPI tools and applications for urban and land use planning*
  v) ***Relevance to urban transport planning*** *- suitability of urbanAPI tools and applications for urban transport planning*
  vi) ***Relevance to policy making*** *- suitability of urbanAPI tools and applications for policy making cycle*
  vii) ***Relevance to technology innovation*** *– potential of innovative approaches for the development of urbanAPI tools and applications*
  viii) ***Relevance to targeted objectives*** *– relevance of urbanAPI tools and applications for issue identification, policy analysis, consultation, decision and evaluation in urban planning and land management policy.*

- **Utility and Benefits** – The potential benefits of the application scenarios in ***local planning process, commonalities, inspirations, comparisons, public engagement***, etc; The objective here is to have an e*stimation of benefits by comparing ex ante and post urbanAPI state of the tools and Applications. More specific sub-criteria for Benefits can be:*

  i) ***Qualitative benefits compared to pre-urbanAPI state*** *– determining any benefits to stakeholders and/or end users by using urbanAPI tools and applications. For example, decision support, conflict management, analysis and visualisation using interactive platforms. Similarly, awareness raising, behavioural change, intervention, improvement and acceptance of planning initiatives, information flow between different levels of governance, etc are few other examples.*
  ii) ***Sustainability and Extensibility*** *– degree to which urbanAPI tools and applications are reusable in different contexts and have the ability to be extensible by allowing to add new features*
  iii) ***Technological innovation*** *– ability to generate new knowledge and innovative solution*
  iv) ***Participatory governance*** *– degree to which urbanAPI supports participatory governance*
  v) ***Evidence based urban planning*** *– ability to collect data evidence to support urban planning initiatives*
  vi) ***Evidence based policy making*** *– ability to collect data evidence to support policy making*
  vii) ***Evidence based decision making*** *– ability to collect data and public participation to support collaborative decision making*
  viii) ***Knowledge transfer*** *– ability to share knowledge and experience, awareness raising for behavioural change etc.*
  ix) ***Efficient resource utilisation*** *– ability to make efficient use of resources (e.g. human resources, hardware, software etc)*
  x) ***Compliance to standards*** *– ability of the software to adopt different standards like EU INSPIRE*

for data harmonisation and local municipality platform (hardware and software) and security standards.

xi) ***Valuation*** *– degree to which urbanAPI can provide economic benefits for sustainable application e.g. low cost and expenditure*

The software quality perspective was also considered using the following criteria:

- **Usability** - The capability of the UrbanAPI tools and applications to be ***understood, learned, used and appreciated by the stakeholders/end-user,*** when used under specified conditions; More specific sub-criteria for usability will be directly derived from ISO 25010 characteristics (please see table 1 below).
- **Functionality** - The capability of the UrbanAPI tools and applications to provide ***functions which meet stated and implied needs, or requirements*** when the software is used under specified conditions. More specific sub-criteria for functionality will be directly derived from ISO 25010 characteristics (please see table 1 below).

And finally,

- **Impact –** The overall effect in ***public engagement in policy decision making processes, stakeholder or public behavioural change, awareness, intervention, improvement and acceptance*** of planning initiatives, etc. The overall impact will be based on the combined assessment of benefits, relevance, usability and functionality criteria.

To follow a more structured approach ISO 25010 standard is adopted and its characteristics are used to derive evaluation criteria and sub-criteria. The mapping of ISO25010 characteristics to above criteria is presented in the following table 1.

Table 1: Mapping of urbanAPI application criteria to ISO25010 characteristics

ISO 25010 Characteristics	UrbanAPI Application Criteria
Quality in Use (refers to system quality)	
Effectiveness	Benefits, Relevance
Efficiency	Usability, Benefits, Relevance
Satisfaction *(Usefulness, Trust, Pleasure, Comfort)*	Usability, Benefits
Freedom from risk *(Economic risk mitigation, Health and safety risk mitigation, Environmental risk mitigation)*	Relevance, Benefits
Context coverage *(Context completeness, flexibility)*	Functionality
Product quality (internal and external quality combined as product quality)	
Functional suitability *(functional completeness, functional correctness, functional appropriateness)*	Functionality
Performance efficiency *(Time behaviour, Resource utilisation, Capacity)*	Functionality and Usability
Compatibility *(Co-existence, Interoperability)*	Functionality, Relevance
Usability *(appropriateness recognisability, Learnability, Operability, User error protection, User interface aesthetics, Accessibility)*	Usability
Reliability *(Maturity, Availability, Fault tolerance, Recoverability)*	Functionality
Security *(Confidentiality, Integrity, Non-repudiation, Accountability, Authenticity)*	Functionality
Maintainability *(Modularity, Reusability, Analysability, Modifiability, Testability)*	Functionality and Usability
Portability *(Adaptability, Installability, Replaceability)*	Functionality and Usability

2.3 The Evaluation Web Portal

A web portal is used to record the evaluation results with the objective to enable wider participation from local stakeholders and automated analysis of the evaluation outcomes. This web tool is accessible from http://cccs09.cccs.uwe.ac.uk/~kamran/urbanapi/ and a basic snapshot is presented in Figure 3 below.

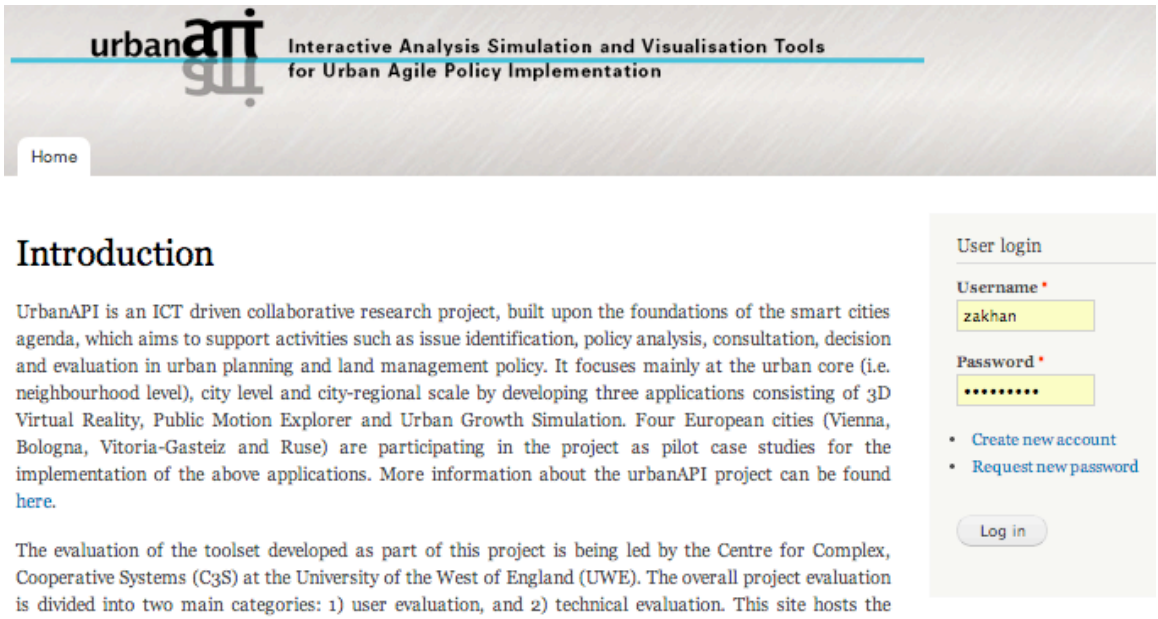


Figure 3: The web tool to collect and analyse evaluation results

3 Application Specific User Evaluation

[This chapter provides summary of evaluation results of individual applications for cities]

This section presents application specific user evaluation results for each case study city. Below in section 3.1 we briefly describe these applications and Figure 4 shows the urban scales at which these applications are applied for each city.

3.1 UrbanAPI applications

3.1.1 3 Dimensional Virtual Reality (3DVR)

[Brief intro about 3DVR application]
UrbanAPI directly addresses the issue of stakeholder engagement and citizen participation in the planning process by the development of enhanced 3D virtual reality (3DVR) visualisation of neighbourhood development proposals. This 3DVR application also enables end users to access and propose amendments to planning proposals using interactive visual interface via the web.

One of the important components used for 3DVR is CityServer3D Admin tool – which is a desktop based application. Rule Editor is developed and integrated with CityServer3D Admin tool and is used to import, export scenario data, crease scenario specific visual scenes, make amendments in scenes, navigate (aerial, pedestrian view) through 3D scenes, compare scenes in 2D and 3D etc.

3.1.2 Public Motion Explorer (PME)

[Brief intro about PME application]
At the city-wide scale, UrbanAPI is developing a public motion explorer (PME) application, a mobile phone location based application using mobile phone location data (also referred as GSM data), that permits the visual representation, and analysis of population distribution and movement patterns across the city which assists planning agencies to explore space attractiveness and carry out mobility analysis.

3.1.3 Urban Growth Simulation (UGS)

[Brief intro about UGS application]
UrbanAPI is developing urban growth simulation (UGS) for city-regions, addressing multiple urban planning challenges including visualisation of planning interventions and assessment of the impact of alternative proposals for urban expansion (and/or shrinkage) in the peri-urban area, and the optimum distribution of residential, employment and associated services.

