Demonstration and Evaluation Methodology

Deliverable D5.1

DATE
30 June 2021

ISSUE
1.0

GRANT AGREEMENT
no 870337

DISSEMINATION LEVEL
PU

PROJECT WEB-SITE
http://cure-copernicus.eu/

LEAD AUTHORS
Zaheer Khan (University of the West of England, Bristol)
David Ludlow (University of the West of England, Bristol)
Kamran Soomro (University of the West of England, Bristol)

CONTIRUBTORS:
Alessandra Gandini (TECNALIA),
Zina Mitraka (FORTH),
Giorgos Somarakis (FORTH),
Christiane Walter (PIK).

CONTENTS

1 Introduction 3
1.1 Acronyms 3
1.2 Overview 3
2 The Evaluation methodology 4
1.1 Background 4
1.2 Evaluation perspectives 5
1.3 The CIM Methodology 5
1.4 Evaluation Criteria 7
3 The Evaluation Process 9
3.1 Derivation and Classification 10
3.2 Verification 10
3.3 Evaluation Implementation 11
3.4 Reporting 11
4 Evaluation Stakeholders 12
5 Evaluation Design 15
6 Summary and Next Steps 18
7 References 20
APPENDICES 22
Appendix – A – Evaluation Design template for CIM2 23
Appendix – B - List of Requirements fulfilled and Reasoning 25
Appendix – C - List of Potential Evaluation Questions 25

LIST OF FIGURES

Figure 1: Evaluation Methodology: Criteria-Indicator-Metrics - 2 6
Figure 2: CIM2 methodology: An Example.....7
Figure 3: Evaluation process..... 10

LIST OF TABLES

Table 1: Mapping of CURE application criteria to ISO25010 characteristics and UTAUT2 8
Table 2: CURE Participation Planning Matrix 13
Table 3: Possible Indicators and mapping to criteria..... 15

1 INTRODUCTION

1.1 Acronyms

Criteria Indicators and Metrics (CIM)
Copernicus Climate Change Service (C3S)
Copernicus Atmosphere Monitoring Service (CAMS)
Copernicus Land Monitoring Service (CLMS)
Copernicus for Urban Resilience in Europe (CURE)
Copernicus Emergency Management Service (EMS)
Data Information Access Service (DIAS)
Drivers, Pressures, States, Impacts, responses (DPSIR)
European Environment Agency (EEA)
Health Impacts (HI);
Local Scale Surface Temperature Dynamics (LSSTD);
Nature-based Solutions (NBS);
Sustainable Development Goals (SDG)
Surface Urban Heat Island Assessment (SUHIA);
Urban Flood Risk (UFR)
Urban Heat Storage Monitoring (UHSM);
Urban Subsidence, Movements And Deformation Risk (USMDR);
Urban Heat Emissions Monitoring (UHEM);
Urban Air Quality (UAQ);
Urban CO2 Emissions Monitoring (UCO2EM);
Urban Thermal Comfort (UTC)
Work Package (WP)

1.2 Overview

This document presents the CURE (Copernicus for Urban Resilience in Europe) demonstration and evaluation methodology. The deliverable based on Task 5.1. provides planning for the demonstration workshops as well as establishing a primary set of criteria based on the user requirements from Deliverable D1.1 – summary of user requirements, as well as criteria related to usability and effectiveness characteristics of ISO 25010 [1]. These criteria are derived by adapting the Criteria, Indicators and Metrics (CIM) methodology [2][3]. This document will provide the basis to carry out Task 5.2 and Task 5.3.

CURE 11 cross-cutting applications address challenges arising from user requirements to more effectively plan sustainable cities. The CURE applications include:

- Local Scale Surface Temperature Dynamics (LSSTD) – AP01
- Surface Urban Heat Island Assessment (SUHIA) – AP02
- Urban Heat Emissions Monitoring (UHEM) – AP03
- Urban CO2 Emissions Monitoring (UCO2EM) – AP04
- Urban Flood Risk (UFR) – AP05

- Urban Subsidence, Movements and Deformation Risk (USMDR) – AP06
- Urban Air Quality (UAQ) – AP07
- Urban Thermal Comfort (UTC) – AP08
- Urban Heat Storage Monitoring (UHSM) – AP09
- Nature-based Solutions (NBS) – AP10
- Health Impacts (HI) – AP11

Before elaborating the evaluation methodology and evaluation design for the CURE project in detail, the following discusses what is meant by evaluation in the project. Shapiro [4] defines evaluation as:

"the comparison of actual impacts against strategic plans. It looks at original objectives, at what was accomplished and how it was accomplished. It can be formative, that is taking place during the life of a project or organization, with the intention of improving the strategy or way of functioning of the project or organisation. It can also be summative, drawing lessons from a completed project or an organisation that is no longer functioning" [4].

The main emphasis in CURE user evaluation will be to have formative evaluation that will be carried out in two stages together with demonstration workshops that provide timely feedback to the CURE application development team. In addition summative evaluation will be conducted at the end of the project mainly to elaborate project development experience, evaluation lessons learnt, and suggested improvements. These evaluation exercises will be performed within a specified context that is directly related to the CURE application requirements defined in Deliverable 1.1 – Summary of User Requirements. Accordingly, the stakeholders and users who provided original requirements will play a critical role in evaluating the CURE applications.

In WP5 the focus is on user evaluation, mainly effectiveness and usability of the CURE applications. The aim to assess the usefulness of CURE applications in decision making processes within the frame of open, integrated and interoperable governance.

