



beAWARE

Enhancing decision support and management services in extreme weather
climate events

700475

D4.3

Social event detection and decision support

Dissemination level:	Public
Contractual date of delivery:	Month 34, 31 October 2019
Actual date of delivery:	Month 35, 30 November 2019
Workpackage:	WP4
Task:	T4.1 - Social media monitoring T4.3 - Semantic representation T4.5 -Reasoning for decision support
Type:	Report
Approval Status:	Final
Version:	V1.0
Number of pages:	69
Filename:	D4.3_Social_event_detection_and_decision_support_2019-11-30_v1.0

Abstract

This deliverable concerns the collection and analysis of social media in the beAWARE system and the decision support components that have been integrated. Starting with social media data, the final workflow is presented, focusing on two techniques that protect the system from fake and irrelevant information. Experiments on text clustering, spatial clustering, and event detection follow, while any additions to the related user interfaces are illustrated. Regarding decision support, additions to the beAWARE Ontology are described. Further, updates in the Knowledge Base Service to support new incident data and new metrics are reported and a two-layer validation system is presented. In addition, two new user interfaces have been integrated into the Knowledge Base, which offer a manual exploration and analysis of the data that is collected and generated during a crisis.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.



Co-funded by the European Union

History

Version	Date	Description / Reason of change	Revised by
0.1	20-Sep-2019	Deliverable outline	CERTH, IOSB
0.2	17-Oct-2019	Contributions in Chapters 3, Appendix A	IOSB
0.3	18-Oct-2019	Contributions in Chapters 2,3,4	CERTH
0.4	22-Oct-2019	Final contributions in the whole document	CERTH, IOSB
0.5	23-Oct-2019	Document submitted for internal review	IOSB
0.6	25-Nov-2019	Content corrections	CERTH
0.7	27-Nov-2019	Internal Review	MSIL
1.0	29-Nov-2019	Final version for submission	IOSB

Author List

Organisation	Name	Contact Information
CERTH	Stelios Andreadis	andreadisst@iti.gr
	Alexandros Koufakis	akoufakis@iti.gr
	Panagiotis Mitzias	pmitzias@iti.gr
	Ilias Gialampoukidis	heliasgj@iti.gr
	Despoina Chatzakou	dchatzakou@iti.gr
IOSB	Jan Blume	jan-wilhelm.blume@iosb.fraunhofer.de
	Philipp Hertweck	philipp.hertweck@iosb.fraunhofer.de
	Tobias Hellmund	tobias.hellmund@iosb.fraunhofer.de
	Jürgen Moßgraber	juergen.mossgraber@iosb.fraunhofer.de

Reviewer List

Organisation	Name	Contact Information
MSIL	Ifat Neumann	ifat.neumann@motorolasolutions.com

Executive Summary

This document constitutes deliverable D4.3 “*Social event detection and decision support*” and focuses on presenting the social media analysis of tweets that is used in the beAWARE system, as well as the decision support components that have been integrated.

The document contains a description of the workflow that involves the collection of social media data, focusing more on the modifications since the previous version. The main additions to the workflow are two techniques that aim to improve the quality of collected data and secure the system from harming material, such as fake news. The first technique estimates whether a tweet is real or fake, while the second inspects tweets for irrelevant emoticons. Next, the clustering of tweets is investigated in two alternative directions with various experiments. Text clustering is based on the content of posts and spatial clustering on their location. The spatial clustering approach was implemented in the frame of beAWARE. Another study concerns the detection of events by using the number of tweets that are collected each day. The fluctuation of social media is compared to incidents reported in news and correlated to weather measurements. The section closes with the updates on the user interfaces that are related to social media, i.e. the annotation tool and the demonstration tool, and with the integration of tweets in the main beAWARE application.

Additions to the beAWARE ontology described in deliverable 4.2 have been made. These includes additions that have been made to semantically represent the extended recognition capabilities of the analysis components. This includes for example new subclasses (meaning more specific entities) of the *VulnerableObject* concept, like bicycles, dogs or dummies, which the object detection can detect now. A further set of changes comes from the integration of a model that allows describing the different stakeholders that will be involved in a crisis as well as the different actions that they can offer to mitigate certain impacts.

The Knowledge Base Service was updated in order to integrate the incident data captured by Unmanned Aerial Vehicles (UAVs) and new metrics calculated by the Crisis Classification component. Moreover, a two-layer validation system that targets to protect the system from malicious or incorrect data is introduced, where the first layer takes place individually at the Visual Analysis and Social Media Analysis modules, while the second layer operates closely to the Knowledge Base Service and Crisis Classification component taking into account the overall situational picture.

In order to enable analysts or decision makers to explore or manually analyze the data further, two user interfaces have been implemented into the Knowledge Base. The first one consists of an incident map and an incident table, which are helping to get a better overview of all incidents as well as to prioritize them. The second one is an analyzation workbench that can be used to search for specific information that is contained in the collected data. It is able to include different data sources, while also offering data source specific visualizations and filters.

Abbreviations and Acronyms

The following abbreviations have been used in this document:

API	Application Programming Interface
CERTH	Centre for Research & Technology Hellas

CRCL	Crisis Classification
EmC	Emergency Classification
FMI	Finnish Meteorological Institute
HRT	Hellenic Rescue Team
HTTP	Hypertext Transfer Protocol
IOSB	Fraunhofer Institute of Optronics, System Technologies and Image Exploitation
JSON	JavaScript Object Notation
KB	Knowledge Base
KBS	Knowledge Base Service
NMI	Normalized Mutual Information
PLV	Policia Local Valencia
PSAP	Public-Safety Answering Point
SMA	Social Media Analysis
SMC	Social Media Clustering
SPI	Service Provider Interface
MTA	Multilingual Text Analysis
UAV	Unmanned Aerial Vehicle
URL	Uniform Resource Locator

Table of Contents

LIST OF FIGURES..... 10

LIST OF TABLES 10

1 INTRODUCTION AND RATIONALE 11

 1.1 What to Expect from this Document 11

 1.2 Relation to other Work Packages 11

 1.3 Relation to other WP4 Deliverables 12

 1.4 Document Structure 12

2 SOCIAL MEDIA MONITORING AND EVENT DETECTION 14

 2.1 Collecting Social Media Data 14

 2.2 Analyzing Social Media Data 15

 2.3 Clustering Techniques 18

 2.3.1 Text Clustering 18

 2.3.2 Spatial Clustering 19

 2.3.3 Spatiotemporal clustering in the beAWARE system 25

 2.4 Event Detection..... 26

 2.4.1 Observation of events in social media data 26

 2.4.2 Correlation of social media and weather data 29

 2.5 Displaying the Data 31

 2.5.1 Annotation tool 31

 2.5.2 Demonstration tool 32

 2.5.3 Social media in PSAP..... 33

 2.6 Chapter Summary 35

3 DECISION SUPPORT 36

 3.1 Additions to the Knowledge Base Ontology..... 36

 3.1.1 Representing Stakeholders and Actions..... 36

 3.1.2 Ontology Additions related to Improvements of Analysis Components..... 37

 3.2 Integration of Incident Data from Drones..... 37

 3.3 New Metrics from Crisis Classification..... 37

 3.4 Two-layer Validation Component 38

 3.4.1 First Validation Layer 38

 3.4.2 Second Validation Layer 39

 3.5 User Interfaces for Additional Data Analysis..... 40

 3.5.1 Analysis Dashboard..... 40

 3.5.2 Incident Map and Table..... 45

 3.6 Chapter Summary 46

4 CONCLUSIONS AND NEXT STEPS 47

REFERENCES 49

APPENDIX A – ONTOLOGY ADDITIONS 51

Classes 51

 Action..... 51

 Action Type 51

 Alert 51

 Asset Representation..... 51

 Climate Condition 51

 Crisis Classification..... 52

Document	52
Domain	52
Feature of Interest.....	52
Guideline.....	52
Input Parameter	52
Issue.....	52
Measurement	53
Parameter.....	53
Parameter Value.....	53
Place of Relief	53
Report	53
Response.....	53
Risk.....	53
Role.....	54
Stakeholder.....	54
Tasking Capability	54
Unit of Measure	54
Forrest Fire Crisis Classification	54
Heatwave Classification.....	54
Hydraulic Crisis Classification	55
Standard Operating Procedure.....	55
HTML Item	55
Decision Maker	55
Domain Expert	55
Organization	55
Person	55
Civil Society.....	56
Private Sector.....	56
Public Sector	56
Municipality	56
Park.....	56
Cellar	56
Forest.....	56
Bycicle	56
Boat.....	57
Bus	57
Truck	57
Motorcycle.....	57
Cat.....	57
Dog.....	57
Dummy	57
Wheelchair User	57
Object Detection.....	58
Voice Analysis	58
Object Properties	58
addressed through.....	58
consists of	58
estimated in.....	58
guideline includes	58
has dataset.....	59
has document	59

has measurements of	59
has parameter	59
has role	59
has tasking capability.....	59
has unit	60
is at risk.....	60
is based on.....	60
is condition for.....	60
is executed by.....	60
is expert in	61
is implemented by.....	61
is input to task	61
is located at.....	61
is measured with	61
is measurement of.....	61
is mitigated by action	62
is prerequisite of.....	62
is quantified in	62
is value for	62
mitigates impact type.....	62
observes asset	63
observes parameter.....	63
organization includes.....	63
performed at.....	63
place of relief is located at.....	63
produced document	63
provides	64
provides data about.....	64
reports Incident Type	64
used by tasking capability.....	64
works for organization.....	64
Data Properties	65
capacity.....	65
capacity used	65
creation time	65
expiresUTC.....	65
flood crisis level	65
forTeam	65
Grid connection	66
has end	66
has originator.....	66
has priority.....	66
has reference code	66
has severity from Crisis Classification.....	66
has start	67
has timestamp	67
instructions.....	67
is spam	67
language	67
level.....	67
radius	68

reason for crisis level	68
reportID	68
SanitationAvailable	68
severity	68
startTimeUTC	68
status	69

List of Figures

Figure 1. Updated flow of SMA.	15
Figure 2. New custom fields in tweet's JSON.	17
Figure 3. Modification of a tweet's place to support 2dsphere index.	18
Figure 4. Clustered tweets and outliers in Valencia.	20
Figure 5. NMI score for different values of eps, for minPts 2, 3, and 4.	22
Figure 6. The flow of SMC in the beAWARE system.	26
Figure 7. Fluctuation of the number of collected tweets per each date of 2018.	27
Figure 8. Updated interface of the annotation tool.	32
Figure 9. Updated interface of the demonstration tool.	33
Figure 10. Twitter reports on PSAP.	34
Figure 11. The HTML file of a Twitter report.	34
Figure 12. Ontology representation of stakeholders that will be involved in a crisis and the actions they can make to mitigate an impact.	36
Figure 13. Architecture of the Analysis Workbench	41
Figure 14. Layout of the Analysis Workbench.	42
Figure 15. The implemented user interface of the Analysis Workbench.	42
Figure 16. The list component of the Analysis Workbench. Reports from the beAWARE mobile app have been selected as data source.	43
Figure 17. The filter component of the Analysis Workbench for tweets.	44
Figure 18. Screenshot of the incident table included in the KB.	45
Figure 19. Screenshot of the incident map.	46

List of Tables

Table 1. Number of collected tweets per use case and language.	15
Table 2. List of emoticons and emojis that are considered irrelevant.	16
Table 3. Estimated number of clusters and NMI score of each clustering index.	21
Table 4. Clustering of tweets in Italy by different indexes.	23
Table 5. Most frequently mentioned words inside the collected tweets.	27
Table 6. Top five keywords for each date that is considered an event.	28
Table 7. Pearson correlation coefficient between social media and weather data.	30

1 Introduction and Rationale

1.1 What to Expect from this Document

This deliverable is divided into two main parts:

Chapter 2 concerns the collection and analysis of social media information, which is relevant to the beAWARE use cases. The chapter begins with a brief description of how the crawling of tweets works, mainly focusing on the changes implemented since the version described in detail in the previous deliverable D4.1. In addition, the current status of the collection is presented as the number of tweets crawled so far per use case and language. The chapter continues with two analysis techniques that aim to protect the system from mischievous posts, such as fake news, and meaningless posts based on emoticons existing in their text. Next, clustering of tweets is explored in two different directions: clustering based on textual information and clustering based on spatial information. Quantitative and qualitative evaluations are included. Another approach to exploit large streams of tweets is the detection of events. To this end, the chapter presents a study of the peaks in the fluctuation of tweets collected per day and a computation of the correlation between social media data and weather data. The final part of this chapter reports all the updates in the annotation tool and the demonstration tool presented in D4.1, while it also introduces the display of Twitter reports (clusters of tweets) in the final system of beAWARE.

Chapter 3 presents the advances that were made regarding the decision support aspect of the beAWARE platform. In particular, some new classes and properties were added to the Knowledge Base (KB) ontology, in order to keep up with the specializations of the analysis modules. Moreover, the Knowledge Base Service (KBS) and KB were upgraded with the capability to integrate incident data from Unmanned Aerial Vehicles video streaming. This process required the establishment of new topics of communication on the internal message bus and new integration methods by the KBS. Another update is the incorporation of metrics and incidents provided by the Crisis Classification (CRCL) component to the KB. The new information is transmitted by two new topics to the KBS, and subsequently KBS enriches the KB with the provided information. Further, a two-layer validation service was implemented in order to protect the system from incorrect or erroneous data passing through. In particular, the service includes a standalone module that communicates (via new topics) with CRCL and KBS and validates the passing incidents using environmental metrics provided by CRCL. In the end of the chapter, two user interfaces are described that have been integrated into the KB in order to enable a manual analysis and exploration of the data.

1.2 Relation to other Work Packages

In addition to the analysis steps described in this deliverable, the tweets will be analyzed by natural language processing that is part of WP3. The CRCL, whose information have been included into the KB, is also part of WP3. Further, the outcomes of the social media analysis and the decision support components are directly affecting the information that is displayed in the PSAP that belongs to WP6. The integration of the components described in this deliverable into the rest of the beAWARE system has been done in WP7. They are evaluated during three pilots that are part of WP2.

1.3 Relation to other WP4 Deliverables

The focus of deliverable D4.1 “Social media crawling and monitoring of multiple sensing platforms” was data aggregation and deliverable D4.2 “Semantic representation and reasoning” described in detail its semantic integration into the beAWARE system.

The aggregation of Tweets described in D4.1 is thereby a direct requirement for their further analysis described in the first part of this document. In addition, this part includes all improvements that have been made to the tweet aggregation since D4.1.

The decision support described in the second part of this deliverable directly builds up on D4.2. Similar to the first part it also includes the changes that have been made to the semantic integration of the data since D4.2.

1.4 Document Structure

This document is structured as follows.

Chapter 2 concerns the collection of social media and all related tasks:

- Section 2.1 makes a brief description of the process of collecting social media data, reports all the updates in the process, and informs about the current status of the collection.
- Section 2.2 describes the two added analyses of social media data, i.e. a verification methodology to detect fake news and a check for irrelevant emoticons.
- Section 2.3 focuses on the text and spatial clustering techniques that were examined and presents the solution adopted in the beAWARE system.
- Section 2.4 presents an investigation on how the fluctuation of the number of collected tweets per day can be used to detect events.
- Section 2.5 reports the changes in the user interfaces that are related to social media monitoring and introduces the display of tweets in the final system.

Chapter 3 presents the advances that were made regarding the decision support aspect of the beAWARE project:

- Section 3.1 presents the additions that were made to the Knowledge Base ontology, such as new classes and properties.
- Section 3.2 reports how the incident data captured by Unmanned Aerial Vehicles were integrated into the Knowledge Base.
- Section 3.3 presents the process of adding new metrics from the Crisis Classification component to the Knowledge Base.
- Section 3.4 presents the two-layer validation system that was deployed, focusing on the second layer that operates closely to the Knowledge Base Service and Crisis Classification component.
- Section 3.5 presents two user interfaces that provide a manual analysis and exploration of the data.