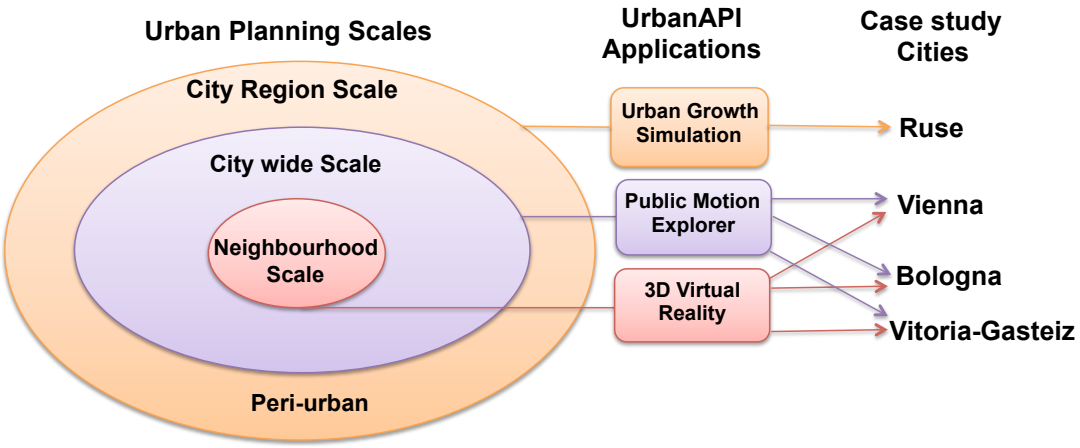


Figure 4: Planning Scales and City specific UrbanAPI Applications

3.2 User Evaluation Cycle-1 Results

[This section presents evaluation results of each application for each city]

In this section we attempt to summarise evaluation results for each application. Detailed evaluation results and analysis is provided in separate city evaluation reports (5 reports) which include detailed scenario based analysis, criteria based analysis, user response/behaviour analysis, technical and user perspectives, qualitative feedback indicating positives, negatives and suggested improvements. Please note that in order to synthesise evaluation results separate justification/reasoning tables are added with scenario based evaluation analysis of all applications in respective city evaluation reports. In these tables it is clarified that which components of the applications (e.g. 3DVR web portal or admin tool) should be used, whether or not functionality exists and users were unable to find it, or what could be the reason of negative response, or why a certain functionality is missing, and what are the future plans for cycle 2.

3.2.1 Evaluation results of 3DVR Application

[This section presents evaluation results for 3DVR application for all participating cities]

This section summarizes evaluation results for all participation cities scenarios. Mainly the final mark values are derived from the evaluation responses, which indicate the areas of strengths and improvements during development cycle 2. The mark value (between 0 and 1) aims to determine whether the overall response is above or below reasonable threshold. The mark values are derived based on the user evaluation responses to variety of criteria based questions such as usability, functionality, benefits, relevance, effectiveness, security, maintainability, etc.

3.2.1.1 Vitoria-Gasteiz Scenarios

[Evaluation results of 3DVR application for Vitoria-Gasteiz scenarios]

Mark Values

Total 6 users participated in the evaluation with different user roles i.e. GIS Experts, Urban Planner, Policy Maker and Other. This section summarizes evaluation results for Vitoria-Gasteiz application scenarios. Both scenarios based and criteria based mark values are derived and presented below in Figure 5.

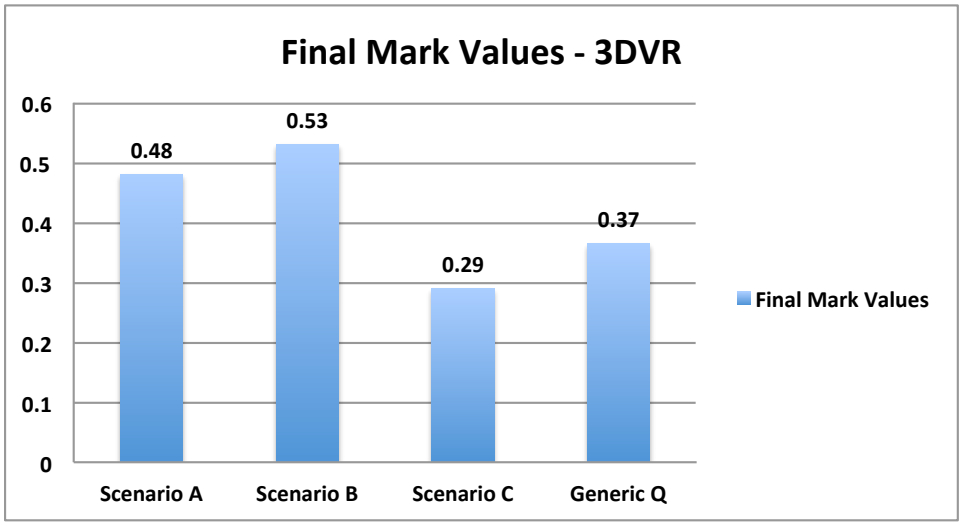


Figure 5a: 3DVR Application Evaluation – Scenario based Final Mark Values for Vitoria-Gasteiz

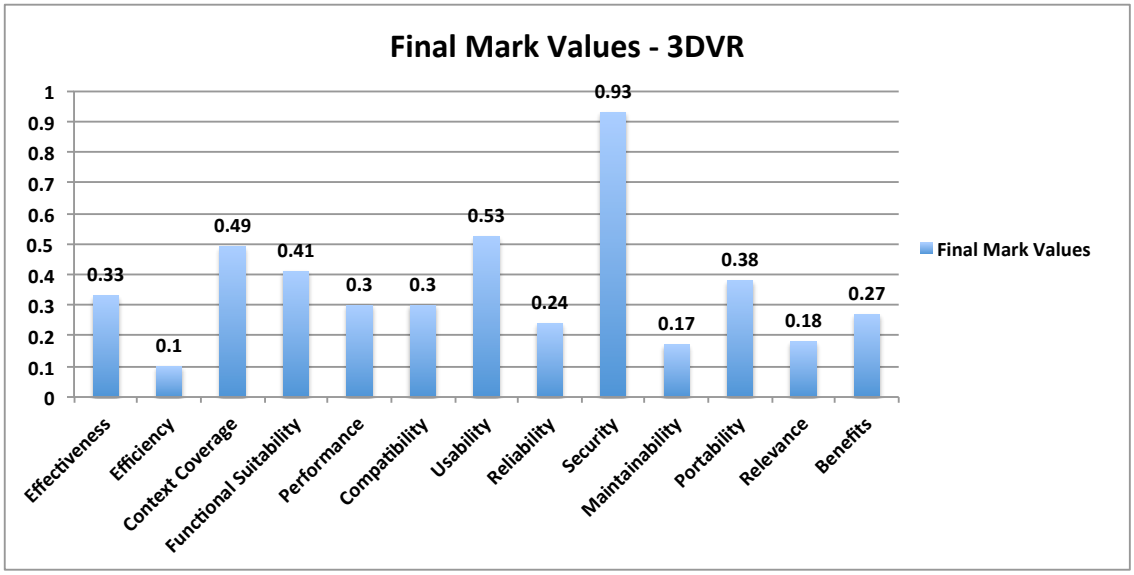


Figure 5b: 3DVR Application Evaluation - Criteria based Final Mark Values for Vitoria-Gasteiz

Figure 5a indicates that results satisfaction is up to 48% for Scenario A, 53% for Scenario B, 29% for Scenario C and 37% for generic questions which can be applied to overall 3DVR application. This clearly indicates that the overall results are about close to the expected threshold (i.e. 50%) for first two scenarios but scenario C needs considerable focus during development cycle 2. Figure 5b indicates that results satisfaction is mostly below the acceptable threshold. For example, functional suitability is only 41%, performance is 30%, compatibility is 30%, context coverage is 49% and this may be the reason that based on the current state of the application evaluators couldn't assess high benefits and relevance 27% and 18%, respectively.

Learning Outcomes

From Technical and Methodological Perspectives, the evaluation results reflect the fact that, especially for the web visualizations, a lot of technical groundwork had to be done which delayed the implementation of functionality for end user. From a technical perspective there are some important issues to consider.

The strongest is obviously the judgement that the 3D visualization is not state-of-the-art as in recent computer games which hinders the citizens to identify their city which may lead to disinterest and frustration. Considering the data situation in Vitoria-Gasteiz there was no sophisticated 3D city model available as for example in Vienna, only basic data like terrain and 2D building footprints. Based on this data situation, where no good data is available the users were taught how to create a simple first "box" 3D city model, which of cause is not as visually compelling as it has no colours, no details like textures etc. Building sophisticated 3D city model is not within the scope of the project as it requires huge investment of time and resources.

The second important aspect is that until now there is only visual and no "calculated" impact assessment possible. The corresponding functionalities are indeed not implemented right now and should be the major focus of the 3DVR application development for the last year.

Another major problem which can be derived from various responses is that the setup of 3D scenarios, especially integrating socioeconomic data, is not satisfying with the current state of development. Especially when several steps have to be followed the users got lost and were not able to find the correct sequence of steps or even were not able to find the appropriate options in the tool. This indicates that the tool should be working more automated and therefore abstract from the technical issues, like uploading files and configuring them afterwards for display. The complexity may be the main reason for the user's negative responses. The complexity should be reduced by common measures found in modern software, for example some simple wizard-like dialogues which guide users could be used for this.

Also the fact that the training sessions took only about 2 hours should be considered, as the experiences show that at least one day could be spent on the usage of the 3D AdminTool. This is supported by the observation that topics that were explicitly demonstrated achieved a better response than those which were not demonstrated which is of course obvious, for example the question if a helicopter or ground navigation is possible.

The above results prove that the overall agile project development methodology is working and we're successfully able to identify limitations in the 3DVR application for Vitoria-Gasteiz scenarios. It is understood that some of the usability related limitations are because of the unavailability of quality 3D data of the city of Vitoria-Gasteiz and others are still work in progress. The above results clearly indicate where development efforts should be spent during cycle 2 to fulfil original user requirements.

From User perspective, the mild responses to the questionnaire –PARTIALLY, NOT SURE, NEUTRAL- are because the users are not sure if the answer to that question is on the City Server or on the online application. Whereas the negative responses to the questionnaire –NO, DISAGREE, STRONGLY DISAGREE- are because users couldn't find that functionality.

3.2.1.2 Bologna Scenarios

[Evaluation results of 3DVR application for Bologna scenarios]

Mark Values

Total 5 users participated in the evaluation with different user roles i.e. GIS Experts, Urban Planner, Policy Maker and General Public. This section summarizes evaluation results for Bologna application scenarios. Both scenarios based and criteria based mark values are derived and presented below in Figure 6.

