beAWARE

Enhancing decision support and management services in extreme weather climate events

700475

D3.2

Empirical study and annotation of multilingual and multimedia material in beAWARE

Dissemination level:	Public
Contractual date of delivery:	31 August 2017
Actual date of delivery:	10 September 2017
Work package:	WP3 Early warning generation
Task:	T3.2, T3.3
Type:	Report
Approval Status:	Final Draft
Version:	0.4
Number of pages:	31

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

Filename:	D3.2_beAWARE_empirical_study_of_multilingual_and_multimedia_material_2017-09-08_v0.4.docx

Abstract

This deliverable reports the results of the empirical study of the multimedia inputs, namely audio, text, audio and visual data, pertinent to the beAWARE pilots, in accordance with the use cases and initial requirements as described in D2.1. As described, the data collection includes: i) historical material made available by the user partners, ii) datasets compiled within beAWARE, and iii) publicly available datasets that are relevant to the beAWARE domains. The outcomes of the study delineate the initial set of specifications and requirements with respect to the targeted information of interest for the content analysis and distillation tasks addressed in T3.2 and T3.3 respectively.

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	Reason	Revised by
0.1	14.07.2017	Deliverable structure, abstract	S. Dasiopoulou
0.2	29.07.2017	Contributions to sections 2, 3 & 4	D. Liparas, K. Avgerinakis, E. Michail, J. Grivolla
0.3	29.08.2017	Updates to sections 3 & 4	K. Avgerinakis, E. Michail, J. Grivolla
0.3.1	08.09.2017	Revision, completion and finalization of sections 2, 3 & 4; executive summary, introduction & conclusion sections	S. Dasiopoulou
0.4	09.09.2017	Internal review	Anastasios Karakostas (CERTH)

Author list

Organisation	Name	Contact Information
UPF	Stamatia Dasiopoulou	stamatia.dasiopoulou@upf.edu
UPF	Jens Grivolla	jens.grivolla@upf.edu
UPF	Leo Wanner	leo.wanner@upf.edu
CERTH	Dimitrios Liparas	dliparas@iti.gr
CERTH	Konstantinos Avgerinakis	koafgeri@iti.gr
CERTH	Emmanouil Michail	michem@iti.gr

Executive Summary

This deliverable reports on the audio, text and visual datasets that have been collected and studied during the period M1-M8 as part of the investigations carried out within T3.2 and T3.3 towards the development of respective content analysis techniques that will enable the distillation of the pertinent semantic information. More specifically, D3.2 explicates the multitude of audio, text and visual information inputs considered within beAWARE, based on the initial requirements and use case scenario descriptions laid out in D2.1, describes the currently acquired datasets, and reports on the findings of their empirical study towards the elicitation of preliminary specifications and requirements for the respective analysis tasks.

Abbreviations and Acronyms

ASR	Automatic Speech Recognition
CCTV	Closed Circuit TeleVision
OWL	Web Ontology Language
RDF	Resource Description Framework
UAV	Unmanned Aerial Vehicles

Table of Contents

1 Introduction

beAWARE advocates an integrated solution to support and enhance crisis management encompassing forecasting, early warning generation, transmission and routing of emergency related information, aggregated analysis of multimodal data, and management of the coordination between the first responders and the authorities.

To achieve this, it capitalizes on a multitude of inputs, including pertinent meteorological and forecasting data (e.g. heat index, forecasted values and duration for wind speed and temperature, emergency prediction models such as the AMICO flood warning system, etc.), sensor readings that provide real-time opportunistic sensing information (e.g. water level sensors, temperature sensors, etc.), cameras affording continuous streams of the visual context for the reference physical vicinity, social media, as well as direct communications between civilians and authorities (e.g. calls to emergency numbers, incidence reports sent via the beAWARE mobile application to the PSAP), and between first responders and authorities (e.g. radio communication, assigned task status updates sent via the beAWARE mobile application). The information extracted through the analysis of the heterogeneous input sources feeds the aggregation and semantic integration techniques for decision support and early warnings generation that are developed within WP4.

In this deliverable, we elaborate on the multimedia content inputs, namely audio, text and visual, and explicate the datasets that have been currently collected as well as the outcomes of their empirical study that lay the groundwork for the semantic analysis tasks T3.2 and T3.3. As described in the following, the collection of the datasets serves a twofold purpose: i) to afford reference material for understanding the characteristics and idiosyncrasies of the material under analysis, thus allowing for adequate technical solutions, and assess the gamut of information of interest to be targeted for extraction; ii) to make available data for training, when and as relevant to the investigated content analysis techniques, as well as for evaluation and benchmarking.

To ensure a comprehensive and adequately insightful approach, a variety of datasets have been collected for consideration, including historical data made available by the consortium, datasets compiled within beAWARE, as well as publicly available benchmark datasets. Naturally, the collection of datasets is a continuous process, driven by the specifications and requirements of the use case scenarios, as further worked out during the course of the project, as well as of the investigations into corresponding analysis tools. Thus, this deliverable only reflects the current state of affairs; further updates, as needed, will be reported in D3.3 (M17) and D3.4 (M34) that will present the basic and advanced techniques for content distillation from multimedia material. Likewise, the observations of the empirical study serve as preliminary guidelines for the development of the T3.2 and T3.3 content analysis techniques, also expected to be revised and further refined, as the corresponding research and development activities progress.

The remaining of the document is structured as follows. Section 2 outlines the use of multimedia data within the targeted use case scenarios, as described in D2.1. Section 3 describes the currently collected datasets, which encompass historical data made available by consortium partners, datasets that have been compiled within, and relevant publicly available benchmark datasets. Following the datasets descriptions, we report the outcomes of the carried out empirical study are in Section 4. Last, Section 5 summarizes the data currently drawn observations with respect to the collection of relevant data and the preliminary observations resulting from their analysis.

2 Multimedia data inputs in beAWARE

beAWARE aims to enhance situational awareness and support decision making during crisis management by capitalizing on the integration and aggregated analysis of meteorological, forecasting and sensor data, as well as of a multitude of pertinent multimedia data inputs. In the context set by the targeted pilots and respective scenario descriptions, as laid out in D2.1, these multimedia inputs pertain to the following information sources: i) incidence reports by civilians via calls to emergency numbers and authorities' call centers, as well as via the beAWARE mobile application, which will enable the sharing of textual messages and also of captured images/videos; ii) communications between first responders and authorities via internal communication channels (e.g. radio communication) as well as via the beAWARE mobile application; iii) information posted on social media, namely textual messages as well as accompanying images and videos; iv) images and videos captured via static and mobile cameras. These act complementary allowing not only for a more complete apprehension and assessment of the reference situational context, but also for cross-checking and validating of the accuracy and trustworthiness of the extracted information, crucial prerequisites in critical applications such as emergency management.

In the following, we briefly review the overall context of usage pertinent to the three types of considered multimedia data, namely audio, text and visual, before continuing with the description of the currently compiled datasets and the outcomes of the empirical analysis, in Sections 3 and 4 respectively.

2.1 Audio data inputs

As noted above, the audio inputs considered within beAWARE consist in incidence reports by civilians to emergency numbers and call centers, which may involve human or non-human receivers, as well as, voice-based communications between authorities and first responders through currently deployed internal communication channels (e.g. radio). In accordance with the planned pilots, the considered languages are Greek (heat wave pilot), Italian (flood pilot), and Spanish (fire forest pilot), as well as English, which will be used for control and demonstration purposes. To distill the conveyed information, Automatic Speech Recognition (ASR) techniques will be first deployed in order to denoise, enhance and eventually transcribe the captured audio streams into text; the semantics of the latter will subsequently be extracted by means of text analysis and fed to the semantic integration and interpretation tools of WP4 for supporting decision making and warning generation. As the conditions pertaining to emergency communications aggravate considerably acoustic and environmental robustness challenges associated with the recognition of spontaneous speech, including high noise levels, fragmented, emotionally charged utterances, hesitations, and so forth, ASR investigations will focus, in addition to ensuring adequate vocabulary coverage, on effective compensation strategies.