Both chapters start with an introduction and end with a recapitulation of their content. In the end, closing remarks are given in Chapter 0.

2 Social Media Monitoring and Event Detection

Social Media is one of the many sources of information to be fused by beAWARE for crisis management in the case of a natural disaster, mainly during the emergency response phase of the crisis. The Social Media Monitoring task includes two components: Social Media Analysis (SMA), responsible to collect and further analyze tweets, and Social Media Clustering (SMC), responsible to collect analyzed tweets and group them by spatiotemporal criteria.

This chapter starts with the changes of how social media data is collected and the current status of the collection. Then, it describes the techniques that were added in the analysis phase and reports experiments on clustering and event detection methodologies. Finally, it presents the updates in the user interfaces that are related to the crawled social media and demonstrates how tweets are displayed in the final system.

2.1 Collecting Social Media Data

The SMA module aims at crowdsourcing information from connected citizens and authorities concerning natural disasters. During the project's lifetime, the platform of Twitter has been selected as a source, due to its high popularity (leading to rich information) and its complimentary real-time streaming API¹. The topics of interest, derived from the project's use cases, are floods, fires and heatwave incidents in English, Greek, Italian, and Spanish. Tweets that concern the combinations of the above topics and languages are stored in separate collections in a database and, after they are analyzed, they are forwarded to the next module of the beAWARE platform i.e. the Multilingual Text Analysis (MTA) Module.

Figure 1 depicts the updated data flow describing the different parts of the SMA. The starting point is a collection of specified key words per use case, stored in a MongoDB database named "Feeds". Those keys are fed to a Java HTTP client, which constantly crawls tweets from the Twitter API that contain any of these terms. When a new tweet is collected, it arrives in JSON format. After using its text content to detect the collection it is targeted to (e.g., a crawled tweet that includes the keyword "flooding" will be stored in the English floods collection), the analysis phase commences, aiming to filter out the non-valuable posts.

Firstly, a verification service, which will be described in section 2.2, tries to identify whether the tweet is real or hoax news. If it is identified as fake, it is not further examined for its relevancy to the topic. If it is considered as real, its content is checked for certain emoticons, which will be listed in section 2.2. In the case that unrelated emoticons are found (e.g. a smiley or a heart), the analysis phase stops again. Otherwise, the relevancy of the tweet to the examined use cases is estimated. If a picture accompanies the tweet, the image classification service, which has been thoroughly described in deliverable D4.1, is called to determine whether it is relevant. If the result is negative or the tweet does not contain an image in the first place, then a text classification service is called to assess. Again, this service has been described in deliverable D4.1. In whichever step the analysis phase concludes, the tweet's JSON is updated in order to include the outcomes of the analyses and it is stored in the respective collection of the Mongo database. In case that the post is finally considered as

¹ <https://developer.twitter.com/en/docs/tweets/filter-realtime/overview>

relevant, then a message is sent to beAWARE’s bus and forwarded to the rest of the system’s modules.

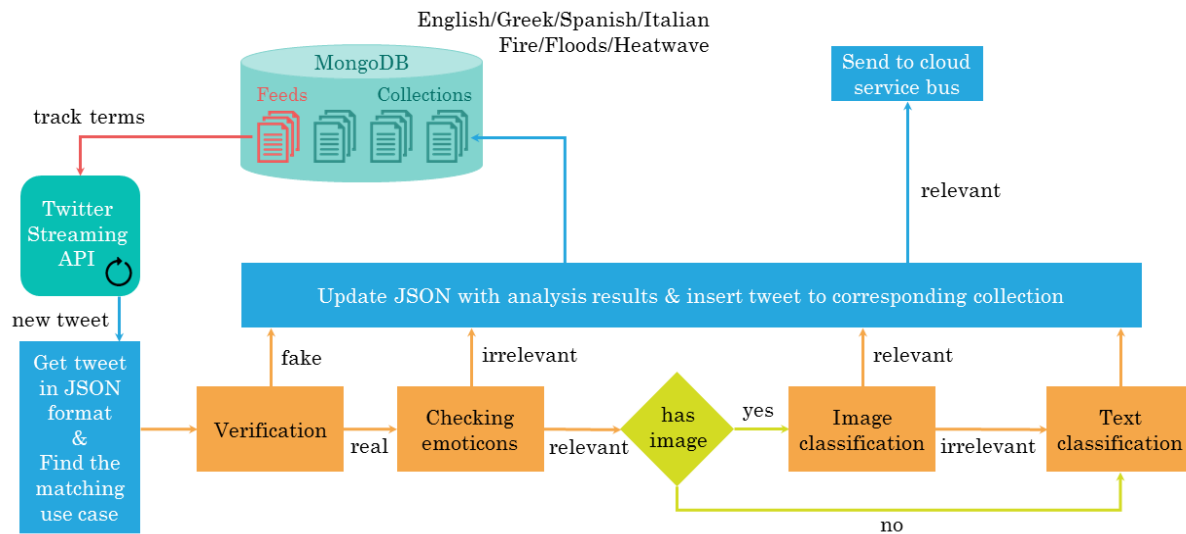


Figure 1. Updated flow of SMA.

In comparison with the previous version of D4.1, the verification procedure and the checking of emoticons have been added, to enhance the quality of the collected information. The source is still Twitter, the same library is used (Hosebird Client²) to connect to the twitter API and there have been no modifications in the lists of search terms for each use case and language. SMA has been actively running since June 2017 (M6) and the status of the collections, as of October 2019 (M34), is shown in Table 1.

Table 1. Number of collected tweets per use case and language.

Language	Use case scenario		
	Floods	Fires	Heatwave
English	5,946,272	81,776	1,396,670
Greek	38,936	237,722	74,255
Italian	72,262	502,541	1,480
Spanish	2,656,237	10,580,761	839,725

2.2 Analyzing Social Media Data

Targeting to solve the recent problem of hoax news circulating online and improve the quality of incoming information into the beAWARE system, a verification methodology has been added as an initial part of the analysis phase. A mature solution, developed by CERTH, has been included instead of a completely new technique, since its development would be outside the scope of this project.

The solution deploys a robust approach for detecting in real time whether a social media post in Twitter is fake or real, by using a variety of content-based and contextual features. The categories and subcategories of the features are:

1. tweet-based:

² <https://github.com/twitter/hbc>

- a. text-based (e.g., how many characters are uppercase)
 - b. language-specific (e.g., slangs)
 - c. twitter-specific (e.g., is retweeted)
 - d. link-based (e.g., existence of an external URL)
2. user-based:
- a. user-specific (e.g., how many followers)
 - b. link-based (e.g., existence of an external URL in a profile description)




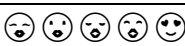
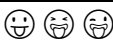
Leveraging part of its own predictions for retraining, the approach follows a semi-supervised learning paradigm, to adapt the model to unseen content, and experiments on a public annotated corpus of multimedia tweets have demonstrated its effectiveness. For details, the reader is referred to “Boididou, 2017” and “Boididou, 2018”.

The second added analysis step that has been mentioned in the previous section is checking inside the tweet’s text for emoticons (a pictorial representation of a facial expression using characters) and emojis (actual picture instead of typographic) that are not expected to be present in posts during a crisis event. The main idea is that important tweets that forewarn or describe the current situation of a natural disaster are not supposed to contain smileys, winks, or hearts.

Based on the official list of emoticons in Wikipedia³, certain categories, which are popular in social media and frequently used, were selected (Table 2) and in the case that a tweet includes any of these, it is labelled as irrelevant. In the JSON structure of the tweet, this estimation is added with the custom Boolean field “emoticon_relevancy” (Figure 2).

Please notice that negative emoticons, such as sad faces or expressions of horror and shock, are excluded from this check, since it is consider normal from them to be used during a crisis.

Table 2. List of emoticons and emojis that are considered irrelevant.

Category	Emoticons	Emojis
smiley, happy face	:-) :) :-] :] :-3 :3 :-> :> 8-) 8) :-} :} :o) :c) :^) =] =)	
laughing, big grin, laugh with glasses, wide-eyed surprise	:-D :D 8-D 8D x-D xD X-D XD =D =3 B^D	
tears of happiness	:'-) :')	
kiss	:-* :* :x	
wink, smirk	;-) ;) *-) *) ;-] ;] ;^) :- ;D	
tongue sticking out, cheeky/playful, blowing a raspberry	:-P :P X-P XP x-p xp :-p :p :-p :p :-p :p :-b :b d: =p >:P	
heart	<3	

³ https://en.wikipedia.org/wiki/List_of_emoticons

⌵ K3 verification	{ 2 fields }	Object
T/F predicted	true	Boolean
## confidence	0.737142857142857	Double
T/F emoticon_relevancy	true	Boolean

Figure 2. New custom fields in tweet's JSON.

Apart from the new additions to the JSON structure of each tweet, a particular modification was necessary, too. In order to support geospatial queries on the MongoDB, which are used by the Analysis Workbench (described in section 3.5.1), the “2dsphere” index had to be created. For this, all geo locations needed to comply with the GeoJSON Format⁴, which is not the case for locations of tweets that are provided in form of a bounding box by the Twitter API. These locations are provided as an array of four coordinates, which are the four corners of the box. However, in order to comply with the geoJSON Format they need to be arrays of five points, where the first and the last points are the same (to close the box). A script was ran on the whole collection of crawled tweets, in order to copy the first point of each bounding box and paste it as a fifth (Figure 3). Afterwards the “2dsphere” index could be created. This modification is now also running when new tweets are crawled, to keep the support of geospatial queries.

⁴ <https://geojson.org>

place	{ 9 fields }	Object	
id	0b30aadf406d13cc	String	
url	https://api.twitter.com/1.1/geo/id/0b3...	String	
place_type	city	String	
name	Hounslow	String	
full_name	Hounslow, London	String	
country_code	GB	String	
country	United Kingdom	String	
bounding_box	{ 2 fields }	Object	
type	Polygon	String	
coordinates	[1 element]	Array	
[0]	[5 elements]	Array	
[0]	[2 elements]	Array	
[0]	-0.457966	Double	
[1]	51.420633	Double	
[1]	[2 elements]	Array	
[0]	-0.457966	Double	
[1]	51.502851	Double	
[2]	[2 elements]	Array	
[0]	-0.243419	Double	
[1]	51.502851	Double	
[3]	[2 elements]	Array	
[0]	-0.243419	Double	
[1]	51.420633	Double	
[4]	[2 elements]	Array	added
[0]	-0.457966	Double	
[1]	51.420633	Double	

Figure 3. Modification of a tweet's place to support 2dsphere index.

2.3 Clustering Techniques

The constant crawling of social media posts in a real-time manner results to a very large amount of individual tweets that can be confusing to the end user. The need to receive information that is more compact can be satisfied by grouping the various posts according to their content or location, expecting that tweets with similar text or sent from a specific region refer to the same incident. Two different directions of clustering have been investigated and are described here: a text clustering based on the content and a spatial clustering based on the location. Furthermore, a straightforward spatiotemporal technique was deployed in the beAWARE system.

2.3.1 Text Clustering

Text clustering algorithms order documents into groups depending on their content. They can be used as a preprocessing step in order to support document summarization, information retrieval and extraction or other text mining algorithms. Like other machine learning algorithms, clustering algorithms work with data in the form of numerical vectors. In the case of text clustering, this means that the text need to be transformed to a numerical vector. The mapping from the text to one vector can thereby be done either per word, sentence or document. To create the vector from the text different approaches can be used. One approach

for example is to use the TF*IDF (term frequency, inverse document frequency (Rajaraman 2011)) score as a measure for the importance of a term. Another one is called word embedding (word2vec algorithm (Mikolov 2013)), which is mainly used with deep learning algorithms for text understanding.

For clustering there are several algorithms available, which are based on different clustering principles. Each algorithm has its own set of parameters, which control their performance. The correct and efficient hyper-parameter (i.e. parameter which are external to the trained model) selection is a major task, to get the optimum from the available input data. For many cluster algorithms, the main parameter to select is the number of clusters under certain boundary conditions.

To test the usefulness of text clustering for the beAWARE system a text clustering system that already had been developed by the beAWARE partner IOSB have been used. This system uses the TF*IDF sentence encoding to create indices that are stored in an Apache Lucene™ index store. The sentences are thereby interpreted as a numerical data point in a high-dimensional vector space model. Since the order of magnitude of the size of a vocabulary will be around 10000, the clustering algorithms will have to be able to deal with sparse vectors and matrices. The following algorithms have been integrated into the text clustering system:

- DBScan (Density-Based Spatial Clustering of Applications with Noise)
- KMeans
- MeanShift
- HClust (Hierarchical Clustering)
- XMeans
- GMeans
- DAC (Deep Adaptive Clustering)
- SIB (Sequential Information Bottleneck algorithm)
- MEC (Minimum Conditional Entropy Clustering)
- Birch (Balanced Iterative Reducing and Clustering using Hierarchies), Specc (Spectral Clustering)
- Clarans (Clustering Large Applications based on Randomized Search)
- SOM (Self-Organized Map)
- NeuralGas

To choose the right algorithm and parameter values turned out to be the main challenge because both are very situation and context depended, i.e. on the tweets that shall be clustered. Therefore, it is not possible to find one setting, which works well in all possible situations that beAWARE tries to cover. Due to this fact, the text clustering was not integrated into the automated social media analysis pipeline and instead other, more stable and predictable, clustering approaches have been analyzed that are described in the next two sections. A possibility that would mitigate this situational dependency would be to use it in form of an explorative tool that can be used by analysts, rather than to integrate it in the automated analysis process. Depending on the situation, the analysts then can adjust the used algorithms as well as its parameters. An example for this is described in section 3.5.1.

2.3.2 Spatial Clustering

The target of the following experiments was to determine which clustering technique is the most suitable for the spatial grouping of tweets in the frame of disaster incidents. A

quantitative and a qualitative evaluation on different datasets were sufficient to find the most suitable methodology, so as to be deployed in the beAWARE system.

Regarding the quantitative evaluation, the dataset consisted of 88 synthetic Spanish tweets about fires, which have been created by PLV specifically for the 3rd pilot of beAWARE in Valencia, Spain and share locations in the regions of El Saler and the natural park of Albufera. The tweets have been assigned to clusters with human annotation (Figure 4), while four of them were considered as outliers and thus were not matched with any cluster (labelled with the white cross in the figure).



Figure 4. Clustered tweets and outliers in Valencia.

Having the dataset and the ground truth, a comparison was realized between 16 clustering indices: CH (Calinski, 1974), Duda (Duda, 1973), Pseudo t2 (Duda, 1973), C-index (Hubert, 1976), Ptbiserial (Davies, 1979), DB (Davies, 1979), Hartigan (Kaufman, 1990), Ratkowsky (Ratkowsky, 1978), Ball (Ball, 1965), McClain (McClain, 1975), KL (Krzanowski, 1988), Silhouette (Kaufman, 1990), Dunn (Dunn, 1974), SDindex (Halkidi, 2000), SDbw (Halkidi, 2000), and DBSCAN (Ester, 1996). The evaluation metric was the Normalized Mutual Information (NMI) score⁵ and the results can be seen in Table 3. The first column is the clustering index evaluated, the second is how this index is called in R language (in which the experiment ran), the third is the estimated number of clusters and the last column is the NMI score. The Hartigan, McClain, KL, and Dunn indexes achieved a very high score, of 0.98 and above, but only a version of DBSCAN managed to predict the correct number of clusters and achieve an

⁵ https://en.wikipedia.org/wiki/Mutual_information

NMI score of 1.0. This is due to its ability to detect outliers and avoid assigning them to clusters.

