beAWARE

Enhancing decision support and management services in extreme weather climate events

700475

D2.2

Evaluation Methodology

Dissemination level:	Public
Contractual date of delivery:	31 August 2017
Actual date of delivery:	10 September 2017
Work package:	WP2: Climate disaster management requirements
Task:	T2.1
Type:	Report
Approval Status:	Final Draft
Version:	1.4
Number of pages:	53
Filename:	D2.2_beAWARE_EvaluationMethodology_2017-09-10_v1.4

Abstract

This document describes the evaluation methodology of the pilot use cases and provides initial implementation plan of uses cases. Furthermore, there are described the literature review of the evaluation, common methodologies and finally the pillars of evaluation.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 700475

Co-funded by the European Union

 This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 700475

History

Version	Date	Reason	Revised by
0.1	10.07.2017	Deliverable structure	All partners
0.4	17.07.2017	Request for contribution from respective partners	AAWA, PLV, FBBR
0.6	24.07.2017	1st Internal Review	Michele Ferri
0.7	22.08.2017	2nd Internal Review	Meliadis Miltiadis
1.0	31.08.2017	3rd Internal Review	Meliadis Miltiadis
1.1	08.09.2017	4th Internal Review	Anastasios Karakostas
1.2	08.09.2017	5th Internal Review	Meliadis Miltiadis, Iosif Vourvachis, Anastasios Karakostas

Author list

Organisation	Name	Contact Information
HRT	Iosif Vourvachis	projects@hrt.org.gr
HRT	Miltiadis Meliadis	m.meliadis@hrt.org.gr
AAWA	Daniele Norbiato	daniele.norbiato@adbve.it
PLV	Carmen Castro Garcés	proyectosplv@valencia.es
PLV	Ruben Fernandez	proyectosplv@valencia.es
FBBR	Amalie Janniche Baneborg	amjan@fbbr.dk

Executive Summary

This deliverable presents the evaluation methodologywhich is going to be used for the evaluation of the pilots that will take place during the implementation of the beAWARE project. This evaluation methodology will be based on the user perspective and will mainly focus on the impact of the beAWARE solution in life saving, protection of infrastructure and property, reducing response time and lowering costs.

In the **first** section of the deliverable, a **specific evaluation plan** is described according to the literature review and the general principles of evaluation methodologies.

In the **second** section, the **initial implementation plans of the three pilots**, flood, fire and heat wave are presented. Each one is analyzed in order to present the purpose of the exercises, stakeholders involved and the target groups and a summary table of each pilot is given.

In the **third** section, we provide an **analysis of beAWARE evaluation** according to the common methodologies that are followed and the proposed evaluation approach is presented, which will focus in three main pillars:

- The impact of the system, which has to be evaluated by comparing the management of an emergency before and after the implementation of the beAWARE system and a series of particular key performance indicators (KPIs).
- The **user interface and the user experience (UI/UX)** based on empirical and heuristic evaluation methodologies.
- The quality of the system, its definition and a list of elements which will be used in assessing it.

Finally, the proposed evaluation procedures for the beAWARE platform are presented, in order to evaluate different factors of the platform. More specifically, the structure of the proposed questionnaires with 5 examples, interviews and hot debriefs is given for further development.

Abbreviations and Acronyms

The following abbreviations have been used in this document:

AAWA	Alto Adriatico Water Authority
CERTH	Center for Research and Technology Hellas
CBFEWS	Community-Based Flood Early Warning System
COC	Coordinated Operation Center
FBBR	Frederiksborg Fire & Rescue Service
DEWS	Digital Early Warning System
EEG	Electroencephalography
EWS	Early Warning System
FMI	Finnish Meteorological Institute
HRT	Hellenic Rescue Team
IBM	IBM Israel – Science and Technology Ltd
IIR	Interactive Information Retrieval
IOSB	Fraunhofer Institute of Optronics, System, Technologies and Image Exploitation
KWS	Key Words spotting System
LID	Language Identification
MSIL	Motorola Solutions Israel Ltd
PLV	Valencia Local Police
PSAP	Public Safety Answering Point
STEP	Systematic Test and Evaluation Process
UI	User Interface
UIQ	User Interface Quotient
UPF	Universitat Pompeu Fabra
UX	User Experience
UXA	User Experience Assessment

Table of Contents

List of Figures

List of Tables

1 Introduction

At the core of the methodology lies the need to identify the critical quality characteristics of the beAWARE solution and test them against real life conditions which will demonstrated through 3 different pilots: flood, fire, heatwave. The evaluation will take place in an iterative cycle to ensure feedback from end-users (user's perspective) and can be incorporated into further design modifications. Software testing and evaluation from user perspective will be based on a method such as STEP (Systematic Test and Evaluation Process), a well-established industry methodology for test and evaluation activities in software projects. beAWARE will be tested through three large scale pilots which are:

- Flood, which will be carried out by AAWA
- Fire, which will be carried out by PLV and FBBR
- Heatwave, which will be carried out by HRT.

This report will provide the proposed evaluation approach and more specifically the impact of the system, the user interface, user experience, the user perspective, its impact and the quality of the produced information from system. beAWARE will explore new methods which are challenging for today's technologies. The new methods will be used in different types of training, theoretical and practical, for example laboratory tests, simulation test on the field, training tests on the field and training exercises on the field, for end users and noise composition will render the speech technologies robust to additive distortions like background noise and reverberation. The language and lexical models will be updated to match the beAWARE needs, and in particular to include any specific terms or expressions used in emergencies. Specific developments will address speech produced under stressed conditions where the speaker may adjust their speech production modifying duration or intensity, or emphasizing some acoustic cues to the listener. Existing language identification (LID) systems must be adapted to speech in noisy environments, possibly containing hesitations, truncated words and code-switching. The existing key-words spotting (KWS) systems must be adapted to the beAWARE languages and to the project specificities, as referred to DoA.

It is important to mention that the combination of modern technologies and operation experience, from PSAP control rooms and from the field, will give the ability for a better evaluation of the platform and its tools. For conclusion, the deliverable will not include any technical evaluation, which will be included in WP7 deliverables, but only an evaluation from the user perspective.

Finally, because some of the details of the technologies are not yet clarified and not yet available and also some of the aspects, such as the indicators, are presented in generic form

therefore all the above will be finalized in the upcoming period and will be presented in deliverable D2.3, which is due to M15.

2 Overview

"Evaluation is the systematic collection of information about the activities, characteristics, and results of programs to make judgments about the program, improve or further develop program effectiveness, inform decisions about future programming, and/or increase understanding" (Patton, 2008). The objective of a literature search is to retrieve as much accurate information on a given subject as possible from suitable sources. The pool of available research literature is huge and only a partial fraction of it, the particular subject of interest, needs to be located. This section presents some basic guidelines to follow when tackling a literature search.

The term evaluation is used to cover a wide range of research activities, undertaken by different stakeholder groups. For some, evaluation might mean a group of staff getting together at the end of a program cycle to reflect on how it went. For others, evaluation could be a complex, multi-year research project with sites all over the province and access to a large team of academic experts. Evaluation can also be undertaken for many different reasons. Sometimes, it is motivated by a desire to hold non-profits accountable for their use of public money. At other times, ongoing program improvement, buy-in from partners, or program expansion might be the goal. Often, a single evaluation project has multiple goals.

According to Zafiropoulos (2005), a general distinction separates the evaluation methodology of research into qualitative and quantitative. The qualitative research method describes, decode, translate the meaning of a phenomenon, explain, analyze factors and causes of behaviors and phenomena. Quantitative research method use numerical measurements, measure frequencies, incidence rates, effect sizes, etc. of phenomena. Both categories of research work autonomously each, but complementary to the scientific investigation, each contributing differently and depending on the subject of investigation, the qualitative or quantitative method precedes.

Furthermore, it is important to mention that inside the UN's International Strategy for Disaster Reduction (2005), there was an Evaluation and Strengthening of Early Warning Systems in Countries Affected by the 26 December 2004 Tsunami. Crucial results have been made for the Ingrate Risk Management, for the Public Awareness and Education, and for community based approaches. The paper was underlining the need to have sophisticated early warning systems for natural disasters in order to allow the authorities and the citizens to be fully prepared to cope with the challenges of the natural event. Furthermore, in the Early Warning Systems for Natural Disaster Reduction by Joseph Zschau and Andreas N. Kuppers (2005), it is described the methodology that has already been used to achieve a reduction in damages from natural disaster. The paper also presents ways to achieve a

better coordination and communication between authorities and the citizens that have been affected by the disaster and are in need of assistance.

Another important early warning system is Copernicus, a European emergency management service which includes EFAS and EFFIS important platforms for early warning notification for floods and forest fires. Additionally, according to Lowe Dianne et al (2011), there are only 12 countries in Europe that use Heatwave Early Warning Systems (HEWS) and each one has developed its own unique system. This divided effort, even inside the EU, is very important to be mentioned because natural disasters, such as heatwaves, can hit many countries simultaneously. As such, a unified and common system can inform organisations, authorities and people faster and more accurately rather than the individual, national-level systems, that sometimes may be even incompatible from each other.

2.1 Other efforts

Early warning and alert systems are today a useful tool and are being used worldwide from different counties, organizations, authorities, scientists and citizens. One of the most global shocking events, the 2004 Tsunami in Indonesia was a breaking point that highlighted the necessity of an organized alert system that will make use of modern technology and information extracted from the social media.