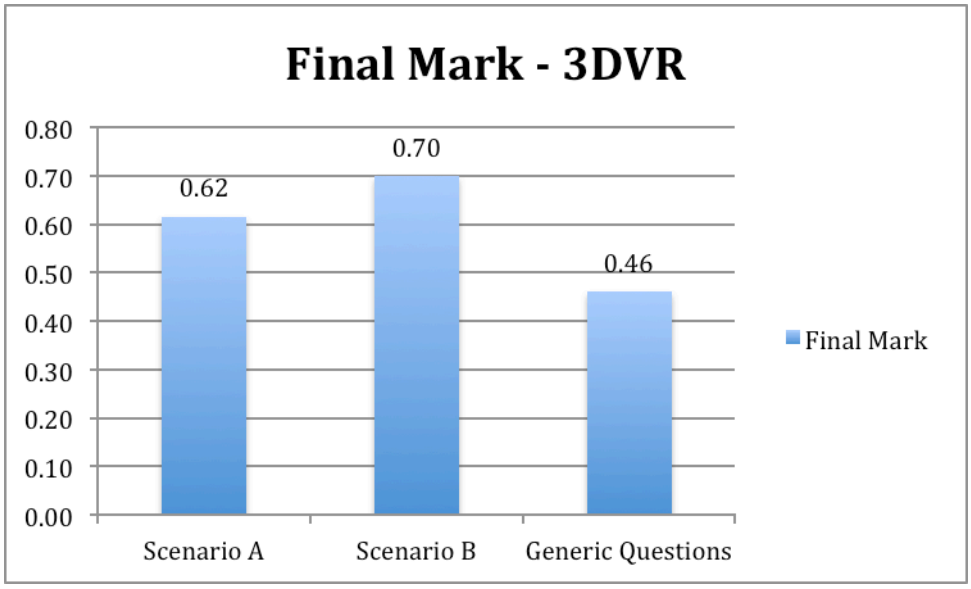


Figure 6a: 3DVR Application Evaluation – Scenario based Final Mark Values for Bologna

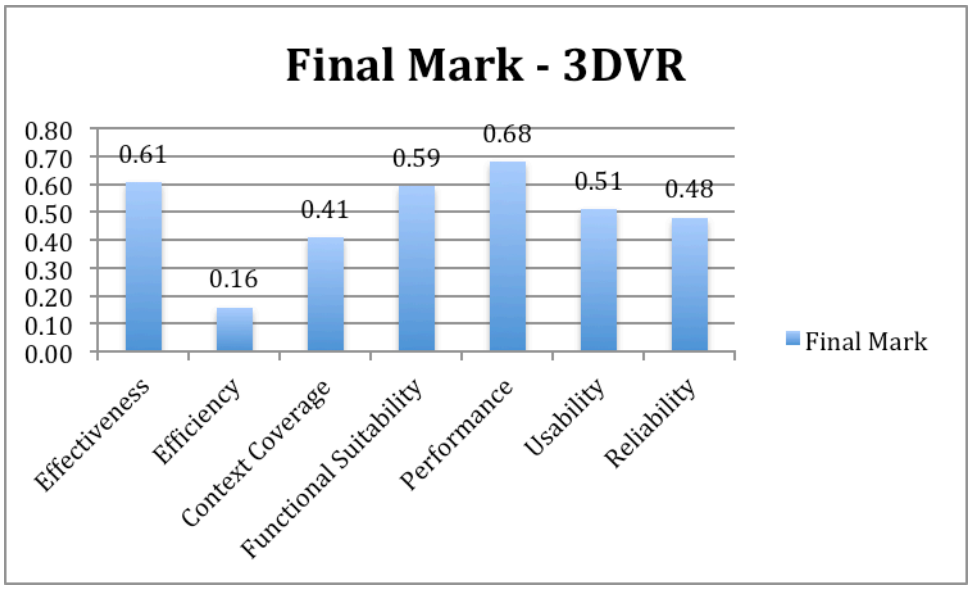


Figure 6b: 3DVR Application Evaluation - Criteria based Final Mark Values for Bologna

Figure 6a indicates that results satisfaction is about 62% for Scenario A, 70% for Scenario B and 46% for generic questions which can be applied to the overall 3DVR application. This clearly indicates that the overall results are above the expected threshold (i.e. 50%).

Figure 6b indicates that results satisfaction is mostly at the acceptable threshold. For example, functional suitability is 59%, usability is 51% and effectiveness is 61%. On the contrary efficiency is only 16% whereas performance is 68%.

Learning Outcomes

From technical perspective, the evaluation results reflect the fact that especially for the web visualizations a lot of technical groundwork had to be done which delayed the implementation of functionality for end user. Generally the feedback seems quite positive and gives the impression that the application is on the right track. Like Vitoria-Gasteiz, the city of Bologna is not maintaining any 3D city model right now. So a basic box like model was created from cadastral data available. It is clear that such a basic model is not as visually compelling as it has no colours, no details like textures etc. Building sophisticated 3D city model is not within the scope of the project as it requires huge investment of time and resources.

An interesting result is about the required training. The fact that the training sessions took only about 2 hours should be considered as one of the reasons of lack of understanding some functionalities. The experiences show that at least one day could be spent on the usage of the AdminTool. This should be further investigated to find out in which way the documentation of tools should be improved as well as how the intuitive use of the application can be improved. This is especially important to citizens and other lay people that should be engaged in public participative planning. It cannot be expected from those people that they spend some efforts in training and learning on such a tool - they rather may not use it at all. The complexity should be reduced by common measures found in modern software, for example some simple wizard-like dialogues that guide users to perform certain functionality could be used for this.

From user perspective, many uncertainties occurred in the answers as it was difficult to evaluate aspects of the applications which where still not available or not fully experienced by using it. These uncertainties are mainly referred to the functionality related questions of the Scenario A. It was difficult for users to find required functionality in the application components in order to respond to specific evaluation questions. Since there were different types of users participating in evaluation, it was difficult to answer certain technical questions as not all the users felt confident in using the application as it was complicated to use. It is suggested to dedicate more time to training, which should be divided into a session with the developers and an internal session (ore more sessions) in which users try the application.

3.2.1.3 Vienna Scenarios

[Evaluation results of 3DVR application for Vienna scenarios]

3.2.1.4 Stakeholder Board Evaluation

[Summary of evaluation results of 3DVR application by Stakeholder Board members]

3.2.2 Evaluation results of PME Application

This section summarizes evaluation results for all participation cities scenarios. Mainly the final mark values are derived from the evaluation responses, which indicate the areas of strengths and improvements during development cycle 2. The mark value (between 0 and 1) aims to determine whether the overall response is above or below acceptable threshold. The mark values are derived based on the user

evaluation responses to variety of criteria based questions such as usability, effectiveness, functionality, benefits, relevance, security, maintainability, etc.

3.2.2.1 Vitoria-Gasteiz Scenarios

[Evaluation results of PME application for Vitoria-Gasteiz scenarios]

Mark Values

Total 6 users participated in the evaluation with different user roles i.e. GIS Experts, Urban Planner, Policy Maker and Other. This section summarizes evaluation results for Vitoria-Gasteiz PME application scenarios. Both scenarios based and criteria based mark values are derived and presented below in Figure 8.

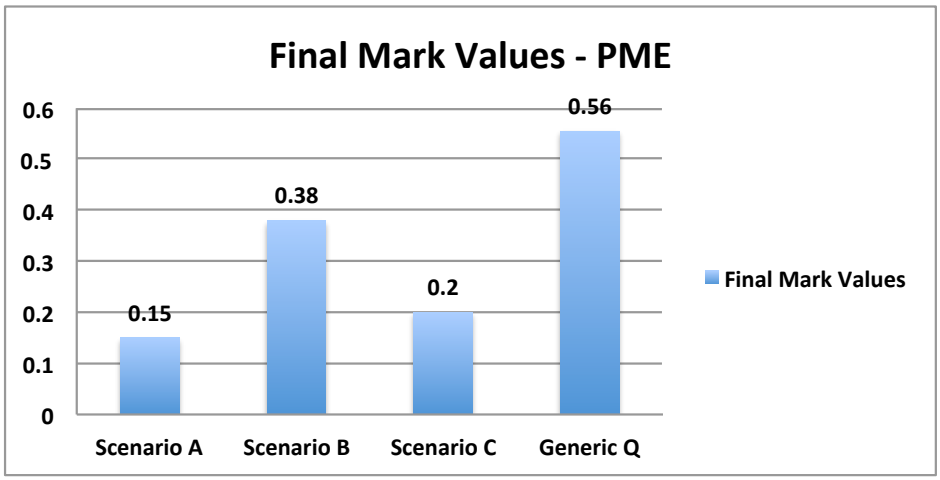


Figure 8a: PME Application Evaluation – Scenario based Final Mark Values for Vitoria-Gasteiz

Figure 8a indicates that results satisfaction is only 15% for Scenario A, 38% for Scenario B, 20% for Scenario C and 56% for generic questions which can be applied to overall PME application. This clearly indicates that the overall results are below the expected threshold (i.e. 50%).