Especially for the case of DBSCAN, several variations were investigated by tuning the *eps* (the local radius for expanding clusters) and the *minPts* (the minimum number of points required to form a dense region) parameters. *Eps* ranged between 0.0001 and 0.0140 in a step of 0.0001, while the integer values of *minPts* were 2, 3, and 4. The fluctuation of the NMI score is displayed in the line chart of Figure 5, where the best scores occurred for *eps* between 0.0006 and 0.0014, for all three values of *minPts*. However, scores for *minPts* = 3 were slightly better than the others.

Table 3. Estimated number of clusters and NMI score of each clustering index.

Index	Index in R	Number of clusters	NMI score
CH	ch	30	0.6647121
Duda	duda	2	0.5180138
Pseudo t2	pseudot2	2	0.5180138
C-index	cindex	3e+01	0.6647121
Ptbiserial	ptbiserial	3	0.5355914
DB	db	8	0.8860461
Hartigan	hartigan	9	0.9853166
Ratkowsky	ratkowsky	2	0.5180138
Ball	ball	3e+00	0.5355914
McClain	mcclain	9	0.9853166
KL	kl	9	0.9853166
Silhouette	silhouette	8	0.8860461
Dunn	dunn	9	0.9853166

SDindex	sdindex	7	0.8650745
SDbw	sdbw	30	0.6647121
DBSCAN minPts = 2 eps = 0.001	dbscan	7	0.9901629
DBSCAN minPts = 3 eps = 0.001	dbscan	6	1.0
DBSCAN minPts = 4 eps = 0.001	dbscan	5	0.9825724



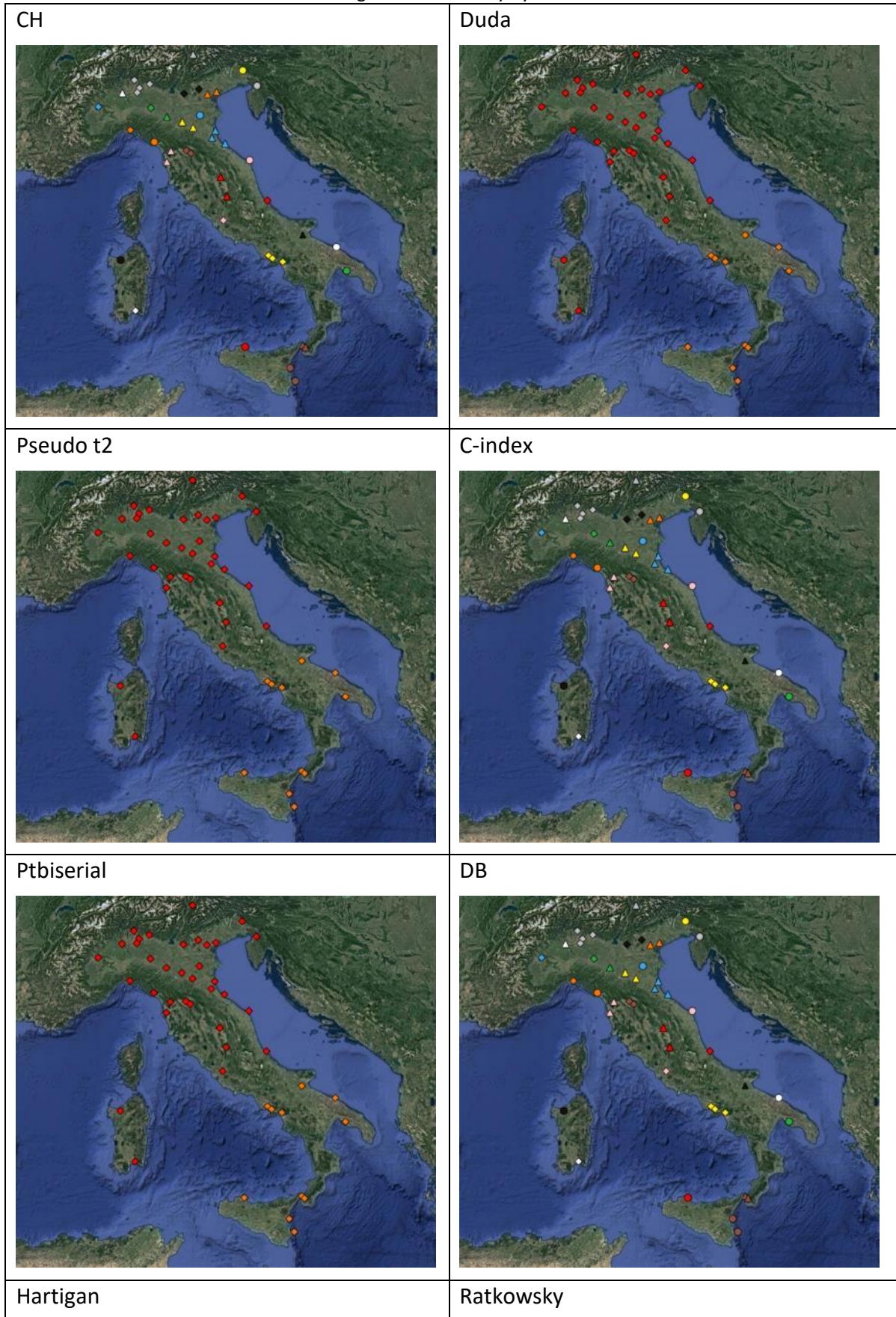
Figure 5. NMI score for different values of eps, for minPts 2, 3, and 4.

Moving on to the qualitative evaluation, the dataset consisted of 13,971 real Italian tweets about floods that mention any Italian city⁶, which have been crawled by SMA. Since the tweets did not contain any specific information on where they originated, the coordinates of the cities mentioned in their text were used instead, to set their location.

All the aforementioned clustering indexes were again compared, but this time the results are visualized as clusters on a map, indicated by color and shape. Table 4 includes the same map of Italy with the grouping of tweets, as it is produced by the different indexes. CH, C-index, DB, Silhouette, Dunn, and SDbw show poor clustering, Duda, Pseudo t2, Ptbiserial, Hartigan, Ratkowsky, Ball, McClain, KL, SDindex perform much better, but only DBSCAN is able to distinguish some points as outliers (white crosses on map).

⁶ <https://www.worldatlas.com/eu/it/cities-in-italy.html>

Table 4. Clustering of tweets in Italy by different indexes.





Ball



McClain



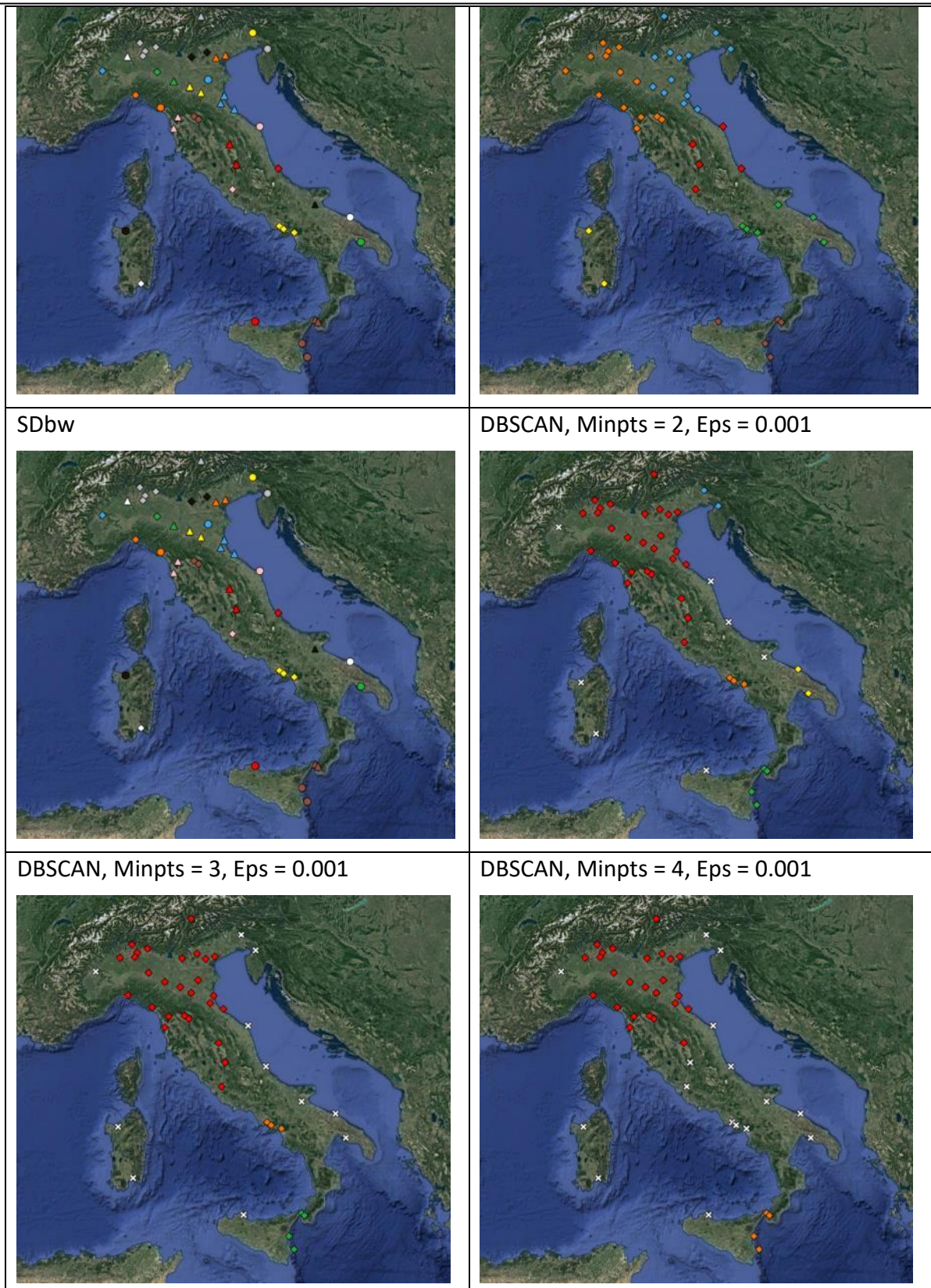
KL

Silhouette



Dunn

SDindex



2.3.3 Spatiotemporal clustering in the beAWARE system

For the actual implementation of clustering tweets in the beAWARE system, the DBSCAN algorithm was selected, because it achieved the best performance in the evaluation described

above. The tweets are crawled in a real-time manner, so time is another important factor that needs to be considered. The final deployment is a spatiotemporal clustering and the respective workflow inside the beAWARE system is shown in Figure 6.

The Social Media Clustering (SMC) component consumes the messages that the Multilingual Text Analysis (MTA) module is producing. The reason that SMC does not consume directly from SMA is that it bases clustering on the location of tweets and the locations are detected by MTA. When x tweets are collected (named bag of tweets) or a timer counts more than y seconds since the last received tweet, SMC performs spatial clustering, empties the bag and resets the timer. The idea behind the first condition is to have a sufficient number of tweets to group and the idea behind the second condition is to be responsive, in case that it takes a long time between incoming tweets. The variables x and y can change; in the current implementation clustering is initiated when 10 tweets are collected or one minute passes. When clustering is completed, the clusters are presented as separate HTML files, which are called Twitter Reports. Each Twitter report contains the list of tweets (text, date) it comprises and the reason that they are produced as HTML is in order to be displayed in the PSAP (see section 2.5.3). Then, a message for each Twitter report is created and sent to the cloud service bus (in topic TOP003_SOCIAL_MEDIA_REPORT), so that the Knowledge Base Service (KBS) creates a corresponding incident.

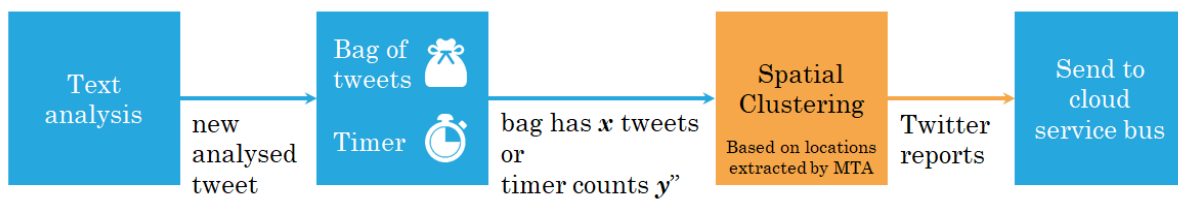


Figure 6. The flow of SMC in the beAWARE system.

2.4 Event Detection

The continuous crawling of tweets for more than two years, in the frame of the Social Media Monitoring task, has led to a very large collection of social media posts, which can be now exploited for examining whether they share beneficial insights and how they can be used to gain further knowledge on crisis incidents. To prove their value in regards to detecting events, two different experiments have been carried out: the first is an investigation of peaks in the fluctuation of daily collected tweets and the second a calculation of the correlation between social media data and weather data.

2.4.1 Observation of events in social media data

The target of this study was to show that peaks in the fluctuation of the number of tweets that have been collected every day for a specific topic can be perceived as incidents in the real world. The methodology comprised three steps: (a) select the use case and the period of examination, (b) create a line chart with the fluctuation of tweets, and (c) manually match the peaks of the line to real events.

The selected topic was tweets that were posted during the whole year of 2018, were in Spanish language and referred to fire incidents, i.e. contained any of the following keywords: “incendio” (fire), “llamasdefuego” (fire flames), or “bomberos” (firefighters). This selection

resulted to 2,773,237 tweets and their fluctuation per each day of the year can be seen in Figure 7. Please notice that the zero number of collected tweets for some periods is due to technical problems and does not mean that there were not any relevant posts shared during these days. The analysis of these streams of tweets is in line with Ethics requirements of beAWARE as reported in D2.9 (M18).

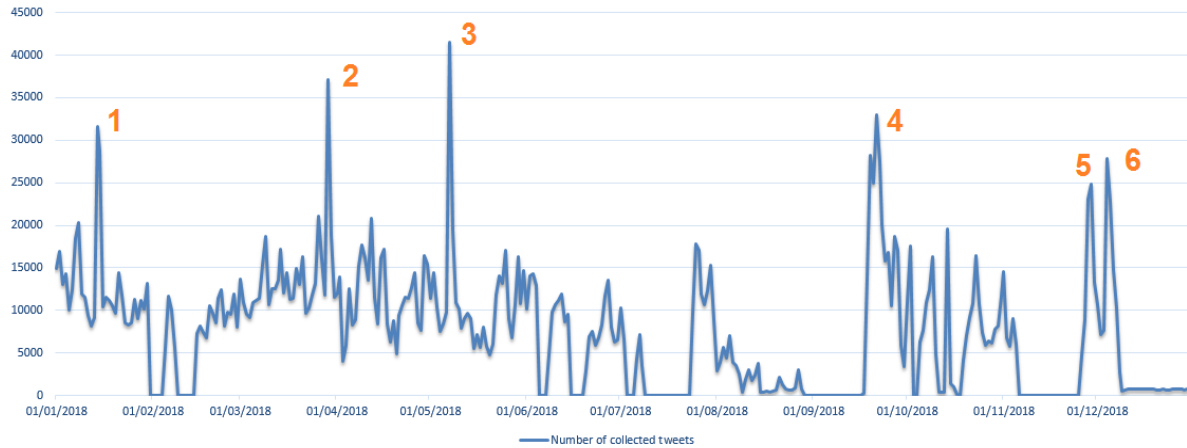


Figure 7. Fluctuation of the number of collected tweets per each date of 2018.

In Figure 7 one can notice that there were six peaks throughout the year (14/01/2018, 29/03/2018, 07/05/2018, 21/09/2018, 29/11/2018, and 04/12/2018) and they could possibly be linked to real events. Apart from the quantity of tweets, their content was expected to be useful too, especially in order to obtain more insights on the detected events. After pre-processing the text of the tweets (i.e., removing punctuation, stop words, and URLs), the most frequently mentioned words were retrieved. Table 5 presents one list with the top ten locations (right) and another list with the top ten keywords that are not locations (left). It should be noted here that the three words that were used a search terms (i.e., incendio, llamasdefuego, bomberos) are not included in the table, since it is obvious that they are mentioned a lot.