Currently, there are several early warning and alert systems which inform citizens and authorities for an imminent disaster. Below the evaluation methodology that was followed in some early warning systems is presented:

1. Community-Based Flood Early Warning System **(CBFEWS)**in Philippines [1] and Himalaya [2] and project Digital Early Warning Systems to Save the Lives and Livelihoods of Communities of Bangladesh (**DEWS**) according to Faizul Kabir and Golam Kabir (2015) the evaluation has been framed to focus on the following objectives:
   - Assess how effective the project was in implementing its activities and reaching its 'output' targets.
   - Analyze how effective the project was in reaching its 'outcome' targets.
   - Assess which approaches, interventions and activities have proved to be most effective and why.

[1] https://www.slideshare.net/IFRCCOMMS/session-4-establishment-of-communitybased-flood-early-warning-system-cbfews

[2] http://lib.icimod.org/record/32318/files/icimodCBFEWS016.pdf

- Articulate and package the lessons learned and recommendations from this innovative initiative to set up an early warning system (EWS) for deep sea fishermen.

The evaluation methodology that was followed is:

- Visits to the DEWS project sites
- Interviews of the beneficiaries
- Workshop attended by project stakeholders
- Discussions with the stakeholders and implementing partners and
- Telephone calls with organizations not covered by the interviews and visits.

2. **SafeLand** (2012)[3], a European Union program, in which the evaluation of methodology that was followed is:
   - Questionnaire in ground basic, airborne and space borne techniques
   - Interviews of the beneficiaries
   - Seminars and Workshops
   - Multilevel testing
   - Training.

3. According to **Neil Dufty** (2015)[4] the National Emergency Warning Principles were used as a general evaluation framework to examine the progress made with early warning systems in Australia since 2005. The data for the evaluation was collected from a variety of sources including:
   - Consultations with emergency services agencies from Australian States and Territories. About half of these agencies responded to the request for data for this evaluation. They provided agency strategies, relevant reports, articles and papers
   - Post-disaster evaluations, including reports from royal commissions, government inquiries and after-action reviews
   - Between-event evaluations, conducted to gauge progress in particular improvements e.g. recommendations from a previous disaster inquiry.

2.2 General principles of evaluation methodologies

Today, there is a great variety of evaluation methodologies. Each method offers advantages and disadvantages. To achieve more validity and objectivity reasons, a combination of one

[3]SafeLand project https://www.ngi.no/eng/Projects/SafeLand
[4]https://ajem.infoservices.com.au/items/AJEM-29-04-09

or more evaluation methodologies is considered to be necessary (Bell 1997). The person/persons who will choose one or more evaluation methodologies must consider some factors such as:

- General principals of the study
- Characteristics of the people that will be involved
- Cost of the evaluation methodology
- Validity of the evaluation methodology
- Reliability of the evaluation methodology
- Accuracy of the data collection
- Time for the data collection and their evaluation
- Availability of personnel

Evaluation involves a judgement of interventions according to their results, impacts and needs that they aim to satisfy. It is a systematic tool which provides a rigorous evidence base to inform decision-making and contributing in making activities more effectively, coherently, usefully, relevantly and efficiently. Evaluation also enhances transparency, learning and accountability. To achieve this, the evaluation standards aim to ensure relevant and timely assessments of high quality. Also they guarantee that the evaluation results are communicated to decision-makers and other relevant stakeholders in a clear and transparent manner to facilitate their use. An evaluation methodology provides a formal opportunity for grantees to document the steps they will take to conduct a program assessment. An evaluation plan typically includes descriptions of the following:

- Purpose of program that will be developed
- Partner assessment
- Evaluation goals
- Evaluation questions
- Data collection plans
- Data analysis plans
- Dissemination and reporting activities
- Other evaluation products
- Timeline and budget
- Staff responsible for each evaluation activity

As stated above, there is a great variety of evaluation methodologies. For this project, a combined evaluation is more preferred for better results and to document as many as possible of the issues defined in the use cases.

Finally, an important issue is to have a concrete plan which is well organized and described in order to present reliable and valid methods of an evaluation approach and generally for the impact of the platform to the society and the users.

3 Initial implementation plan of the pilots

In order to better develop and evaluate the program, especially in its initial stages, some pilot use cases are selected. Those use cases intend to highlight key steps and aspects of pilots based on a specific timetable.

Below are described the use cases of the 3 Pilots, which will be carried out by end-user partners, their proposed initial timetable steps, the way they will be performed, the preparation steps, the methodology to be followed, the participants and the equipment that will be used per pilot.

3.1 Pilots description and execution

There are three different pilots; each focused in a different emergency (flood, forest fire and heat wave). The main purpose of these pilots is to test and evaluate beAWARE system during the management of an emergency incident caused from extreme weather conditions. Another important aspect will be the evaluation of the beAWARE system and the given response to the emergency. beAWARE system will provide support in all phases of the emergency, improving the response of disaster planners and first responders.

Flood scenario will be set in Vicenza (Italy), and beAWARE system will receive inputs from the already installed water level sensors and the weather forecast in order to detect any dangerous situation related to floods.

In the case of forest fire, the pilot will be set in La Devesa del Saler (Valencia, Spain), and its aim will be to detect dangerous situations, not only according to weather forecast inputs (pre-emergency), but also when a fire spot has already started (emergency). Concerning the heatwave, the pilot will be set in an urban area of Thessaloniki, and the beAWARE system will provide an early warning of a heatwave event, as well as managing the heatwave event more adequately.

The initial Use Cases that will be tested, as they are presented and analyzed in D2.1, are:

- UC_101: Declaration of the attention status and continuous monitoring of flood forecasting
- UC_102: Management of new flood emergencies
- UC_204: Evacuation management during an emergency
- UC_305: Management of Places for relief

Moreover, on the following table the initial steps that will be followed during an emergency operation are presented.

TIME	INPUT	ACTION	OUTPUT	ACTOR	Definition of Actors
0.00	Forecast in the next 30h, max. water level in the river	Check with Thresholds	Crisis Classification	beAWARE System	
	Crisis Classification	Send warning to the Authority	Notification	beAWARE System	
	Notification	Activate Crisis Management (CMO) Standard Protocols - beAWARE	Decision to mobilize resources (Flood: Control room opened)	Authority	
12.00	Decision of the Authority	Activation of first responders - PSAP through beWARE	Activation message: "be ready for the next 24h" to all first responders	Authority	
12.05	Message received from authority to first responders	Push " CONFIRM" to declare availability in beAWARE app.	YES/NO Availability of first responders Classification	First responders	
	Crisis Classification	Through beAWARE app./System inform citizens/Media	Messages: "We are in emergency 1"		

Table 1.Initial Steps on an emergency situation, flood example.

3.1.1 Flood

The pilot is foreseen to take place in the city of Vicenza in Italy, with the participation of the entire project and other relevant partners. In detail it will consist of Civil Protection volunteers' exercises during which flood event crisis will be simulated.

The exercise for the pilot of flood will involve AAWA team and the other members of the "COC - Municipal Operative Center", supporting the Mayor in the emergency phase:

- Municipal group of Vicenza Civil Protection
- National Association of Carabinieri
- National Alpine Trooper Association
- Coordination of Voluntary Associations of Civil Protection, Province of Vicenza
- Italian Red Cross
- Fire Brigades
- Local Police
- Public authorities at different levels

The PSAP will be located in the Municipality building of Vicenza. Volunteers First Responders and citizens will be recruited for practical activities in the field; they will be equipped with mobile phones and trained to use the beAWARE technologies. In particular the volunteers will be invited to take part in 2 collective trainings with duration of half a day each one; the program of the course is described as follows:

Day 1:

- introduction to the beAWARE project
- explanation of the use of citizen' app and the first responders' smartphone apps
- description of the Flood Exercise

Day 2:

- registration of the volunteers involved in the exercise and delivery of the smartphones
- installation of the smartphone apps
- assignment of tasks
- Internal test of the apps

At the end of these meetings each of them will receive a user guide for platforms and Smartphone applications and a list of activities to be performed during the exercise. Technological partners will have to support AAWA in such initiatives attending the meeting and preparing in advance handbooks and presentations.
The purpose of the exercise will be to test the beAWARE system by implementing all the use cases described in D.2.1:

- UC_101: Declaration of the attention status and continuous monitoring of flood forecasting
- UC_102: Management of new flood emergencies
- UC_103: Monitoring river water level and assignment of tasks to first responders
- UC_104: Evaluation of the execution of tasks
- UC_105: Monitoring rainfall
- UC_106: Monitoring river breaking/overtopping and assignment of relative tasks
- UC_107: First responders monitoring
- UC_108: Flood forecasting alerts

In detail an historical flood event, happened in Vicenza, will be considered as the reference scenario.

The exercise is composed by 2 parts, the first one before the flood and the second one during the flood.

First part of the exercise

In the first part, the exercise will start with the simulation of an imminent flood. Based on the last flood forecast, the system will issue an alert, in order to warn decision makers and authorities to start preparing for proactive measures. Thanks to this message the Mayor will activate the structure for the emergency management (COC) in order to be ready to manage the incoming emergency.

Additionally, an alert will be send to citizens through the mobile application along with information on what to do and don't do, so that people know how to behave and take proactive actions as well.