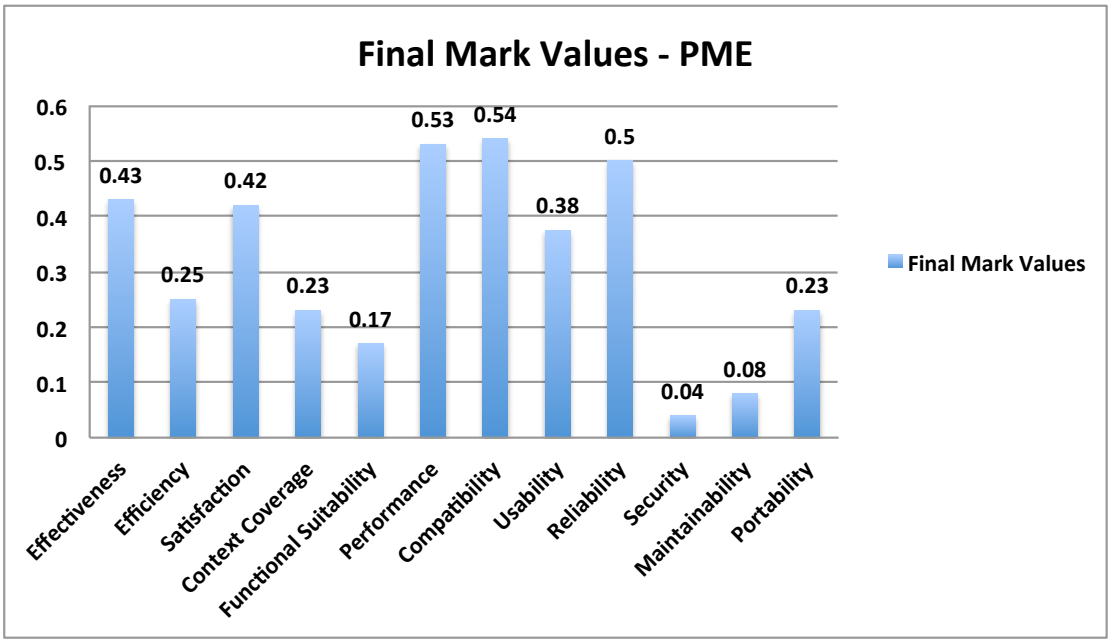


Figure 8b: PME Application Evaluation – Criteria based Final Mark Values for Vitoria-Gasteiz

Figure 8b indicates that results satisfaction is mostly below the acceptable threshold. For example, functional suitability is only 14%, usability is 34%, satisfaction is 42% and this may be the reason that based on the current state of the application evaluators couldn't assess high effectiveness and efficiency, 43% and 25%, respectively. On the contrary, performance, compatibility and reliability are above 50%.

Learning Outcomes

The data provided by Telefonica for the Vitoria-Gasteiz area were data on mobile phone activity events (mobile phone call starts, text message deliveries) but no call end or hand over location and time stamp information between telephone cell towers exist. This dataset allows only aggregating activity event locations by summing up the activity observation numbers within hourly time spans. Investigations of the cell tower location pattern (provided by Telefonica) show that the cell towers are not evenly distributed over Vitoria-Gasteiz. There is a high density in the centre and a low density in the Western residential areas with some concentration along the mountain ridges in the North which makes it necessary to define larger cells for aggregating the mobile phone user distribution in the residential areas and the outskirts. In the centre area smaller cells have been defined to allow documenting a more detailed mobile phone user distribution.

This limitation – only observing the static distribution of mobile phone users during the day allows only limited use of the functionalities of the PME tool for Vitoria- Gasteiz. Functionalities allowing exploring motion pattern dynamics for every origin area as applied by using the Vienna data could not be used in this case. The test users in Vitoria- Gasteiz recognised this big disadvantage which makes the tool less appropriate for traffic planning purposes.

Thus in the meantime negotiations between the City of Vitoria- Gasteiz and the mobile telecom provider Telefonica have taken place and AIT has been informed that a new dataset including phone call end and cell-handover location and time stamp will be provided for the area. This will allow – not now but in the future when this data will be provided - to identify and aggregate motion patterns aiming to use the full functionality of the PME tool.

From technical perspective, the developer of the PME is well aware of the above limitations because the GSM data of Telefonica were not as detailed as the GMS data provided by A1 in Austria. For example, the Telephonica data included only user data for completed and uncompleted calls, SMS traffic and data traffic each of the files holding information of start and end point of a call e.g., so following a trace of a user over a whole day was not possible (as it was the case for Vienna e.g.). Additionally the density of GSM cell antennas in Vitoria-Gasteiz is quite low (again compared to Vienna) which may not allow for a much finer resolution of the map. Nevertheless in cycle 2 it is planned to integrate more detailed data from Telephonica which the city of Vitoria-Gasteiz is trying to collect from Telephonica in order to try to improve the possibility of mapping diurnal movements within the city.

Regarding the functionality of the PME Vitoria-Gasteiz: Some of the functionality present in the Vienna application were not present, since the data provided by Telephonica did not contain the information needed to e.g. map motion traces, but since the development process is not finished, yet, the missing functions like download of O-D matrices and the possibility to load additional data are planned to be implemented in cycle 2 – provided that better data will be available.

From methodological perspective, the above results prove that the overall agile project development methodology is working and we're successfully able to identify limitations in the PME application for Vitoria-Gasteiz scenarios. It is understood that the major limitation is due to unavailability of quality and detailed GSM data received from Telefonica and the size of the city itself where cell towers density is not sufficient to capture accurate mobility patterns, present accurate population distribution in smaller cell sizes. The above results clearly indicate where development efforts should be spent during cycle 2 to fulfil original user requirements with the assumption that more detailed GSM data is available.

From User perspective, the negative/unsure answers are provided because users couldn't find the required functionality for Vitoria-Gasteiz, though available for other cities. In order to synthesise evaluation results separate justification/reasoning tables are added with scenario based evaluation analysis of in city evaluation reports, where it is clarified whether or not functionality exists and users were unable to find it, and what are the future plans for cycle 2.

3.2.2.2 Bologna Scenarios

[Evaluation results of PME application for Bologna scenarios]

Mark Values

Total 4 users participated in the evaluation with different user roles i.e. GIS Experts, Policy Maker and Other. This section summarizes evaluation results for Bologna PME application scenarios. Both scenarios based and criteria based mark values are derived and presented below in Figure 9.

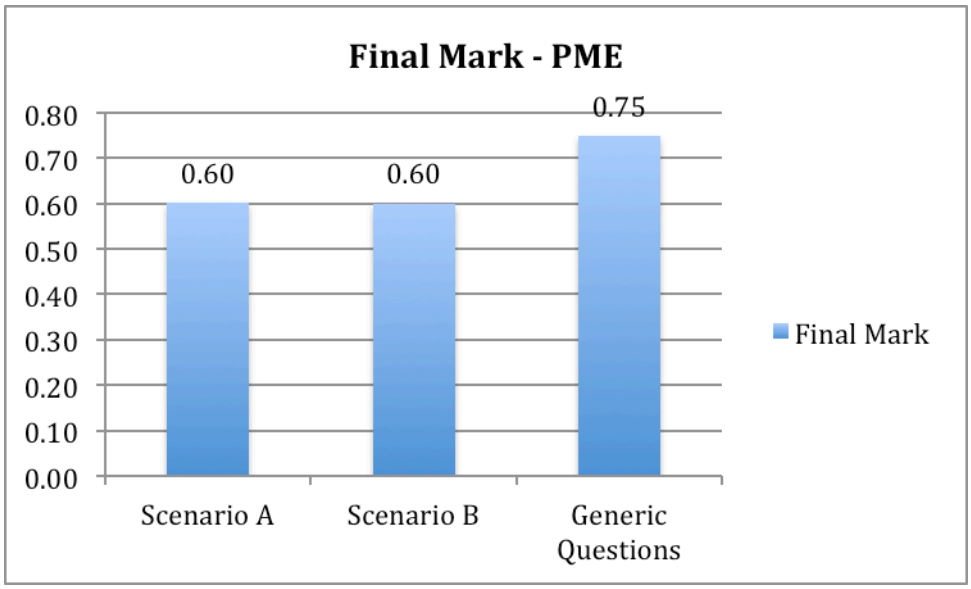


Figure 9a: PME Application Evaluation – Scenario based Final Mark Values for Bologna

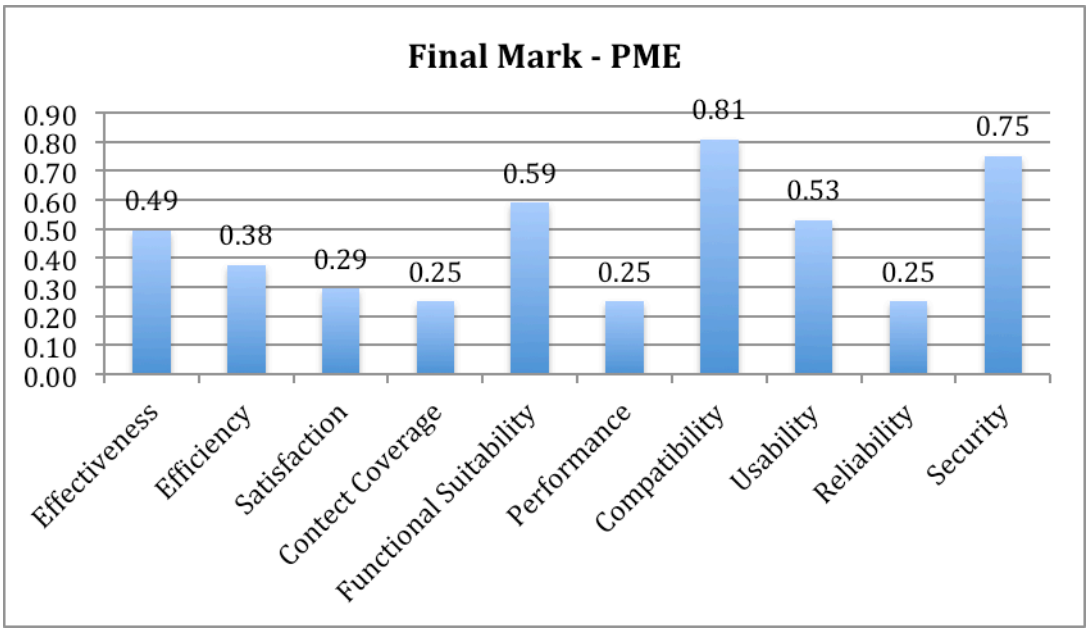


Figure 9b: PME Application Evaluation – Criteria based Final Mark Values for Bologna

Figure 9a indicates that results satisfaction is about 60% for Scenario A, 60% for Scenario B, 75% for generic questions which can be applied to the overall PME application. This clearly indicates that the overall results are above the expected threshold (i.e. 50%).

Figure 9b indicates that results satisfaction is mostly below the acceptable threshold. For example, functional suitability is 59%, usability is 53% and effectiveness is 49%. On the contrary efficiency is only 38% whereas performance is 25%.