2 THE EVALUATION METHODOLOGY

1.1 Background

One of the challenges in the development of a collaborative and integrated project is to collect and analyse the requirements from different stakeholders to develop the required software and applications. Further, the evaluation of outcomes against the stated objectives and requirements of a project requires a structured and coherent evaluation methodology. Undertaking the evaluation of a collaborative project is not straightforward due to a variety of factors, for instance, the timely delivery of different components of the system when: i) there are dependencies between various components; ii) several stakeholders from different countries are involved who typically have diverse technological and/or application domain backgrounds and objectives; iii) there exist multiple and sometimes conflicting application specific and user-defined requirements; iv) there are multiple and often conflicting technological research and development objectives, and, v) the real impact of the project may not be realisable within the lifetime of the project. All these aspects require a coherent

evaluation methodology and planning to adapt to the diverse requirements of collaborative integrated research projects.

In the above context, notwithstanding the fact that there are many research and system evaluation approaches, methods and techniques [5-12], the assessment of the outcomes of an integrated and collaborative project using ad-hoc evaluation approaches may not fully reveal the limitations of the underlying system. Also, a single approach or technique may not be fully appropriate in assessing the various evaluation aspects of the system. The following introduces the overall evaluation methodology, defining and/or selecting the best-fit evaluation methodology, techniques and process for evaluating CURE systems.

1.2 Evaluation perspectives

According to Juristo N. et. al. [10][11], 'the evaluation of a system can be divided into four basic types:

- Checking the correctness of the system structure for design and coding errors (i.e. *verification*).
- Checking the correctness of the content or semantics of the system (i.e. *validation*).
- Checking the system externality for its operational success (i.e. *system usability*).
- Checking what improvements the current system has brought to the organisation (i.e. *usefulness*)'.

Based on the above types and nature of the development of the CURE applications which are both user-driven (based on actual user requirements) and technology-driven (based on a state of the art analysis), different perspectives can be addressed in the evaluation design of the project e.g., a more technology-oriented i.e. "perspective 1: software quality" and a more user and context-oriented "perspective 2: benefits". This may also be referred as technical evaluation and user evaluation. Only perspective 2 is within the scope of WP5. Perspective 2 aims to assess the outcomes of the CURE applications from the perspective of benefits to potential users and the relevance to the stated or implied objectives of the project. This perspective is particularly important for the city requirements and should answer the main question: "To what extent do the CURE cross-cutting applications achieve the stated and implied objectives of the project such as providing decision making support, data integration, service openness, service compatibility, etc.?".

1.3 The CIM Methodology

In order to perform the above evaluation in the CURE project, the Criteria Indices Metrics – Ver 2 (CIM2) methodology has been adapted which primarily defines the means to secure evaluation results as shown in Figure 1. CIM2 is extended from the experience developed in the FP6 HUMBOLDT project [13] [14] and applied in the FP7 projects UrbanAPI [15], and DECUMANUS [16] as well as H2020 Smarticipate project [17]. Basically, this approach reflects on the design of the evaluation to be carried out, and defines a set of criteria based on specific aspects e.g. user functional, non-functional or contextual requirements, which need to be considered, for instance aspects related to benefits, usability, functionality, performance, efficiency, compatibility, deployment, etc. Each criterion may have additional associated sub-criteria to address various other aspects. Each sub-criterion is made operational by defining one or more indicators which address the evaluation criteria in question. In order to better understand the context of the evaluation for a particular aspect, each indicator is represented by one or more questions or statements. Also, for each indicator (and associated questions)

some metrics are defined to judge whether or not the result is regarded as good or bad. Additionally, qualitative assessment is also included in CIM2 to enable evaluators to provide subjective (and/or objective) assessment mainly contributing to benefits, relevance and the overall impact of the CURE applications. These qualitative outcomes can be in the form of subjective statements. Techniques like system usability scale [18] or its variations [19] may be utilised to give a global view of the subjective assessment.

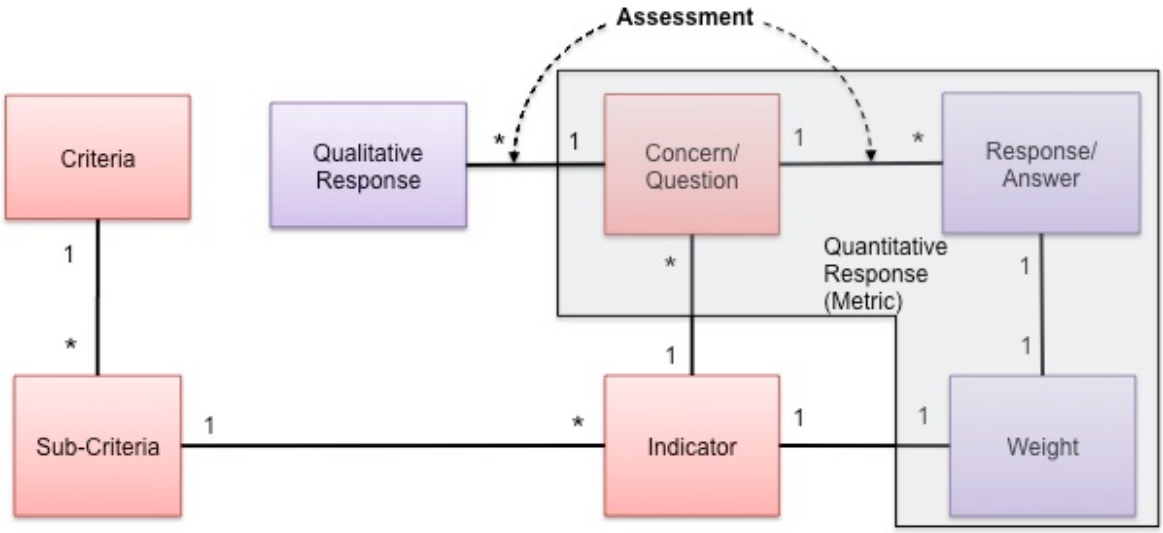

Figure 1: Evaluation Methodology: Criteria-Indicator-Metrics - 2

As an example, we illustrate in Figure 2 'usability' as one of the main elements derived from ISO 25010 [1], and selected as a main criterion with two further sub-criteria defined as: 'learnability' and 'understandability'. Each sub-criterion identifies indicators, also considered as concerns to be evaluated, and in order to define metrics for indicators, specific questions are defined which address the aspects to be evaluated for each indicator. These questions can have specific response options each defined with quantitative and measurable weights. *Appendix A* provides an evaluation design template derived from CIM2 to be used for indicating overall criteria, indicators, metrics and available resources.