Second part of the exercise

During the emergency, incident flood reports can be sent:

- By citizens from their Mobile App(with attached video, audio and images)

- By fixed cameras recording the rainfall intensity, the water level in a river section, the flooded areas and the status of the embankments

In the meantime the Authority will assign to First Responders predefined tasks that are well specified inside the Risk Management Plan of the city based on well-established triggers. In addition the Authority will have to check the incident flood reports, previously sent by citizens or cameras, by assigning to first responders new tasks. Later the Decision Makers, who are represented by the Mayor of the City, will be interested to check the status of the execution of the tasks previously assigned.

Additionally, new alert will be send to citizens, approaching previously verified dangerous areas, through the mobile application along with information concerning alternatives safe places and ways.

For the first pilot Use Case 101 and Use Case 102 are selected. The first one, as stated on D2.1, concerns the declaration of the "ATTENTION" status and the continuous monitoring of the forecasted level of risk. The second one, as also stated on D2.1, concerns the management of new emergencies reported by citizens and first responders. Furthermore, two block diagrams are presented below for those two use cases.

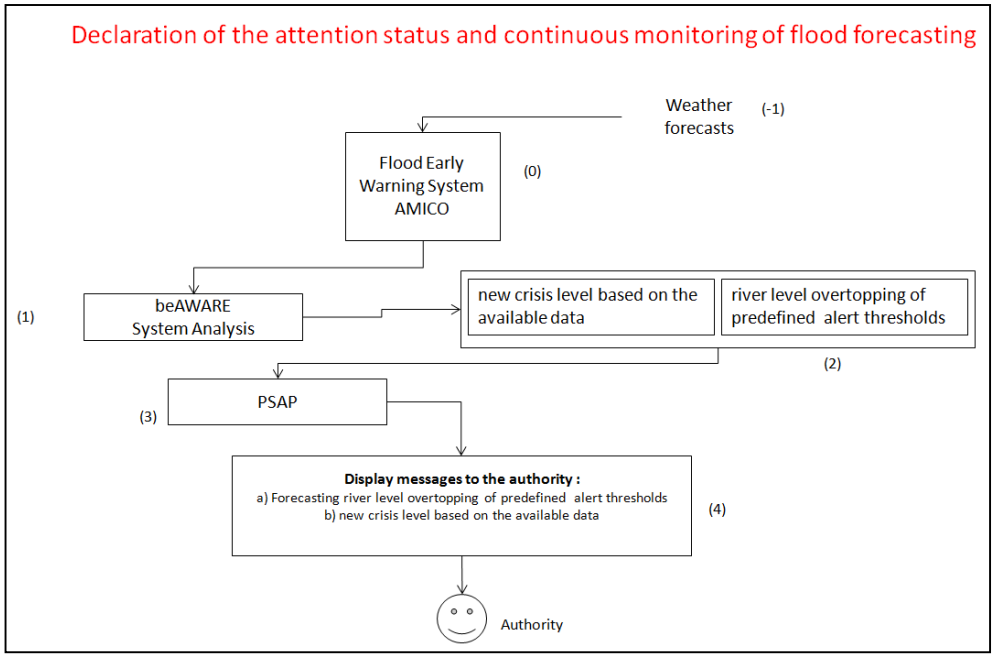

Figure 1.Block Diagram of the UC_101

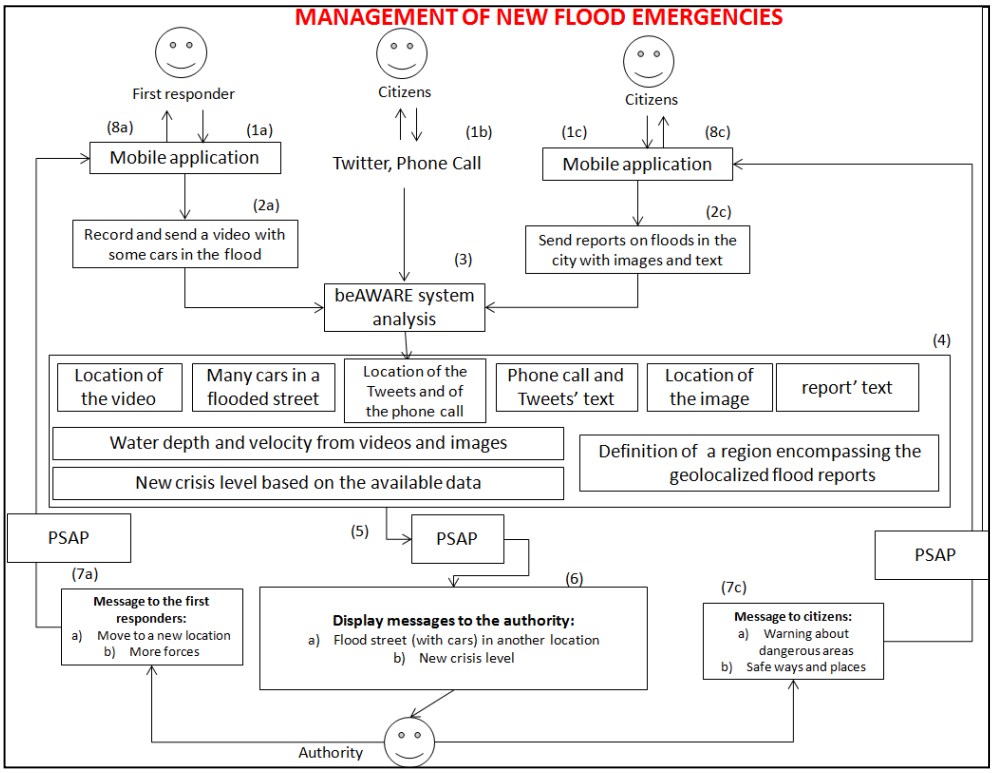

Figure 2.Block Diagram of the UC_102

For the implementation of flood pilot, a categorized table is followed with the majority of the information needed.

Pilot	Flood
Responsible partner	AAWA

Pilot	Flood

Participants	• Volunteers:
	  ○ First Responders:
	  ○ Citizens
	• Decision Makers: AAWA team and the other members of the "COC - Municipal Operative Center", supporting the Mayor in the emergency phase:
	  ○ Municipal group of Vicenza Civil Protection
	  ○ National Association of Carabinieri
	  ○ National Alpine Trooper Association
	  ○ Coordination of Voluntary Associations of Civil Protection, Province of Vicenza
	  ○ Italian Red Cross
	  ○ Fire Brigades
	  ○ Local Police
	• Public authorities at different levels

Potential Target Group	• Regional and National Civil Protection • Provincial Fire Department • Italian Environmental Ministry • Citizens
Command and Control Room place	City Hall of Vicenza
When will the pilot take place	• M18: A laboratory test to test the first prototype (June 2018) • M23-26: A field trial to test the 2nd prototype (December 2018- February 2019) • M34-36: A field trial to test the final system (October 2018-December 2020)2 Demonstrations will be organized at M28 (April 2019) and M36 (December 2020) to demonstrate the 2nd prototype and the final system

Pilot	Flood
Communication infrastructure/ Computing infrastructure displays, workstations, servers/ End devices available for the pilots (types and quantities):	• Vicenza Municipal LAN • 2 laptop • personnel smartphones and tablets(>30) • 2 multi-purpose monitors
Pilot parts	During each exercise the following activities will be carried out: simulation of an imminent flood and consequent test of the system's functionalities during the emergency

Table 2. Pilot summary table: Flood

3.1.2 Fire

This exercise will involve personnel from Valencia Local Police, Valencia Fire and Rescue Service, and Civil Protection. Besides these end users, some citizens will participate in the exercise in order to test the beAWARE early warnings data input and overall functioning. The PSAP will be located at Local Police headquarters.

The exercise will take place in La Devesa, Valencia. The PSAP will receive a fire warning, preferably from a citizen through the beAWARE system, and the fire protocol will be activated, according to its severeness, informing and mobilizing the correspondent personnel. An advanced Post will be established close to the affected area, and it will be in constant coordination with the PSAP.

The main tasks to be done by the involved personnel will be the following:

- Valencia Local Police will provide its PSAP and will perform the required emergency tasks.
  Responsibilities: Spotting/locating the fire, clearing the zone and supporting firefighters in their tasks, especially with regards to allow them open access to the zone, traffic incidents and rescue operations.
- Firefighter personnel will perform all required fire and rescue operations.
  Responsibilities: Extinguishing the fires, rescue of people in danger. Using the developed technologies through the team leaders and incident commanders.
- Civil Protection Volunteers will support the firefighters in their tasks, providing extra material, catering etc.

PREPARATION OF THE EXERCISE

Training

PLV will work together with local stakeholders and volunteers, registering the personal information of the involved personnel, providing them with the use guidelines of the beAWARE system, as well as a list of activities to be performed during the exercise.

Potential contents of a list of activities form (for first responders):

- GROUP LEADER
- BRIEF DESCRIPTION OF THE ACTIVITY TO BE CARRIED OUT: Sends reports about the status of the assignments containing the following messages: "…"
- STARTING TIME OF THE ASSIGNED TASK
- ENDING TIME OF THE ASSIGNED TASK
- FIRE SPOT: Fire spot will be established around 39°22'34.3"N 0°19'36.9"W
- PSAP LOCATION (CISE at Local Police Headquarters): 39°28'04.0"N 0°23'56.8"W

For the first pilot Use Case 204is selected and some specific success indicators are:

- People saved
- Cooperation
- Response time of the system

As stated on D2.1, this use case concerns the evacuation management during a forest fire. In case of population at risk, the authority will order the evacuation and first responders will be given instructions to facilitate the evacuation.