Learning Outcomes

From technical perspective, the data provided by TIM for the Bologna Centre Area contains pre-aggregated data by summing up the mobile phone call observations within 15-minute time spans. Further investigations of the cell tower distribution (not provided by TIM but available via www.opensignal.org) show that the cell towers in and around Bologna centre turn out to be distributed quite unbalanced which requires spatial interpolation of the active mobile phone locations. There exists no information on this, but this must be carried out either through observed signal strength or/and through signal direction.

This interpolation steps let assume particular uncertainties in the mobile phone user location information which cannot be evaluated as we do not have access to the original data. The limitation – only observing the static distribution of mobile phone users during the day - allows only limited use of the functionalities of the PME tool for Bologna. Functionalities allowing to explore motion pattern dynamics for every origin area as applied by using the Vienna data could not be used in this case. Obtaining more useful data from TIM might be not possible due to strict data protection laws in Italy which do not allow citizen related data to leave the country.

From User perspective, this application was perceived as more easy to use, and its functionalities were clearer to the users. Some uncertainties in the answers (mainly the ones belonging to Category A, Scenario B) are due to the lack of knowledge of the requirements or to the technology behind the application. In this case, the second cycle evaluation should focus on technical specs related to the requirements met and not dealt during the training.

3.2.2.3 Vienna Scenarios

[Evaluation results of PME application for Vienna scenarios]

3.2.2.4 Stakeholder Board Evaluation

[Summary of evaluation results of PME application by Stakeholder Board members]

3.2.3 Evaluation results of UGS Application for Ruse

[Evaluation results for UGS application for Ruse scenarios]

3.2.4 Evaluation results of Rule Editor

Rule Editor is an integrated component of CityServer3D Admin Tool. Deliverable *D3.3: Rule User Interface Elements* - that covers most of the user interface specifications. However, these specifications are mostly derived based on the state-of-the-art literature review and do not have explicit user defined requirements for functionality and usability. As a result, the Rule Editor evaluation report of user workshop in Darmstadt, Germany on 2-3 July 2013 (D3.6 – Beta Applications 1-3) recommended defining explicit functional and non-functional requirements so that a more structured evaluation could be carried out for functionality and usability.

It is understood that various recipes are developed according to user requirements specifications - defined in Deliverable D2.1: User requirements definition. However, there is no traceability matrix

from requirements to Rule Editor features/recipes defined in D3.3 or in any other WP3 deliverable, which makes it difficult to test or validate functional suitability and/or completeness of Rule Editor against specified requirements. For instance, without identifying such requirements, we cannot test which conditional operators should be implemented and whether or not such operators are already available in the current release of Rule Editor. In the absence of such a matrix or explicit rule editor requirements, the user evaluation mostly covers usability aspects (Category B and generic questions) and to a limited extent functionality (Category A - based on the contents of deliverable D3.3 – version 2 - available on urbanAPI intranet on 1st September 2013) for Vitoria-Gasteiz and D3.3 – version 3 – available on urbanAPI intranet on 1st October 2013) for other cities. The detailed analysis is covered in city-based reports. Below we summarize the evaluation results.

Final Mark Values

[Evaluation results of Rule Editor for Vitoria-Gasteiz]

Total 6 users participated in the evaluation with different user roles i.e. GIS Experts, Urban Planner, Policy Maker and Other. For Vitoria-Gasteiz, the final mark values (between 0 and 1) for all categories are depicted in Figure 11. It indicates 75% satisfaction for category A questions, 60% for category B and 58% for generic usability questions.

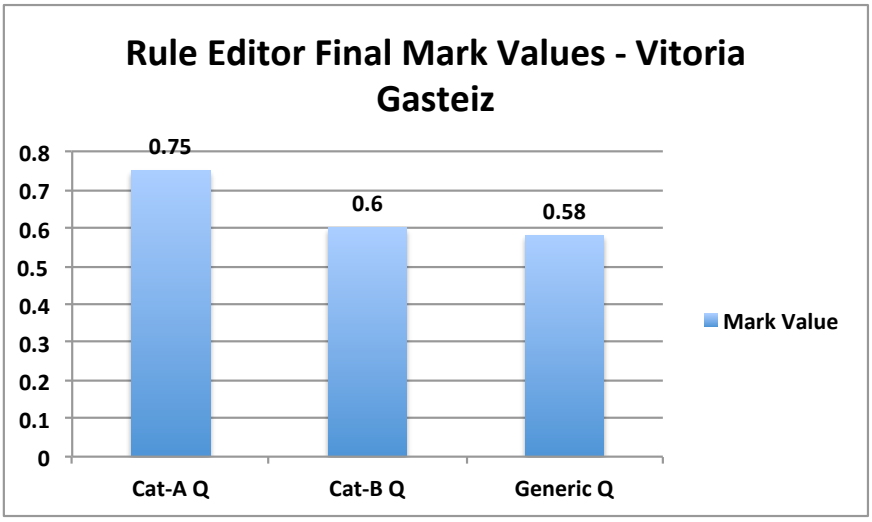


Figure 11: Rule Editor Final Mark Values – Vitoria-Gasteiz

[Evaluation results of Rule Editor for Bologna]

Total 2 users participated in the evaluation with different user roles i.e. GIS Experts and Policy Maker. The final mark values (between 0 and 1) for all categories are depicted in Figure 12. It indicates 76% satisfaction for category A questions, 53% for category B and 45% for generic usability questions.

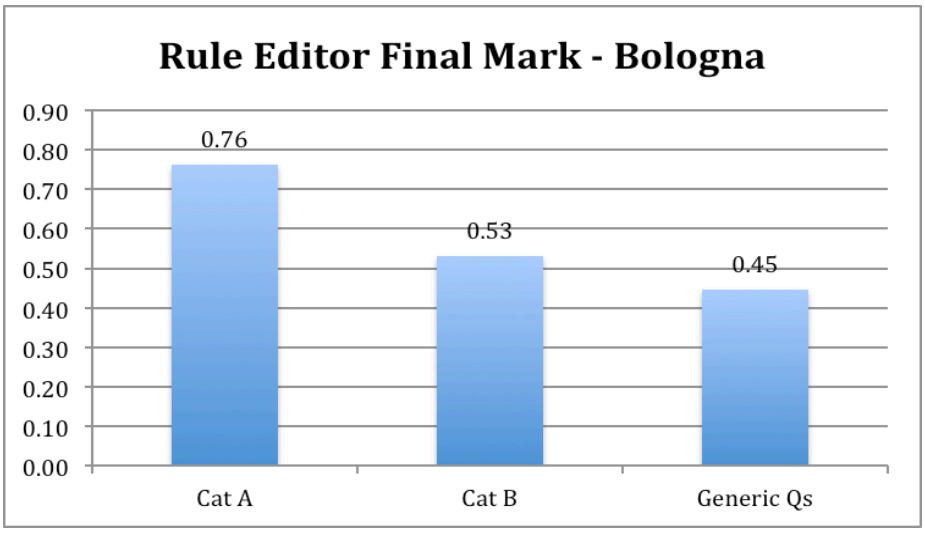


Figure 12: Rule Editor Final Mark Values – Bologna

[Evaluation results of Rule Editor for Vienna]

Learning outcomes

From Technical perspective, the evaluation shows that the rule editor helps users automate processes and that it is relatively easy to use. However, at the current state the editor only provides a small set of recipes that are targeted to use cases in urbanAPI specifically. Due to that, the editor is not yet flexible enough to be used in more advanced scenarios. Consequently, one of the next steps will be to implement more recipes that target a wider range of geospatial/urban use cases.

Additionally, users pointed out that there has to be some introductory material for the rule editor in order to make it easier for new users to understand its concepts and functionality. In the future we will include tutorials based on the 'first steps' feature that we already have in the AdminTool. These tutorials will guide the user through a simple example in order to make them familiar with the rule editor. Finally, the usability of the individual recipes can be further improved by implementing additional features such as greater/less comparison for metadata, selection of extents out of a 2D map, etc.

From Vitoria-Gasteiz user perspective, the answers to questions in category A are more distributed and are more variable as compared to category B or generic usability questions. It is found that the end users understand the functionalities and the purpose of the rule editor (creation of automated processes, specify different data sources…) but they are not sure how these functionalities can meet their needs. The final evaluation about satisfaction is near neutral perhaps because rule editor is fine but this was not one of the requirements defined by the users at the beginning of the project. From the user perspective, there is a need for more training to understand the rule editor and to learn to use it.

From Bologna user perspective, the rule editor is perceived as something behind the application and users were neither aware nor prepared to answer to the questionnaire, especially general public type users. And, therefore, the results are based on the responses of two expert users (one GIS expert and

one Policy maker). Users found the questions difficult to understand and relate to functionalities shown during training. A further training is strongly needed and it should explain more about the rule editor regarding both the theoretical and practical aspects.

4 User Analysis, Evaluation Issues and Lessons Learned

[Evaluation issues and user response analysis. Also, this chapter presents lessons learned and any future directions e.g. new requirements]

This chapter briefly presents the evaluation and training issues and summarises user behavioural analysis as lessons learned. With respect to suitability of CIM based evaluation methodology and mechanism used to evaluate the urbanAPI application, the following issues are identified regarding evaluation outcomes:

i) The main issue raised was whether or not the participation of different types of stakeholders having different levels of expertise can effectively respond to evaluation questions. In this regard, Vitoria-Gasteiz evaluation results are analysed based on level of expertise of users in different domains to test the hypothesis: '*the CIM based user evaluation can identify limitations and strengths of the urbanAPI applications regardless of different types of evaluators with different level of expertise*'.

ii) The second issue identified was related to relevance of the evaluation questions to specific application features and the number of questions presented to evaluators.

iii) The third and final issue is related to appropriate alignment of evaluation responses on to aspects beyond to control of application implementation e.g. quality and/or unavailability of data to implement certain feature that can satisfy certain requirements.

These issues are briefly discussed and answered using different examples of city evaluation exercises and a detailed analysis is provided in individual city evaluation reports.