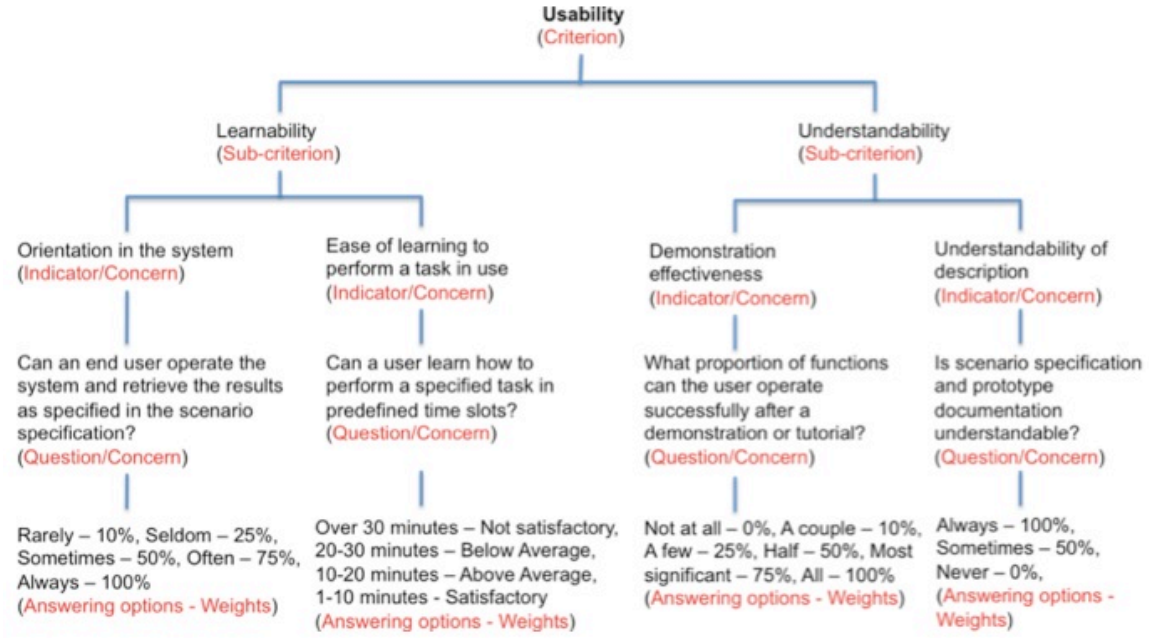

Figure 2: CIM2 methodology: An Example

1.4 Evaluation Criteria

The main purpose of the evaluation is based on the following criteria from both evaluation perspectives :

- Utility and Benefits – The potential benefits of the CURE applications in ***effective use of resources, new information for local planning and decision making***, etc.; This also includes the relevance of the CURE applications in achieving the project goals related to the development of cross-cutting applications for urban resilience;
- Usability - The capability of the CURE applications to be ***understood, learned, used and appreciated by the stakeholders/end-user,*** when used under specified conditions.

To support a more structured approach ISO 25010 standard [1] is adopted and its characteristics are used to derive evaluation criteria and sub-criteria. The mapping of ISO25010 characteristics to the above criteria is presented in the following Table 1. i.e. using both *'quality in use'* and *'product quality'* models. Furthermore, CURE user groups can be mapped on to ISO25010 stakeholder types i.e. primary user (*person who interacts with the system to achieve the primary goals e.g.planning officer, climate change officer, etc.*), secondary user (*who provide support, for example a) content provider, system manager/administrator, security manager; b) maintainer, analyser, installer*) and indirect user (*person who receives output, but does not interact with the system*).

To make evaluation criteria more robust other approaches were also considered such as Venkatesh et al 2003 and 2012 [20] [21]. Vekatesh et al presented the Unified Theory of Acceptance and Use of Technology (UTAUT) model with strong theoretical underpinnings from behavioural science. According to UTAUT2 behavioural

intentions and behaviour are determined by certain key constructs which can be related to CURE application evaluation including:

- *Performance expectancy* - degree to which using CURE applications will provide benefits to users in performing certain activities
- *Effort expectancy* – degree of ease associated with users' use of CURE applications
- *Social influence* – extent to which CURE application users perceive how important others believe they should use a CURE application
- *Facilitating conditions* – how CURE applications users believe that technical infrastructure exists to help them to use the applications
- *Price value* – trade-off between the perceived benefits of using CURE applications and monetary costs of using technology
- *Habit* – results of past experiences in using similar applications or going through similar process

These UTAUT2 constructs have similarities to different ISO25010 characteristics and hence are also included in Table 1. Table 1 indicates that UTAUT2 constructs may complement some ISO25010 characteristics and be used together to derive a wide range of indicators and evaluation questions.