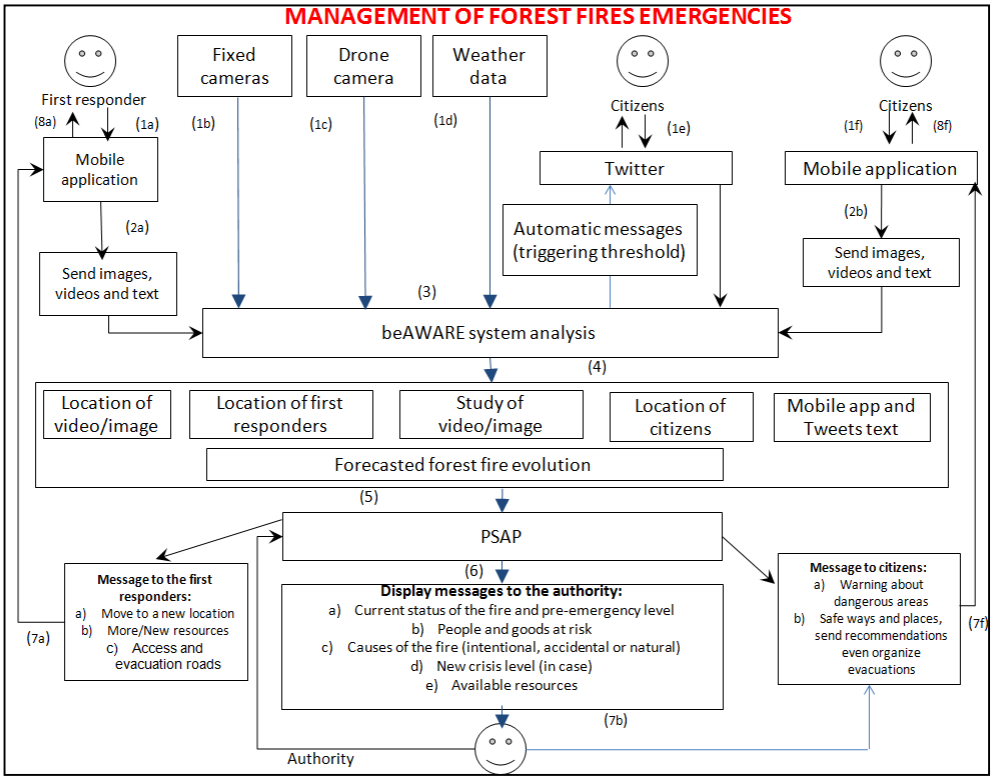

Figure 3.Block Diagram of the UC_201

For the implementation of fire pilot, a categorized table is followed with the majority of the information needed.

Pilot	Forest Fire
Responsible partner	PLV & FBBR

Pilot	Forest Fire
Participants	• Valencia Local Police • Valencia Forest Fires • Civil Protection • Technical staff (Albufera Natural Park) • First responders at national scale • Citizens • Cross Red
Potential Target Group	• Neighbourhood associations • Civil Protection • Fire Department • Citizens (holidaymakers, visitors) • Cross Red • Technical staff of Valencia City Council • Companies in emergencies research • Universities • Authorities
Command and Control Room place	Valencia Local Police Headquarters

Pilot	Forest Fire
When will the pilot take place	• **Laboratory test:** M18 (June 2018) to test the first prototype. • **2 Fields trials:** ○ M21 (September, 2018) to test the 2nd prototype ○ M30 (June, 2019), to test the final system • **2 Demonstrations:** ○ M23 (November, 2018) M35 (November ,2019)
Communication infrastructure/ Computing infrastructure displays, workstations, servers/ End devices available for the pilots (types and quantities):	To be confirm with Fire Department & our Command Control
Pilot parts	1) Simulation of a forest fire & evacuation 2) Test the system's functionalities during the forest fires

Table 3. Pilot summary table: Fire

3.1.3 Heatwave

The pilot is foreseen to take place in the city of Thessaloniki, with the participation of the entire consortium and other relevant partners.

The exercise for the pilot of heatwave will involve HRT members and members from other Civil Protection organizations. An effort will be made to also involve:

- Municipal social services, that are responsible for the management of places for relief during a heatwave
- Civil protection, on both municipal and regional level
- Citizens

The purpose of the exercise will be to test the beAWARE system and especially those functionalities relative to the use cases that will be chosen to be tested during the pilot. The exercise will consist of 2 parts. The first part will take place before the event and the second will take place during the event.

In the first part, the exercise will start with the simulation of an imminent heat wave. Based on the meteorological data, the system will issue an alert, in order to warn decision makers and authorities to start preparing for proactive measures

Additionally, an alert will be send to citizens through the mobile application along with information about what to do and don't do, so that the people know how to behave and take proactive actions as well.

Furthermore, based on the meteorological data, the system will issue a warning of a high fire risk in the nearby forest.

In the second part of the exercise, the scenario will be to test the system's functionalities during the heat wave. People will be seen to be in distress and trapped in elevators and in houses without A/C (e.g. elder people). They will declare their emergency through the system, and first responders will be sent to respond. The managing of the first responders will be done through the platform.

Moreover, the municipalities will provide dedicated places for relief especially for specific target groups such as elders, young children with their parents etc. The occupancy will be monitored through the platform and relative information will be displayed.

Finally, by monitoring the traffic conditions, the system will show jammed areas so that the PSAP will be able to direct first responders more efficiently to an event.

Because heat wave is a complex and multidimensional situation, many organizations and authorities are involved and for this reason HRT, has identified a list with the proposed target groups that will be involved in the specific pilot.

For the first pilot, Use Case 305 Management of Places of Relief is selected. Some success indicators are selected which are presented below.

- Number of places managed
- Number of people to use the app
- Comparison of estimated occupancy vs real occupancy
- Even (or uneven?) distribution of people who used the places

As stated on D2.1, this use case concerns the monitoring of the places of relief offered to people as a shelter during the day, in the period of a heatwave. Moreover, a block diagram is presented below for the use case.

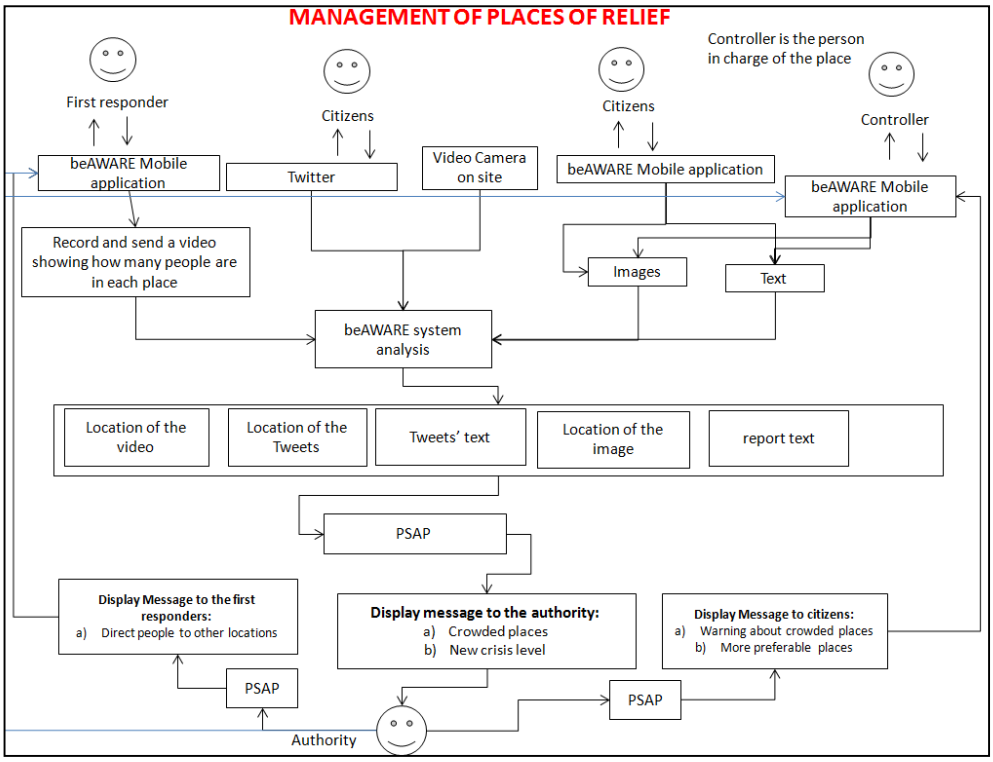

Figure 4.Block Diagram of the UC_305

For the implementation of heat wave pilot, a categorized table is followed with the majority of the information needed.

