



# Improving Mapping With the integration of Social Media and Crowdsourcing

Barbara Pernici,  
Politecnico di Milano

[WWW.E2mC-PROJECT.EU](http://WWW.E2mC-PROJECT.EU)



*The E2mC project is co-funded by the European Commission /  
H2020 Programme Grant Agreement No.730082".*

**PSCE Conference, May 23, 2018**



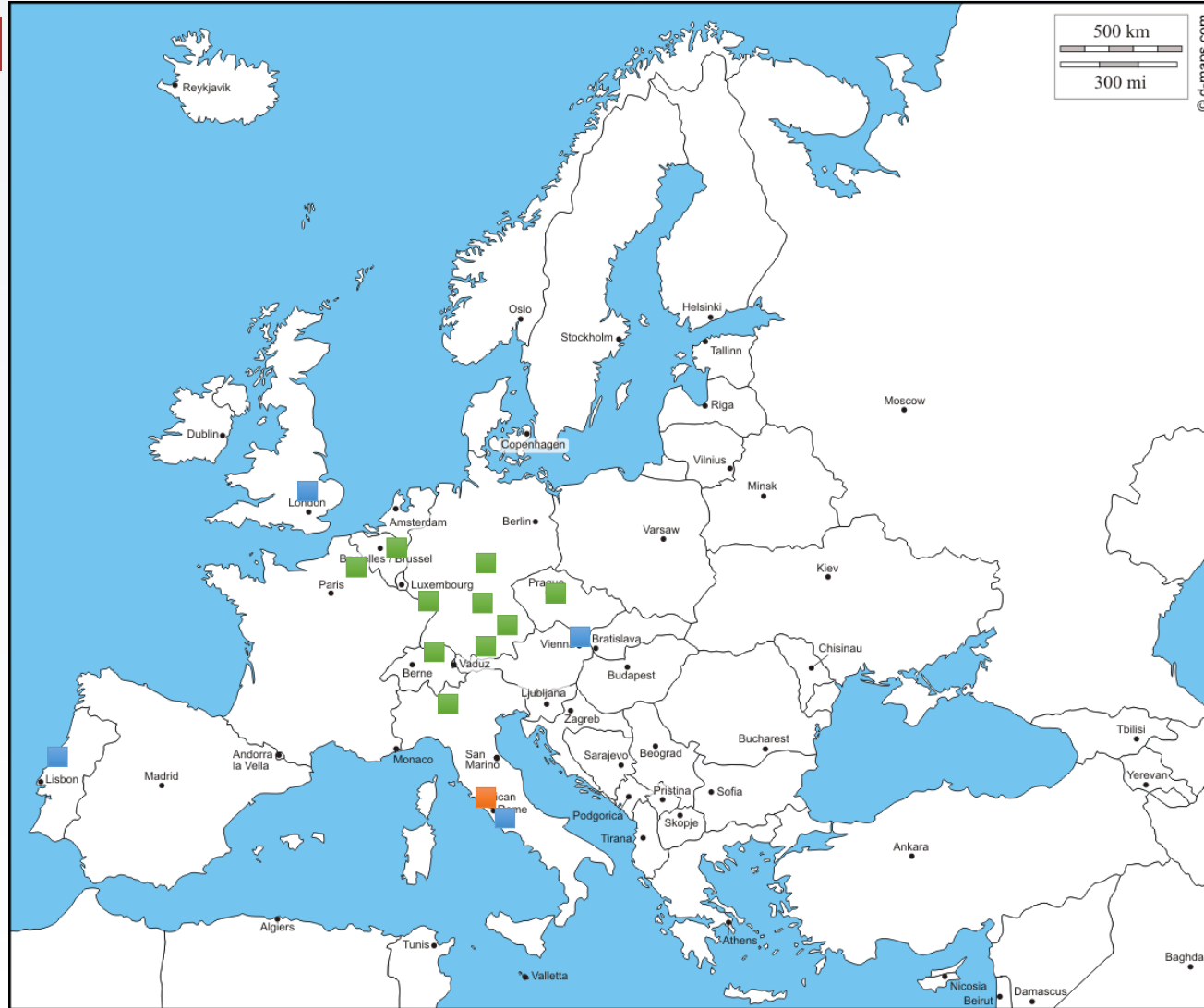


# EU H2020 Project E2mC – Evolution of Emergency Management Services in Copernicus The Project Team

## Project Coordinator



## Project Partners



## Users



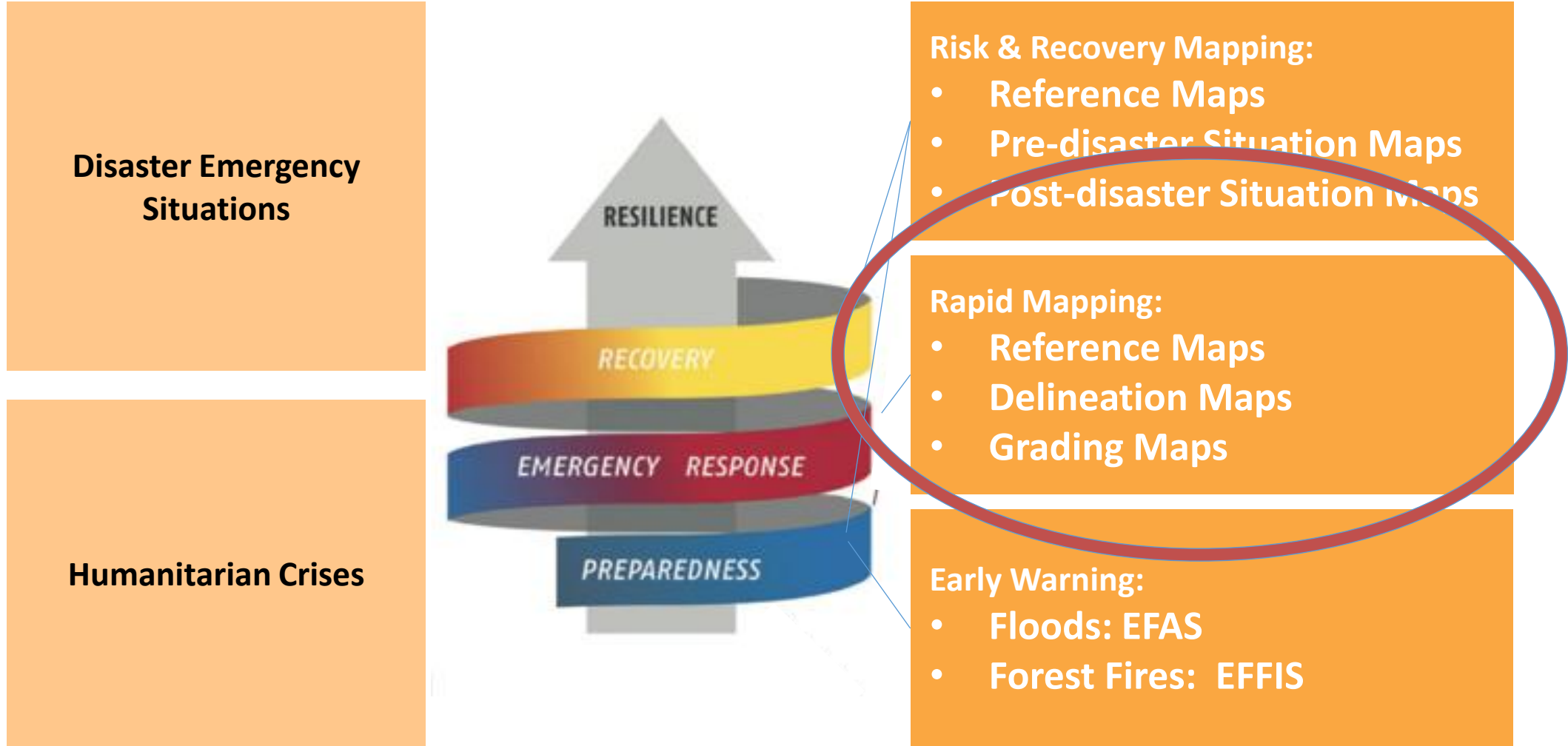
Orange square: Coordinator

Green square: Partner

Blue square: User



# Copernicus Emergency Managements Services



*EFAS = European Flood Awareness System;  
EFFIS=European Forest Fire Information System*





# Copernicus Emergency Management Service

## Rapid mapping

### Southern England floods 2014



<http://emergency.copernicus.eu/> (detail)





## Copernicus EMS - Mapping

...timeliness...not yet fully achieved...

...most of the **delay** is concentrated in the availability of the first usable post event satellite image...

...map production **throughput**...