Table 1: Mapping of CURE application criteria to ISO25010 characteristics and UTAUT2

C#	ISO 25010 Characteristics (C)[1]	UTAUT2 [20][21]	Application Criteria
	Quality in Use (refers to system quality)		
1	Effectiveness	Performance expectancy	Utility and Benefits (relevance)
2	Efficiency	Performance expectancy, Price value, Habit	Usability, Benefits
3	Satisfaction *(3.1 Usefulness, 3.2 Trust, 3.3 Pleasure, 3.4 Comfort)*	3.1 performance expectancy, 3.3 & 3.4	Usability, Benefits
4	Freedom from risk *(4.1 Economic risk mitigation, 4.2 Health and safety risk mitigation, 4.3 Environmental risk mitigation)*		Benefits (relevance)
5	Context coverage *(5.1 Context completeness, 5.2 flexibility)*	Not relevant	Not relevant
	Product quality (internal and external quality combined as product quality)		

[1] Brief introduction to ISO25010 characteristics is available from here: https://iso25000.com/index.php/en/iso-25000-standards/iso-25010?start=0

6	Functional suitability *(6.1 functional completeness, 6.2 functional correctness, 6.3 functional appropriateness)*	facilitating conditions, 6.3 effort expectancy	Feature coverage (cross-cutting)
7	Performance efficiency *(7.1 Time behaviour, 7.2 Resource utilisation, 7.3 Capacity)*	7.1 and 7.2 performance expectancy, 7.3 facilitating conditions	Functionality, Usability and Benefits
8	Compatibility *(8.1 Co-existence, 8.2 Interoperability)*	Facilitating conditions	Functionality
9	Usability *(9.1 appropriateness recognisability, 9.2 Learnability, 9.3 Operability, 9.4 User error protection, 9.5 User interface aesthetics, 9.6 Accessibility)*	Effort expectancy, 9.5	Usability
10	Reliability *(10.1 Maturity, 10.2 Availability, 10.3 Fault tolerance, 10.4 Recoverability)*	Facilitating conditions	Functionality
11	Security *(11.1 Confidentiality, 11.2 Integrity, 11.3 Non-repudiation, 11.4 Accountability, 11.5 Authenticity)*	Facilitating conditions	Functionality
12	Maintainability *(12.1 Modularity, 12.2 Reusability, 12.3 Analysability, 12.4 Modifiability, 12.5 Testability)*	Facilitating conditions	Functionality
13	Portability *(13.1 Adaptability, 13.2 Installability, 13.3 Replaceability)*	Facilitating conditions	Functionality and Usability

3 THE EVALUATION PROCESS

The evaluation process has the following stages as depicted in Figure 3.

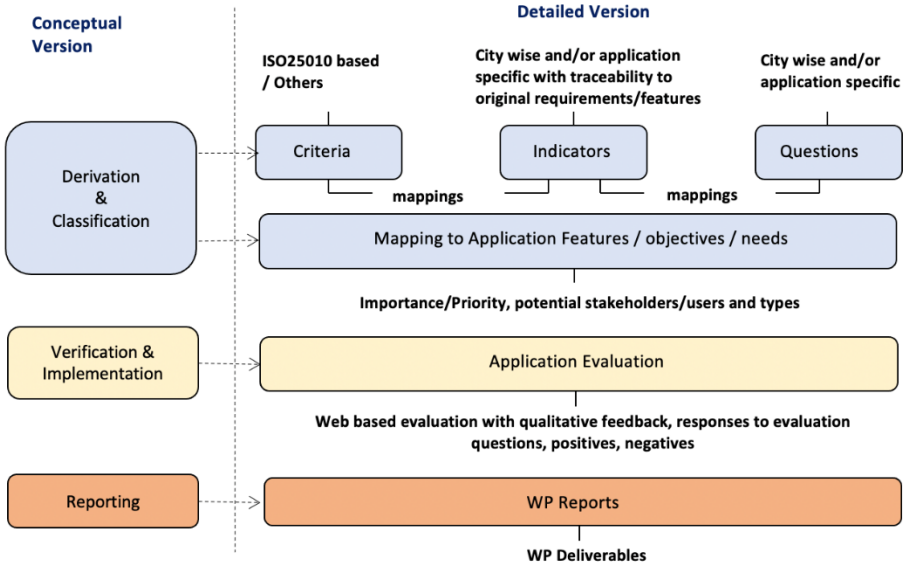

Figure 3: Evaluation process

3.1 Derivation and Classification

This stage is used to prepare the evaluation contents and consists of the following activities:

- Identification of main set of criteria based on the requirements specification, application objectives and stakeholder needs;
- Deriving sub-criteria and identifying indicators to achieve the objectives of the project and its applications;
- Deriving questionnaire and response options with associated weights to measure the outcomes. In order to facilitate different types of stakeholders (domain experts, end users);
- Specifying the means to perform specific evaluation exercises to answer the evaluation questions.

3.2 Verification

The purpose of this stage is to ensure that evaluation questions are appropriate for different stakeholders and are appropriately prioritised at different evaluation stages. Also, this stage can be used to check whether or not the CURE application features being evaluated have been implemented and are accessible to evaluators. This stage will also identify any training needs for using the CURE system and applications. A rigorous verification stage is necessary to identify known gaps and/or limitations, which can be documented prior to evaluation implementation. The overall evaluation design will be shared with project partners, and in particular with CURE application leads to propose any changes. For example,

- Review of requirements specification and reasons for unfulfilled requirements
- Verify that the evaluation design contents (i.e. criteria, questionnaire, etc.) capture important assessment elements and are in line with stakeholder needs and requirements.

- Highlight the importance (high, medium, low) of specific questions. This will mainly be mapped to priority and/or importance of corresponding requirement/feature.
- Indicate improvements in the questionnaire, response options, metrics etc.
- Identify additional criteria, indicators, questionnaire and metrics.