4.1 Evaluation Issue 1: User Response Analysis and Training

As stated in chapter 1 that the main objective of the cycle 1 user evaluation is to assess the extent to which user requirements are fulfilled by the 3DVR and PME applications at the end of cycle 1 so that any limitations or new requirements can be identified and dealt with during cycle 2. This means the right intended audience for the evaluation could be those users who are either interested or aware of urbanAPI project and somehow have the basic understanding of user requirements of city scenarios. This can help all end users to make evaluation responses in the right context.

In all the evaluation exercises there has been participation from different types of users (e.g. Urban planners, GIS experts, Policy makers, General Public, Other etc.) with level of expertise as beginner, intermediate and experts – belonging to primary, secondary or tertiary stakeholder groups.

Such variation of roles and level of experience is useful for the evaluation as it helps in performing user response analysis from different perspectives. For instance, in Vitoria-Gasteiz evaluation exercise, the evaluators have different level of expertise (beginner, intermediate, expert) and roles that may result in variations of evaluation questions responses The overall responses to different scenarios by different users help to analyse the extent to which the level of expertise of evaluators affects the overall evaluation results. In this regard, all the responses are analysed for each scenario of 3DVR and PME applications and Rule Editor. Ideally, a greater sample set by involving more evaluators could have been useful to generalise the results but evaluators (2 experts, 1 intermediate and 3 beginners) can be considered as indicative analysis depicting the variation of responses based on the level of skill set. It is important to note that the level of expertise are covered by different types of stakeholders and domain of expertise (i.e. roles) that helps in collecting evaluation responses from different domains

which makes it easy to represent results in three groups – i.e. expert, intermediate and beginners as below:

- For *Expert* group, the domain is 'Urban Planner' and 'GIS Expert' and all participants are primary stakeholders i.e. main operators of the application.

- For *Intermediate* group, the domain is 'Policy maker' and participant is as tertiary stakeholder i.e. functional or political beneficiaries of the system.

- For *Beginner* group, the domain of expertise is 'GIS Expert' and 'Other' and all participants are primary stakeholders.

- There is no secondary stakeholder (i.e. indirect beneficiaries of the output of the applications) and general public, transport planner, Architect, IT expert etc involved in the cycle 1 evaluation process.

In summary, the following charts Figure 14 and Figure 15 represent average sum of all the responses to category A (functionality and benefits) and B (usability, benefits and relevance) questions, provided by different evaluators. The analysis of the results indicate that evaluators with level of skillset identified as 'Beginners' responded to questions positively as compared to 'Experts' and/or 'Intermediate' category users. This indicates that for beginners such an application can be highly useful where as for experts features of the application can be further improved. The variations in responses can be attributed to variety of reasons e.g. 'Neutral' response can be due to not fully understanding the functionality during training, or 'strongly disagree' response due to not been able to perform certain functionality, or 'Agree' response because users fully agree on the usefulness of the application based on specific features.

Based on the analysis of user responses it can be concluded that the level of expertise for evaluation of urbanAPI applications for Vitoria-Gasteiz does not affect much the overall scenario based and criteria based results but some variations are noticed such as peak 'Neutral' response in Figure 15b. This proves the hypothesis that *the CIM based user evaluation can identify limitations and strengths of the urbanAPI applications regardless of different types of evaluators with different level of expertise* to a certain extent.

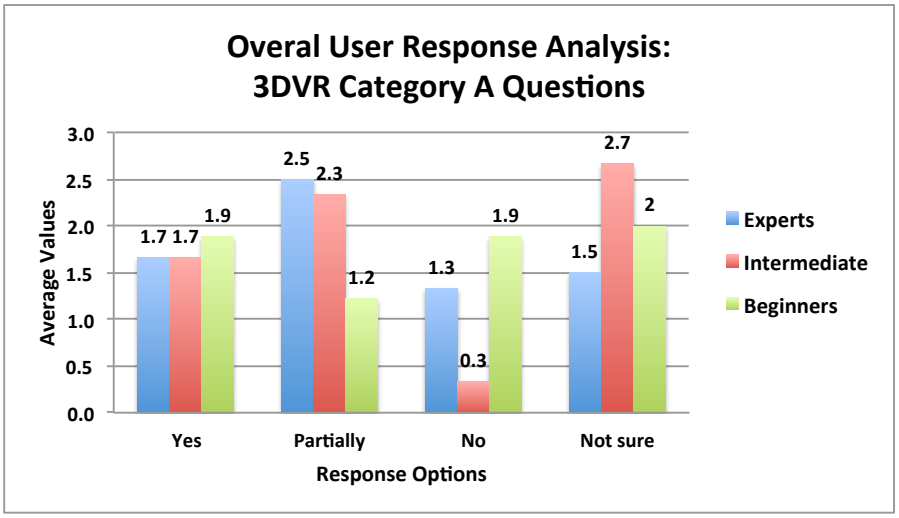


Figure 14a: Overall User response analysis: for all Category A Questions

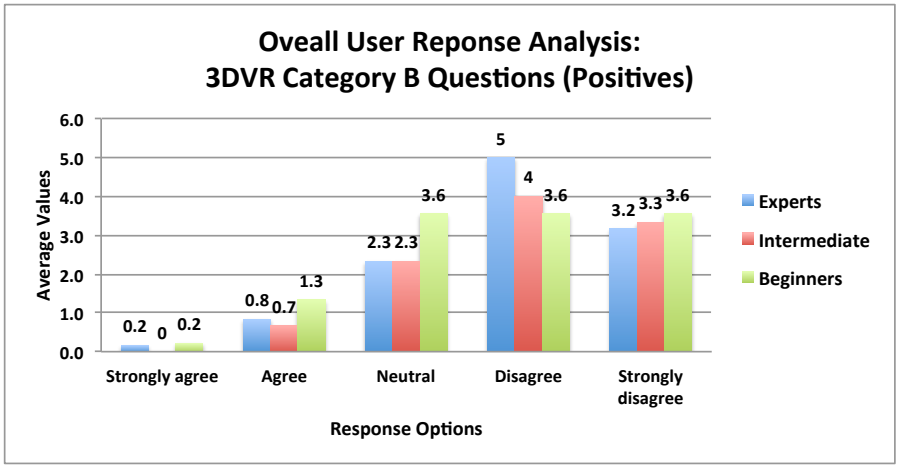


Figure 14b: Overall User response analysis: for all Category B Questions (Positives)

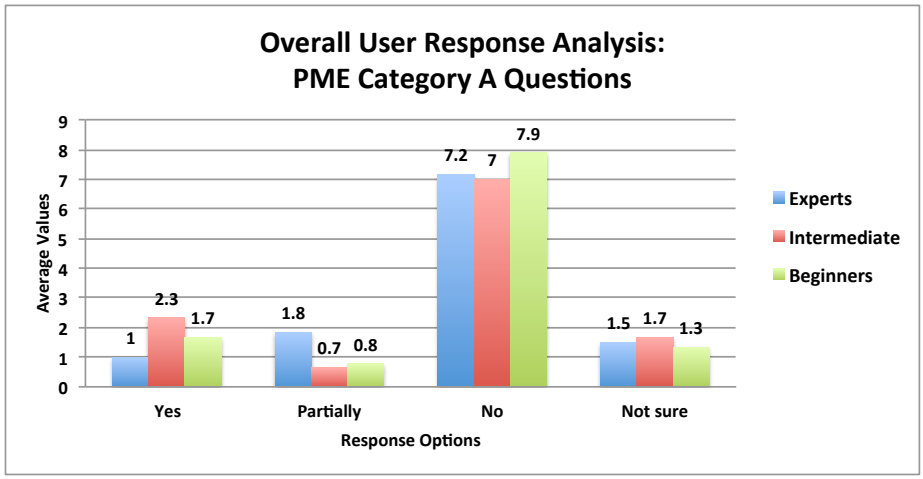


Figure 15a: Overall User response analysis: for all Category A Questions

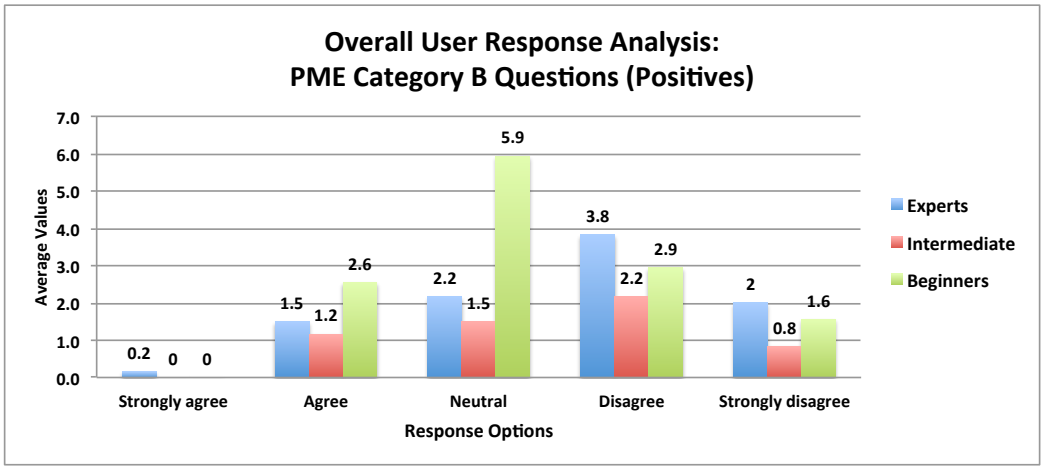
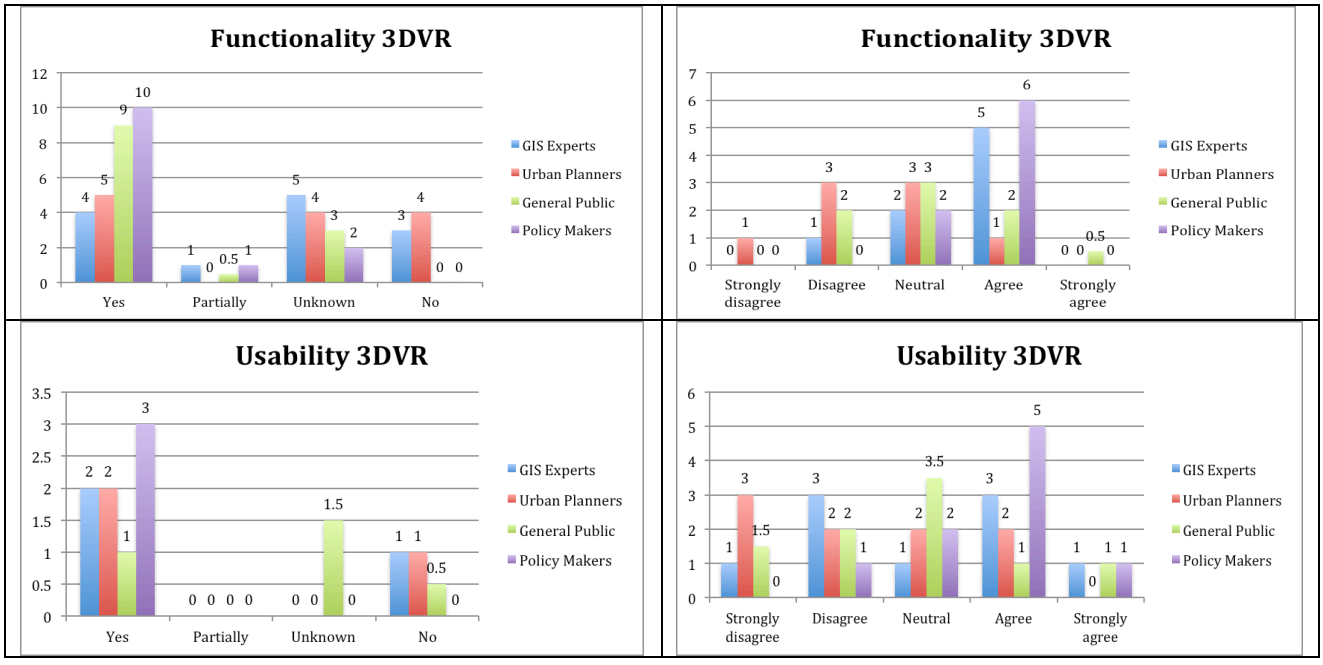