2.2 Textual data inputs

The considered textual inputs include: i) posts on social media, in particular Twitter, crawled via the beAWARE social media monitoring service (T4.1), ii) messages fed to the system by civilians and first responders via the beAWARE mobile application, as well as iii) transcriptions of the audio inputs described above. Leaving aside processing variations pertinent to the textual inputs provenance (e.g. twitter messages need to undergo normalization to account for the use of abbreviations and idiosyncratic expressions, emoji, etc.), the distillation of the conveyed information consists in the extraction of the mentioned entities and the relations between so as to capture the communicated events/situations (e.g. a car being carried away in the flooded river, traffic interrupted, etc.). These will be represented in a structured, formal manner, namely RDF/OWL, that abstracting away from language variations facilitates their semantic integration and aggregated interpretation, tasks addressed in WP4. The key challenge in the analysis and understanding of the considered textual inputs lies in the idiosyncratic nature of all three types of textual inputs, compared to proper written language. Naturally, as for audio inputs, the considered languages are Greek, Italian, Spanish and English.

2.3 Visual data inputs

The visual inputs utilized within beAWARE include images and videos obtained from static and aerial (e.g. UAV, satellite) cameras deployed on the pilot sites, crawled Twitter messages, as well as reports sent to the PSAP by first responders and civilians using the beAWARE mobile application. The captured visual material will contribute to the detection of emergencies and the enhancement of the real-time contextual understanding, through the recognition of emergency indicative situations, such as smoke, traffic bottlenecks, flooded areas, elements at possible risk (e.g. people, cars, and buildings), etc. Key challenges involve the need for real-time processing of high-discriminating quality in order to extract information that effectively complements and enriches that of the audio and text modalities. This challenge will be tackled by deploying a powerful server on the cloud, which will be able to run online state-of-the-art deep learning algorithms and extract from the visual content highly accurate conclusions in a reasonable near real time computational cost. Additionally, embedded systems, such as the Jetson TX2 module and Raspberry Pi 3 model B, could also be used in some special cases, where the computer vision algorithms do not require a large amount of memory and energy resources, deductions could be made on-site, where the crisis event occurs, and send the results to the server instead of the media.

3 Multimedia data collection

In the following we describe the current datasets and the underpinning acquisition process, which, as aforementioned followed a three-pronged approach, involving: i) historical data already in the disposal of the beAWARE user partners, ii) data compiled within beAWARE, and iii) publicly available datasets for benchmark and development purposes. The compiled datasets form the basis for the empirical study towards the elicitation of preliminary requirements and specifications for the targeted multimedia content analysis techniques (T3.2, T3.3), and also provide reference, train and evaluation, material for supporting their development. Although the datasets that were compiled, and will continue being compiled, within beAWARE are clearly the ones of higher interest, the historical and publicly available ones are of equally important, especially for kick starting insights and investigations, since, as described in the following, the opportunity to collect actual relevant data varies considerably depending on the modality considered.

3.1 Historical datasets provided by user partners

The consortium's user partners possess and started making available, a considerable volume of historical text, image and video data, from the early onset of the project, namely starting in M2 in response to the first discussions during the kick-off meeting, in order to assist the technical partners with the understanding and elicitation of specifications pertinent to the analysis of such data. Table 1 summarizes the datasets, their media types and the user partner that has provided them.

Table 1. Audio, text and visual historical data provided by beAWARE user partners

Partner	Data Description	Media Type
AAWA	~1G of images (UAV and satellite ones) and videos, collected from YouTube and other public web sources, that depict the flood that occurred in the region of Vicenza in November of 2010	Image, Video
AAWA	Reports sent through the mobile app, collected during trials organized within the FP7 WeSenseIt project (2012-2016)	Text, Image
FBBR	~3MB of fire images from their private recordings Pictures of fire events	Image
HRT	~1.67G of image and videos that were recorded by first responders in real fire incidents in Greece during the last decade	Image, Video
PLV	~25MB of flood, fire and heat wave related images and videos, recorded by first responders and citizens in the Valencia region (2014-2016)	Image, Video
PLV	Tweets related to flood, fire and heat wave related events in the Valencia region, posted by Valencia police (2014-2016)	Text

As can be seen, the historical data include in their majority textual and visual material, and though very useful for preliminary investigations, additional material is required to ensure the adequate coverage of the scope of the pilots and specific use cases scenarios targeted by beAWARE (e.g. audio inputs, descriptive reports enabled through the beAWARE envisaged mobile application, etc.). Examples of the provided material are showing in Figure 1 and Figure 2.

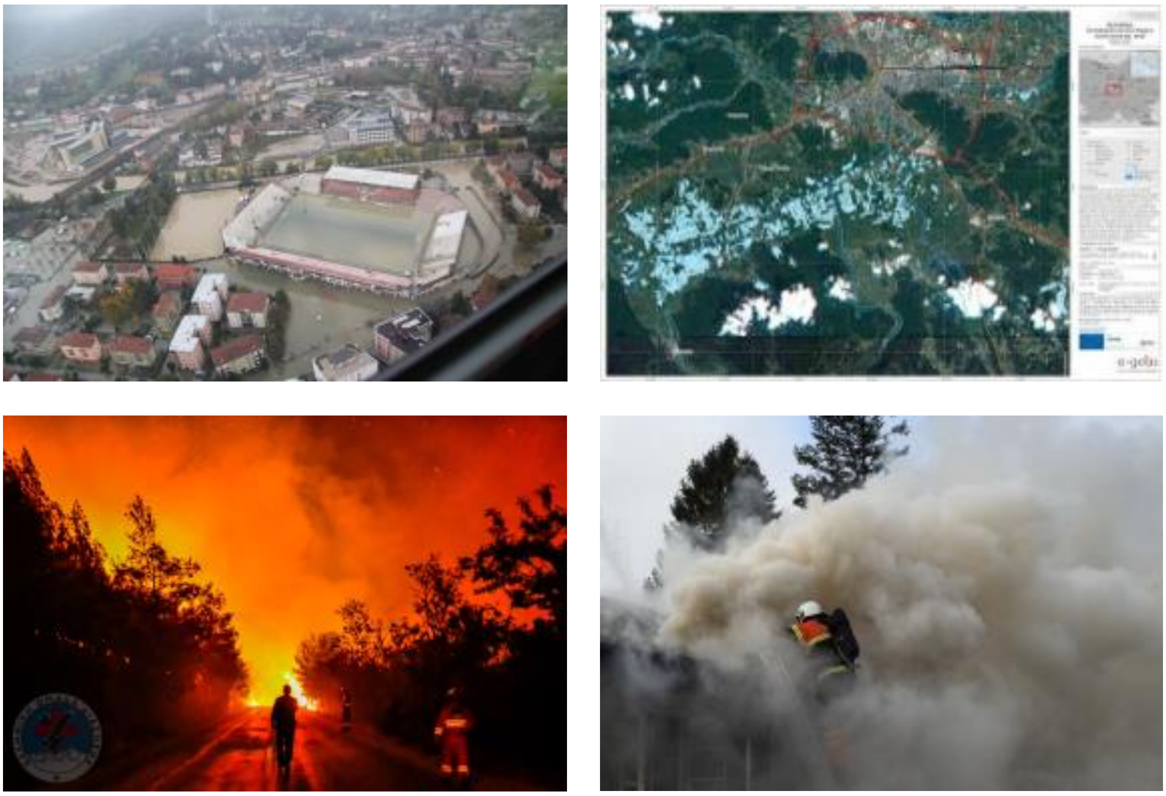
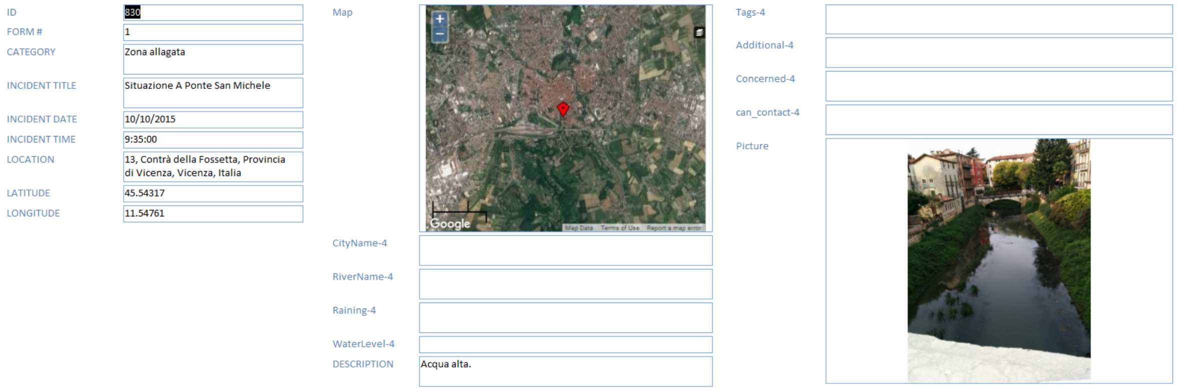