Pilot	Heatwave
Responsible partner	HRT
Participants	• HRT personnel/volunteers • Local decision makers involved during the heatwave emergencies • Citizens • Other first responders (apart from HRT) • Research community • Civil Protection Authorities at National scale

Pilot	Heatwave
Potential Target Group	• Regional Civil Protection • Municipality of Thessaloniki – Civil Protection Office • Fire Department • Other volunteer first responder teams • EKAV (National Ambulance Service) • Citizens
Command and Control Room place	HRT Headquarters
When will the pilot take place	• 1st part can take place in real time – M18-M21, test 1st prototype (July-October 2018) – M30-M33, test final system. (July-October 2019) • 2nd part can take place – M19 as a lab test (August 2018) – M31 as field pilot (August 2019) • Demonstrations – M26 (December 2018) – M36 (December 2019)

Pilot	Heatwave
Communication infrastructure/ Computing infrastructure displays, workstations, servers/ End devices available for the pilots (types and quantities):	• VDSL line (50 MBps), Cellular, HRT's radio Network • 10 Laptops / smartphones / >30 Radios • 2 Workstations • "Hermes" mobile communication center • 3 shortwaves (HF) transceivers • 4 transceivers VHF/UHF mobile • 2 transceivers VHF marine • 1 transceiver VHF Air-band • 1 transceiver CB • RF Scanner 1,6 MHz- 3 GHz all mode • RF direction finding system • Aircraft transponder mode S receiver • 2 multi-purpose monitors • 2 laptop
Pilot parts	1) Test the system's functionalities **before** an imminent heatwave 2) Test the system's functionalities **during** a heatwave

Table 4. Pilot summary table: Heatwave

4 Evaluation

4.1 Common Methodologies

There is a large variety of methodologies that can be followed such as: Questionnaire, Interview and Experimental method.

Questionnaire is a research tool that is used successfully and effectively in sample surveys and not just in them. It is a series of questions that the respondent is asked to answer and which measures the social, economic and general characteristics of the team he is investigating. The quality and accuracy of the survey of the characteristics surveyed depends on the questionnaire itself, the rating scale it uses, the way it is filled in, the process of completing it (Zafiropoulos, 2005).

The questionnaire is designed for each survey and is only applicable to it. Its design and construction is a particularly difficult and critical phase of the research process since it has been selected as one of the methods of collecting information. The key questions the investigator has to answer before planning are:

- How to fill in the questionnaires and
- What kinds of questions will be there?

The answers on those questions will determine the extent, format and processing of the questionnaires.

Moreover, one important issue is how the questionnaires will be sent to the respondents. It can be sent, by e-mail, by post, by interview scheduled questionnaire and on the field of practice questionnaire.

There are certain types of questions for the questionnaires.

1. Based on their content
    a. Filters of questions
    b. Direct or Indirect questions
    c. Knowledge questions
    d. Queries or intentions questions
    e. Explanatory or interpretative questions
2. Based on the type of the answer
    a. Closed type questions
        i. Simple choice scale
        ii. Multiple choice scale

iii. Evaluation scale
iv. Likert scale
v. Importance

b. Open type questions
c. Prefabricated questions or questions with prefabricated answers

Questionnaires will be given to end users, PSAP operators, citizens and other organizations that want to learn about the platform. Some examples of them are:

- “Is there a good accessibility and usability of the platform?”
- “Is there availability of the data provided”
- “What type of information do you believe should be added?”
- “The platform helped you to take a better decision?"
- “Is there a good communication between PSAP and rescuers?”
- “Do you need as PSAP more information to collect and from whom?”
- “As PSAP what type of information you send to rescuers and what to citizens?”
- “As a citizen what type of information you want to receive and how to be demonstrated?”

Another methodology that is followed is an Interview. **Interview** is a qualitative method of data collection. It is oral, dynamic communication between an interviewer and an interviewer, between a researcher and an "object of research", that is, an organized debate. Its key element is the experiential approach by the interviewee, the recording of his / her experience, his / her perceptions, the exploration of the attitudes and behaviors of the subject / group under investigation and the methodologically correct implementation of it. It is possible to investigate the causes and structural elements of Behavior. It presupposes a methodological strategy, flexibility and immediacy. It is also necessary to organize in advance both the subject under investigation and the preparation of the interviewer. In more detail the preparation contains (Filias 1996).

- Converting the aims of the survey into individual questions
- Converting the questionnaire according the interviewees for a better understanding
- Informing the interviewers to clearly present the questions to the persons and to predispose them to spontaneously transmit the information / knowledge of interest.

The interview may take several forms:

- Structured interview with predetermined questions about number, content and order. Responses are recorded faithfully, and the researcher at the end of the discussion should make his / her notes in such a way as to indicate the particular circumstances of each interview that help in the interpretation of the answers.
- Unstructured interview with a Distinction of “Free Interview” that means discussion on a generally referenced topic with a lesser degree of orientation, and a focused interview that aims to draw attention to a living experience and results that have brought some special stimuli. This is a loose discussion, in which the respondent can respond as the investigator asks the researcher to skillfully re-establish the issue when there are plagiarisms. The localized interview can also be made to more than one person at the same time as a focus group, and it concerns the people who took part in the phenomenon or the situation being investigated.
- Direct or Indirect Interviews are forms of interviews where questions are asked directly in relation to the subject investigated or indirectly so that the actual value of the discussion is given by the information indirectly provided by the respondent.
- Repeated Interview, which is the repeated conversation with the same people (random choice) in the various stages of evolution of a phenomenon / situation, so that through the repetition, the progressive influence / change of attitudes and behaviors can be explored.

Furthermore, interviews can be divided in two phases. The first one is with end users and organizations that are interested to use the platform and will give their opinion about the platform, what tools would they want to have and what information would they want to be provided to them. In the second phase of interviews, there will an evaluation of the platform and a deeper discussion with the users in order to improve the platform.

Another methodology that is followed is the **experimental method**. This method can be divided in two sub methods, the theoretical and the practical sub methods. Another approach is that experimental methods are divided in laboratory and on the field tests. The laboratory test / training is more theoretical and will be given in the initial steps of the platform, for a first evaluation and report, in the middle of the period for a change report and in the end for training of more people before the field test. The field test methodology will be able to be done in pre-test training, in order to have a better understanding of the interface, the use of the platform and its tools, to note down and potential problems, changes or even ideas and finally on the pilots.

As already described in section 2.1, there various types of evaluation methodologies are used in different project, such as visits, workshops, multi-training, telephone calls and others. In the relative section, the questionnaires, the interview and the experimental method are analyzed. However, there is not an optimal and ideal evaluation methodology that will be followed therefore it is proposed to use more than one, in order to have more subjective and multidimensional answers.

4.2 Proposed evaluation approach

4.2.1 Impact of the system

The focus of the evaluation methodology will be given primarily to the impact of the system, i.e. how it affects the management of the emergency once the system of beAWARE technologies is completely operative in the pilot.

It will have to be evaluated by comparing the management of an emergency before and after the implementation of the beAWARE System. The starting scenario is the management of an emergency without the developed technology (baseline). The same exercise is planned to be replied by considering the availability of developed technology, in order to evaluate the following indicators:

- **Time of tasks execution**. This will be evaluated by comparing the time that is required to perform a task first without and afterwards with the use of beAWARE. The objective will be to show that the use of beAWARE reduces the required time to perform certain tasks.
- **Number of assigned tasks completed during an emergency** (e.g. number of completed task at a fixed time). The objective of beAWARE is to optimize the management of the first responders' teams in order to perform the maximum number of the activities with the minimum number of first responders. It is expected a greater number of completed task in less time.
- **Number of warned citizens by the beAWARE system during an emergency at a fixed time**: beAWARE aims at diffuse quality alert messages to an ever increasing number of people during an emergency in order to reduce their exposure to the risk, thanks to the developed technologies. Every person, registered to the beAWARE platform, is in fact expected to receive a warning every time a dangerous situation is recognized by the system).
- **Number of saved elements at risks during the emergency** (people, cars, buildings, monuments involved in the scenario).

Asan extra approach, Borland (2000) is proposed, an extensive user-cantered evaluation and trials. User-centered evaluation emphasizes on the role of the user rather than the system and considers the needs and limitations of the end-users. The focus lies in testing the system or specific modules in a near-real-life scenario by giving test persons realistic tasks in a staged environment. All this procedure is important to be carefully planned and methodologically executed, which means it will take a considerable amount of time and effort. Also, a user-centered approach of the whole project is necessary in order to develop and produce genuinely user-friendly and ultimately exploitable systems. There is no standard user-centered evaluation method; in fact, the respective methodology needs to be chosen according to the specific use case scenario and to the specific functionality that needs to be tested.

Borland (2000)also propose, an “Interactive Information Retrieval (IIR) Evaluation Model”. This is centered on the idea of defining a realistic scenario for the user, called a “simulated work task situation”. A “short cover story with situational relevance” will be created for an easier simulation. The scenario may change over time and may be context-dependent according the needs. Borland defines three parts of the IIR evaluation model:

1. A functional, realistic system, with realistic data and realistic users (as subjects);
2. Simulated work task situations that are empirically-based applications (i.e., realistic);
3. Performance measures that allow for non-binary scores, i.e., not just "relevant / not-relevant" but including also partially relevant options and possibly ranked output and evaluation.

As Stone et al. 2005 stated, the focus of the evaluation should be on the design of the platform in order to have more user friendly interfaces. Moreover, there should be a clear and well-justified decision of which features to include. A well-developed interface will be proven crucial for the success of the whole process.