Figure 15b: Overall User response analysis: for all Category B Questions (Positives)

The main reason for such results can be attributed to the amount of training provided prior to evaluation and allowing evaluators sufficient time (about a week) to respond to evaluation questions.

The online evaluation portal for urbanAPI applications proved to be really useful as it allowed evaluators to record their evaluation results in their own time after testing the application features. This also helped evaluation team to manage the evaluation process and compile evaluation results.

Based on the experience of Vitoria-Gasteiz evaluation and lessons learned, Bologna evaluation was planned with the objective to cover basic training needs and then allow end users to perform the evaluation in their own time and respond to evaluation questions using online portal. In case of Bologna, we also performed user analysis based on their roles rather than level of expertise. In addition, we also performed user analysis for different criteria. The different types of users participated in the Bologna evaluation were i) one GIS expert, ii) one urban planner (only for 3DVR application), iii) one policy maker, and iv) two members of the general public. All in all that makes two primary users and three tertiary users but no secondary users as defined in the deliverable D5.1 – Assessment Methodology. The results from the evaluation responses for the various applications including the rule editor are prepared and analysed. Even though the results shown in Figure 16 for 3DVR and Figure 17 for PME give us some idea about the usefulness of the applications for various types of users, the number of users participating in the various role categories is not enough to generalise these results. Therefore, caution must be exercised when drawing conclusions from the results.



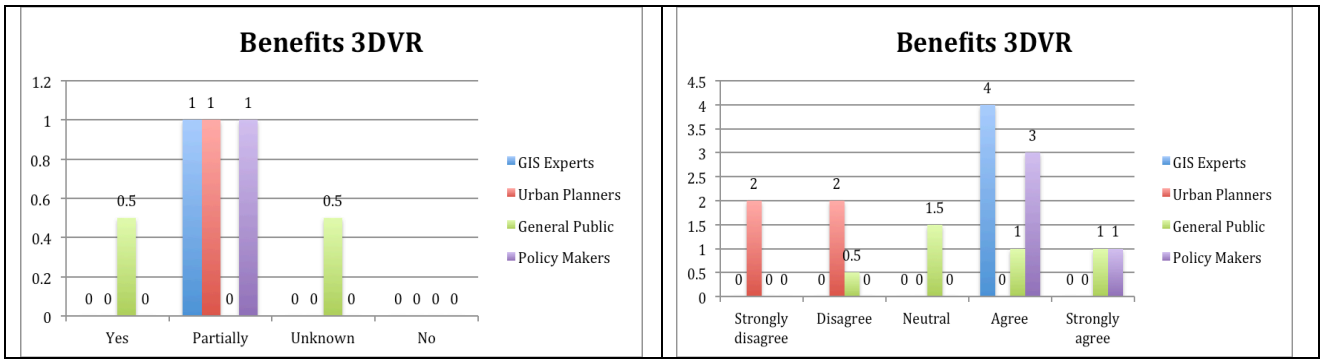


Figure 16: Overall User response analysis: Criteria based 3DVR Application

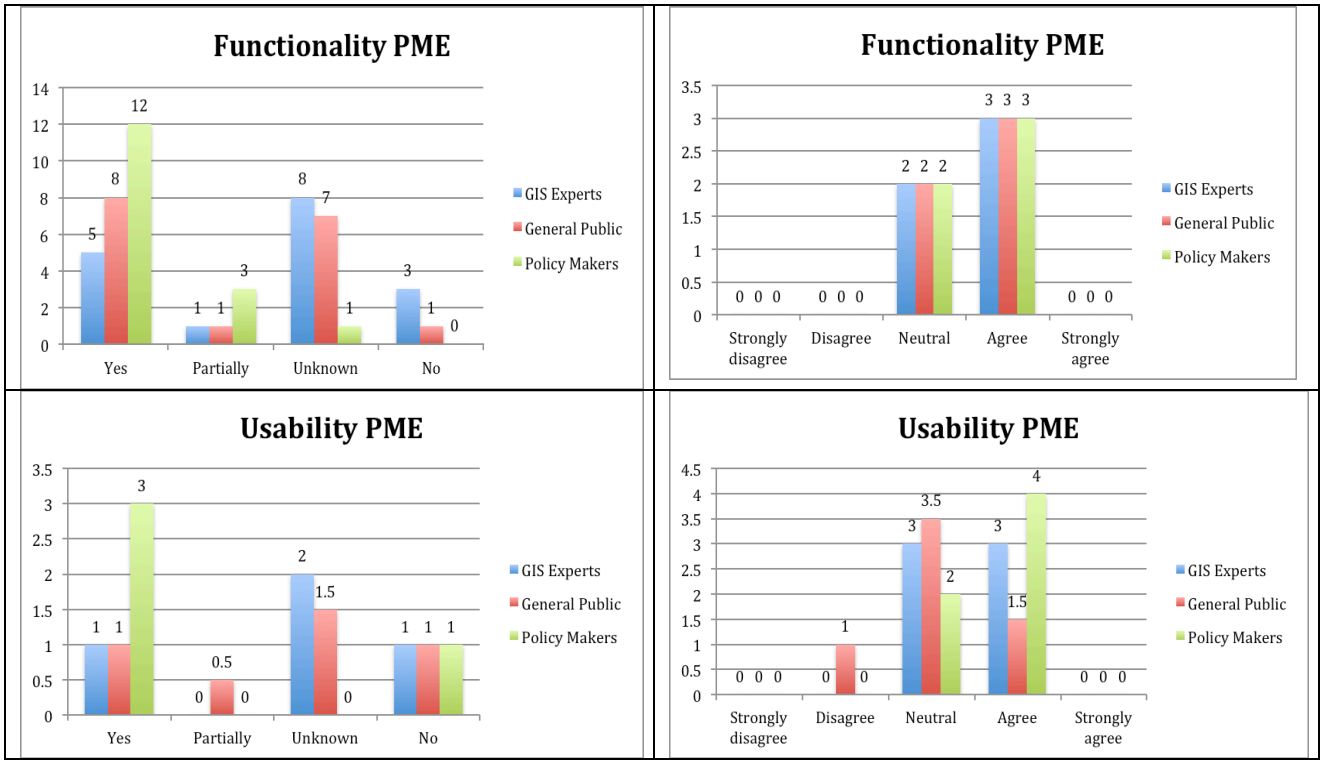


Figure 17: Overall User response analysis: Criteria based PME Application

In order to fully support the Bologna evaluation exercise and not fully aware of the skillset of the all end users the training material (videos) and hand-on training was delivered with aim to cover maximum possible features of the applications so that end users can respond to evaluation questions properly. However, this raised two issues. First, not all the participants were the right users to get training of detailed and complex features of the applications e.g. Admin tool and Rule editor is not intended for general public type of users rather these tools are for domain expert users. For general public there is limited number of features available (e.g. web portal – scene navigation, annotation and object placement) in the applications and hence dealing with more technical features such as loading data, changing properties, applying rules were somehow beyond their capacity and expertise. The second issue is related to the detailed extent of training covered in limited time (2 hours per application), which seems to have overwhelmed end users and consequently not all the application operations could be performed, though are implemented and available in the applications. Similarly, the visualisation of the PME application is more intended for domain experts such as urban and transport planners, policy makers or business organisations such as estate agents. This suggests that end users should be encouraged to get familiar with the applications and features by using the training material (e.g. videos) prior to such hand-on training and evaluation exercise.

In the above context, general public type user category was not in a position to provide meaningful feedback for all functionality-related technical questions. This fact is evident by the statements like, *"Questions are too technical",* and *"I don't know the application in detail".* One user also commented, *"Questions are not so clear, some are technical and too specific".* However, expert users also felt not confident due to lack of training time and not able to find required application features and answer evaluation questions. This clearly suggests that evaluation users, application features and training and evaluation, need to be properly aligned in order to handle the above issues in a better way.

One important lesson learned this exercise was that to make a clear distinction between i) evaluation, ii) training, and iii) availability of application features. The sole purpose of the evaluation is to discover that whether or not applications fulfil the required objectives and satisfy the user needs and requirements. Therefore, the questions asked are directly related to original requirements.