Figure 1 - Example of flood and fire events images from UAV, satellite and mobile cameras.



ID	830
FORM #	1
CATEGORY	Zona allagata
INCIDENT TITLE	Situazione A Ponte San Michele
INCIDENT DATE	10/10/2015
INCIDENT TIME	9:35:00
LOCATION	13, Contrà della Fossetta, Provincia di Vicenza, Vicenza, Italia
LATITUDE	45.54317
LONGITUDE	11.54761
Map	
CityName-4	
RiverName-4	
Raining-4	
WaterLevel-4	
DESCRIPTION	Acqua alta.
Tags-4	
Additional-4	
Concerned-4	
can_contact-4	
Picture	

Figure 2 - Screenshot from example WeSenseIt report informing about high-water level.

3.2 Datasets compiled within beAWARE

The datasets that have been compiled, and that are to be compiled, within beAWARE during its lifetime, fall into two categories. The first refers to datasets that will be collected during the field trials planned for M18-M21 and M30-33 for the three pilots, namely:

- PUC1 – *flood*: to be deployed and implemented by AAWA in Vicenza, Italy;
- PUC2 – *fire*: to be implemented and deployed by PLV in Valencia, Spain;
- PUC3 - *heat wave*: to be implemented and deployed by HRT in Thessaloniki, Greece.

Collected within field trials that simulate the unfolding of emergency events, these datasets not only will provide modality-correlated information, but will also afford temporally- and spatially-associated inputs, thus allowing the validation of the analysis and interpretation outcomes for both the distinct modalities and their aggregation.

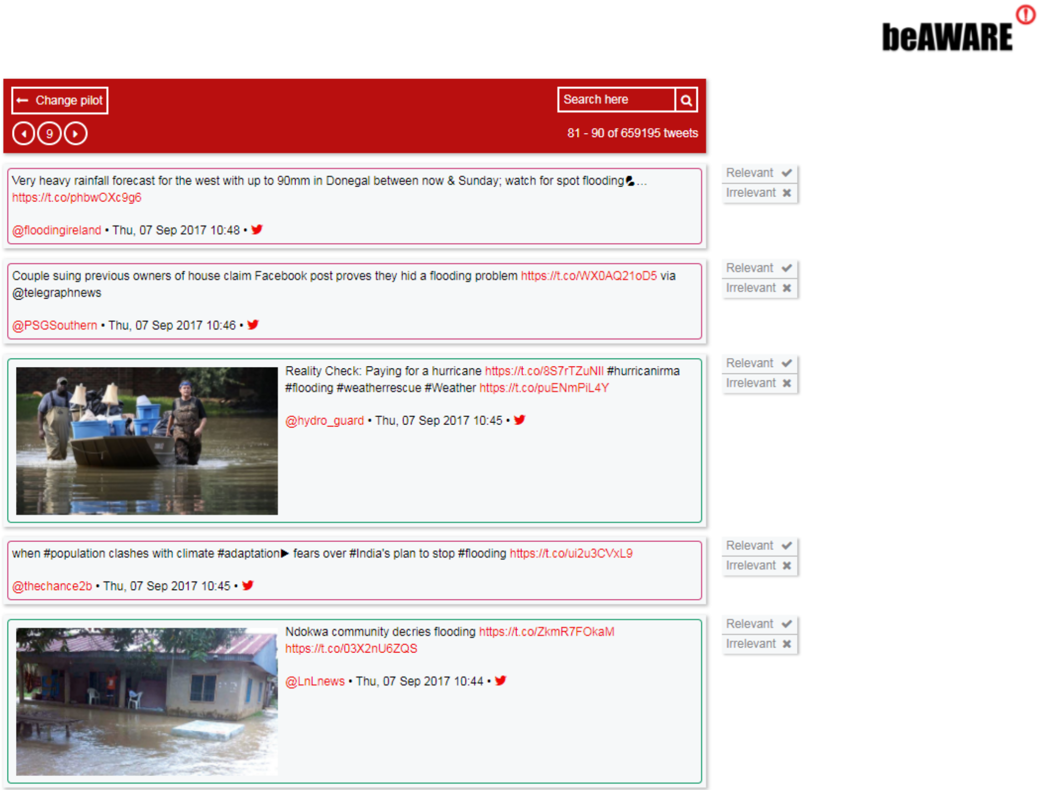
The second category refers to datasets that have been and will be compiled by consortium members, technical and user partner ones, to support the analysis of distinct multimedia content types and/or specific analysis objectives, e.g. smoke detection in aerial images, smoke detection from static camera videos, extraction of information about element at risk from twitter messages, and so forth. The respective, currently underway and planned, activities of user partners are listed in the following.

- **AAWA** is responsible for supporting the collection of audio, text and visual data related to flood emergencies in the area of Vicenza. To this end, they contribute to the definition of criteria for the crawling of flood-related tweets in Italian and their annotation for development and evaluation purposes. Furthermore, they have started the compilation of simulated emergency calls recordings, currently amounting to 20 and including both “clean” and noisy data, based on the simulation trials carried out within the FP7 WeSenseIt project. AAWA will also provide surveillance videos from already available static cameras that monitor regions of interest in the city of Vicenza (http://www.bacchiglione.it/); these will be used to build reference patterns for the recognition of non-flood scenes and for flood prediction.
- **FBBR** is responsible for supporting the collection of visual data related to the forest fire use case scenarios, serving complementary to the data collection by PLV. This involves images and videos from mobile devices and static cameras in the region of Denmark that monitor regions of danger. As Danish is not among the languages targeted for the planned pilots, no efforts for the collection of audio and textual material has been currently scheduled.
- **HRT** is responsible for supporting the collection of audio, text and visual data related to the heat wave use case scenarios. To this end, they have already provided a list of social media account references from involved authorities and contribute actively to the definition of criteria for crawling heat wave related tweets in Greek as well as to their annotation; moreover, within D2.1’s scenarios definition, they have provided a preliminary list of exemplar textual messages that could be sent by first responders and civilians via the mobile application. Furthermore, likewise to the AAWA initiative, they have prepared an initial set of emergency call recordings in Greek, affording both “clean” and noisy data. HRT will also provide access to close circuit surveillance cameras (CCTV)

that monitor highway roads in Greece in order to support the implementation of the traffic congestion related use case scenarios aspects. Last, being operational and involved in fire emergencies management, HRT will also provide videos related to fire events from mobile and static cameras.