3.3 Evaluation Implementation

The evaluation will be carried out using an online web-based or paper-based questionnaire in the demonstration workshops as previously used in EC funded projects [15][16][17]. These will be specific questions (up to 15 maximum) to answer usability, benefits and relevance aspects. Most of these questions addressed to demonstration workshop participants will be qualitative (open-ended) type questions. The CURE project team members will act as moderators for the discussion groups in order to elicit as many details as possible. After the demonstration and discussions, the participants will be invited to complete the evaluation questionnaire. For quantitative questions Likert scaling will be used, where the objective is to get qualified responses for the evaluation questionnaire. Such a qualified response assists in understanding why an evaluator responded to a certain question in a specific way that will help in deriving overall impact assessment and evaluation conclusions. Additional fields will collect qualitative responses, which will help in understanding the reasoning behind specific responses. The questionnaire will be customised to suit city and application specific needs, and hosted on a professional survey platform. For the web-based questionnaires different platforms are under consideration: a) **EU Survey**[2] which is the official platform supported by the European Commission for hosting surveys; b) **Bristol Online Survey** (BOS) tool[3] with which UWE has data protection agreement; and c) other easy to access online service such as **Microsoft Forms**.

Other techniques presented in [22] can be adopted to perform the evaluation. More specifically, *role-play* technique to ensure that interests of under-represented or difficult to access stakeholders are captured and represented; *purpose network* technique to identify common and conflicting evaluation purposes that may be considered as concerns or criteria to be used in CIM2. Also, *evaluation recommendations support versus opposition grids* and *recommendation attractiveness versus stakeholder capability grid* may be considered for evaluation analysis purposes. Similarly, *recommendation implementation strategy development grid* can be used to depict overall assessment of limitations, strengths and resources needed to accomplish stakeholders' interests and goals based action agenda.

Prior to evaluation implementation, ethical approval will be sought from the UWE Faculty Research Ethics Committee. This ethical approval process will lay out all ethical procedures to be followed during the demonstration workshops as well as evaluation exercises.

3.4 Reporting

All the above stages including evaluation results will be properly documented, and analysis will be provided in different deliverables.

[2] EU Survey service - https://ec.europa.eu/eusurvey/home/welcome
[3] BOS tool - https://www.onlinesurveys.ac.uk/

4 EVALUATION STAKEHOLDERS

Bryson J, et. al. [22] emphasised properly identifying different categories of evaluation stakeholders for evaluation planning, design, implementation and decision making process. These stakeholders can be represented via the 'onion model' [23] or in other categories [22][24] to depict how close interaction each stakeholder group/role should have with the system.

CURE stakeholders refer to an individual or an organisation or any legal entity that have vested interest i.e. planning and/or policy development, public services, and community engagement etc. in using applications for organisational benefit e.g., awareness raising, knowledge exchange, policy development, decision making, etc.

Stakeholders can be individual users who are involved in development of a system i.e. operators, developers or maintenance staff, or users of the system i.e. functional or operational beneficiaries e.g. city policy makers e.g., urban planners, climate change experts etc., or organisations who are mainly interested in the outcomes including political or societal beneficiaries.

In CURE project, stakeholders play a key role in deriving requirements for the development of applications, local data provision and evaluating the outcomes of the project. CURE stakeholders and user groups include:

- **City Planners (CP):** city planners including policy makers from city department(s) such as urban planning, climate change, air quality, public engagement etc. that are direct beneficiaries of CURE applications. Both the front-runner and the follower city partners are included in this category.
- **Environment Agencies (EA):** Environment agencies e.g. EEA can be direct beneficiaries of the CURE data and applications outcomes as they can reuse data and outcomes of the CURE applications in their processes.
- **Commercial (CO):** SMEs/Industry users who would be interested in CURE applications for developing further commercial applications.
- **Researchers (RE):** Individuals and/or research organisations interested in one or more CURE application theme(s) related data and/or application outcomes so that they can reuse and/or extend their research.
- **Other Stakeholders (OS):** selected organisations (e.g.Climate-KIC etc), which are interested in CURE research and outcomes and can use CURE services as an input in their work

Bryson J, et. al. [22] provide a composition of techniques to engage stakeholders at different stages of the evaluation process. The participation planning matrix technique (originally from International Association for Public Participation) appears to be useful to indicate the level of participation from various stakeholders at different stages of the evaluation process. In **Table 2:** CURE Participation Planning Matrix, the *type of involvement* indicates benefits to the project of different types of stakeholders based on their level of engagement. Furthermore, for *the overall evaluation process* **Table 2** identifies the expected level of engagement by different user groups at different stages of the evaluation process i.e. evaluation design – derivation, classification and test case verification, implementation and documenting results.

Table 2: *CURE Participation Planning Matrix*

	Levels of Engagement						
	Do not Engage	**As Data Source**	**Inform**	**Consult**	**Involve**	**Collaborate**	**Empower**
Types of Involvement							
Commitment from Project and privileges	No promises	Fulfil data usage protocols; quality assessment; identification of missing data;	Inform project outcomes and evaluation results	Consider input for evaluation design and inform evaluation output e.g. through requirements	Include your interests and inputs in evaluation design; consider you part of the evaluation decision-making process	Include advice and suggestions to greater extent possible; consider you part of the evaluation decision-making process; suggests system improvements	Fully engaged in the evaluation planning, design, implementation and decision-making process; Influence system improvement
Benefits of engagement	Not involved	Provides necessary data; assessment of fulfilment of application specific data needs	Dissemination to wider related user/scientific community	To get different perspectives, issues, needs, concerns	Attracting attention, identifying key issues and establishing credibility of evaluation results	Act as primary intended users and have high interest and availability and can influence evaluation outcomes	Capacity development; to be able to participate in evaluation and have sense of ownership of the evaluation and making impact in improvement of system
The Overall Evaluation Process							

Derivation & Classification		Optional	Main source of evaluation design is user requirements – mainly referred to D1.1 – Summary of User Requirements where all user groups have contributed
Design Verification		CURE application developers	Here mainly CURE application leads/developers will be involved
Evaluation Implementation		This will be open to all especially CA, SB	This will be open to all during demonstration workshops. Stakeholder board and both front-runner and follower cities will be involved.
Results		Outputs from Implementation stage and analysis of results will be prepared by WP5 working group	