The direct and indirect impacts of disasters can be multiform and extend over various time scales. During the event, human lives and properties are at risk, and authorities’ primal goal on the short term is to protect people, their homes, and any critical infrastructure. Damage to transport networks, including harbors and airports, as well as energy grid infrastructure, can produce medium-to long-term interruptions with negative consequences for the competitiveness of the local, regional, and national industries. For long term, there is the protection and restoration of natural environment and large affected areas.

In these situations, time is one of the biggest enemies and many natural disasters can occur with devastating effects. A system like beAWARE is designed with the view to have a successful impact on the society in the most efficient and reliable way.

The most important factors that will indicate the positive impact of the system are:

i. The reliability and the accuracy of the data,
ii. The ability to identify a potential disaster,
iii. The production of an alert coming both from the platform and the citizens, and
iv. The capacity to inform accurately the authorities and rescuers for a potential threat or incident.

This impact is difficult to be measured in general terms, but as stated above, there are some aspects that are of high importance. It goes without saying that loss of life is the number one priority and concern for authorities. In the floods of August 2017 in Texas, early warning systems like FEMA's informed local authorities, rescuers and citizens to be prepared[5] and thanks to the mass social media alerts, like #Harvey, #HarveyStorm, #HarveySOS and #HarveyRescue and danger notifications of citizens in the area like in figure 5, it was easier to track any citizens in need and to rescue them faster.

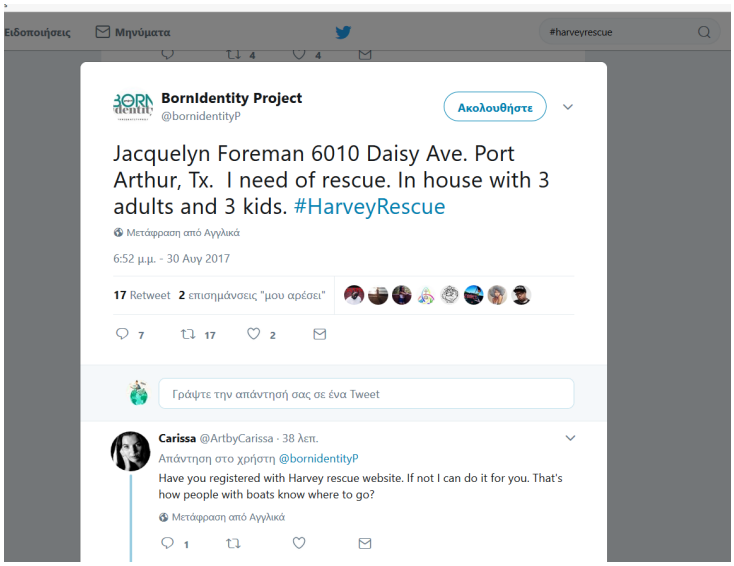

Figure 5. Rescue notification in social media

An early warning system, which connects authorities and citizens, receives accurate data and filtered information from social media, and coordinates the rescuing efforts, can have a positive impact as a support tool both for the authorities and for the citizens in need.

[5]FEMA alert notification https://www.fema.gov/news-release/2017/08/24/fema-encourages-residents-follow-directions-state-local-and-tribal-officials

4.2.2 User interface – User experience

The design of a user interface (UI) is the major factor that determines the user experience and the user's decision on whether to keep on using a certain product or abandon it. Evaluating the usability of a user interfaceand user experience can be done in *a subjective way* by depending on the opinions of users and experts to determine the quality of a system or in *an objective way* by using certain rules and metrics to decide on the quality of a given system (Mosqueira-Ray). Both the opinions of users and data used by rules to evaluate the system can be collected either *manually or automatically*. Subjective evaluation of a user interface is done by analyzing the opinions of users and experts to give a sound judgment accordingly.

One type of the subjective evaluation of the user interface is the *heuristic evaluation* which is an inspectional method, where a certain number of experienced evaluators determine to which extends the design of the UI follow an established guidelines (Heuristics). Those heuristics are set of rules defined by J. Nielsen who is the author of the technique (J. Nielsen). Another type of subjective evaluation like usability scale method uses well defined and standardized questionnaires filled by the users post using the targeted user interface, and then results are analyzed to evaluate the user interface according to those users' experience.

Observing the behaviours of the user while interacting with a UI can reflect the actual feedback of that user, and in turn give *objective evaluation* of the interface. This can be achieved during laboratory and on the field tests and evaluate his ability of multitasking, to understand the given data and to give the needed data to the system. All those can be measured with a questionnaire with indicators such as time to complete a task, facility to complete a task. If enough users are involved in such evaluation, an index can be generated from their experiences to compare different user interfaces. Such an evaluation can be used as an indicator of the level of the machine intelligence since it tests the ease of use in terms of user interactions. In the work done by Ahmet et.al; they have compared the results obtained from user interactions evaluation with the results of survey based evaluation. They have collected data from users with similar background and experience whiles the tested systems doing the same tasks. Then, they have used a fuzzy logic system to evaluate the systems. The fuzzy logic inference system designed was a function of:

- Complexity of each subtask in the main task
- UIQ (User Interface Quotient) data
- Total number of subtasks
- Difficulty of data transfer between the machine and human

Based on this, they got an index for each UI objectively and they compare it with the results from a survey that was given to the same users involved in the first test. Their results showed that their methodology matched the survey in 70% of the cases. The researchers declared that the coverage of the methodology can be improved if the user interactions logging is done automatically, and more users are involved in the process.

Furthermore, according to Cognitive, evaluating user interface is becoming an important skill. The two main techniques for evaluating a user interface are:

1) **Empirical Evaluation** (testing with users) and
2) **Heuristic Evaluation** (based on a set of rules).

While empirical evaluation is by far the best technique, mastering **Heuristic Evaluation** is also a necessary skill.

Heuristics are rules developed over time by trial and error that have shown to work. For example, when you play chess, you are better off using established initial moves. Those initial chess openings are heuristics that have been proven to work over time.

In user interface design, one of most frequently referred lists of heuristic is the Jacob Nielsen's ten Heuristics.

- **Visibility of System Status** (The system should always keep users informed about what is going on. For example, the progresses during a file transfer. Provide immediate feedback, etc).
- **Match between system and the real world** (The system should speak the user's language, with words, phrases, metaphors and representations familiar to the user).
- **User control and freedom** (Supports undo and redo actions. Allow user to override the system).
- **Consistency and standards** (Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow behavior user except and platform conventions).
- **Error prevention** (Prevents a problem from occurring in the first place. Be permissive to various entry formats).
- **Recognition rather than recall** (Making objects, actions, and options visible. Make navigation visible and make it easy to go back, go to home page and quit).
- **Flexibility and efficiency of use** (Speed up the interaction for the expert users with accelerator. Experts prefer few screens with a lot of information and a lot of flexibility in the methods, while novice users prefer going through step by step sequences with little information).

- **Aesthetic and minimalist design** (Dialogues should not contain information that is irrelevant or rarely needed. In other words, present only necessary information for the task, rarely used information should be in secondary screens or windows).
- **Manage errors** (Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.
- **Help and documentation** (Use prompt and contextual help related the task, allow easy search).

As Law E. et al (2009) stated User experience (UX) evaluation or user experience assessment (UXA) refers to a collection of methods, skills and tools utilized to uncover how a person perceives a system (product, service, non-commercial item, or a combination of them) before, during and after interacting with it. It is non-trivial to assess user experience since user experience is subjective, context-dependent and dynamic over time.

An individual method can collect data about a set of specific constructs of user experience. For instance, usability testing is used to collect data about usability construct. Methods also differ if they are to measure a momentary or episodic experience (i.e., assessing how a person feels about a specific interaction episode or after executing a task) or an experience over time, also known as a longitudinal experience. UXA methods can be classified in three categories: *implicit, explicit and creative methods*. The measures and methods for these categories are described below.

Implicit methods

Implicit methods of UX research focus not just only on what the users say, but also on what the user cannot express verbally. Many available tools can assist in the implicit evaluation, in particular to gather implicit or objective data. When available, UX researchers utilize state of the art equipment to uncover all aspects of the experience.

Examples of implicit evaluation methods and tools:

- Eye Tracking
- Attention Tracking
- User Tracking
- Task and reaction measurement Galvanic Skin Response or skin conductance
- Electroencephalography EEG
- Observation Studies: participant observation where observers monitor the participant's reactions such as facial and other gestures, the tone of voice or other body language cues.

Explicit methods

Explicit methods of UX research explore what the user is consciously aware of getting them to reflect on their own feelings or thoughts, and gather their views and opinions. An important aspect of explicit methods includes usability testing and emotion evaluation.

Creative methods

Equally important to implicit and explicit methods are the creative methods that the user researcher can utilize in order to bring together the design team's view, as well as the target market's dreams, aspirations and ideas of optimal design. These activities are more open and allow people to either co-create with the engineers/designers, or to use their imagination to express their ideal system.

Examples of creative assessment methods:

- Co-design activities
- Creativity Workshops
- Paper Prototyping/ Wire-framing mind mapping
- Card sorting
- Personas

To be more specific there can be some procedures/steps which will help PSAP, rescuers on the field and citizens during an event:

- Log in to the system,
- Data incoming,
- Upload data, images and information,
- Send notifications and alerts report,
- Map view of all reports and important information,
- Marked zones,
- Situation report,
- Provide updated capacity,
- Provide alert notifications etc.