Table 5. Most frequently mentioned words inside the collected tweets.

#	Non-locations		Locations	
	Keyword (with English translation)	Appearances	Keyword	Appearances
1	forestal (forest)	160,732	Valencia	19,711
2	persona (person)	93,219	Zaragoza	18,236
3	control (control)	91,658	Madrid	18,188
4	cuerpo (body)	83,597	Barcelona	5,898
5	reportar (report)	81,540	Getafe	5,790
6	policía (police)	79,872	Malaga	5,639
7	viviendo (living)	76,660	Bilbao	5,230
8	apagar (put out)	74,377	Cartagena	2,860

9	rescatar (rescue)	73,952	Huelva	2,836
10	sector (sector)	72,976	Granada	2,780

The two top ten lists cannot offer concrete details on the various events during the year, but it is a first step to confirm that there is valuable information in them. Applying this method to single dates led to much more specific keywords. The top five terms for each of the six detected events are displayed in Table 6 and they were used as a guidance for the verification of these events. Indeed, on January 14 there was a fire in the park of Ávila⁷, on March 23 a fire during a prison riot in Carabobo, Venezuela⁸, on May 7 an accusation against firefighters of trafficking⁹, on September 21 a scandal with president Maduro¹⁰, on November 29 a big strike¹¹, and on December 4 a viral video with a fire department turning its back on the authorities¹². In this way, it can be verified that the peaks in the chart line are actually related to real incidents, which are successfully captured by the monitoring of social media.

Table 6. Top five keywords for each date that is considered an event.

#	Date	Keyword (with English translation)	Appearances
1	14/01/2018	atropellar (outrage)	11,456
		manguera (hose)	8,275
		cabecal (chief)	7,584
		Ávila	4,594
		arrollar (overwhelm)	3,181
2	29/03/2018	PoliCarabobo (Carabobo Police)	13,122
		muerto (dead)	11,883
		comando (command)	11,454
		motines (riots)	8,569
		Venezuela	6,624
3	07/05/2018	Lesbos	11,993
		proem (Proem-Aid)	10,322
		español (Spanish)	8,526
		sevillano (Sevillian)	8,254
		juzgar (judge)	8,194

⁷ <https://www.laprensalarara.com.ve/nota/-7140/2018/01/controlado-incendio-producido-en-parque-el-avila>

⁸ https://en.wikipedia.org/wiki/2018_Valencia,_Venezuela_fire

⁹ <https://www.elperiodico.com/es/internacional/20180507/juicio-bomberos-proem-aid-en-lesbos-6803882>

¹⁰ <https://www.theguardian.com/world/2018/sep/18/venezuela-nicolas-maduro-salt-bae-istanbul>

¹¹ <https://www.thelocal.es/20181130/catalonia-doctors-nurses-teachers-civil-servants-strike-over-austerity>

¹² <https://www.mibrujula.com/cuerpo-bomberos-da-espalda-autoridades-ceremonia-oficial-francia/>

4	21/09/2018	calendario (calendar)	6,550
		años (years)	6,435
		Maduro	6,193
		venezolano (Venezuelan)	5,258
		video	5,072
5	29/11/2018	médico (medic)	6,616
		parlamento (parliament)	5,119
		estudiante (student)	3,139
		Generalitat	2,764
		manifiesto (manifest)	2,600
6	04/12/2018	cuerpo (body)	14,983
		espalda (back)	14,804
		oficial (official)	14,631
		ceremonia (ceremony)	14,625
		andremoreaul (André Moreau)	14,584

2.4.2 Correlation of social media and weather data

The second study in order to prove that the fluctuation of collected tweets could be used to identify real-world events was an investigation on the possible relation between social media data and weather observations/predictions. For example, if the rise and decrease of tweets about heatwave for a specific region follows the variation of temperature in that region, then it means that social media are able to reflect incidents that are connected to weather measurements, such as heatwaves or heavy rains.

To this end, a set of nine time series, four of social media and five of weather data, was defined to measure the correlation between pairs of the two groups. The period of examination was the months of July and August of 2017 and the area of interest was Greece, since there were many heatwave incidents. The time series created were:

1. Social Media
 - a. Greek Heatwave Collected: number of tweets per day that have been crawled
 - b. Greek Heatwave Annotated: number of tweets per day that have been manually annotated as relevant or irrelevant
 - c. Greek Heatwave Relevant: number of tweets per day that have been manually annotated as relevant
 - d. Greek Heatwave Images: number of tweets per day that include an attached image
2. Weather Data
 - a. Observed Max Temperature: the maximum measured temperature per day
 - b. Observed Mean Temperature: the mean measured temperature per day
 - c. Observed Mean Wind Speed: the mean measured wind speed per day

- d. Predicted Max Temperature: the maximum predicted temperature per day
- e. Predicted Mean Temperature: the mean predicted temperature per day

The above weather measurements resulted from the mean average of all stations in Greece and of all recorded times during the day. It should be also mentioned here that these weather data were provided by FMI and the human annotation of tweets was performed by HRT.

To measure the association between pairs of social media and weather data, the Pearson correlation coefficient¹³ was selected, also referred to as Pearson’s *r*. Its value ranges between -1 and +1, where -1 is total negative linear correlation, 1 is total positive, and 0 is no correlation. To verbally describe the strength of the correlation, the guide in (Evans, 1996) was followed, which suggests for the absolute value of *r*:

- 0.0 – 0.19 very weak
- 0.2 – 0.39 weak
- 0.4 – 0.59 moderate
- 0.6 – 0.79 strong
- 0.8 – 1.0 very strong

The complete results are presented in Table 7, where strong correlation is indicated with green and moderate with light green. In July 2017, there is a strong correlation between the measured mean temperature and all social media time series, while tweets with images are strongly associated with all types of temperatures. On the other hand, the mean wind speed shows a very weak association with social media. In August 2017, there is an even stronger correlation between collected tweets and temperatures, a moderate correlation between tweets with images and temperatures, but a very weak correlation for annotated and relevant tweets. This can be easily explained, because the human annotation for August was much less rich than for July.

The general conclusion that arises from the table is that the quantity of posts about heatwaves is affected by the measured or predicted temperature, thus social media are able to reveal incidents that are linked to the weather.

Table 7. Pearson correlation coefficient between social media and weather data.

Examined Period	Weather	Social Media			
		Greek Heatwave Collected	Greek Heatwave Annotated	Greek Heatwave Relevant	Greek Heatwave Images
01/07/2017 – 31/07/2017	Observed Max Temperature	0.5979	0.5986	0.6211	0.6562
	Observed Mean	0.6427	0.6380	0.6623	0.7001

¹³ https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

	Temperature				
	Observed Mean Wind speed	-0.2109	-0.2272	-0.2078	-0.2371
	Predicted Max Temperature	0.5417	0.5409	0.5823	0.6117
	Predicted Mean Temperature	0.5715	0.5530	0.5275	0.6117
01/08/2017 – 31/08/2017	Observed Max Temperature	0.7255	0.2493	0.2489	0.4786
	Observed Mean Temperature	0.7303	0.2223	0.2207	0.4983
	Observed Mean Wind speed	-0.2613	-0.0175	-0.0244	-0.0304
	Predicted Max Temperature	0.6860	0.2786	0.2777	0.5186
	Predicted Mean Temperature	0.6743	0.1768	0.1759	0.5133

2.5 Displaying the Data

During the project’s lifetime, different needs have prompted the implementation of different user interfaces, which have been described in D4.1. This section presents the changes of the two tools that show crawled tweets, as well as the display of Twitter reports on the PSAP.

2.5.1 Annotation tool

The beAWARE annotation tool serves a twofold purpose: (i) it can be used to label tweets as relevant or irrelevant, so as to create an annotated dataset for classification training, and (ii)

it can be used to see and search the complete collection of tweets that have been crawled so far by SMA.

The main change from the previous version is the addition of tags. The estimation of the verification is shown as “Real”/“Fake” tags, the estimation of emoticons checking as “Emoticons” tags with green or purple color, and the estimation of text and image classification as “Relevant”/“Irrelevant” tags. Some example of tags can be seen in Figure 8. Another change is that the authors of the tweets are not displayed anymore, to further protect their privacy.

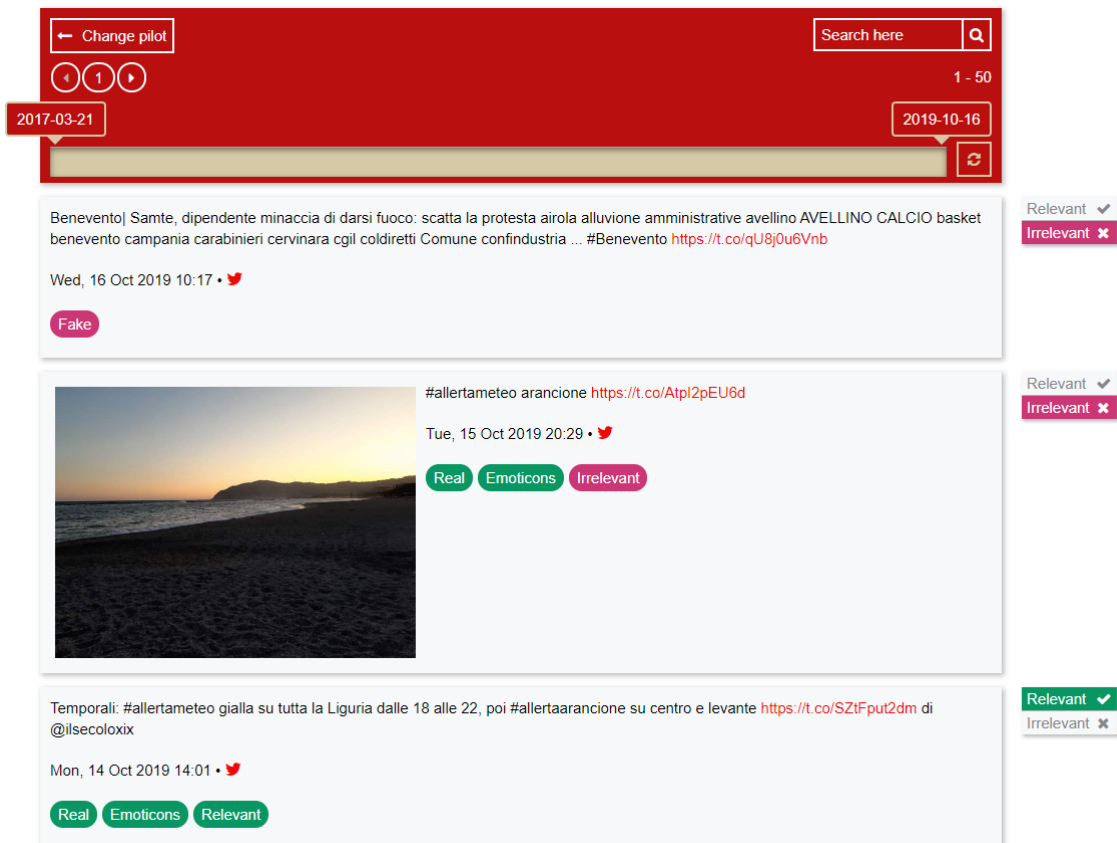


Figure 8. Updated interface of the annotation tool.

2.5.2 Demonstration tool

The second interface is the beAWARE demonstration tool, which is used during the pilots and shows the tweets at the same instant they are collected. Compared to the previous version, there have been two additions: the relevancy tags mentioned also in 2.5.1 and the option to send simulated sets of tweets per pilot (e.g., a set of relevant tweets about smoke in Valencia). The tags are visible at the right of Figure 9, while the option to select simulated tweets can be seen at the upper left.

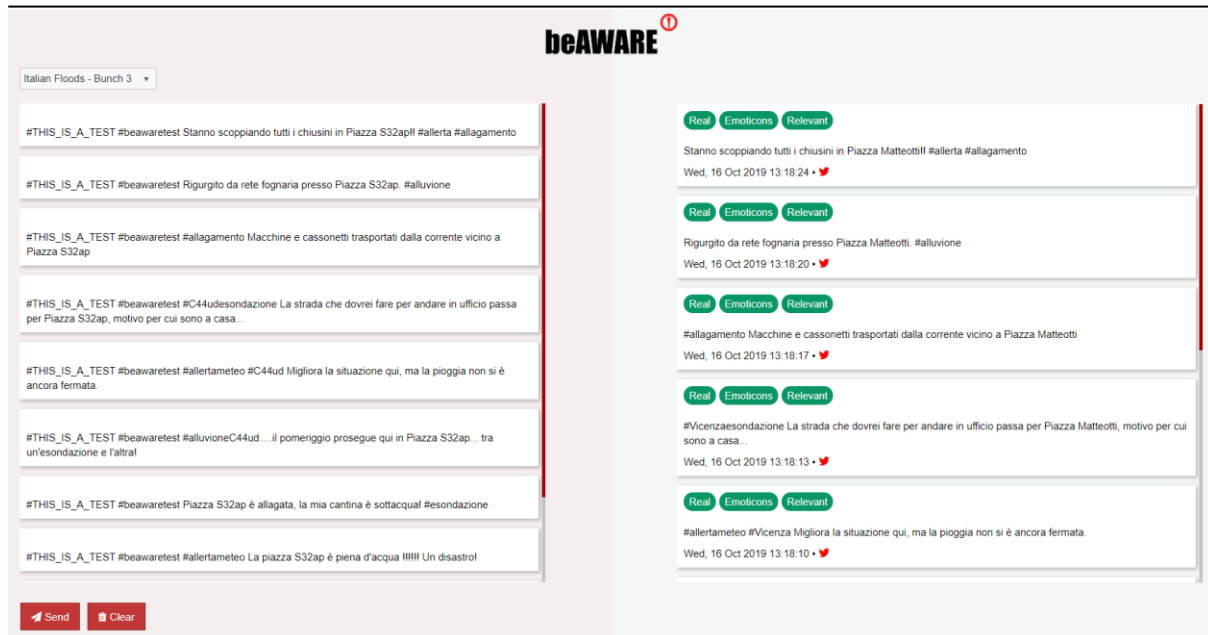


Figure 9. Updated interface of the demonstration tool.

2.5.3 Social media in PSAP

As explained in section 2.3.3, the clusters that are produced by SMC are converted to HTML files, i.e. Twitter reports, and KBS creates respective incidents to be visualised in PSAP, with the Twitter reports as attachments. In Figure 10, the occurring incidents are seen in the Monitor at the bottom left and, by clicking one incident, details and attachments appear at the bottom right. Clicking on the URL of an attachment, a new tab opens in the browser with the Twitter report (Figure 11). Moreover, the incidents are shown on the map, where the location of a Twitter report is the center of the locations of the tweets it consists of.

It should be noted here that there is no need to show in PSAP the verification or the classification results, because fake and irrelevant tweets are already filtered out in SMA and only real and relevant tweets are sent to MTA and used in SMC.