The training of applications is a separate activity that helps end users to fully understand the functionality of the applications and indirectly helps them to provide evaluation responses. However, the amount of training needed to fully understand these applications vary between different types of users. Similarly the mechanisms used for training is also subject to different time constraints for city partners and the suitability of training material provided by technical partners. On the one hand technical partners prepared number of training videos explaining how to use specific features of applications and shared with city partners well in advance (two weeks) of evaluation exercise but

these training videos could be further improved by presenting how to perform a specific application scenario which could be useful for city partners and evaluation purpose. In addition, hand-on training sessions were provided but there was not enough time to go through all features of the applications and explain their relevance and benefits in the context of urban planning, decision making, policy making and stakeholder engagement. On the other hand it is understood that training videos shared prior to evaluation exercise could be used by local stakeholders to establish basic understanding of the urbanAPI applications and get them prepared for hands-on training which could be more focused on city specific scenarios. The amount of training material/information about application features and how to use them appeared to be of great interest but was too much to handle in one day and remember later. This suggests that the training approach needs to be improved with continuous but flexible engagement between technical and city partners.

Nevertheless, the responses to evaluation questions mainly depend whether or not the actual functionality is implemented and easy to perform by the evaluators by their own. This is what exactly the user evaluation cycle 1 was intended to do. And, consequently, evaluation results can be interpreted from different perspectives –i.e. scenario based and criteria based. But these results should not be interpreted without going through technical perspectives and justification provided in associated tables in each city evaluation report.

4.2 Evaluation Issue 2: Quality and Number of Questions

Regarding the second issue, the fine granularity of questions and high number of questions in multiple choice questions format enable evaluators to validate directly whether or not the original requirements have been fulfilled by the application. The number of questions for each application are directly proportional to number of application scenarios and functional/non-functional requirements identified by city stakeholders in deliverable D2.1 – User requirements definition. Due to the complex nature of urbanAPI applications and requirements, there is no simple way to test/validate application features against stated requirements. Other approaches such as qualitative survey type evaluation would have missed many aspects which these applications were supposed to implement. Therefore, both quantitative and qualitative aspects are covered. For quantitative evaluation, the high priority questions (using original requirement priority and its respective evaluation question(s) priority identified by city stakeholders) were selected and lower priority questions are eliminated to keep the number of questions within a reasonable limit that can effectively help in identifying the application strengths and limitations. The results based on such a technique provide enough data to perform quantitative analysis as provided in the city based evaluation reports. Furthermore, qualitative responses are also invited from the evaluators (category C questions) where explicit details of application strengths (positives), weaknesses or limitations (negatives) and improvements are provided by the evaluators. These qualitative feedbacks help to interpret the evaluation responses. After Vitoria-Gasteiz evaluation, we introduced special text fields for each scenario evaluation questionnaire requesting evaluators to provide any additional information that can help in interpretation of their negative responses. The comprehensive and thorough nature of CIM2 helped to identify the application limitations and strengths and this is exactly what the CIM2 was intended to do.

However, users qualitative feedback indicates that some users experienced difficulties in relating questions to specific application features/functionalities. For example, end users found it difficult in figuring out whether the required functionality exists in 3D Admin Tool, Rule Editor or Web portal. This may be attributed to lack of detailed training of the application and hence suggests to provide an index table of all the application features and appropriate user manual explaining how to perform a certain functionality.

4.3 Evaluation Issue 3: Synthesizing Evaluation Responses with Known Limitations - Mainly Due to External Factors

The final issue is not straightforward to handle because user evaluation cycle-1 attempts to clearly identify that what city administrations actually need and what can be certainly fulfilled by the urbanAPI applications against specified requirements. The benefit of applying applications on two cities somehow allows handling the issue by demonstrating that feature can be implemented if required data is available. For instance, in PME application, Origin-Destination matrices are available for Vienna but not for Vitoria-Gasteiz and Bologna due to limitation in GSM data from different mobile service providers. The real concern is that such limitation can not be appropriately reflected in the evaluation responses of each city. For instance, majority of functionality for PME application is subject to availability of fine grained and detailed GSM data. Similarly, high quality 3D model results for a city is subject to quality, fine granularity and details of original city data provided by city administration. For example in Vitoria-Gasteiz and Bologna, data is 2D and without texture/façade of buildings and hence not a limitation of the application itself. But the evaluation responses indicate that such a limitation is recognised as limitation of the application itself. The evaluation results for Vitoria-Gasteiz and Bologna do not implicitly accommodate such limitations which are not due to implementation features. Whilst evaluation approach is attempting to identify the extent to which the original requirements of Vitoria-Gasteiz have been fulfilled, it does not implicitly identify and reflect such data related issues and limitations in the evaluation results. A workaround used in the city evaluation reports is to detail such limitations explicitly as *user perspective* and *technical perspective* for each application. In addition, justification/reasoning for each evaluation question response and future plan is covered explicitly and presented with evaluation graphs/charts. This clearly suggests that the evaluation results (charts) should be seen in the context of the above perspectives to have synthesised outcome of the evaluation. This will, to some extent, help all concerned stakeholders in identifying what additional considerations are needed during cycle 2.

[Insert User Analysis and Evaluation issues for Vienna]

[Insert User Analysis and Evaluation issues for Ruse]

5 Summary and Conclusions

[This chapter concludes this report and briefly indicates future work]

This report presents the results of user evaluation of urbanAPI applications i.e. 3DVR (including rule editor), PME and UGS, performed by different evaluators from all the participating cities: Vitoria-Gasteiz, Bologna, Vienna and Ruse. The objective of the user evaluation cycle 1 is to get an early interim assessment of the current development status of the urbanAPI applications against predefined assessment criteria and application requirements as defined in Deliverable 2.1 – User requirements definition. Using the CIM2 methodology - Deliverable 5.1 – Assessment Methodology, selected sets of questions were directly derived from the user requirements specification and refined and prioritized by going through a thorough user consultation process.

The overall approach mostly followed for the evaluation was to provide training material (videos) two weeks in advance. Also, a hands-on training session is conducted so that the evaluators can understand specific functionality and learn how to use the applications and then respond to specific questions. Based on the experience it is wise to say that this training based evaluation approach was successful to the extent that evaluators got first-hand experience of using and getting better

understanding of the specific functionalities these applications can support. It is understood that without training sessions it would have been difficult to fully understand the functionality and especially how to perform a specific operation. The results reflect positive evaluation responses for the aspects covered in the hands-on training sessions.

The overall evaluation participation rate was very promising and different types of users – urban planners, policy makers, GIS experts, General public etc participated in the evaluation. That is all local stakeholders who attended training sessions responded to evaluation questions. However, the size of user sample set for individual cities is too small to generalise the results. After training sessions, evaluators were allowed to reuse application for a week and then respond to evaluation questions through online evaluation portal. This allowed evaluators to fully understand strengths and limitations of these applications in their own time and respond to evaluation questions based on educated knowledge of urbanAPI application scenarios. Based on the evaluation results a detailed analysis is performed and presented in city based evaluation reports. These analyses are provided from two perspectives: i) scenario based and ii) criteria based. In addition, evaluation results of rule editor are presented.

The overall assessment results are projected in the form of final mark values, which indicate whether or not the outcome is above acceptable threshold for application scenarios or criteria. It is found that many features were already implemented in the application but end users couldn't use them due to lack of detailed training and hence responded negatively. This suggests the need for better and detailed training needs during cycle 2. Also, some features couldn't be implemented for different cities due to lack of fine granularity of data available but similar features are implemented for other cities where data is of high quality. Evaluation responses and charts presented in city reports do not reflect such limitations which are beyond the control of application implementation capacity. Rather these limitations are explained separately in tables and covered in technical perspectives as learning outcomes.

Nevertheless, the overall purpose of the evaluation exercise is fulfilled, as it is able to identify the aspects, which require development focus during cycle 2. In addition, it is also understood that some of the limitations revealed are directly related to availability of quality and detailed data set e.g. high quality 3D city model or detailed GSM data to capture mobility patterns, or UGS socio-economic data etc. These aspects will be further explored during cycle 2 of the project development.

The evaluation results are also analysed from users level of expertise and roles point of view and the extent to which it affects the evaluation results. The results show that different types of evaluators (urban planner, policy maker, GIS experts, others) contributed to evaluation results more or less equally regardless of their level of expertise. However, role based user analysis indicate different types of users provide evaluation responded based on their domain knowledge. This suggests that the overall limitations identified in the scenario based and criteria based evaluation results are not fully dependent on the evaluators level of expertise. Rather, these limitations can be attributed either to level of training provided (including training videos) to evaluators or required applications features were not found by evaluators. For interpretation purposes, explicit user perspective is included in city reports. Similarly, technical perspective is also provided that explains limitations beyond control of application implementation capabilities. These perspectives help to synthesise the evaluation results. In general, this is what exactly the CIM 2 based user evaluation cycle 1 was intended to do. And, consequently, evaluation results identified number of aspects in the urbanAPI applications which should be dealt with during development cycle 2.

References

[Any citations will go here]

[1] Deliverable D2.1 – User Requirement Definition – FP7 urbanAPI project, 2013.
[2] Deliverable D5.1 – Assessment Methodology – FP7 urbanAPI project, 2013.
[3] Khan Z, Ludlow D, Rix J, (2012), "Applying the Criteria Indicators and Metrics Evaluation Method on ICT Research: The HUMBOLDT Project, Research Evaluation Journal, Vol. 22, No. 1, 2013, p.p. 30-40.
[4] Khan Z, Ludlow D, Caceres S, (2013), "Evaluating a collaborative IT based research and development project", Evaluation and Program Planning (2013). DOI: 10.1016/j.evalprogplan.2013.04.004, Vol. 40, p.p. 27-41.
[5] Description of Work, FP7 urbanAPI project, 2011.

Annex 1: Evaluation Design for All Cities and Applications [CONFIDENTIAL]

Annex 2: City Evaluation Reports and Stakeholder Board Evaluation Report [CONFIDENTIAL]