- **PLV** is responsible for supporting the collection of audio, text and visual data related to the forest fire use case scenarios. To this end, they contribute to the definition of criteria for the crawling of fire-related tweets in Spanish and their annotation and have started the compilation of simulated emergency calls recordings, also including both “clean” and noisy data. Furthermore, PLV will provide videos from static cameras monitoring the forest of the Devesa region, as well as flood-related videos to which they have access.

Parallel to the aforementioned user partners-centered data collection activities, further datasets are being gathered and compiled by technical partners, within the corresponding content analysis tasks.



Very heavy rainfall forecast for the west with up to 90mm in Donegal between now & Sunday; watch for spot flooding… https://t.co/phbwOXc9g6

@floodingireland • Thu, 07 Sep 2017 10:48 •

Couple suing previous owners of house claim Facebook post proves they hid a flooding problem https://t.co/WX0AQ21oD5 via @telegraphnews

@PSGSouthern • Thu, 07 Sep 2017 10:46 •

Reality Check: Paying for a hurricane https://t.co/8S7rTZuNIl #hurricanirma #flooding #weatherrescue #Weather https://t.co/puENmPiL4Y

@hydro_guard • Thu, 07 Sep 2017 10:45 •

when #population clashes with climate #adaptation▶ fears over #India's plan to stop #flooding https://t.co/ui2u3CVxL9

@thechance2b • Thu, 07 Sep 2017 10:45 •

Ndokwa community decries flooding https://t.co/ZkmR7FOkaM https://t.co/03X2nU6ZQS

@LnLnews • Thu, 07 Sep 2017 10:44 •

Figure 3 - Screenshot of the crawling and annotation tool developed for Twitter messages.

More specifically, and as part of the work on social media monitoring carried out within T4.1, CERTH has developed a tweets crawler and annotation tool that continuously gathers tweets of possible relevance to the domains of the three pilots, while enabling manual relevance annotations; a screenshot for the flood domain is shown in Figure 3. Furthermore, CERTH has been collecting relevant visual material from video-sharing websites, including Dailymotion[1] and YouTube[2].

[1] http://www.dailymotion.com/

3.3 Benchmark datasets

Complementing the historical data collections and the datasets compiled within beAWARE, several public available datasets have been identified. In addition to contributing to the development of respective multimedia analysis solutions by providing training data, these will enable to evaluate and compare the beAWARE approaches not only with respect to the project goals, but also with respect to the current state of the art. The complete list follows.

3.3.1 Audio datasets

Generic **audio** datasets, containing annotated speech recordings, are being collected for three purposes: i) the development of the ASR algorithms and the initial training of the ASR tool, ii) the realization of the empirical study regarding the most important aspects and information for the management of a crisis event and iii) the final adaptation of the language model and the refinement of the ASR dictionary. The collected datasets include general annotated speech recordings as well as general emergency data, as described in the following.

As far as general annotated speech recordings are concerned, the following datasets of conversations and phone calls, in both "clean" and noisy environments have been identified:

- TalkBank (English, Greek, Spanish) **http://talkbank.org/browser/index.php**
- VoxForge (English, Greek, Italian, Spanish) **http://www.voxforge.org/el/Downloads**
- ELRA (English, Italian, Spanish) **http://www.elra.info/en/catalogues/free-resources/free-lrs-set-1/** **http://www.elra.info/en/catalogues/free-resources/free-lrs-set-2/**
- CMU Census Database (English) **http://www.speech.cs.cmu.edu/databases/an4/**
- CMU_SIN database (English) **http://www.festvox.org/cmu_sin/**
- Santa Barbara Corpus of Spoken American English (English) **http://www.linguistics.ucsb.edu/research/santa-barbara-corpus**
    - EMOVO Corpus (Italian) **http://voice.fub.it/activities/corpora/emovo/index.html**
    - LibriSpeech ASR corpus (English) **http://www.openslr.org/12/**
    - Speech/Song Database – RAVDESS (English) **http://smartlaboratory.org/ravdess/**

With respect to general emergency event data, which form the basis for the conducted empirical study and will serve as reference for the initial adaptation of the ASR module, the list of relevant datasets includes speech recordings of real life emergency calls to 911 reporting general incidents, online inspected on YouTube. Examples, representative examples include:

- **https://www.youtube.com/watch?v=q-Tr0u35Tek**
- **https://www.youtube.com/watch?v=qSywoZAO2SE**

[2] https://www.youtube.com/

- https://www.youtube.com/watch?v=P8t_weHgrJw

In addition, we have considered emergency call speech corpora from published studies, including the NineOneOne project's Spanish to English code-switching corpora. A license agreement document is being prepared in order to enable the acquisition and use of real recordings from actual emergency call centers, as well as from published studies of research groups and institutes; examples of the latter include the database of emergency calls used in Baicerek's[3] study as well as Lefter's[4] database of spontaneous emotional speech from an emergency call-centre.

3.3.2 Text datasets

The growing adoption of information posted on social media during disasters as means for assisting in crisis management and recovery has given rise to a plethora of research studies on tools and methodologies, underpinned by respective datasets. The diversity through of the pertinent objectives, scope and domains renders many of no direct interest for the textual analysis pursuits within beAWARE, the main issue being that datasets for languages and/or domains that do not match those of beAWARE cannot be used for addressing the linguistic phenomena (synctactic, idiomatic, etc. constructions) and vocabulary coverage pertinent to the beAWARE languages and pilot domains. This is not the case however for CrisisLex[5], a large-scale repository of crisis-related social media data media that includes collections of crisis data as well as lexicons of crisis terms. The compiled collections span various languages, albeit not of equal breadth, and likewise various domains, ranging from human-induced crisis events (e.g. protests, accidents, attacks) to climate-related and natural disasters (e.g. hurricanes, floods, earthquakes), and are organized in six sub-collections[6] with varying levels of annotation. Of interest to beAWARE are the relevant climate-related collections afforded in English and Italian.

3.3.3 Visual datasets

Based on the initial use case requirements and scenario descriptions explicated in D2.1, four key visual contexts, namely fire, smoke, flood and traffic scenes, have been identified. As such, the search and selection of publicly available datasets have been driven by the need to effectively detect as well as distinguish such contexts from others with visually similar characteristics (e.g. fire from water reflections, smoke from cloud formations, etc.). In the following, we list the identified relevant datasets for each category.

Fire datasets

- The *Mivia fire* dataset (http://mivia.unisa.it/datasets/video-analysis-datasets/fire-detection-dataset/) contains several video samples (31) that depict fire cases that occur in both indoor (i.e. office scenario) and outdoor (i.e. forest fire

[3] http://ieeexplore.ieee.org/document/5943720/citations

[4] http://ii.tudelft.nl/~iulia/papers/ijidss_il.pdf

5 http://crisislex.org/

[6] http://crisislex.org/data-collections.html

scenario) environments. The dataset also contains non-fire situations in order to evaluate the discrimination power of the fire detector.

- The *FIRESENSE fire detection* dataset (**http://signal.ee.bilkent.edu.tr/VisiFire/**), which was compiled within the FIRESENSE project, includes both forest fires as well as human-induced ones, recorded by static cameras. The dataset was broadly used the last decade to evaluate fire detection algorithms. The demo fire clips (**http://signal.ee.bilkent.edu.tr/ VisiFire/Demo /FireClips/**) contain some further video samples, depicting fire situations recording with a static camera.
- The *Rabot 2012* dataset (**http://multimedialab.elis.ugent.be/rabot2012/**) is a rather simple dataset that contains, among others, video samples of fire scenarios in indoor environments.