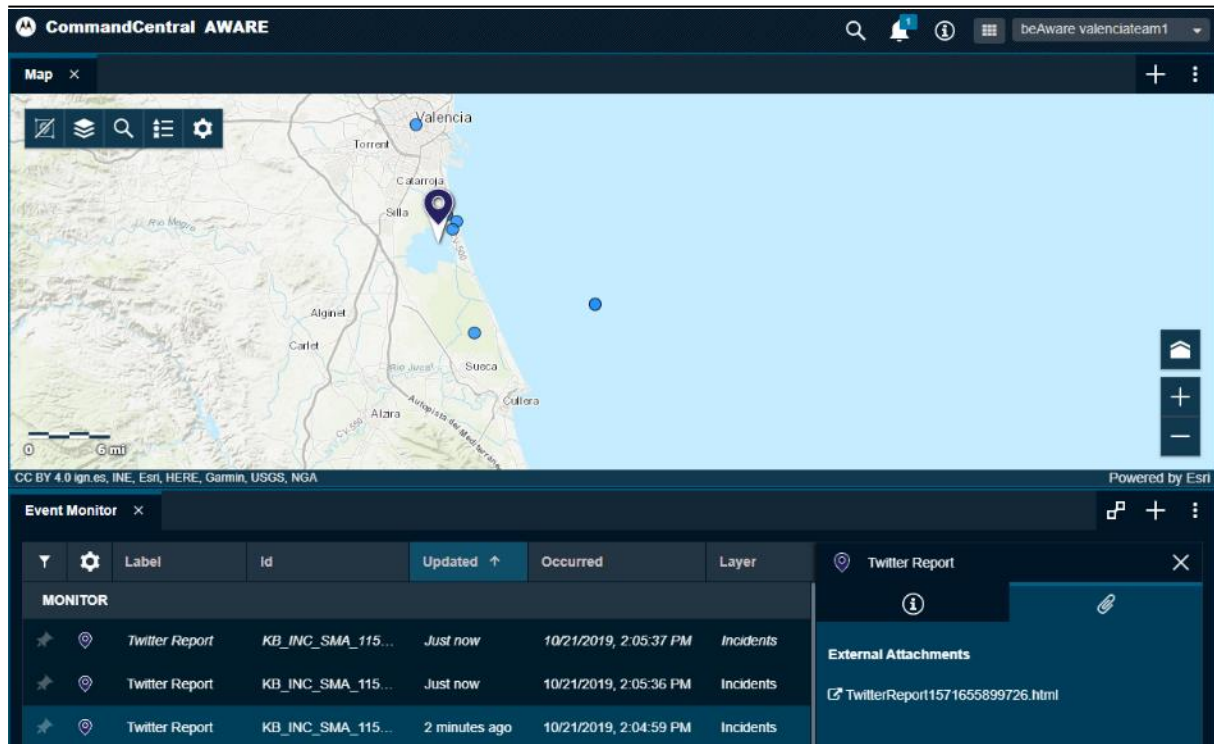


Figure 10. Twitter reports on PSAP.

Fuerte viento en la zona de la Devesa, hay que extremar las precauciones por peligro de incendio

Posted at Mon Oct 21 11:04:18 +0000 2019

[Show on Twitter](#)

Riesgo muy alto de incendio en la Devesa debido a las altas temperaturas

Posted at Mon Oct 21 11:04:22 +0000 2019

[Show on Twitter](#)

Mucho viento y calor en la Devesa, hay mucho riesgo de que se prenda fuego

Posted at Mon Oct 21 11:04:25 +0000 2019

[Show on Twitter](#)

Hay peligro de que se prenda fuego la Devesa, altas temperaturas y fuerte viento, estad alerta

Posted at Mon Oct 21 11:04:32 +0000 2019

[Show on Twitter](#)

Hay riesgo extremo de incendio forestal en la Devesa debido a la baja humedad, el fuerte viento y las altas temperaturas

Posted at Mon Oct 21 11:04:39 +0000 2019

[Show on Twitter](#)

Figure 11. The HTML file of a Twitter report.

2.6 Chapter Summary

This chapter made a short description on the collection and analysis of social media, focusing mainly on the updates since the previous corresponding deliverable. Two techniques have been added to improve the quality of incoming information: one that detects fake news in tweets and one that checks for irrelevant emoticons. The large number of crawled tweets has offered the capability to identify ways to gain further knowledge, so clustering and event detection methodologies were investigated and evaluated. The chapter continued with the modifications made on the annotation and demonstration tools and concludes with the display of Twitter reports in the PSAP interface.

An update on the decision support-related modules follows in the next chapter.

3 Decision Support

Decision support related modules were updated in order to support the increased functionality of the beAWARE analysis modules. In detail, the ontology was extended to fit the needs for the three pilots that have been conducted since the first version of the beAWARE ontology have been described in deliverable D4.2. Moreover, the Knowledge Base Service was updated in order to integrate the valuable information from UAVs to the Knowledge Base. Additionally, new metrics from Crisis Classification were integrated into the Knowledge Base. Further, a two-layer validation technique was developed to protect the system from misuse and incorrect data. Finally, two user interfaces have been integrated into the KB, which enable users to manually analyze and explore the data.

3.1 Additions to the Knowledge Base Ontology

A couple of additions have been made to the beAWARE Ontology described in deliverable D4.2. In the following, the two main sources of the additions will be described, while a full list of the changes is provided in Appendix A.

3.1.1 Representing Stakeholders and Actions

In order to be able to model the different stakeholders that will be involved in a crisis, several concepts, which are depicted in Figure 12, have been added to the ontology. Each stakeholder can thereby have a certain role, be an expert in a certain domain and can execute different actions. Each action will further have the goal to mitigate a certain impact, while a response to an incident can consists of one or more actions.

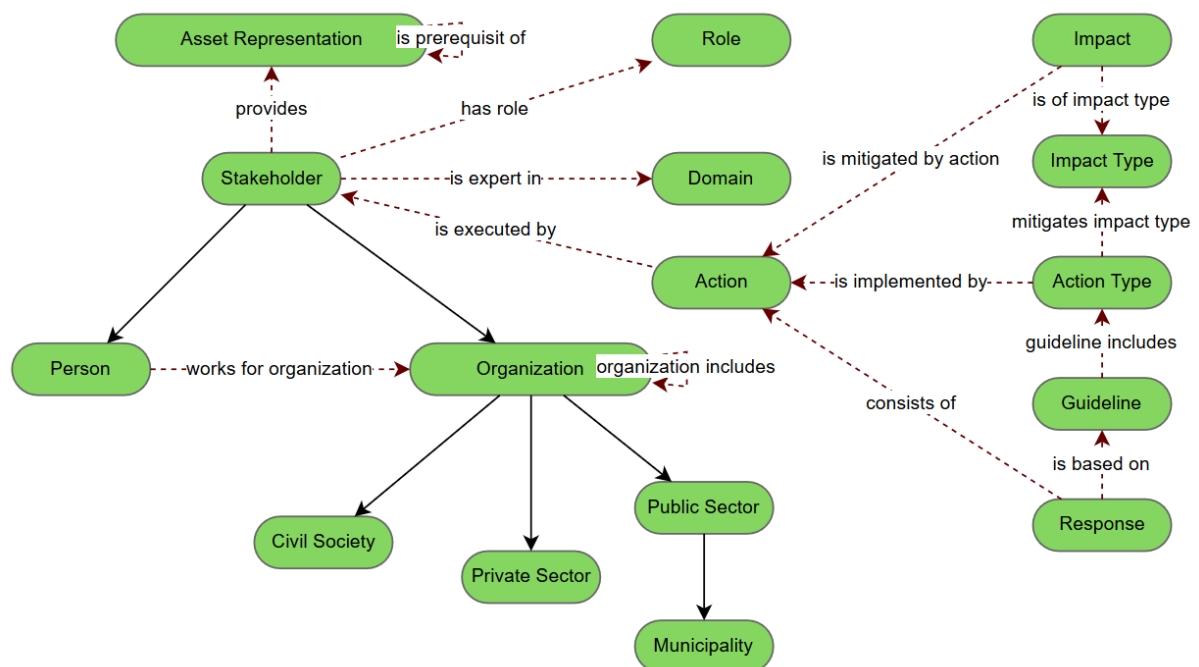


Figure 12. Ontology representation of stakeholders that will be involved in a crisis and the actions they can make to mitigate an impact.

3.1.2 Ontology Additions related to Improvements of Analysis Components

Several additions to the ontology have been made in order to adapt it to improvements of the analysis components. Among others this includes the new subclasses of the *VulnerableObject* concept that have been added, which reflects the different objects that now can be detected by the image and video analysis. These are the concepts *Bicycle*, *Boat*, *Bus*, *Motorcycle*, *Truck*, *Cat*, *Dog*, *Dummy* and *WheelchairUser*. Further three new datatype properties have been added to the *IncidentReport* concept. The *isSpam* data type property thereby will be used to store the result from the social media analysis described in section 2.2. With the *hasSeverityFromCRCL* data type property the severity level calculated by the CRCL component can be assigned to an *IncidentReport*. The third data type property *hasOriginator* is used to preserve the originator of an incident report.

3.2 Integration of Incident Data from Drones

In a nutshell, UAV footage is subjected to video analysis by the respective beAWARE component, as is any other footage harvested by the beAWARE sensors (application, social media, etc.). The main difference lies in the fact that drones provide a live video stream, rather than a clip of finite duration. Thus, in order to achieve live processing and take full advantage of the provided aerial overview towards precise decision support, the incoming footage is segmented into clips of pre-defined duration, which are then processed by the established video analysis pipeline.

Since a UAV video stream might contain redundant parts, like take-off, landing and flying towards the target zones, a filtering mechanism has been established to avoid flooding the Knowledge Base (KB) with incident reports that do not contain any actionable knowledge (i.e. detected incidents, vulnerable objects, etc.). As a result, UAV media analysis output – communicated to the message bus by topic *TOP019_UAV_MEDIA_ANALYZED* – contains a relevant flag to facilitate the Knowledge Base Service (KBS) with the distinction of useful detections that need to be populated to the KB. Consequently, semantic reasoning is applied to the new, filtered knowledge and the appropriate actions take place for the notification of the PSAP for new incidents, along with locations, participants and detected severity levels.

In conclusion, it has been achieved to seamlessly integrate UAV media in the thoroughly tested beAWARE workflows, while at the same time managing the peculiarities of the media's nature effectively.

3.3 New Metrics from Crisis Classification

Since the Crisis Classification component (CRCL) preserves a higher overview on multiple data sources, including weather forecasts and elaborate detection algorithms, it was considered crucial to provide mechanisms for the direct reporting of detected incidents from the CRCL to the KBS. To this direction, two new message bus topics have been defined and monitored by the KBS:

- *TOP006_INCIDENT_REPORT_CRCL* – Allows the broadcasting of incident reports identified by the CRCL itself. The KBS, as the recipient of such messages, manages the semantic fusion by populating the new knowledge to the KB, and the application of semantic reasoning mechanisms. These steps are identical to the way the KBS

handles all reported incidents from the other beAWARE modules (application, social media, etc.). Finally, the incident reports are transmitted accordingly to the user interface.

- *TOP007_UPDATE_INCIDENT_RISK* – As opposed to the previous case, this message type is intended for the CRCL to inform the KBS of a new severity value that needs to be assigned to an already populated incident. The CRCL component is charged with monitoring reported incidents, combining data from multiple sources and, based on thorough crisis classification algorithms, identifying potential risk levels by the augmented knowledge. Respectively, the KBS assumes to update the new metrics to the KB, and inform the PSAP.

3.4 Two-layer Validation Component

Incorrect or malicious information could affect negatively the effectiveness of the system, thus it is important to take measures against this happening. A two-layer validation system that filters out such incorrect information was deployed to ensure the seamless operation of the system even in the case of misuse or even malware attacks. The first layer of protection takes place individually at the Visual Analysis and Social Media Analysis modules, while the second layer of protection is an independent validation component that works on later stages of the pipeline and regardless the provenance of the input as an extra protection level.

3.4.1 First Validation Layer

The first layer of the validation system is an integral part of three components: the Image Analysis, the Video Analysis and the Social Media Analysis (SMA). The filtering capability of these modules serves as a mechanism to protect the main system from malicious or simply false information.

Image and Video Analysis are responsible to examine every image/video item that is forwarded to the system through the Mobile App and discard them if the result after the analysis is irrelevant for the beAWARE system in general. With the exception of some operations defined with the help of exception flags, the components always filter the results through the internal validation mechanism. The exception flags are:

- Accreditation documents from first responders are forwarded through the Visual Analysis system without examination, using the appropriate flag in the Mobile App's report form.
- Other items sent from inside the beAWARE system are not filtered, since they are considered already validated. One such case was the traffic analysis on the Angeli Bridge in Vicenza during the 2nd pilot. These videos were sent from the Visual River Sensing component, in order to further analyze the traffic on top of the bridge whenever the water underneath was found to exceed a predefined level.

Every request for visual analysis that does not carry any of the above exception flags is analyzed and the results filtered through the internal validation mechanism.

The validation mechanism was first developed and integrated for the 2nd pilot, where the flood use case was under examination. Therefore, in the first version each image or video that was found by the Emergency Classification module (EmC) as flood-related was forwarded through

the filter. If an image or video did not contain any flooded area, the filter discarded the request and no further analysis was performed or results forwarded to the system. The beAWARE system will be able to recognize and handle two other incident types by the time the final prototype is finished: fire and smoke incidents. Accordingly, our EmC module has been updated to also recognize images and videos that show fire or smoke. Since it is considered dangerous for a person, a vehicle or an animal to be located not only on the actual burning site but even at a nearby location, it is not given that an image or video that contains targets but not actual flames or smoke is not relevant to the system. Consequently, the validation mechanism has been updated for the final prototype in order to forward items that show either an emergency taking place or targets that are possibly in danger (i.e. people, animals and vehicles) even if no flames or smoke is detected.

Regarding Social Media, SMA integrates three steps in its analysis to filter out dangerous or unnecessary information coming into the system. The first step is to throw away fake news with the verification technique introduced in 2.2. The second step is to exploit emoticons in the post, again described in 2.2, in order to discard tweets that seem unimportant (e.g., a tweet about a flood that contains a smiley face cannot be of great significance). The third step is use case-related, since text and image classification techniques estimate how relevant the tweets are to the examined natural disasters.

3.4.2 Second Validation Layer

The second validation layer (or Validator, VAL) works alongside with Knowledge Base Service (KBS) and the Crisis Classification module (CRCL) to detect implausible instances. VAL is a standalone module that communicates with KBS and CRCL via the message bus using newly established topics. VAL consumes asynchronously the output of KBS and crosschecks it with CRCL in order to determine if the output is valid. In detail, what is checked is whether the declared incident type (e.g. Fire, Flood, Precipitation, Heatwave ...) is in accordance with the metrics provided by CRCL for the location and time of the particular instance. The validation results are propagated back to KBS and consequently to PSAP when necessary.

Relevant metrics provided by CRCL:

- Weather data (precipitation, humidity, temperature, wind).
- Water level (case of flood), both real-time values and predictions.
- Fire weather index (case of fire).
- Discomfort index (case of heatwave).

The new topics are:

- *TOP802_WEATHER_REQUEST*: From VAL, to CRCL. This topic is used whenever a new instance is under examination by VAL and its purpose it to probe CRCL to return relevant metrics for the particular location and time of the incident.
- *TOP803_WEATHER_REPORT*: From CRCL, to VAL. This topic is the reply of the previous topic and consists of the relevant metrics for the particular location and time, in order for the validation to be completed.
- *TOP801_INCIDENT_VALIDATION*: From VAL, to KBS. This topic is the result of validation and denotes if a particular incident is erroneous or not, using the Boolean field *spam*.

The algorithm of the second layer of validation:

1. VAL consumes *TOP030_REPORT_REQUESTED* and processes the included incidents.
2. In case more information from CRCL is necessary:
 - a. Sends *TOP802_WEATHER_REQUEST* to request additional information for a specific incident given its parameters (position, time).
 - b. Receives *TOP803_WEATHER_REPORT* for the corresponding request.
3. Performs validation given the incident information and the metrics.
4. Sends *TOP801_INCIDENT_VALIDATION* to inform KBS with the validation result.
5. KBS receives *TOP801_INCIDENT_VALIDATION* and updates PSAP with the new incident information if necessary by sending a *TOP101_INCIDENT_REPORT* event.

3.5 User Interfaces for Additional Data Analysis

In addition to the automated analysis, processes of the beAWARE system different user interfaces have been integrated into the KB, which shall enable analysts or decision makers to conduct a further analysis and exploration of the data either during the event or afterwards. The interfaces for navigating through the semantic content together with the access to the river level measurements have been already described in deliverable D7.5. For the final system two other views the Analysis Workbench and the Incident Map and Table were integrated.