As it stated above, user interface and user experience play vital roles for the platform. The tests on the platform, its tools and the feedback that will be given, will help for a better understanding of the users' needs.

Furthermore, the User Interface depends on the evaluation of the participants. These could be people of the project, colleagues or other contacts that are sympathetic to the project

and its aims. This will help to minimize the mistakes. In later stages of the prototype, it will be able to test it in the wider public audience.

Additionally, a "User test plan" will be created. This is a brief document that specifies the objectives of the particular user study, what will be evaluated, how, where, when, who the participants are, who the test team is, and how the test results will be documented.

Bellow, there are presented, the proposed 2 steps for a successful user test plan.

1. **User studies in context:** The aim is to ensure that there is a good match between the evaluation study and the application of the system in the intended situation. The main points to consider include:
    - **Involve the right participants**: Involve participants that are either current users or likely future users.
    - **Choose the right situations:** Choose situations where the system will be used according aspects like environment, timing, interests and availability etc.
    - **Set relevant user tasks:** Choose user tasks that make the participant seek information and are in accordance with situations that have been identified. The tasks should be realistic and close to users' activities.
    - **Document results in the situation:** Results are most accurate if recorded in the situation. For a better and more accurate procedure, results and observations will be noted and participants will fill the needed papers.
    - **Document the context:** A separation within the results and between the partners will help additionally in order to retrieve and recall specific details.
    - **Use relevant evaluation approaches and measures:** Each stage of iterative development and evaluation may have a slightly different sub-goal within an overall objective.
2. **Iterative development of the system and its evaluation:** User-centered evaluation, in common with other evaluation approaches, should not be considered towards the end of the project when there will be no time to benefit from the results.
    - **Iterative development:** Improve both the system and the information in it, based on the results of each study conducted for a better and more accurate evaluation.
    - **Iterative studies:** Improve and redesign the evaluation studies iteratively. This is important and will ensure that everything is tracked with a total respect to the overall objectives and expectations.

- **Scale up the number of participants between the experiments:** In the first study there will be a small number of participants and carefully this number will increase according the development of the platform.
- **Shorten the time spent per participant between the experiments:** Carefully scale down on the amount of time that is spent with each participant.
- **Specify user test plans:** Since each study can reveal problems, the plan should be recorded as also any change of improvement or added information to the system is important to track in order to have a clear trail of reasoning.

As each evaluation is completed, the results should be shared with the partners.

Finally, it is important to mention that from the demonstration and field trial users experience will be an excellent asset in order to have a more sophisticated product with more specific and targeted results and all people of the project should learn the results of each evaluation step.

4.2.3 Quality of the system

The following categories of information will be subject of evaluation:

1) Maps and general visualization of the provided information from the system, such as assigned tasks, crisis classification, etc.

2) Textual information

"Information quality" is a measure of the value which the information provides to the user of that information. "Quality" is often perceived as subjective and the quality of information can then vary among users and among uses of the information. Nevertheless, a high degree of quality increases its objectivity or at least the inter-subjectivity.

A list of elements will be used in assessing Information Quality in beAWARE:

- Accurate and Believable (validity of some information has to do with the degree of obvious truthfulness which the information carries)
- Concise and complete (few words allowing to understand the type of the emergency, its timing and its location)
- Consistently Represented (information has to be easily localized in a map and its content has to be clearly understood by adding a legend in a map for example)
- Timely (Timeliness refers to information that is current at the time of publication)

The quality of the information provided by the system is of highly importance. Quality check can be performed both in the data provided and in the information given. Furthermore, it

goes without saying that a very important aspect that will determine the quality of a system has to do with its cost. As David Rogers and Vladimir Tsirkunov (2010) stated, in order to fully appreciate cost-benefit of early warning systems, the overall operational cost of the system, the societal and economic losses due to false alarms and savings due to timely action should be considered. These data are needed to properly assess whether and where early warning systems should be established.

To have an effective quality control, four elements of the warning system must co-exist:

i. Risk knowledge
ii. Monitoring and warning service
iii. Dissemination and communication
iv. Response capability.

If any of these elements is missing or poorly developed, the overall system fails.

The quality of the system is determined by the following factors:

i. What information is needed,
ii. How predictions and alerts will be most efficiently used,
iii. How reliable the prediction must be to produce an effective response,
iv. How to communicate this information with the respective authorities, rescuers and citizens, and finally
v. How much is the tolerable prediction uncertainty.

Decision-makers should also know the expected consequences of taking action, in terms of probability of false and missed alerts, the cost saving due to mitigation actions and the cost of a false alert (Grasso 2007).

Quality control procedures begin as soon as the central control computer receives the required information based on automated data (provided from the system) and oral information (provided from the rescuers and citizens). An additional quality control can be performed innon-automated data such as textual, image or video data provided by citizens.

It is worth mentioning that the quality control can be performed both automatically from the system and manually, from people who operate the system.

4.3 Evaluation procedure

For a better evaluation of beAWARE platform it is proposed to make time and usability tests on the use cases with and without the platform in order to evaluate different factors. There will be an indicator comparison that will be chosen based on User Cases (UCs) and User

Requirements (URs). Those factors can be, time response of the involved authorities and other organizations, hazard time notification, platform usability, quality of information, faster information etc.

Furthermore, for a more complete evaluation methodology, a procedure of specific stages is proposed. First of all, there will be an evaluation of the system, its tools and the manual. Secondly external observers and bystanders who will participate in the pilots and finally indicators will be defined based on the use cases and the user requirements that will be tested. The three stages are analyzed below:

- Evaluation of the system, its tools and the manual which will be done by the users and the observers. A systematic evaluation will be done by the users in different steps of the development of the system. This evaluation will be done with the help of questionnaires, interviews, hot debriefs and written feedbacks. The questionnaires will be given after the finalization of each pilot in order to adjust the reported changes in the platform.

-External observers to evaluate, the usability of the system, how the users operate the system, if they encounter any problem or malfunction as well as to evaluate the system itself. Bystanders who will give their feedback and some of their key points from a questionnaire. Moreover, it is proposed, those two groups to be sub-grouped as followed:

a. Those who have knowledge about Rescue applications and systems
b. Those who have no knowledge or limited one.

- Definition of indicators based on the use cases and the user requirements that will be tested. Each member responsible for a specific pilot after the selection of the user cases and the user requirements will define the indicators that will be used. When those indicators will be defined, a final table of the indicators for the three pilots will be created and uploaded in the Wiki.

4.3.1 Proposed Questionnaires

In order to have better results and get the required information from participants and observers there should be different types of approach. One of them as discussed is with questionnaires. Some examples are described below.

- ***Example 1***

Which of the following would best describe your role in the pilot? (Please mark as many as apply)

- ❑ Command and Control Room operator
- ❑ Field commander
- ❑ Rescuer
- ❑ Field (para)medic/ Firefighter/ Police officer
- ❑ Transport officer
- ❑ Citizen in need
- ❑ Observer/Bystander

Other (Please Specify) _____

- ***Example 2***

Please answer the following questions regarding the functionalities performed via beAWARE during the pilot.

Action	Q1 **Did you complete the required function on beAWARE?**	Q2 **[Answer If Q1 = Yes]** **Please rate the ease with which this function was performed on beAWARE?**	Q3 **[Answer If Q1 = Yes]** **Did you ask for or receive assistance in completing this task?**	Q4 **[Answer If Q1 = No, I could not]** **Which of the following would best describe why you did not complete the function?**
Log into the Platform	☐ Yes ☐ No, I could not ☐ No, I was not assigned this function	☐ Very Easy ☐ Easy ☐ Difficult ☐ Very Difficult	☐ Yes ☐ No	☐ I could not find the relevant option ☐ The system was not responsive ☐ Other, please specify:
Reporting Arrival on the incident area	☐ Yes ☐ No, I could not ☐ No, I was not assigned this function	☐ Very Easy ☐ Easy ☐ Difficult ☐ Very Difficult	☐ Yes ☐ No	☐ I could not find the relevant option ☐ The system was not responsive ☐ Other, please specify:
Uploading information, pictures	☐ Yes ☐ No, I could not ☐ No, I was not assigned this function	☐ Very Easy ☐ Easy ☐ Difficult ☐ Very Difficult	☐ Yes ☐ No	☐ I could not find the relevant option ☐ The system was not responsive ☐ Other, please specify:

- ***Example 3***

Overall, how would you rate the beAWARE system in terms of the following functionalities?

	Poor	Fair	Average	Good	Excellent
Passing incident information to dispatched response teams	☐	☐	☐	☐	☐
Getting needed data and information from the system	☐	☐	☐	☐	☐
Informed on time about a potential or eminent hazard	☐	☐	☐	☐	☐

- ***Example 4***

If any, please list and summarize up to three major problems you encountered while using the platform.

1st most major problem	
2nd most major problem	
3rd most major problem	

- ***Example 5***

We would appreciate it greatly if you could provide any comments that you think would be beneficial for the improvement of the beAWARE platform

4.3.2 Interviews

Another type of approach is Interviews. In order to have a good result and take if not all the needed information, at least the maximum information that the interviewee will provide, there is a very good procedure and structure that is proposed according to Information Services and Technology[6].