Smoke datasets

- The *Mivia smoke* dataset (**http://mivia.unisa.it/datasets/video-analysis-datasets/smoke-detection-dataset/**) is composed by 149 videos, each lasting approximately 15 minutes and is widely used for smoke detection evaluation. It is a very challenging dataset, since it contains scenes and elements red houses in a wide valley, mountains at sunset, sun reflections in the camera, and clouds.
- The *FIRESENSE smoke detection* dataset, also compiled within the FIRESENSE project (**http://signal.ee.bilkent.edu.tr/VisiFire/**).
- The Visor smoke detection dataset (**http://www.openvisor.org/video_videosInCategory.asp?idcategory=8**) contains a wide number of smoke video samples and is used to evaluate the discrimination power of smoke detectors. Several motions also exist in the videos, so the separation between smoke and non-smoke regions is determined as a quite challenging task for state-of-the-art smoke detectors.
- The *VisiFire smoke* (**http://signal.ee.bilkent.edu.tr/VisiFire/Demo/Smoke Clips/**) and *forest smoke* (**http://signal.ee.bilkent.edu.tr/VisiFire/Demo/Forest Smoke/**) datasets were recorded by the Bilkent University and are broadly used to evaluate smoke detection algorithms. The video samples are recorder by using static cameras in both forest and office environments.

Flood datasets

- *The Video Water database* (**https://staff.fnwi.uva.nl/p.s.m.mettes/**) consists of 260 high-quality videos that contain water depictions of predominantly 7 subcategories, namely canals, fountains, lakes, oceans, ponds, rivers, and streams, as well as non-water depicting samples that contain objects with similar spatial and temporal characteristics, such as clouds/steam, fire, flags, trees, and vegetation. The dataset is used for modeling water dynamics and can be used to evaluate flood detection algorithms.
- *The Dyntex* database (**http://dyntex.univ-lr.fr/**) is a diverse collection of high-quality dynamic texture videos that consists of more than 650 sequences, which can be used in order to model water, smoke, fire and other texture dynamics. It is broadly used for evaluating water, fire and smoke detection algorithms.
- The *Moving Vistas* dataset (**http://www.umiacs.umd.edu/users/nshroff/DynamicScene.html**), broadly used for dynamic scene understanding, comprises

10 videos for each of the following 13 categories: Avalanche, Iceberg Collapse, Landslide, Volcano eruption, Chaotic traffic, Smooth traffic, Tornado, Forest fire, Waterfall, Boiling water, Fountain, Waves and Whirlpool, thus making it relevant for several beAWARE pertinent detection purposes, such as traffic, water, and fire recognition. Large variations in the background, illumination, scale and view render it a very challenging dataset that can be used for modeling and evaluating dynamic texture algorithms.

Traffic datasets

- The *UA-Detrac* dataset (**http://detrac-db.rit.albany.edu/**) is a challenging real-world multi-object detection and multi-object tracking benchmark. The dataset consists of 10 hours of high-quality videos captured from a static camera at 24 different locations leading to more than 140 thousand video frames consisting of 8250 vehicles. The dataset is used broadly to evaluate object detection and multi-object tracking methods.

4 Empirical Study

In this section, we report the findings of the empirical study carried out with respect to the three types of multimedia data inputs considered in the project, using as reference the datasets described in the previous Section. As noted in the beginning of the document, the observations drawn are preliminary and are expected to be further refined and elaborated; this is not only attributed to the fact that the majority of current datasets are external to beAWARE, but also to the early phase of the tasks pertinent to the development of tools for the harvesting and analysis of such data. In the following, we list the key observations.

4.1 Audio data study

As aforementioned, audio inputs in beAWARE include emergency calls by civilians to authorities, and voice-based communications between first responders and authorities. Given the currently available audio datasets, the empirical study drew mainly upon the collected speech corpora of cross-domain emergency calls recordings so as to afford a comprehensive understanding of the types of information that authorities and first responders seek for decision making during emergencies. In addition, the empirical study took into account the simulated calls compiled by user partners within beAWARE; being currently limited though in extent, as aforementioned they amount to 20 recordings for each of the three emergency domains, they need to be considerably extended before allowing for comprehensive insights into the actual coverage required for the ASR lexicons and the language models expressivity for the purposes of beAWARE. Yet, we could already determine the key information sought after during such calls:

- Type of event, so as to mobilize the appropriate type of personnel.
- Severity and extent, in order to classify the event, the urge for help, as well as the required personnel and equipment.
- Address of the reported incidence; if the person calling is not able to provide the address, a general description of the surrounding area is requested.
- Number of people in danger and their condition, as well as overall number of people involved.
- Identification details and characteristics of the people involved such as name, age, sex, phone numbers
- Information about bystanders that could be of assistance; e.g. if there is any doctor or specialized personnel around.
- In case of an emergency taking place within a building, key information includes the type of the building, the floor where the incident takes place, its spatial configuration, the existence of toxic or inflammable materials etc.

Similar information is sought after during communications among first responders and first responders and authorities, along with more specialized updates, such as the progress of assigned tasks and the current severity of the incident.

4.2 Text data study

The inputs in beAWARE include Twitter posts collected in real-time through the Twitter Streaming API, reports sent by civilians and first responders to the PSAP via the beAWARE mobile applications, as well we transcriptions of the captured audio communications. As Twitter posts comprise the most populous of the three and given the limited historical report-relevant material made available, compared to the already considerable volumes of tweets that have been crawled using beAWARE developed technologies, the empirical study has focused on the examination of the collected Twitter data and their particularities.

At the time of writing, the beAWARE tweet collection amounts a total to 5,797,042 posts. Table 2 shows the distribution across the languages and emergency domains (fire, flood, heat wave) addressed by the pilots.

Table 2 – beAWARE tweet collection distribution

Language/Domain	Fire	Flood	Heat wave
English	81,693	1,022,967	367,839
Greek	27,526	2,988	50,540
Italian	141,745	12,003	685
Spanish	3,158,257	712,556	217,442

As the collected tweets have been crawled based on predefined lists of domain-specific keywords, as opposed to more semantic criteria, they are subject to lexical ambiguity (e.g. "flooded with messages", "he's on fire", etc.), and hence not all of them refer necessarily to actual fire, flood and heat wave. Since within the overall beAWARE pipeline, the Twitter messages that are fed to textual analysis have been first filtered so that mostly relevant ones are kept, the empirical study was based on a subset of tweets that had been manually annotated with relevant judgments, and which amounted, at the time, to 7,914 tweets, most of which in Italian, Spanish and English. Interestingly enough, slightly over half the manually annotated tweets were judged to be relevant, i.e. refer to actual fire, flood or heat wave events.

4.2.1 Tweet location metadata

A key aspect of interest for the analysis of crisis situations is the location of the message, whether it is the origin of the message, or a location that the message is about.

Twitter allows users to share their location as metadata with their tweets; however, this option is not widely used. In our sample, only 88,287 out of 5,797,042 have any location information associated with them (or around 1.5%), and in many cases the location is too coarse, indicating only the country or city of origin. Only 4,244 tweets (or 0.07% of them) have exact geo-coordinates which would allow the system to know the exact location, where the tweet originated from.

This makes it very important to be able to extract the location from alternative sources, and in particular, from the text itself, assuming that relevant information is provided (e.g. name of a street, name of bridge or monument which can be subsequently geo-localized, etc.), or through more integrated approaches, as needed further down the overall processing pipeline, e.g. making use of preceding and following tweets by the same user and the pertinent spatiotemporal correlations or through the aggregated interpretation of multimodal information.

4.2.2 Tweets provenance and information content

In order to determine what information can (potentially) be extracted from the tweets, we proceeded to manually annotate a sample of messages, starting from the subset of the relevant marked ones for the following language-domain combinations:

- English - floods (536 of 1,015 were marked as relevant)
- Spanish - fires (615 / 973)
- Greek - heat wave (755 / 897)

For each of these collections, 200 tweets were manually annotated in more detail as described in the following. The reason for not including an Italian dataset at the moment, was that the English dataset, chosen for demonstration purposes, already covered the domain associated with the Italian pilot, and as in the current stage of analysis intra-language differences within the same domain are out of scope (being pertinent to development choices during the implementation of the respective multilingual text analysis module, they are planned for once a first baseline version is first made available).