3.5.1 Analysis Workbench

While the focus of the PSAP lies on displaying the current situation to the emergency managers, the analysis workbench was developed to enable analysts to search for specific information in the data that have been collected since the beginning of a crisis. This could for instance be useful if a change of the situation raises questions or a deeper understanding of the data (since the PSAP shows the aggregated and processed information) is needed. This might be useful either during or after an event.

Since it is impossible to anticipate all possible situations that might arise during a crisis, it is important to give analysts as much freedom as possible when filtering the data. The high time pressure that exists during a crisis also makes a good visualization of the data very important in order to help analysts to get a faster overview of the data. These two points combined with the facts that the data collected during a crisis is heterogeneous and that the data, in case of the beAWARE platform, is not stored in a central location have been the main challenges in the development of this tool.

Design Considerations

Following assumptions have been made for the kind of data that shall be analyzed in the workbench. The workbench shall be able to analyze data from different data sources like for example Twitter, the beAWARE mobile app or also something like the FROST server. While data from different data sources are assumed to be heterogeneous, data from the same data source, like for example tweets or mobile app messages, are assumed homogenous. In order to enable a high freedom concerning the analysis and visualization of the data, it is therefore important to include data source specific features into them. For example, this could be data source specific filters.

In order to make it easier to adapt the workbench to new data sources that might arise in the future or that depend on the needs of the end users, the base structure of the workbench was developed independently from the data sources. This was achieved by creating an application-programming interface (API) and service provider interface (SPI) over which a data source can be added to the analysis workbench in form of a plugin. The base structure of the workbench is responsible for the general layout and the available visualization and analysis components. The data source plugins on the other hand will take care of the data retrieval from the respective database and will provide data source specific information for the visualization and analysis components that are included in the base structure. The architecture of the analysis workbench is shown in Figure 13.

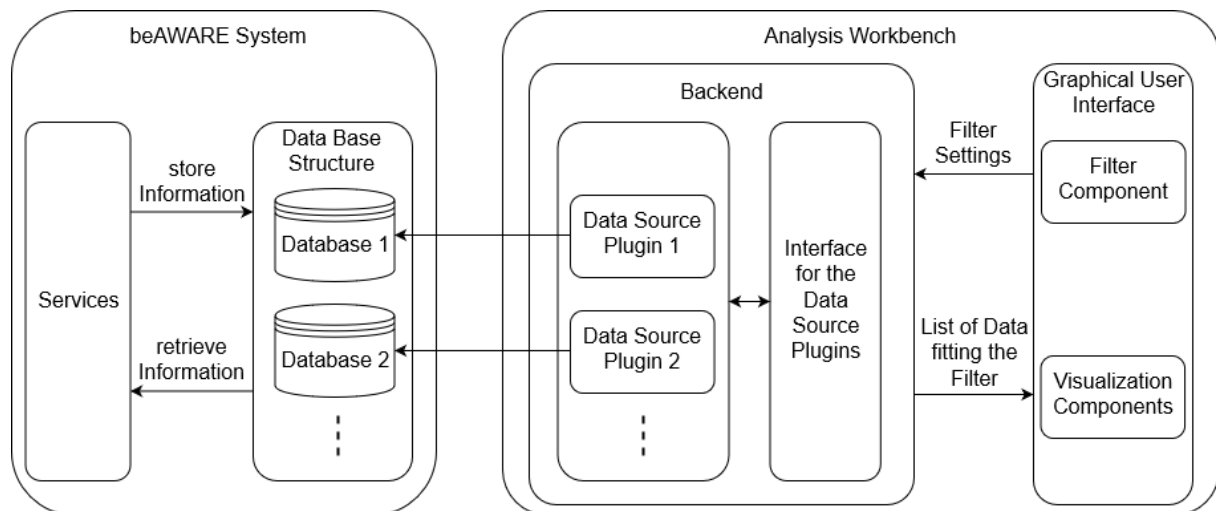


Figure 13. Architecture of the Analysis Workbench

Implementation of the Analysis Workbench

For the implementation of the base structure, we decided to divide it into three parts. These parts are a map and a list for the visualization of the data as well as a filter component for the analysis of the data. In order to be able to include data source specific features we decided that the list and the filter component uses only one data source at a time. This allows optimizing the sort options in the list and the available filters to each data source separately. A location-based overview of the data of all data sources and the ability to find location-related connections between different data sources is thereby still available over the map.

For the development of the workbench, we further took principles from the field of interactive visual analysis into account. Heer and Shneiderman (2012) gave a taxonomy of those. One of these principles is the so-called brushing and linking of data. This refers to the connection of different views by highlighting the selected data in one view also in the other views (brushing) and by transferring the filter results from one view to others (linking). We decided to use a tab system in which the list and the filter component are integrated, in order to realize the linking between the map and the list. While for each tap one data source needs to be chosen, which will be used for the two components, it is possible to select different data sources for different tabs. The linking between the map and the list can then be achieved by displaying the union of the list entries from all tabs in the map. A draft of this layout is shown in Figure 14.

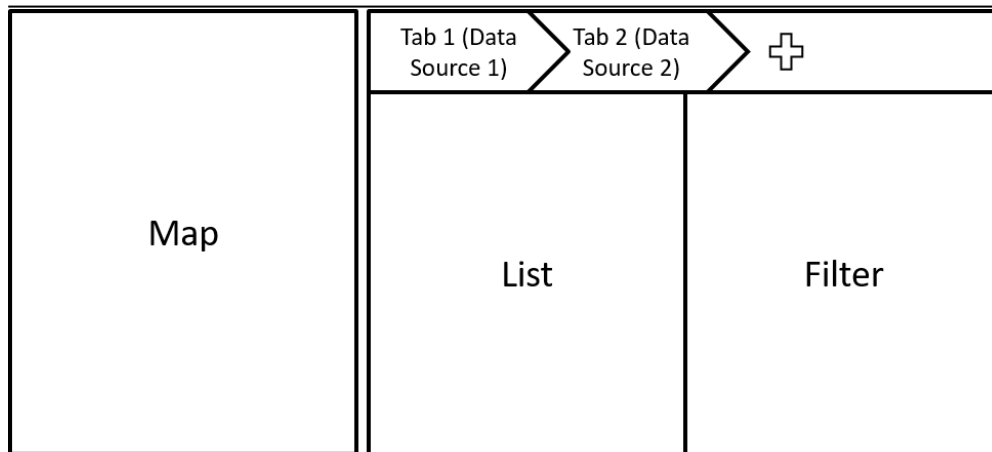


Figure 14. Layout of the Analysis Workbench

With the condition that a list entry also has a location associated with it, the brushing between the map and the list was realized by changing the color of a marker and its associated list entry after a user clicks on either of them. In addition to the change of the color, the map will zoom to the marker and the tab system will switch to the according tab in which further the list will be scrolled to the associated entry respectively.

The analysis workbench has been built into the WebGenesis platform, which is also used to host the Knowledge Base (KB) that has been described in deliverable D4.2. The message workbench is therefore implemented as a web application, which enables it to be used without a specific installation on multiple devices. The user interface is shown in Figure 15.

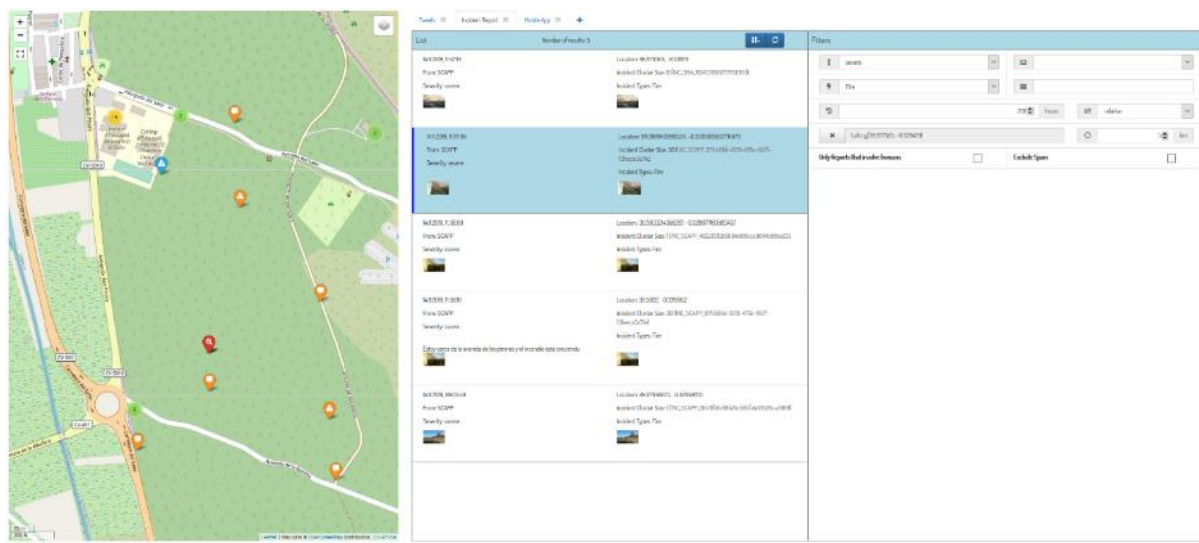


Figure 15. The implemented user interface of the Analysis Workbench.

Map Component

The map has been realized with the map service of the WebGenesis system. Additionally, new map layers can be created, which the user on demand can display. This has already been described in deliverable D7.5. For a better visualization, the map will cluster the markers in it depending on its zoom level. The layout and the color of the markers can thereby be defined in the data source plugins for each data source separately.

List Component

An enlarged picture of the list is shown in Figure 16. The available sort options and the layout of the single list entries are defined in each data source plugin separately. With the button next to the sort options, the list can be updated manually.

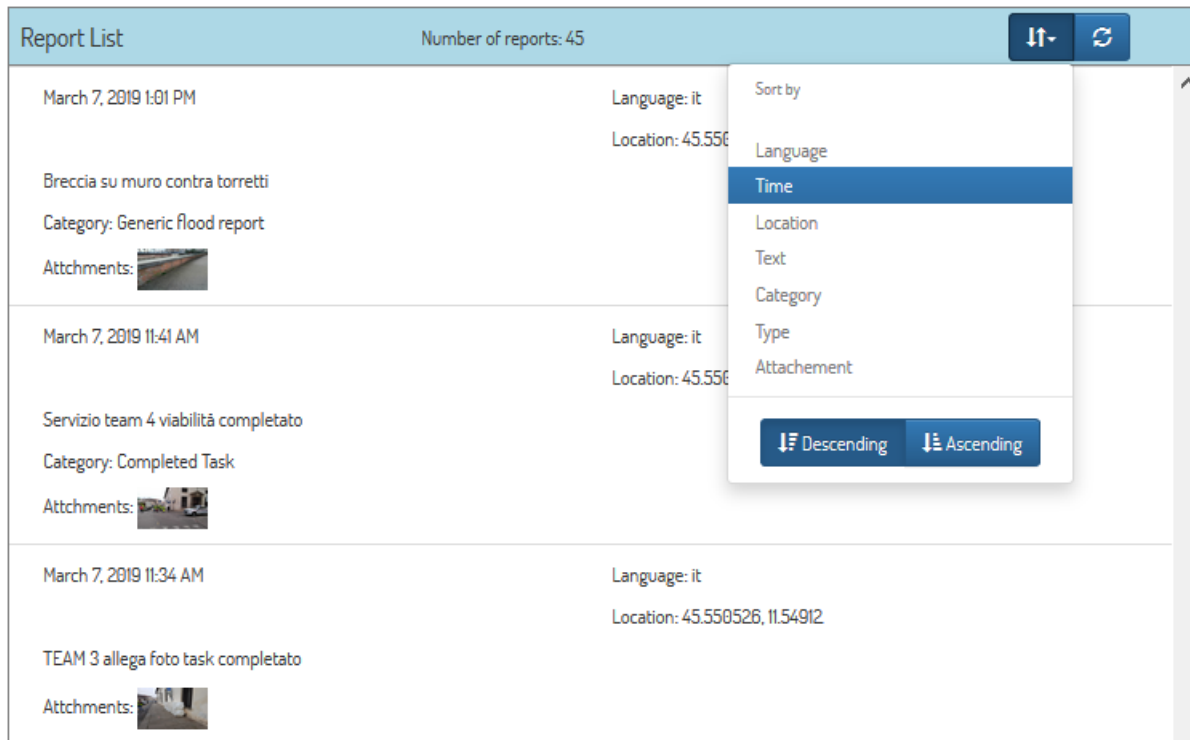


Figure 16. The list component of the Analysis Workbench. Reports from the beAWARE mobile app have been selected as data source.

Filter Component

The entries displayed in the list depend on the selected filters of the filter component. The available filters are also defined in the data source plugins. As an example, the filter component for tweets is shown in Figure 17. A short description of the single filters will be provided in form of a tooltip when hovering over the symbol in front of the filters. For the filters all available information from the data source can be used, which also includes results from automated analysis processes, as long as the analysis result are reported back to the data source. For the tweets, this is for example the case for the checkbox filter to include relevant tweets only, which is a result of the automated analysis described in section 2.2.

Report Filters	
Spanish	Fire
24 hours	relative
LatLng(39.533174, -0.359116)	20 km
Only Relevant Tweets	Only Geotagged Tweets <input checked="" type="checkbox"/>
Exclude Retweets <input checked="" type="checkbox"/>	Only Tweets with Attachements <input type="checkbox"/>

Figure 17. The filter component of the Analysis Workbench for tweets.

Data Source Plugins

For the data sources that are used in the beAWARE platform, the tweets, which are stored in the MongoDB, the reports from the beAWARE mobile app, which are stored in the KB and the incident reports generated by the KBS have been identified as the data sources that benefit the most from the Analysis Workbench. Therefore, for these three a data source plugin has been realized. The inclusion of the incident reports into the Analysis Workbench thereby provides an additional analysis option of the incident reports to the Incident Map and Incident List, which will be described in the next section (3.5.2). Another possible data source that could be included into the Analysis Workbench is for example a specific sensor data streams stored in the FROST server. The usefulness of creating a data source plugin for the FROST server would however depend a lot on the sensors that the end users are using, i.e. whether the ability to filter on their data in combination with the map view would be helpful or not.


Possible Additions to the Analysis Workbench

There are several possibilities how the Analysis Workbench could be further improved in the future, which due to time reasons have not been pursued during this project. The biggest improvements could probably be achieved by integrating additional visualization and analysis components. A time-based visualization like a time-line for example could be used as a second option for the combined visualization of all data sources. Further visualizations like pie charts, line graphs, steam graphs, etc. could be used for a better visualization of certain aspects of data source specific information.

While the high situational dependency of the text-clustering component that have been described in section 2.3.1, prevented it from being used in the automated analysis pipeline of the beAWARE system, it could still be useful as integration into the analysis workbench. For example could it be used by integrating it as another data source specific analysis component that would cluster the list entries and mark the different clusters in the list and the map with different colors. The text, which is used for the clustering, would thereby need to be defined in the data source plugins. By including the options of choosing different text cluster algorithms and their parameters that will be used for the clustering, trained analysts will be able to adapt the text clustering to their current needs.

3.5.2 Incident Map and Table

To get a better overview of all incidents, the KB offers two visualization options: as a table and on a map. In the table, all incidents with their respective number of detected humans, animals and objects are displayed. A screenshot of the table is shown in Figure 18. To help prioritizing the different incidents, the table can be sorted by any of its columns. Furthermore, the incidents are linked to their entries in the KB so that more information about an incident can be obtained by clicking on it.