Typically interviews are for 30 or 60 minutes. The key is that the interviewer will control the conversation and receives more information than gives. A good interview will be like a (guided) conversation; however the interviewee will be doing most of the talking. A general rule of thumb is the 80/20 rule; the interviewee will be doing 80% of the talking, and the interviewer, 20%. Interview questions will ONLY be subject related. Below is presented the 2 types of interviews structure:

- The 30-minute interview structure:
  - Opening: 3 minutes
  - Providing Info: 5-10 minutes
  - Gathering Info: 15-20 minutes
  - Closing: 2 minutes
- The 60-Minute interview
  - Opening: 5 minutes
  - Providing Info: 10-20 minutes
  - Gathering Info: 30-40 minutes
  - Closing: 5 minutes

Furthermore, Interview is divided in three phases: The Pre-interview, the Interview and the Post-Interview phase.

In the Pre-Interview phase the interviewer will create a plan based on some questions, before he go to the interview. “What am I really looking for? What information and details I am expecting to gather from the conversation?” The interviewer must be sure he knows those answers before the meeting.

[6]Information Services and Technology: https://www.google.gr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwizkoPU_rPVAhUHCcAKHef3DUEQFgglMAA&url=https%3A%2F%2Fist.mit.edu%2Fsites%2Fdefault%2Ffiles%2Fhr%2Finterviewing_for_success%2FInterviewing%2520Techniques%2520and%2520Structure.doc&usg=AFQjCNE6p6clsqV-HBV3VEa2ubPvNQjN8Q

The interview phase is analyzed in 4 steps: The opening, the purpose and providing need information, during the interview and closing the interview. Below, according to Information Services and Technology those 4 steps are analyzed:

1. Opening (or greeting): Greet the interviewee with a smile, and thank them for agreeing to meet with you. Be sure to say something to help build rapport - just as talking with a co-worker. It is important that the interviewee will feel being more in a conversation (vs. an "interview") as possible.

2. Purpose and Providing Info: Explain the purpose of the interview. Specifically, explain the reason of the interview with this person, the expected result of the conversation, and cover any "ground rules" that need to be covered. It is good to mention to the interviewee that some notes will be taken during the interview.

3. During the interview (gathering information):Ask the question, and then listen carefully. Acknowledge what the person is saying with facial expressions, a head nod, or "uh-huh", "yes I understand", etc. Listen for clues that there is more to what they're saying and be sure to probe for more information. A probing question could be as simple as "Really? Well that is interesting. Could you please elaborate on that?" or "Interesting…can you give me some examples (of when that happened)?" Take notes - write as much down (that is useful) as you can. It is good the interviewer to review his notes after the interview and likely adding to them. It's good to get as much down during the conversation.

4. Closing: Thank the interviewee for their time, and their input. Also be sure to ask for permission to follow up if needed.

In the Post-Interview phase the interviewer will review his notes and add to them as needed. It is important to do this immediately after the interview, as thoughts are still fresh in mind.

4.3.3 Hot debriefs

The last proposed type of proposed evaluation procedure is the hot debrief. According to Bethany Smith there are 10 critical questions to be answered for a successful debrief. Those questions are:

1. What were our original event objectives?
2. Did we meet them?
3. Were there any problems encountered as we tried to meet our event objectives? (think registration or tech issues, budgetary constraints, revenue goals, overall

performance, communication issues, and general attendee engagement) If so, what were they?

4. Did those problems get solved? How? Was the provided solution effective?
5. Was your individual role in the event production process clear to you from the outset?
6. Did you find that the information you needed to do your job was readily available to you?
7. What were some triumphs at our event? Who or what was responsible for them? How can we replicate that success in the future?
8. How effective and efficient was our registration process?
9. How did we utilize technology at this event? Was the tech we used easy to implement and analyze?
10. What would you like to see happen at similar events in the future?

As a last step, an Event Debrief Template will be made in order to keep everything recorded and noted.

Debriefing is a crucial step in gathering the necessary feedback. Conducting a detailed debrief will ensure that the debriefer knows exactly how everyone involved feels about. He will also have a recorded knowledge of what worked, what didn't, and how he can build on that information for better improvement.

To sum up, the evaluation procedure will be used to have more efficient and better results on the pilots and overall for the beAWARE platform.

5 Conclusions

In this document an analysis of the pilots and a literature review was presented. Because of the great number of different methodologies the literature review was based and targeted to procedures that will support the beAWARE platform.

As stated above, according to the literature review, the main and more important methodologies that will be followed in beAWARE are questionnaires, interviews and the experimental method.

Questionnaires will focus on each group, such as citizens, rescuers and PSAP operators, with specific methodological questions in order to exploit their opinions and needs for the beAWARE platform and how to improve it in order to be more useful to them.

Interviews will exploit from the perspective of each group, their experience and their needs for the platform, the tools and the overall structure that they already know and use from other platforms and how beAWARE will help them take faster and better decisions.

The experimental method is also vital for the beAWARE platform development, because and from the laboratory testing and the field testing will give the ability to have better and reliable understanding of the platform, monitor the results for corrections and further development of specific assets and tools of the platform.

The most important part of the evaluation methodology is the evaluation of the impact of the system. The indicators that will play vital role are:

- Time of task execution
- The number of assigned tasks completed on a specific time
- The number of the warned citizens
- The number of lives saved
- The economic values of properties.

To summarize from above, time and the number of assigned tasks are very important during the crisis period and one of the goals of beAWARE is to give, the ability to rescuers and PSAP controllers to take decisions, after multiple and complex evaluations, faster and more accurate.Also, the evaluation of the interface of the platform will be done or in a subjective or an objective way which will give the ability to all users express their opinion, with questionnaires, interviews and in experiments for the platform and make it more user-friendly and will be tested in different stages. Of course, a lot depend by the identity of the users who will be tested in the pilot cases.

Finally, it is important to mention that in order to assess the needed information from all the evaluation methodologies that will be done a list of elements will be used in order to exploit

the quality of the given answers, such as to be accurate and believable, which mean to validate in the degree of obvious truthfulness, to be concise and complete, which mean to exploit if the answer will give unnecessary information and erase them and to be consistently represented, which mean to give information easily localized on a map and timely which mean, the information that is current at the time of the situation.All those are presented in some questionnaires examples, interviews guideline procedure and debrief key point questions for a better understanding of the evaluation methodology that will be followed.

To sum up, evaluation methodology is one important step for the better development of the platform and for the continuous and correct evaluation of the system. It should explore correctly all the potential areas of interested, in order to be more efficient, reliable and precise with the results that are goaled.

6 References

Al Zarqa, Ahmet, Tarik Ozkul, and A. R. Al-Ali(2014): "A Study toward Development of an Assessment Method for Measuring Computational Intelligence of Smart Device Interfaces." INTERNATIONAL JOURNAL OF MATHEMATICS AND COMPUTERS IN SIMULATION.

Bethany Smith. The Planner's Process. http://www.eventmanagerblog.com/how-to-debrief-your-event

Bell J, (1997): Methodological Planning of Pedagogy and Social Research, Gutenberg, Athens.

Borland P (2000) Experimental components for the evaluation of interactive information retrieval systems. Journal of Documentation 56(1):71-90,

Cognitive: http://cognitivegroup.com/10-things-to-pay-attention-in-user-interface/

Grasso V.F., J.L. Beck, G. Manfredi (2007): Automated Decision Procedure for earthquake early warning. Journal of Engineering Structures.

Law, E., Roto, V., Hassenzahl, M., Vermeeren, A., Kort, J. (2009): Understanding, Scoping and Defining User Experience: A Survey Approachin Proceedings of Human Factors in Computing Systems conference, CHI'09. 4–9 April 2009, Boston, MA, USA

Lowe Dianne, Kristie L. Ebi and Bertil Forsberg (2011): Heatwave Early Warning Systems and Adaptation Advice to Reduce Human Health Consequences of Heatwaves

E. Mosqueira-Rey, D. Alonso-Ríos, A. Vázquez-García, B. B. del Río, and V. Moret-Bonillo, (2009): "A multi-agent system based on evolutionary learning for the usability analysis of websites," in Intelligent Agents in the Evolution of Web and Applications, ed: Springer, pp. 11-34

J. Nielsen (2014): "10 usability heuristics for user interface design," Fremont: Nielsen Norman Group.[Consult. 20 May 2014]. Disponívelna Internet.

Patton, Michael. (2008): Utilization-focused Evaluation, 4th Edition. Thousand Oaks, CA: Sage.

Pelt, Mason (2006): "Stop overthinking UX and try the coffee shop test". venturebeat.com

ZafiropoulosK. (2005): How to write down a scientific paper. Athens. Kritiki.

Patton, Michael (2008): Utilization-focused Evaluation, 4th Edition. ThousandOaks, CA: Sage.

Rogers David and Vladimir Tsirkunov (2010): Costs and benefits of early Warning Systems. Global assessment report on disaster risk reduction.

Stone, P., Sutton, R. S., and Kuhlmann, G. (2005). Reinforcement learning for RoboCup-soccerkeepaway. Adaptive Behavior, 13(3):165-188.

Filias B. (1977): Introduction to Methodology and Techniques of Social Research Collaboration: Sociology School.Athens: Gutenberg

United Nations (2005): International Strategy for Disaster Reduction, International Strategy for Disaster ReductionEvaluation and Strengthening of Early Warning Systems in Countries Affected by the 26 December 2004 Tsunami

Zschau Josephand Andreas N. Kuppers (2003): Early Warning Systems for Natural Disaster Reduction, Springer New York