5 EVALUATION DESIGN

For user evaluation, indicators provide the basis to derive evaluation questions. The indicators are derived from the above criteria and common categories used for requirement features. **Table 3** presents a mapping from criteria to indicators:

Table 3: Possible Indicators and mapping to criteria

	ISO 25010 Characteristics [1]	UTAUT2 [20][21]	CURE Application Criteria	Possible Indicators
	Quality in Use (refers to system quality)			
1	Effectiveness	Performance expectancy	Benefits, Relevance	Achieving performance expectancy, awareness raising
2	Efficiency	Performance expectancy, Price value, Habit	Usability, Benefits, Relevance	Achieving performance expectancy, evidence collection for planning applications/policy-making/decision-making, efficient resource utilisation, openness, Level of technological innovation,
3	Satisfaction *(3.1 Usefulness, 3.3 Trust, 3.3 Pleasure, 3.4 Comfort)*	3.1 => performance expectancy; 3.3 & 3.4 => Hedonic motivations	Usability, Benefits	Achieving performance expectancy, Level of technological innovation, knowledge transfer, Awareness raising
4	Freedom from risk *(4.1 Economic risk mitigation, 4.2 Health and safety risk mitigation, 4.3 Environmental risk mitigation)*		Relevance, Benefits	Not relevant
5	Context coverage *(5.1 Context completeness, 5.2 flexibility)*	5.1 => performance expectancy and facilitating conditions; 5.2 => effort expectancy	Not relevant	Not relevant
	Product quality (internal and external quality combined as product quality)			

6	Functional suitability *(6.1 functional completeness, 6.2 functional correctness, 6.3 functional appropriateness)*	6.3 => effort expectancy	Functionality (cross-cutting)	Integrated and/or cross-cutting
7	Performance efficiency *(7.1 Time behaviour, 7.2 Resource utilisation, 7.3 Capacity)*	7.1 and 7.2 => performance expectancy; 7.3 => facilitating conditions	Not relevant	Not relevant
8	Compatibility *(8.1 Co-existence, 8.2 Interoperability)*	Facilitating conditions	Functionality	Perceived sustainability and extensibility, level of integration/interoperability, openness
9	Usability *(9.1 appropriateness recognisability, 9.2 Learnability, 9.3 Operability, 9.4 User error protection, 9.5 User interface aesthetics, 9.6 Accessibility)*	Effort expectancy; 9.5 => Hedonic motivations	Usability	Minimal effort expectancy, platform accessibility, appropriateness, ease of learning/operability, user error protection, user interface aesthetics
10	Reliability *(10.1 Maturity, 10.2 Availability, 10.3 Fault tolerance, 10.4 Recoverability)*	Facilitating conditions	Not relevant	Not relevant
11	Security *(11.1 Confidentiality, 11.2 Integrity, 11.3 Non-repudiation, 11.4 Accountability, 11.5 Authenticity)*	Facilitating conditions	Not relevant	Not relevant
12	Maintainability *(12.1 Modularity, 12.2 Reusability, 12.3 Analysability, 12.4 Modifiability, 12.5 Testability)*	Facilitating conditions	Functionality	Perceived sustainability and extensibility, Facilitating conditions
13	Portability *(13.1 Adaptability, 13.2 Installability, 13.3 Replaceability)*	Facilitating conditions	Not relevant	Not relevant

Possible indicators will be grouped for analysis purposes according to the 3-dimensions of the Open, Integrated and Interoperable model for urban governance and land use planning (Horizon 2020 Smarticipate project 2019 [17]).

- ***Level of integration*** *– Degree to which the CURE system is connected with existing systems in the pilot cities. It is also the degree to which cross-cutting aspects (e.g., green areas, heat, air quality, health, etc.) are implemented.*
- ***Minimal effort expectancy*** *– degree of ease in using CURE applications by different users*
- ***CURE system accessibility*** *– Degree to which the CURE system and applications can be accessed through different hardware/software (e.g. screen sizes, etc.) for people with diverse abilities*
- ***User interface aesthetics*** *– Ability to present CURE applications through easy to use graphical user interface*
- ***Efficient resource utilization*** *- ability to make efficient use of resources (e.g. human resources, hardware, software, data, urban spaces, plans, etc.)*
- ***Level of technological innovation*** *– ability to generate new knowledge and innovative solution through CURE system and application data*
- ***Evidence collection for policy making*** *– ability to collect data evidence to support new policy development*
- ***Evidence collection for decision making*** *– ability to collect data evidence to support decision making*
- ***Awareness Raising*** *– degree to which the CURE system is able to facilitate awareness raising about environmental issues and urban sustainability in cities.*
- ***Level of interoperability*** *– Degree to which the CURE system can handle importing and exporting different datasets and/or formats*
- ***Perceived sustainability and extensibility*** *– degree to which CURE system and applications are reusable in different contexts (e.g. new cities) and have the ability to be extensible by allowing to add new features (e.g., cross-cutting data integration)*
- ***Openness -*** *Degree to which CURE data and algorithms are accessible/available for reuse.*
- ***Achieving performance expectancy*** *- degree to which using CURE applications will provide benefits to stakeholders in performing certain activities e.g. awareness raising for behaviour change, evidence based decision making, etc.*
- ***Ease of learning*** *– Degree to which the CURE users can easily learn and use the system*
- ***User error protection*** *– Degree to which the CURE can avoid user input error*
- ***Ease of operability*** *– Degree to which the CURE users can easily operate the system*
- ***Facilitating conditions*** *– Degree to which technical infrastructure exists to help users to use the CURE system and applications whenever necessary*

Please note that the proposed list indicators are indicative and may be amended/extended after verification stage. A list of initial evaluation questions is presented in Appendix - C. Please note that new questions will be added as soon as specific CURE application features are implemented and made available for user evaluation.