4.2.2.1 Authorship

Tweets can originate from a variety of sources. Since provenance is an important factor for assessing the credibility and trustworthiness of the considered information, we selected to differentiate between messages from authorities, such as the police or the fire department ("official"), messages from news outlets ("media") and others ("personal/unknown"). The distributions of the author types for the three considered collections are shown in **Error! Reference source not found.**.

As observed, messages from media outlets make up a significant proportion of the collected tweets, whereas official accounts of authorities are much less prominent. The second big group consists of personal accounts (or accounts for which no professional affiliation could be established). However, going through the messages, it becomes clear that, in their majority, the "personal" messages originate from respective news media and are frequently generated by readers clicking on the "share on Twitter" buttons on the respective news sites. While we did not manually annotate this aspect, it can be somewhat be approximated by looking how many of the tweets contain a URL (which most commonly points to a news article). For the "English flood" collection, 80% of tweets contain a URL, 84% for "Spanish fire", and 95% for the "Greek heat wave".

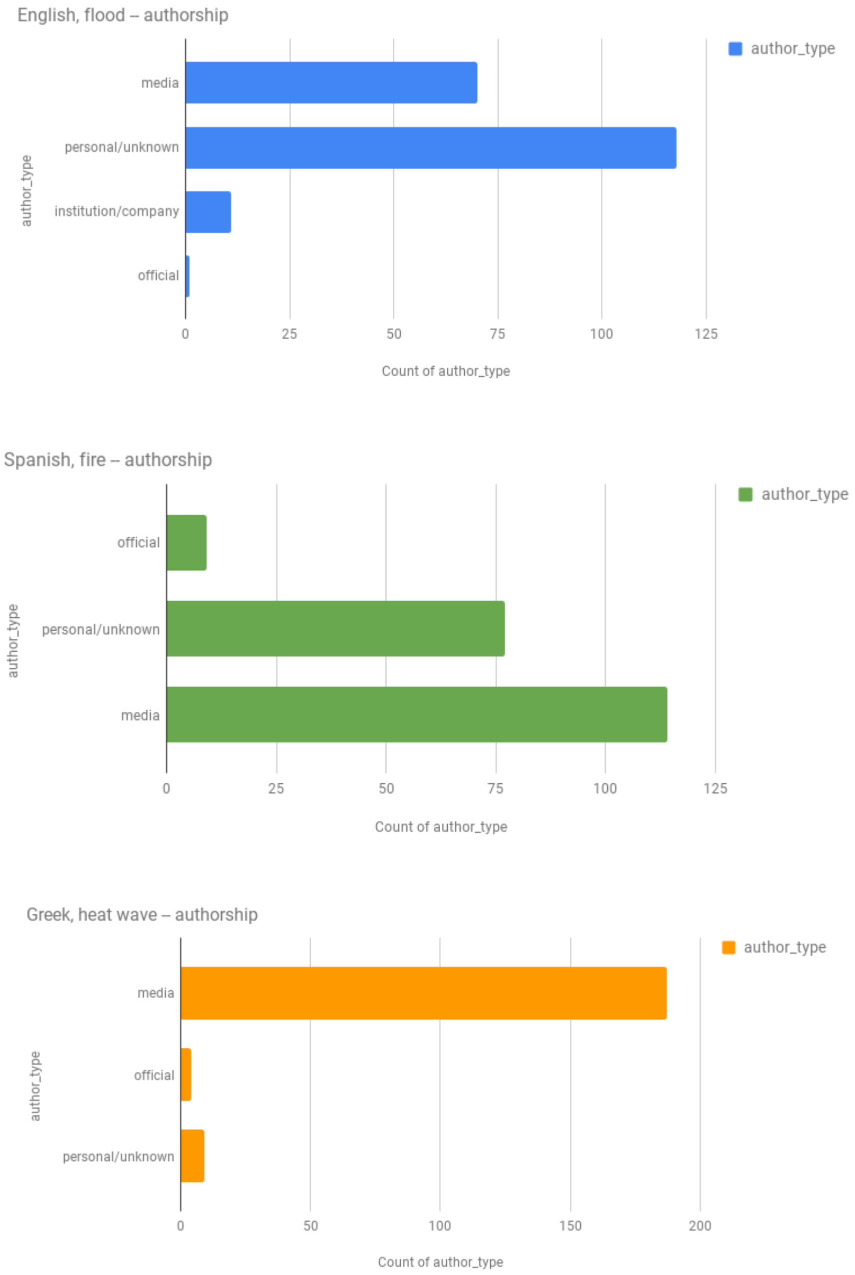


Figure 4 - Distribution of author types for the considered collections

A second interesting observation is that here are important differences between the collections, with media outlets accounting for 57% of the Spanish Tweets, compared to 35% of the English ones. This is also reflected in the other dimensions analyzed in the following sections. Whether there are underlying cultural and/or emergency type related correlations, remains to be found once the basic version of the multilingual text analysis is set up and such analysis will be fully automated.

4.2.2.1 **Categories**

Table 3 lists the categories that emerged during the manual inspection and annotation of the collected tweets and that have been used in order to create a preliminary categorization of the types of conveyed information.

Table 3 – List of tweets categories

Category	**Description**
Advice	advice on what to do, in a specific crisis situation or in general
Comment	general comments about a situation
Enquiry	questions about a situation
Warning	a warning about a potential crisis situation
Forecast	predictions about the development of the forecasted emergency
Miscellaneous	commentary that may or may not be related to a crisis situation[7]
News	news about an ongoing or past crisis situation (usually not actionable / not aimed at persons affected by the situation)
Places/routes	advice on safe/relief places or escape/blocked routes for affected persons
Status_update	an update on an ongoing situation in a way that is potentially relevant / actionable for affected persons

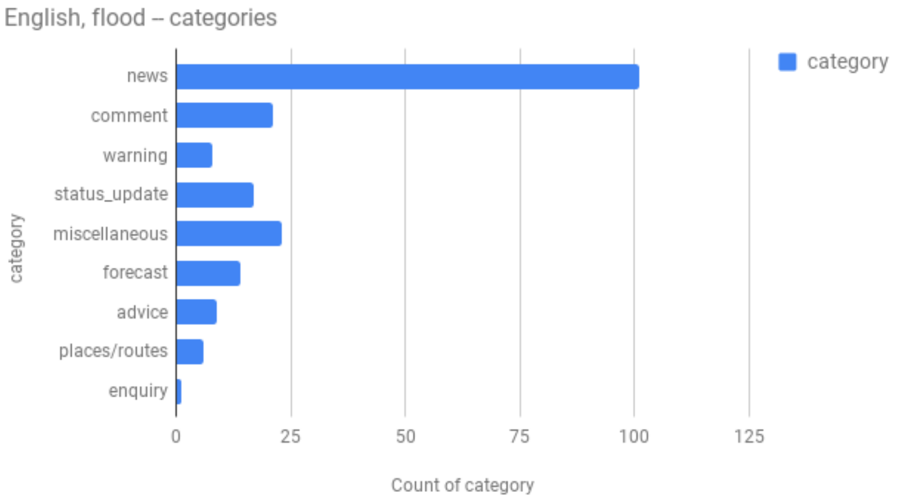
The differences between these categories are not always clear-cut, as it often depends on the (implicit) intention of the sender as well as the receiver. As such, a message can be a warning to people unaware of a situation that may affect them, a status update to others, and merely "news" to people unaffected by the event, and our categorization is somewhat subjective as to the most likely reading of any given message. Some examples:

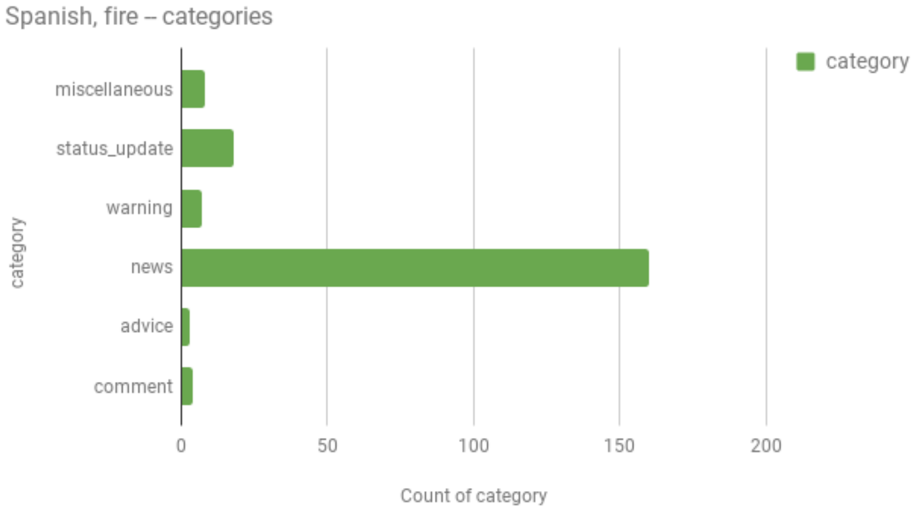
- News: *"Millions under flood threat in Carolinas https://t.co/halp4PiAPw"*. (The majority of the recipients of this message are unlikely to be affected by the situation, and it is too vague to be actionable.)
- Status update: *"Pockets of heavy rain continue to fall between the Roanoke area and the Southside. Flooding issues persist. https://t.co/0L6xABKZvz"* (Appears to be aimed mainly at people affected by the situation.)
- Forecast: *"By noon we should have over an inch of rain added to the already saturated soils. Flooding still an issue Tuesday. https://t.co/gTQEDr4BpA"*
- Warning: *"#PrepperNews https://t.co/ZdeEkHcqQV Warning: Flooding possible this week - Times Daily https://t.co/Hv2y69Zaz9"*

[7] Note that while we only examine tweets previously marked as "relevant", this does not mean that they all refer to the same emergency event, only that the use of crisis-related terms is non-metaphorical.

- Advice: *"#Flood tip: Avoid driving through pools of standing water. Water could be covering fallen power lines & other debri… https://t.co/7LnBySuBPU"*
- Comment: *"Crap! I hope the rain stops before anymore flooding! What a day already! I hope it gets lots better! https://t.co/4vIm58i7O1"*

The distributions of the tweets' categories for the considered collections are shown in Figure 5.





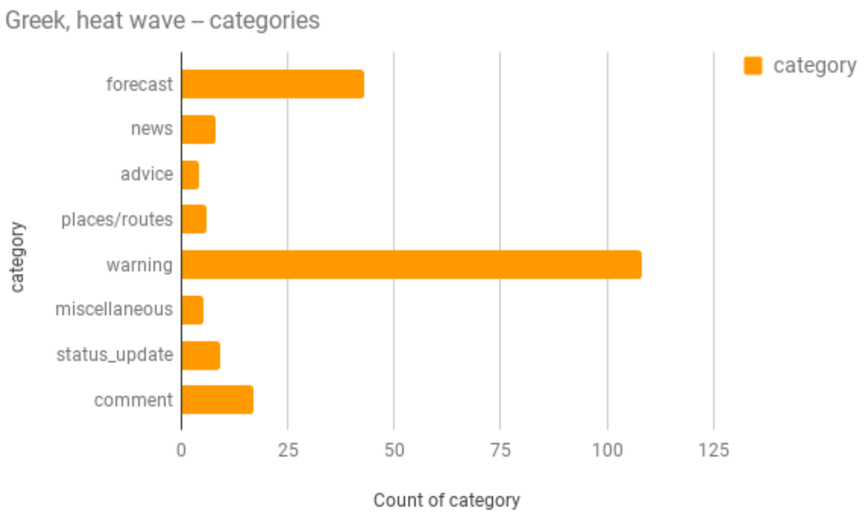


Figure 5 – Distribution of categories, Greek, heat wave collection

Overall, we see that many messages are general news, with little practical impact or actionable information. This is especially true for the "Spanish fire" collection where news represent 80% of all messages. The "English flood" collection is somewhat more varied, with much more practically relevant information such as warnings, forecasts, status updates, information on places and routes to use or avoid, etc. It also contains more personal comments, including messages of empathy. The "Greek heat wave" collection, on the other hand, contains a considerable number of warnings and forecast messages, as well as advices and information on places of relief, as during the reference period, Greece experiences continuous heat waves.

The differences between the collections are largely due to the way the data is collected (selection of keywords, etc.), but also the types of emergency events taking place during the time frame of collection. As such, the Spanish language tweets are dominated by news about the London Grenfell Tower fire, with little practical impact to people getting their information from Spanish language sources, which may have significantly biased the collection. The flood-related English collection, on the other hand, contains much more information aimed at local residents worldwide (in the US, UK, Australia, Sri Lanka, etc.), be it from local news media or from personal accounts. Table 4, Table 5 and

Table 6 show a further breakdown of the respective tweet categories distribution, by indicating the respective author type.

Table 4 - Categories and authors, English, flood collection

	Institution/ company	Media	Official	Personal/ unknown	Total
advice	3	2	1	3	9
comment	1	0	0	20	21
enquiry	0	0	0	1	1

forecast	3	6	0	5	14
miscellaneous	1	7	0	15	23
news	0	39	0	62	101
places/routes	0	1	0	5	6
status_update	1	10	0	6	17
warning	2	5	0	1	8
Total	***11***	***70***	***1***	***118***	***200***

Table 5 - Categories and authors, Spanish, fire collection

	Media	Official	Personal/ unknown	Total
advice	2	0	1	3
comment	1	0	3	4
miscellaneous	4	1	3	8
news	96	0	64	160
status_update	7	7	4	18
warning	4	1	2	7
Total	***114***	***9***	***77***	***200***

Table 6 - Categories and authors, Greek, heat wave collection

	Media	Official	Personal /unknown	Total
advice	3	1	0	4
comment	17	0	0	17
forecast	41	1	1	43
miscellaneous		0	5	5
news	8	0	0	8
places/routes	5	1	0	6
status_update	9	0	0	9
warning	104	1	3	107

Total	***187***	***4***	***9***	***200***

4.2.2.2 **Location**

As already mentioned, though an option offered by Twitter, location metadata tend to not be commonly used. In the "English flood" collection, only 15 out of 537 tweets (2.8%) contained location information; the “Spanish fire” collection follows a similar distribution, with 15 out of 600 (2.5%), while for the "Greek heat wave" collection, only 4 out of 756 tweets (0.005%) contained location metadata.

According to the manual annotation, the number of tweet, whose text contained some kind of location mention was as follows:

- English, flood: 150 / 200 (75%)
- Spanish, fire: 178 / 200 (89%)
- Greek, heat wave: 29 / 200 (14.5%)

Despite the high percentage, in the majority of cases the location information provided is very coarse (at the city, state, or even country level) and thus of limited use, when it comes to extracting actionable information. Interestingly, the same can apply when very specific location information is provided, when more general information that would allow to make sense of it is lacking, e.g. providing a street name without mentioning the city the street is in, *"Chester Bridge reopens after being closed due to flooding"*. In those cases, additional context is needed, e.g. from surrounding tweets by the same user, or other tweets in the same thread (if there are responses).

Delving into the individual collections, the "English flood" ones contains many more useful location mentions (i.e. specific and relevant to affected people), while the "Spanish fire" collection contains mostly very generic location information (e.g. London). This might be a reflection of the much greater variety of message types in the former collection, whereas the latter consists in a much greater proportion of (often non-local) news. The “Greek heat wave” collection, not only includes a considerably lower percentage of tweets with some of sort location information mention, but, when such information is present, it is either at city or prefecture level; yet, this is largely attributed to the country-wide impact of the heat wave events mentioned in the collection.