FORREST FIRE VALENCIA | FLOOD VICENZA | RISK ANALYSIS | MESSAGE ANALYSIS | INCIDENT ANALYSIS | SEMANTIC CONTENT ▾ |  LOGOUT |

beAWARE > beAWARE+Ontology > Queries > IncidentListAndCount

List incidents and impact-count

Search
Edit
Export Data

Show entries

incident 	humancount 	animalCount 	objectcount 
Example: Flooding of Vicenza city centre	6	0	6
Example: Flooding of Vicenza Stadium	6	0	0
Incident_bd0b83ea-0bd4-4097-abad-35ebcfda0fff	6	0	0
Example: Flooded Via Carlo Scarpa	2	0	2
Incident_2	2	0	0
Incident_4859b70a-4b18-43c5-92c3-e3e9b1201a14	1	0	3
Incident_36f6d944-f04f-4aba-bba2-54c54af62fbf	1	0	0
Incident_6a8c9a45-5c34-4885-9f36-f95aa534867e	1	0	0
Incident_83216e8e-e954-482c-bacd-ef3478d31506	1	0	0
GreenLightIncident	0	0	2

Showing 1 to 10 of 13 entries

Previous
1
2
Next

Figure 18. Screenshot of the incident table included in the KB.

In the incident map all incident clusters are displayed, which are obtained from the spatiotemporal clustering method described in section 2.3.3. A screenshot of the incident map is shown in Figure 19. For each incident cluster in the map a link is provided, which will lead to a similar table as the incident table described above. In this table instead of displaying all incidents, only the incident belonging to the selected cluster will be shown. In contrast to the data shown in the PSAP, which is used for time critical decision-making, this offers the possibility to analyze the underlying data in more detail.



Incident Map

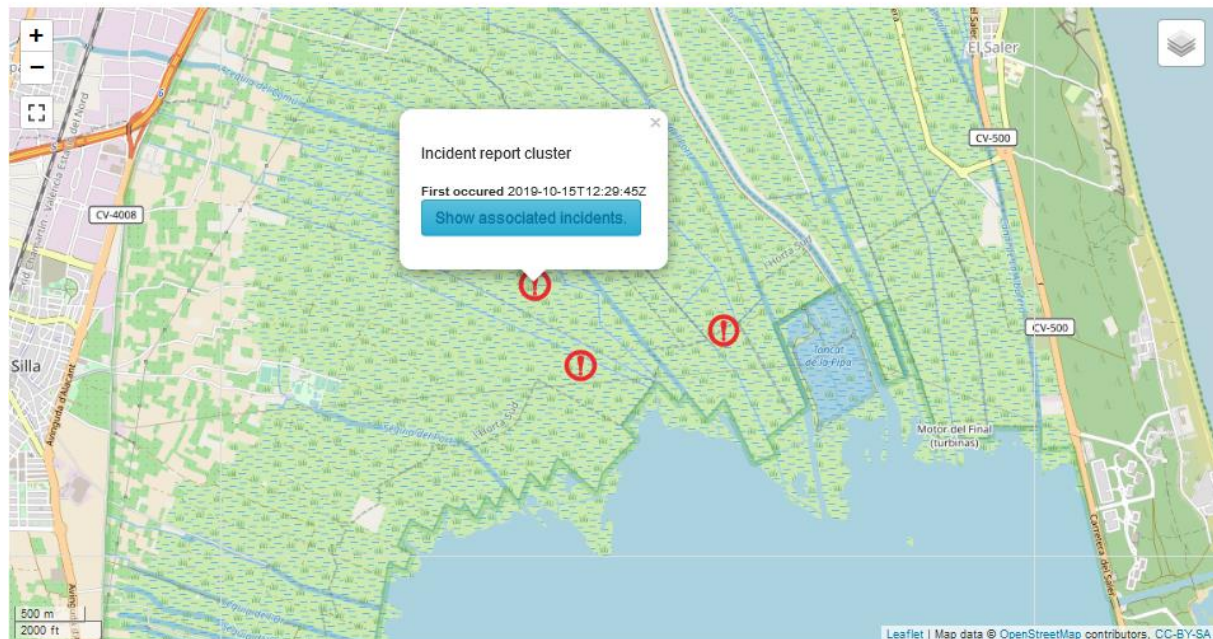


Figure 19. Screenshot of the incident map.

3.6 Chapter Summary

This chapter presented the latest additions to the KB ontology that correspond to the requirements for the third pilot. Additionally, the integration of UAV data into the KB was shown, along with the necessary adaptations in comparison to regular video analysis data. Moreover, the process of incorporating CRCL incidents and metrics into the KB was illustrated. Further, a two-layer validation component was introduced, along with details about its implementation. In the end of the chapter two user interfaces have been presented that enable a manual analysis and exploration of data, collected and generated during a crisis. In the following chapter the document is concluded with the final remarks and future steps.

4 Conclusions and Next Steps

The workflow of crawling social media data was shortly described, concentrating more on the modifications that have been made in order to improve the quality of collected data. A verification technique was presented, which examines various characteristics of an incoming tweet to determine whether it carries fake news, so as to discard it. A supplementary technique followed that checks for inappropriate emoticons (e.g., smileys and winks in tweets about a crisis incident) to filter out insignificant information. The indexing of the location of the tweets in the database was also mentioned as part of the changes in social media analysis.

The constant crawling of social media data during the project's lifetime led to a big amount of tweets, allowing their exploitation for additional techniques. First is the clustering of the numerous posts in order to have groups of similar items and thus more compact information. Two types of clustering were studied: one that is based on text content and another that is based on location. Both quantitative and qualitative evaluations were exhibited, while the best approach drove the implementation integrated in the beAWARE system, which was also presented.

The next technique that was explained is the detection of events. Using the fluctuation of the number of tweets collected per day for a specific use case (Spanish posts about fires in 2018), it was proven that peaks in the fluctuation can be perceived as real-world events, by matching the peaks to news articles. To enhance this statement, a second study was described that showed the strong correlation between the number of tweets of an alternative use case (Greek posts about heatwave in July and August of 2017) and weather data (e.g., predicted or observed temperature) in the same region at the same period.

The addition of new analysis techniques forced alterations in the tools that display the collected social media data, so the latest versions of the annotation and the demonstration user interfaces were illustrated. Moreover, the deployment of clustering in the beAWARE system produces incidents with Twitter reports (groups of relevant tweets) as attachments; these incidents in the PSAP application were also illustrated here.

Additions to the beAWARE ontology have been made in order to adapt it to improvements in the analysis components. Further a model have been integrated into the ontology that allows to describe the different stakeholders that will be involved in a crisis as well as the different actions that they can offer to mitigate certain impacts. Video data from UAV are a big asset toward crisis avoidance, so the KBS were modified to accept information from UAVs. While the UAV analysis is in essence video analysis, however, the continuous streaming video of UAVs requires a partially different approach than regular video analysis. A new pipeline was created within the KBS to accommodate for the produced information that includes a new message bus topic, filtering methods and adapted semantic integration. Finally, the resulting information that is propagated to the PSAP includes information about new incidents, their locations, participants and detected severity levels.

In regards to the CRCL advances, two topics were defined that facilitate the communication between CRCL and the KBS. The first topic (*TOP006_INCIDENT_REPORT_CRCL*) allows the CRCL component to provide information about newly detected incidents. Subsequently KBS integrates the information to the KB, in a similar way to information from other components. The second topic (*TOP007_UPDATE_INCIDENT_RISK*) is a communication channel that allows CRCL to update information about the already stored incidents. In detail, CRCL determines

new metrics about the already stored incidents and the KBS does the necessary updates to the KB.

Validation of data that passes through the beAWARE system is important to avoid misdirection and waste of resources. For that reason a two-layer validation component was deployed, which checks the incoming data (e.g. images, videos, tweets) to determine if they are erroneous/malicious and discard them.

The second layer of validation (VAL) works concurrently to the regular pipeline of the system and validates the incoming data in the background. The validation is done by crosschecking the incoming data with metrics from the CRCL. For the purpose of retrieving the metrics from the CRCL the topics *TOP802_WEATHER_REQUEST* and *TOP803_WEATHER_REPORT* were defined. The first topic carries the requests for CRCL metrics for specific time and place, while the second is the resulted metrics. Finally, when the VAL has all the necessary information, the validation is performed and the result is sent to the KBS (via topic *TOP801_INCIDENT_VALIDATION*). If the validation results are positive, (i.e. the data are incorrect) then the PSAP is alerted via topic *TOP101_INCIDENT_REPORT*.

In order to enable analysts or decision makers to explore or manually analyze the data further, two user interfaces have been implemented into the Knowledge Base. These are the Analysis Workbench and the Incident Map and Table. The first one shall help to find specific information that is contained in the data, collected or generated during a crisis. The second one gives a good overview over all incidences and helps in prioritizing them. Since this is the final deliverable of WP4 there are no next steps that have to be conducted in its scope. However, in the frame of this work package the AAWA team has annotated so far more than 30,000 Italian tweets as relevant or irrelevant to the use case of floods. This rich annotation could really serve as a training data set for classification techniques and as a means to evaluate a methodology. Due to its significance to the scientific community, we will consider its publication, while following all the regulations that Twitter has defined for sharing with third parties, such as providing the IDs of the posts instead of the tweets themselves in JSON.

References

- Ball, G. H., & Hall, D. J., 1965. ISODATA, a novel method of data analysis and pattern classification. *Stanford Research Institute* (NTIS No. AD 699616).
- Boididou, C., Papadopoulos, S., Apostolidis, L. and Kompatsiaris, Y., 2017, June. Learning to detect misleading content on twitter. In *Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval* (pp. 278-286). ACM.
- Boididou, C., Papadopoulos, S., Zampoglou, M., Apostolidis, L., Papadopoulou, O. and Kompatsiaris, Y., 2018. Detection and visualization of misleading content on Twitter. *International Journal of Multimedia Information Retrieval*, 7(1), pp.71-86.
- Calinski, T., & Harabasz, J., 1974. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1), 1-27.
- Davies, D. L., & Bouldin, D. W., 1979. A cluster separation measure. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, (2), 224-227.
- Duda, R. O., & Hart, P. E., 1973. *Pattern classification and scene analysis* (Vol.3). New York: Wiley.
- Dunn, J. C., 1974. Well-separated clusters and optimal fuzzy partitions. *Journal of cybernetics*, 4(1), 95-104.
- Ester, M., Kriegel, H.P., Sander, J. and Xu, X., 1996, August. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd* (Vol. 96, No. 34, pp. 226-231).
- Evans, J.D., 1996. *Straightforward statistics for the behavioral sciences*. Thomson Brooks/Cole Publishing Co.
- Halkidi, M., Vazirgiannis, M., & Batistakis, Y., 2000. Quality scheme assessment in the clustering process. In *Principles of Data Mining and Knowledge Discovery* (pp. 265-276). Springer Berlin Heidelberg.
- Hubert, L. J., & Levin, J. R., 1976. A general statistical framework for assessing categorical clustering in free recall. *Psychological bulletin*, 83(6), 1072.
- Heer, J., & Shneiderman, B. (2012). Interactive dynamics for visual analysis. *Communications of the ACM*.
- Kaufman, L., & Rousseeuw, P. J., 1990. Finding groups in data. An introduction to cluster analysis. *Wiley Series in Probability and Mathematical Statistics*. Applied Probability and Statistics, New York: Wiley, 1990, 1.
- Krzanowski, W. J., & Lai, Y. T., 1988. A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics*, 23-34.
- McClain, J. O., & Rao, V. R., 1975. Clustisz: A program to test for the quality of clustering of a set of objects. *Journal of Marketing Research* (pre-1986), 12(000004), 456.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. CoRR.
- Rajaraman, A., & Ullman, J. (2011). In *Data Mining: Mining of Massive Datasets* (S. 1–17).

Ratkowsky, D. A., & Lance, G. N., 1978. A criterion for determining the number of groups in a classification. *Australian Computer Journal*, 10(3), 115-117.

Appendix A – Ontology Additions

This appendix lists the ontology classes, object properties and data properties that have been added to the ones listed in deliverable D4.2.

Classes

Action

Name	Action
Definition	Represents an actual action that has to be executed or has been executed in order to protect life and property. Implements a potential action defined in a standard operation procedure or in ad-hoc procedures.
Instance of	owl:Class

Action Type

Name	ActionType
Definition	Types of potential actions that could be helpful to mitigate risk and prevent damage or repair assets.
Instance of	owl:Class

Alert

Name	Alert
Definition	An alert is a message sent out to the public to warn addressees of specific and urgent hazards (e.g. flood, heat wave ...).
Instance of	owl:Class

Asset Representation

Name	AssetRepresentation
Definition	An Asset Representation can be regarded as an entity that enables access to the available data about an asset, for example, temperatures in a building, pictures and measurements of the building obtained in a measurement campaign or the results from a structural model. For online sensors producing data points, it is equivalent to a Thing as defined in the SensorThings API specification: a thing is an object of the physical world (physical things) or the information world (virtual things) that is capable of being identified and integrated into communication networks.
Instance of	owl:Class

Climate Condition

Name	ClimateCondition
Definition	A climate condition that increases the likelihood that a certain natural disaster occurs.
Instance of	owl:Class

Crisis Classification

Name	CrisisClassification
Definition	The Crisis Classification gives a quantitative rating of a predicted or recent crisis.
Instance of	owl:Class

Document

Name	Document
Definition	A generic Document class to store document in the KB, e.g. an Excel/Word/Image file with results from a measurement.
Instance of	owl:Class

Domain

Name	Domain
Definition	A knowledge domain that is relevant to beAWARE.
Instance of	owl:Class

Feature of Interest

Name	FeatureOfInterest
Definition	An Observation's FeatureOfInterest can be the Location of the SensorThing. For example, the FeatureOfInterest of a wifi-connected temperature sensor can be the Location of the sensor (i.e., the room where the sensor is located in). In the case of remote sensing, the FeatureOfInterest can be the geographical area or volume that is being sensed.
Instance of	owl:Class

Guideline

Name	Guideline
Definition	A guideline contains information intended to advise people on what and/or how something should be done.
Instance of	owl:Class

Input Parameter

Name	InputParameter
Definition	Input parameter accepted/required by a tasking capability. Records the description of the parameter and the definition of the data type and unit of measurement.
Instance of	owl:Class

Issue

Name	Issue
Definition	An issue represents a problem that needs to be addressed, for instance: the risk level associated to a property item has trespassed a certain threshold or

	damage has been observed.
Instance of	owl:Class

Measurement

Name	Measurement
Definition	A measurement refers to the act or process of measuring in the context of a measurement campaign.
Instance of	owl:Class

Parameter

Name	Parameter
Definition	A parameter refers to an observed property of an asset or its surroundings, e.g. temperature or humidity.
Instance of	owl:Class

Parameter Value

Name	ParameterValue
Definition	An acceptable value for the parameter and a human-readable description of the value.
Instance of	owl:Class

Place of Relief

Name	PlaceOfRelief
Definition	Places of relief offer climate controlled places for civilians to recover during a heatwave.
Instance of	owl:Class

Report

Name	Report
Definition	Report generated by users using the mobile app.
Instance of	owl:Class

Response

Name	Response
Definition	A response aggregates the actions that have been / are being / shall be carried out to deal with an issue.
Instance of	owl:Class

Risk

Name	Risk
Definition	The probability of harmful consequences, or expected losses, deaths, injuries, property, livelihoods, economic activity disrupted or environment damaged) resulting from interactions between (natural, human induced or

	man-made) hazards and vulnerable conditions in a given area and time period (UNISDR (2009)).
Instance of	owl:Class

Role

Name	Role
Definition	The position or purpose that stakeholders have.
Instance of	owl:Class

Stakeholder

Name	Stakeholder
Definition	A person or an organization that is involved in the disaster and crisis management.
Instance of	owl:Class

Tasking Capability

Name	TaskingCapability
Definition	The tasking capability allows other devices or users to remotely request the execution of tasks that shall be carried out by connected devices and models.
Instance of	owl:Class

Unit of Measure

Name	UnitOfMeasure
Definition	Unit of measure as in the Ontology of units of Measure and related concepts (OM). A unit of measure is a definite magnitude of a quantity, defined and adopted by convention and/or by law, that is used as a standard for measurement of the same quantity. Any other value of the quantity can be expressed as a simple multiple of the unit of measure. For example, length is a quantity. The metre is a unit of length that represents a definite predetermined length. When we say 10 metre (or 10 m), we actually mean 10 times the definite predetermined length called "metre".
Instance of	owl:Class

Forrest Fire Crisis Classification

Name	ForrestFireCrisisClassification
Definition	There are two different emergency levels: one with preventive tasks in which specific emergency has not started (levels 1 to 3), and another one in which the emergency (forest fire in this case) has already started (levels 0 to 3).
Subclass of	CrisisClassification

Heatwave Classification

Name	HeatwaveClassification
Definition	Crisis classification for a heatwave.