6 SUMMARY AND NEXT STEPS

As indicated above the main emphasis in CURE user evaluation will be to have formative evaluation that will be carried out in two stages together with demonstration workshops that provide timely feedback to the CURE application development team. In addition summative evaluation will be conducted at the end of the project mainly to elaborate project development experience, evaluation lessons learnt, and suggested improvements. These evaluation exercises will be performed within a specified context that is directly related to the CURE application requirements defined in Deliverable 1.1 – Summary of User Requirements. Accordingly, the stakeholders and users who provided original requirements will play a critical role in evaluating the CURE applications.

In WP5 the focus is on user evaluation, mainly effectiveness and usability of the CURE applications. The aim to assess the usefulness of CURE applications in decision making processes within the frame of open, integrated and interoperable governance. Specifically to assess the outcomes of the CURE applications from the perspective of benefits to potential users and the relevance to the stated or implied objectives of the project. This perspective is particularly important for the city requirements and should answer the main question: "To what extent do the CURE cross-cutting applications achieve the stated and implied objectives of the project such as providing decision making support, data integration, service openness, service compatibility, etc?".

In order to perform the above evaluation, the Criteria Indices Metrics – Ver 2 (CIM2) methodology has been adapted which primarily defines the means to secure evaluation results. Basically, this approach reflects on the design of the evaluation to be carried out, and defines a set of criteria based on specific aspects e.g. user functional, non-functional or contextual requirements, which need to be considered. In order to better understand the context of the evaluation for a particular aspect, each indicator is represented by one or more questions or statements. Also, for each indicator (and associated questions) some metrics are defined to judge the results. Additionally, qualitative assessment is included in CIM2 to enable evaluators to provide subjective (and/or objective) assessment mainly contributing to benefits, relevance and the overall impact of the CURE applications.

Next steps in the context of Tasks 5.2 and 5.3 will focus on Demonstration workshops that provide the means to carry out the evaluation. Two demonstration workshops are planned, the first will present CURE application outcomes to the front-runner cities, and the second for the follower cities. Despite reducing Covid-19 restrictions it is still not clear whether a physical workshop can take place in 2021, and the first workshop plans for September/October 2021 may need to be completely online. The date and venue for the 2nd workshop will be decided in consultation with WP3 and WP4 in 2022. Workshops will be divided into CURE cross-thematic areas, and associated applications will be presented by the application leads supported by dialogue between application leads and participants. This dialogue will be followed by an evaluation questionnaire which will be based on the above evaluation methodology.

The following stages will be part of the Demonstration Workshops process:

- Ethics approval: An ethics application will be submitted to UWE Faculty Research Ethics Committee to seek ethics approval. This application will specify consent forms, participant information sheet, data governance plan, privacy notice etc. Data collection will only commence once the ethics approval is available.

- **Invitations:** Workshop participation invitations will be sent to CURE city partners and the Advisory Board to participate in the demonstration workshops. City partners and Advisory Board will extend the invitation to relevant organisations/departments.
- **Post-workshop analysis:** Dialogue summary and the analysis of the evaluation questionnaire will be used to feed into Deliverables D5.2 and D5.3.

7 REFERENCES

[1] ISO/IEC 25010:2011, Systems and Software Engineering – System and Software Quality Requirements and Evaluation (SQuaRE) – System and Software Quality model, 2011.
[2] Khan, Z., Ludlow, D. and Caceres, S. (2013) Evaluating a collaborative IT based research and development project. Evaluation and Program Planning, 40. pp. 27-41. ISSN 0149-7189
[3] Khan, Z., Ludlow, D. and Rix, J. (2012) Applying the criteria indicators and metrics evaluation method on ICT research: The HUMBOLDT project. Research Evaluation Journal, 22 (1). pp. 30-40. ISSN 0958-2029
[4] Shapiro, J., Monitoring and Evaluation, URL http://www.civicus.org/view/media/Monitoring%20and%20Evaluation.pdf, Last Accessed: 17 June 2021
[5] Andras P. (2011), 'Research: metrics, quality, and management implications', Research Evaluation, 20/2: 90–106.
[6] Babar MA, Zhu L, Jeffery R. (2004), 'A Framework for Classifying and Comparing Software Architecture Evaluation Methods', Proceedings of the 15th Australian Software Engineering Conference (ASWEC), IEEE Computer Society, pp. 309-319.
[7] Basili V, Caldiera G, Rombach D. (1994), 'Goal Question Metric Approach, Encyclopedia of Software Engineering', pp. 528-532, John Wiley & Sons, Inc.
[8] Dobrica L, Niemela E. (2002), 'A Survey on Software Architecture Analysis Methods', IEEE Transactions on Software Engineering, 28/7: 638-653.
[9] Jong S, Arensbergen P, Daemen F, Meulen B and Besselaar P. (2011), 'Evaluation of research in context: an approach and two cases', Research Evaluation, 20/1: 61–72.
[10] Juristo N, Morant J. (1998), 'Common framework for the evaluation process of KBS and conventional software', Elsevier Knowledge-Based Systems, 11/2: 145-159.
[11] Juristo N. (1997), 'Structuring the Knowledge-Based Systems Evaluation Process', Proceedings of Intelligent Information Systems'97, pp. 266-270.
[12] Sommerville I (2006), Software Engineering, ISBN # 9780321313799, Addison-Wesley, Wokingham, UK.
[13] The HUMBOLDT Project, 2011, http://www.esdi-humboldt.eu/home.html
[14] HUMBOLDT project, 2010, Deliverable A10.6-D1 – Scenario Evaluation Report, September 2010, http://www.esdi-humboldt.eu/press/public_deliverables.html
[15] Soomro, K., Khan, Z. and Ludlow, D. (2016) Participatory governance in smart cities: The urbanAPI case study. International Journal of Services Technology and Management. ISSN 1741-525X
[16] Ludlow, D., Khan, Z., Soomro, K., Marconcini, M., Metz, A., Jose, R., Perez, J., Malcorps, P. and Lemper, M.(2016) From top-down land use planning intelligence to bottom-up stakeholder engagement for smart cities – a case study: DECUMANUS service products. International Journal of Services Technology and Management. ISSN 1460-6720
[17] H2020 Smarticipate project – (2016-2019). https://www.smarticipate.eu/resources/
[18] Brooke J, SUS – A quick and dirty usability scale, In: Jordan P, Thomas B, Weerdmeester B (eds) Usability Evaluation in Industry, pp. 189–194, Taylor and Francis, 1996.