4.3 **Visual data study**

As aforementioned, as far as image and video inputs are concerned, and based on the currently available use case scenarios and respective specifications laid out in D2.1, key types of information to be targeted by visual content analysis involve scenes that depict flooded areas, smoke, fire and traffic congestions. The goal and challenge is to develop algorithms, of low computational cost, that can accurately discriminate in real-time such situations, and eventually enable to detect and predict elements at risk (people, cars, monuments, etc.). In addition to aforementioned content-wise observations with respect to the visual information of interest that drove the collection of respective datasets, we looked into the characteristics entailed by the mainly considered capture devices, namely Close Circuit TeleVision (CCTV) surveillance systems and mobile cameras, as well as the potential added value that could be brought through the incorporation of UAV and wearable cameras:

- **CCTV surveillance systems** are deployed in remote areas (i.e. across a forest or embankment) or regions of interest (i.e. across a city's bridge, over a highway), so as to monitor the area over regular time intervals. The used frame per second recording rate and frame resolution are low due to storage and energy considerations, as the capturing procedure involves particularly large volumes of data (e.g. 5 days of recording could easily amount to 100G of visual data). The captured data are transferred regularly in a central server, from where the acquisition and subsequent feeding to the beAWARE system and image and video contents will take place.
- **Mobile visual data** include image and video contents captured by civilians and first responders, posted on social media, i.e. Twitter and video-sharing websites (e.g. YouTube, Dailymotion, etc.) or fed directly to the system via the beAWARE mobile application. This data are usually accompanied with textual information that could possible help an analyst to deduct more meaningful conclusions. The great amount and variability of such data, during a crisis scenario, requires a powerful server and a well-designed big data strategy (i.e. targeted crawling, accurate data filtering, multimodal fusion) to be deployed and deal with the nature of this data.
- **UAV visual data**, captured by drones, usually comprise videos of short duration. Within our investigations, the main motivation would be that of first responders, trained on UAV flight technology, operating such systems to capture the unfolding situation in the affected area and uploading them to the beAWARE system for online processing. It is worth mentioning that the stabilizers that have been deployed on the latest UAVs afford high flexibility and thus applicability in even challenging emergency situations.
- **Wearable cameras** are a relatively new technology that has been adopted to record extreme sports and since then has been expanded to more fields of interest, such as crisis scenario. First responders could use this technology to capture small videos or image samples and transfer them to a central center in order to deduct further conclusions. The data usually contain a great deal of noise that is produced from the moving camera and require stabilization algorithms to acquire a safe conclusion.

Last, and in collaboration with the user partners, the use of satellite images that capture the region of interest before and after the crisis incident will be investigated in order to determine whether there is a potential for added value within the targeted pilots. Figure 6 shows some example taken from **https://www.planet.com/disasterdata/datasets/**

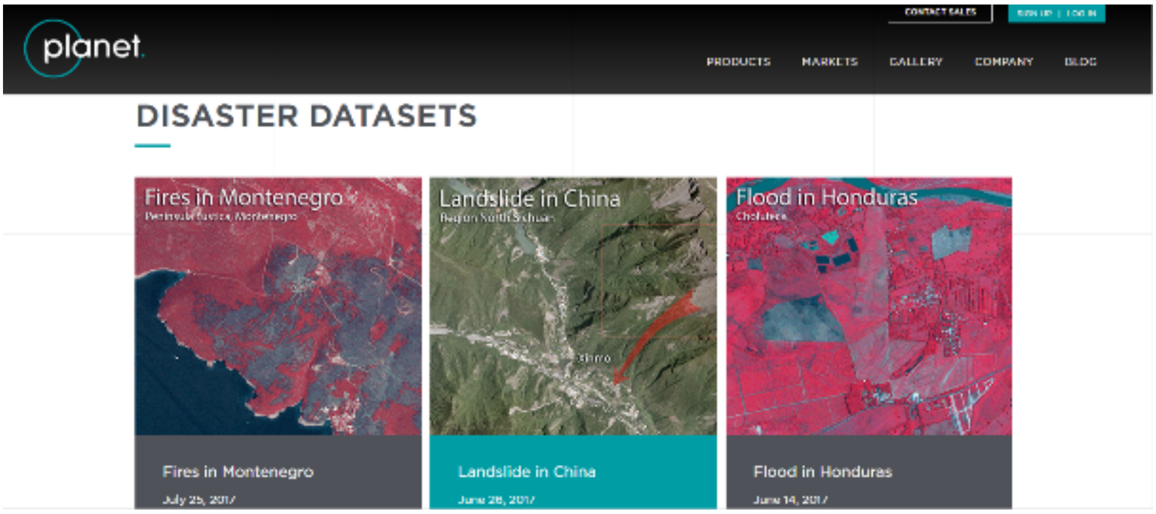


Figure 6. Disaster dataset for compensating crisis events

5 Summary

Data collection is an important step towards the development of accurate and robust content analysis techniques that can adequately meet the pertinent, functional and non-functional, specifications and requirements. Addressing the multimedia inputs considered in beAWARE, namely audio, text and visual data that originate from the multitude of involved sources (Twitter posts, reports sent by civilians and first responders via the beAWARE mobile application, calls to emergency numbers, internal communications between first responders and authorities, static cameras, etc.), this deliverable described the currently collected datasets and the findings of their empirical study.

As outlined, data collection activities have been initiated from the early onset of the project, starting with historical data provided by the user partners and continuing with the search for relevant publicly available ones; in parallel, tasks for the collection of data within beAWARE have been launched and also planned during the course of the project, in accordance with the scheduled field trials.

An overall observation is that as far as text and visual data are concerned, a wealth of them can be acquired quite straightforwardly, both within beAWARE (Twitter messages are continuously crawled and annotated, user partners already deploy such data, e.g. AAWA is using static cameras across bridges to monitor the water level and velocity during floods, etc.) as well as from third parties. Audio data, on the other hand, involving emergency calls and communications between first responders and authorities, present certain challenges due to imposed privacy and security restrictions. However, as aforementioned, we are already in touch with emergency call centers from both user partners and third parties, as well as with research groups that possess such data in order to sort out respective license agreement documents; in parallel, and in collaboration with the user partners, we have already initiated the compilation of recordings of simulated calls to ensure the availability of beAWARE-specific audio inputs.

The empirical study of the collected material resulted to a number of interesting findings with respect to the types of information of interest with respect to the development of corresponding analysis techniques, as well as the subsequent integration and aggregated interpretation. The study of the collected audio datasets outlined the key information aspects that authorities and first responders seek, during the report of an incidence. The examination of the crawled tweets revealed preliminary content and provenance/authorship categories, while highlighting the need for contextualization, as individual tweets, in their majority, do not provide much actionable information (e.g. location information may be too concise or abbreviated to the point of needing additional external information to be mapped to real-world locations). In addition, the types of differences observed in the distribution of the contents of tweets for the three emergency domains (fire, flood, heat wave), suggest the existence of correlations between the type of emergency and what people post, thus allowing to delineate, even coarsely, the underlying domain's scope and coverage. A final observation, to be further investigated as more data are been crawled, is to what extent the keywords used for filtering, result in skewed conclusions with respect to the wealth of the information that is actually being posted. As far as the visual data are concerned, empirical study showed that the video analysis modules depend significantly from the nature of the capturing device and the way that data are acquired. Fast and lightweight processes can run online, but most of the data should be transferred into a powerful server and analyzed offline. Recent developments on wearable and UAV devices pave new grounds for visual context analysis and understanding, yet, with a great deal of challenges that need to be addressed before safe conclusions could be deducted.

Last, and are already noted, the deliverable reports on the datasets and empirical analysis findings that correspond to the time of writing. As this is a continuous process, updates and further refinements are expected and will be reported in the upcoming deliverables D3.3 and D3.4 that present the basic and advanced, respectively, techniques for the analysis of the audio, text and visual beAWARE inputs.