Subclass of	CrisisClassification
--------------------	----------------------

Hydraulic Crisis Classification

Name	HydraulicCrisisClassification
Definition	Crisis classification for a flood.
Subclass of	CrisisClassification

Standard Operating Procedure

Name	StandardOperatingProcedure
Definition	A standard operating procedure, or SOP, is a set of step-by-step instructions compiled by an organization to help workers carry out routine operations. SOPs aim to achieve efficiency, quality output and uniformity of performance, while reducing miscommunication and failure to comply with industry regulations.
Subclass of	Guideline

HTML Item

Name	HTMLItem
Definition	A media file in form of a HTML document.
Subclass of	MediaItem

Decision Maker

Name	DecisionMaker
Definition	A user of the system that con decides which actions should be carried out. Decision makers may require authorization to execute the proposed actions.
Subclass of	Role

Domain Expert

Name	DomainExpert
Definition	Domain experts, such as university researchers and professional figures, have the know-how to analyze the more specialized technical/scientific information provided by the system in order to provide information for the support to the decision makers.
Subclass of	Role

Organization

Name	Organization
Definition	A stakeholder such as an institution, company or department.
Subclass of	Stakeholder

Person

Name	Person
Definition	A person stakeholder that usually will be part of an organization.

Subclass of	Stakeholder
--------------------	-------------

Civil Society

Name	CivilSociety
Definition	A person stakeholder that usually will be part of an organization.
Subclass of	Organization

Private Sector

Name	PrivateSector
Definition	The private sector is the part of the economy which is run by private individuals or groups, usually as a means of enterprise for profit.
Subclass of	Organization

Public Sector

Name	PublicSector
Definition	The public sector is the part of the economy concerned with providing various governmental services.
Subclass of	Organization

Municipality

Name	Municipality
Definition	A municipality.
Subclass of	PublicSector

Park

Name	Park
Definition	A park.
Subclass of	Structure

Cellar

Name	Cellar
Definition	Refers to the cellar/basement of a building, which is below ground level.
Subclass of	Building

Forest

Name	Forest
Definition	A forest.
Subclass of	EcologicalAsset

Bycicle

Name	Bicycle
Definition	A bicycle.

Subclass of	Vehicle
--------------------	---------

Boat

Name	Boat
Definition	A boat.
Subclass of	Vehicle

Bus

Name	Bus
Definition	A bus.
Subclass of	Vehicle

Truck

Name	Truck
Definition	A truck.
Subclass of	Vehicle

Motorcycle

Name	Motorcycle
Definition	A motorcycle.
Subclass of	Vehicle

Cat

Name	Cat
Definition	A cat.
Subclass of	Animal

Dog

Name	Dog
Definition	A dog.
Subclass of	Animal

Dummy

Name	Dummy
Definition	This class represents a human-like dummy that will be used in the Vicenza pilot (drone scenario)
Subclass of	Human

Wheelchair User

Name	WheelchairUser
Definition	A person who uses a wheelchair.
Subclass of	Human

Object Detection

Name	ObjectDetection
Definition	The object detection that is used for images and videos.
Subclass of	DataAnalysis

Voice Analysis

Name	VoiceAnalysis
Definition	The analysis that is used on calls to the call center.
Subclass of	DataAnalysis

Object Properties**addressed through**

Name	addressedThrough
Definition	The issue that will be addressed through a response.
Instance of	owl:ObjectProperty
Domain	Issue
Range	Response
Inverse of	Targets

consists of

Name	consistsOf
Definition	The action that a response consists of.
Instance of	owl:ObjectProperty
Domain	Response
Range	Action
Inverse of	isPartOf

estimated in

Name	estimatedIn
Definition	The risk that is estimated by the crisis classification.
Instance of	owl:ObjectProperty
Domain	Risk
Range	CrisisClassification
Inverse of	estimates

guideline includes

Name	guidelineIncludes
Definition	The action types that are included in a guideline.
Instance of	owl:ObjectProperty
Domain	Guideline
Range	ActionType
Inverse of	isIncludedInGuideline

has dataset

Name	hasDataSet
Definition	A data set that is connected to an asset.
Instance of	owl:ObjectProperty
Domain	AssetRepresentation
Range	Dataset
Inverse of	belongsToRepresentation

has document

Name	hasDocument
Definition	A document that describes a feature of interest.
Instance of	owl:ObjectProperty
Domain	FeatureOfInterest
Range	Document
Inverse of	refersToFeatureOfInterest

has measurements of

Name	hasMeasurementsOf
Definition	A data set that contains measured data of a parameter.
Instance of	owl:ObjectProperty
Domain	DataSet
Range	Parameter
Inverse of	measurementsInDataset

has parameter

Name	hasParameter
Definition	A parameter that is used to describe a climate condition.
Instance of	owl:ObjectProperty
Domain	ClimateCondition
Range	Parameter
Inverse of	isParameterForCondition

has role

Name	hasRole
Definition	The role of a stakeholder.
Instance of	owl:ObjectProperty
Domain	Stakeholder
Range	Role
Inverse of	playedBy

has tasking capability

Name	hasTaskingCapability
Definition	The tasking capabilities that an asset has.
Instance of	owl:ObjectProperty

Domain	AssetRepresentation
Range	TaskingCapability
Inverse of	isOfferedByAssetRepresentation

has unit

Name	hasUnit
Definition	The unit in which a parameter is measured.
Instance of	owl:ObjectProperty
Domain	Parameter
Range	UnitOfMeasure
Inverse of	isUnitOf

is at risk

Name	isAtRisk
Definition	The risk that a vulnerable object is exposed to.
Instance of	owl:ObjectProperty
Domain	VulnerableObject
Range	Risk
Inverse of	isRiskOf

is based on

Name	isBasedOn
Definition	The guideline that a response is based on.
Instance of	owl:ObjectProperty
Domain	Response
Range	Guideline
Inverse of	isFollowedBy

is condition for

Name	isConditionFor
Definition	The natural disaster that depends on a certain climate condition to occur.
Instance of	owl:ObjectProperty
Domain	ClimateCondition
Range	NaturalDisaster
Inverse of	hasClimateCondition

is executed by

Name	isExecutedBy
Definition	The stakeholder that executes an action.
Instance of	owl:ObjectProperty
Domain	Action
Range	Stakeholder
Inverse of	executes

is expert in

Name	isExpertIn
Definition	The domain a stakeholder is an expert in.
Instance of	owl:ObjectProperty
Domain	Stakeholder
Range	Domain
Inverse of	hasDomainExpert

is implemented by

Name	isImplementedBy
Definition	The action that implements an action type.
Instance of	owl:ObjectProperty
Domain	ActionType
Range	Action
Inverse of	implements

is input to task

Name	isInputToTask
Definition	The task that includes a parameter value.
Instance of	owl:ObjectProperty
Domain	ParameterValue
Range	Task
Inverse of	hasParameterValue

is located at

Name	isLocatedAt
Definition	The location of an asset.
Instance of	owl:ObjectProperty
Domain	AssetRepresentation
Range	Location
Inverse of	isLocationOfAssetRepresentation

is measured with

Name	isMeasuredWith
Definition	The sensor that is used for a measurement.
Instance of	owl:ObjectProperty
Domain	Measurement
Range	Sensor
Inverse of	usedToMeasure

is measurement of

Name	isMeasurementOf
Definition	The Parameter that is measured in a measurement.
Instance of	owl:ObjectProperty

Domain	Measurement
Range	Parameter
Inverse of	measuredIn

is mitigated by action

Name	isMitigatedByAction
Definition	An action that mitigates the impact.
Instance of	owl:ObjectProperty
Domain	Impact
Range	Action
Inverse of	mitigatesImpact

is prerequisite of

Name	isPrerequisiteOf
Definition	An asset representation that depends on another asset representation.
Instance of	owl:ObjectProperty
Domain	AssetRepresentation
Range	AssetRepresentation
Inverse of	dependsOn

is quantified in

Name	isQuantifiedIn
Definition	The parameter value that quantifies a parameter.
Instance of	owl:ObjectProperty
Domain	Parameter
Range	ParameterValue
Inverse of	quantifies

is value for

Name	isValueFor
Definition	An input parameter of a parameter value.
Instance of	owl:ObjectProperty
Domain	ParameterValue
Range	InputParameter
Inverse of	assignedToInputParameter

mitigates impact type

Name	mitigatesImpactType
Definition	An impact type that is mitigated by the action type.
Instance of	owl:ObjectProperty
Domain	ActionType
Range	ImpactType
Inverse of	isMitigatedByActionType

observes asset

Name	observes
Definition	An asset that is observed in a measurement.
Instance of	owl:ObjectProperty
Domain	Measurement
Range	Asset
Inverse of	isObservedIn

observes parameter

Name	observes2
Definition	A parameter that is observed by the sensor.
Instance of	owl:ObjectProperty
Domain	Sensor
Range	Parameter
Inverse of	isObservedBy

organization includes

Name	organizationIncludes
Definition	An organization that is included in another organization.
Instance of	owl:ObjectProperty
Domain	Organization
Range	Organization
Inverse of	isIncludedInOrganization

performed at

Name	performedAt
Definition	The location of a measurement.
Instance of	owl:ObjectProperty
Domain	Measurement
Range	Location
Inverse of	isLocationOfMeasurement

place of relief is located at

Name	placeOfReliefLocated
Definition	The location of a place of relief.
Instance of	owl:ObjectProperty
Domain	PlaceOfRelief
Range	Location
Inverse of	locatesPlaceOfRelief

produced document

Name	producedDocument
Definition	A document that is produced by the measurement.
Instance of	owl:ObjectProperty

Domain	Measurement
Range	Document
Inverse of	isPartOfDataSet

provides

Name	provides
Definition	An asset representation that is provided by the stakeholder.
Instance of	owl:ObjectProperty
Domain	Stakeholder
Range	AssetRepresentation
Inverse of	isProvidedBy

provides data about

Name	providesDataAbout
Definition	The asset that is represented by the asset representation.
Instance of	owl:ObjectProperty
Domain	AssetRepresentation
Range	Asset
Inverse of	hasRepresentation

reports Incident Type

Name	reportsIncidentType
Definition	The incident type that is reported by a report.
Instance of	owl:ObjectProperty
Domain	Report
Range	IncidentType
Inverse of	isReportedBy

used by tasking capability

Name	usedByTaskingCapability
Definition	A tasking capability that uses the input parameter.
Instance of	owl:ObjectProperty
Domain	InputParameter
Range	TaskingCapability
Inverse of	hasInputParameter

works for organization

Name	worksForOrganization
Definition	An organization that a person works for.
Instance of	owl:ObjectProperty
Domain	Person
Range	Organization
Inverse of	engagesPerson

Data Properties

capacity

Name	capacity
Definition	The maximal capacity of persons for a place of relief.
Instance of	owl:DatatypeProperty
Domain	PlaceOfRelief
Range	xsd:int

capacity used

Name	capacityUsed
Definition	The current capacity of persons for a place of relief.
Instance of	owl:DatatypeProperty
Domain	PlaceOfRelief
Range	xsd:int

creation time

Name	creationTime
Definition	The date time to which a media item have been created.
Instance of	owl:DatatypeProperty
Domain	MediaItem
Range	xsd:dateTime

expiresUTC

Name	expiresUTC
Definition	The date time to which an alert expires.
Instance of	owl:DatatypeProperty
Domain	Alert
Range	xsd:dateTime

flood crisis level

Name	floodCrisisLevel
Definition	The current flood crisis level calculated in CRCL.
Instance of	owl:DatatypeProperty
Domain	HydraulicCrisisClassification
Range	xsd:integer

forTeam

Name	forTeam
Definition	The first responder team name that a task is assigned to.
Instance of	owl:DatatypeProperty
Domain	Task
Range	xsd:string

Grid connection

Name	GridConnection
Definition	The closests public transportation stops of a place of relief.
Instance of	owl:DatatypeProperty
Domain	PlaceOfRelief
Range	xsd:string

has end

Name	hasEnd
Definition	The date time to which an action has been completed.
Instance of	owl:DatatypeProperty
Domain	Action
Range	xsd:dateTime

has originator

Name	hasOriginator
Definition	The name of the beAWARE component that created the incident report.
Instance of	owl:DatatypeProperty
Domain	IncidentReport
Range	xsd:string

has priority

Name	hasPriority
Definition	The priority assigned to a media item. Can be high, medium or low.
Instance of	owl:DatatypeProperty
Domain	MediaItem
Range	xsd:string

has reference code

Name	hasReferenceCode
Definition	The reference code of a thing from the SensorThings API.
Instance of	owl:DatatypeProperty
Domain	Thing
Range	xsd:string

has severity from Crisis Classification

Name	hasSeverityFromCRCL
Definition	The severity of an incident report calculated by CRCL. Can be extreme, high, medium or low.
Instance of	owl:DatatypeProperty
Domain	IncidentReport
Range	xsd:string

has start

Name	hasStart
Definition	The date time to which an action shall start.
Instance of	owl:DatatypeProperty
Domain	Action
Range	xsd:dateTime

has timestamp

Name	hasTimestamp
Definition	The date time to which a climate condition was last fulfilled.
Instance of	owl:DatatypeProperty
Domain	ClimateCondition
Range	xsd:dateTime

instructions

Name	instructions
Definition	The instructions that are provided in an alert.
Instance of	owl:DatatypeProperty
Domain	Alert
Range	xsd:string

is spam

Name	isSpam
Definition	Boolean that is true in case that a tweet was identified as spam by the social media analysis.
Instance of	owl:DatatypeProperty
Domain	IncidentReport
Range	xsd:boolean

language

Name	language
Definition	The language in which a report is written.
Instance of	owl:DatatypeProperty
Domain	Report
Range	xsd:string

level

Name	Level
Definition	The risk level.
Instance of	owl:DatatypeProperty
Domain	Risk
Range	xsd:integer

radius

Name	radius3
Definition	The radius around a coordinate in which an alert is send to the beAWARE mobile app users.
Instance of	owl:DatatypeProperty
Domain	Alert
Range	xsd:double

reason for crisis level

Name	reasonForCrisisLevel
Definition	The reason for the current flood crisis level.
Instance of	owl:DatatypeProperty
Domain	HydraulicCrisisClassification
Range	xsd:string

reportID

Name	reportID
Definition	The id of a report.
Instance of	owl:DatatypeProperty
Domain	Report
Range	xsd:string

SanitationAvailable

Name	SanitationAvailable
Definition	True if a place of relief includes sanitation.
Instance of	owl:DatatypeProperty
Domain	PlaceOfRelief
Range	xsd:Boolean

severity

Name	severity
Definition	The severity of an alert.
Instance of	owl:DatatypeProperty
Domain	Alert
Range	xsd:string

startTimeUTC

Name	startTimeUTC
Definition	The date time to which an alert is sent.
Instance of	owl:DatatypeProperty
Domain	Alert
Range	xsd:dateTime

status

Name	Status
Definition	The current status of an action.
Instance of	owl:DatatypeProperty
Domain	ActionType
Range	xsd:string