[19] Finstad K, (2010), The Usability Metric for User Experience, Interacting with Computers, 22(2010), pp. 232-327.
[20] Venkatesh, V., M. G. Morris, G. B. Davis, and F. D. Davis. 2003. User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly 27 (3): 425-478.
[21] Venkatesh, V., J. Y. Thong, and X. Xu. 2012. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS quarterly 36 (1): 157-178.
[22] Bryson J, Patton M, Bowman R, (2011), 'Working with evaluation stakeholders: A rationale, step-wise approach and toolkit', Evaluation and Program Planning, 34/1, pp. 1-12, ISSN 0149-7189, 10.1016/j.evalprogplan.2010.07.001.
[23] Alexander, I., (2005), A taxonomy of stakeholders: Human Roles in System Development, International Journal of Technology and Human Interaction, Vol 1, 1, 2005, p.p.23-59.
[24] Sutcliffe, A., (2002), User-Centred Requirements Engineering: Theory and Practice [book], Springer.

APPENDICES

Appendix – A – Evaluation Design template for CIM2

Table 4: Evaluation Design Template

Criterion ID	Main Criterion/ Concern (ISO25010 characteristics [1])	Sub-criterion	Possible User Groups (e.g., City Planners, Environment agencies, Commercial, Researchers, Others)	Indicator	Description/ Questionnaire	Metrics/ Answering Options	Level of Importance for Stakeholders - High (H), Medium (M), Low (L), Not Relevant (NR)	Remarks
C8	Compatibility	Interoperability	CP, EA, C, R, O	Openness	Are you able to access and reuse CURE data? Are you able to access and reuse CURE application processing algorithms?	Yes, No, Partially, Not Applicable	H	None
C1	Effectiveness	Effectiveness	CP, EA, C, R, O	Evidence collection for	CURE applications provide useful	Strongly agree, Agree,		Likert-scale

				policy making	information for policy evaluation or development	Neutral, Disagree, Strongly disagree	H	

Key:

Criterion ID: uniquely identifies a criterion. **Main criterion or concern:** refers to aspect to be evaluated. **Sub-criterion:** further categorisation of the main criterion. **Evaluation stakeholders:** who will participate in the evaluation of a specific aspect. **Indicator:** description of expected outcome. **Description or Questionnaire:** question(s) to be answered based on the assessment/evaluation. **Metrics:** indicates answering options with associated weights. **Level of importance:** indicates how important this particular aspect is for a specific stakeholder? **Remarks or resources:** indicates which resources (data, documents, services) are available to perform evaluation for this particular aspect or are there any remarks related to resource limitation and/or availability? Two examples C1 and C8 are shown.

Brief introduction to ISO25010 characteristics is available from here: https://iso25000.com/index.php/en/iso-25000-standards/iso-25010?start=0. Please note and as indicated above this full list may not be applicable for the assessment of all CURE applications and only selected criteria suitable for the application needs will be used.

Appendix – B - List of Requirements fulfilled and Reasoning

Deliverable D3.1 – Urban Cross-cutting Applications Preparation [available from: http://www.cure-copernicus.eu/deliverables.html - published 31 Dec 2020] covers which specific requirements are being fulfilled by CURE applications. Please refer to the above deliverable.

Appendix – C - List of Potential Evaluation Questions

A list of initial questions is presented below. Please note that this list will be modified/updated according to the implementation status of the CURE applications.

Do CURE applications provide integrated assessment supporting evidence-based decision-making in urban planning?
Do CURE applications assist in promoting open governance process in urban planning?
Do CURE applications define frameworks promoting interoperable protocols linking governance agencies in urban planning?
Do CURE applications support the definition of common urban planning solutions applicable to cities globally?
Do CURE applications support the definition of sustainable urban development solutions in the context of climate change mitigation and carbon neutral cities?
Do CURE applications support the delivery and implementation of planning strategies for net - zero neighbourhoods?
Do CURE applications allow city governments to gain deeper insight into citizen preferences?
Do CURE applications help to raise awareness amongst citizens about ongoing neighbourhood developments that might affect them?
Do CURE applications allow cities to engage with citizens efficiently about planning proposals e.g. more effective consultation with the public on a proposals compared with traditional means, and reduced cost of engaging with the public etc?
Do CURE applications support the specification of city development visioning process in relation to strategic planning over 20+ years?
Do CURE applications provide sufficient urgency and responsiveness in supporting policy responses in conditions of future uncertainty and rapid socio-economic change?

Do CURE applications effectively address options for behavioural change in support of the planning strategy?
Are CURE applications more effective at particular urban scales, for example, neighbourhood level or citywide?
Do CURE applications support integration with other local data specifications including socio-economic and in situ data?
Are CURE applications compatible with other technologies you use?
Are CURE applications easily accessible from various platforms such as mobiles, tablets and computers?
Are interactions with CURE applications clear and understandable?
Are CURE applications supported by guides and wizards to explain their purpose and use to new users?
Will you use or continue to use CURE applications in the future?