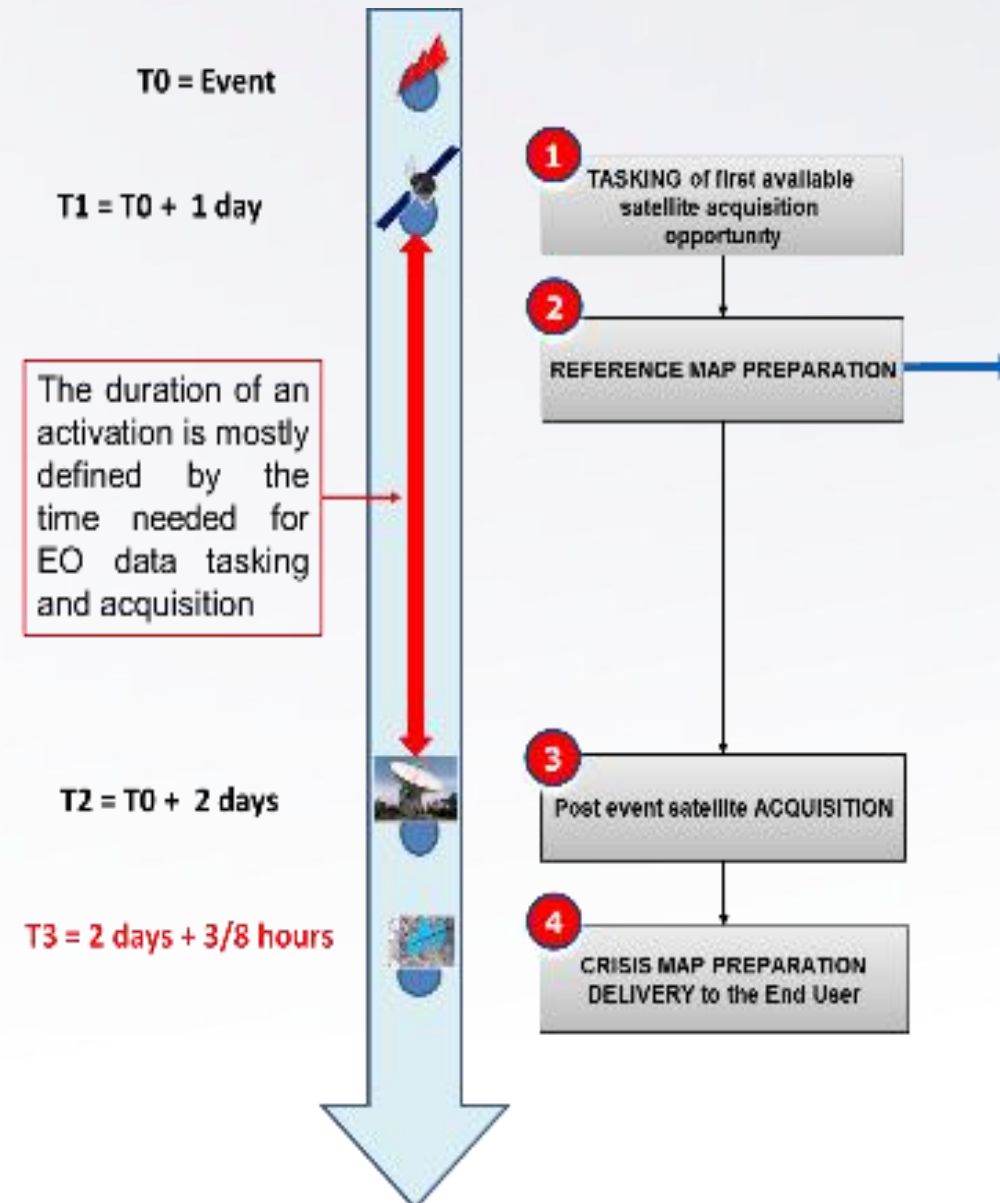
...**quality**...

...mapping of urban areas...

...broadening the current **scope** of the EMS service...

...enlarging the Copernicus EMS **range of use**...

# E<sup>2</sup>mC Vision – The Problem





OBJECTIVE







# Key Points

- **Combined use of Social Media and crowdsourcing** to derive actionable information supporting satellite emergency mapping, in particular in those cases where satellite data show limitations (e.g.. SAR based flood mapping in urban and vegetated areas)
- Innovative approach for the **precise geolocation of social media contents** and for **the automatic identification of relevant/not relevant posts** based on Artificial Intelligence combined with crowdsourcing



# E<sup>2</sup>mC vision – The Solution

## Before E<sup>2</sup>mC

User as «a User»



Professional  
Emergency  
Responders



## After E<sup>2</sup>mC

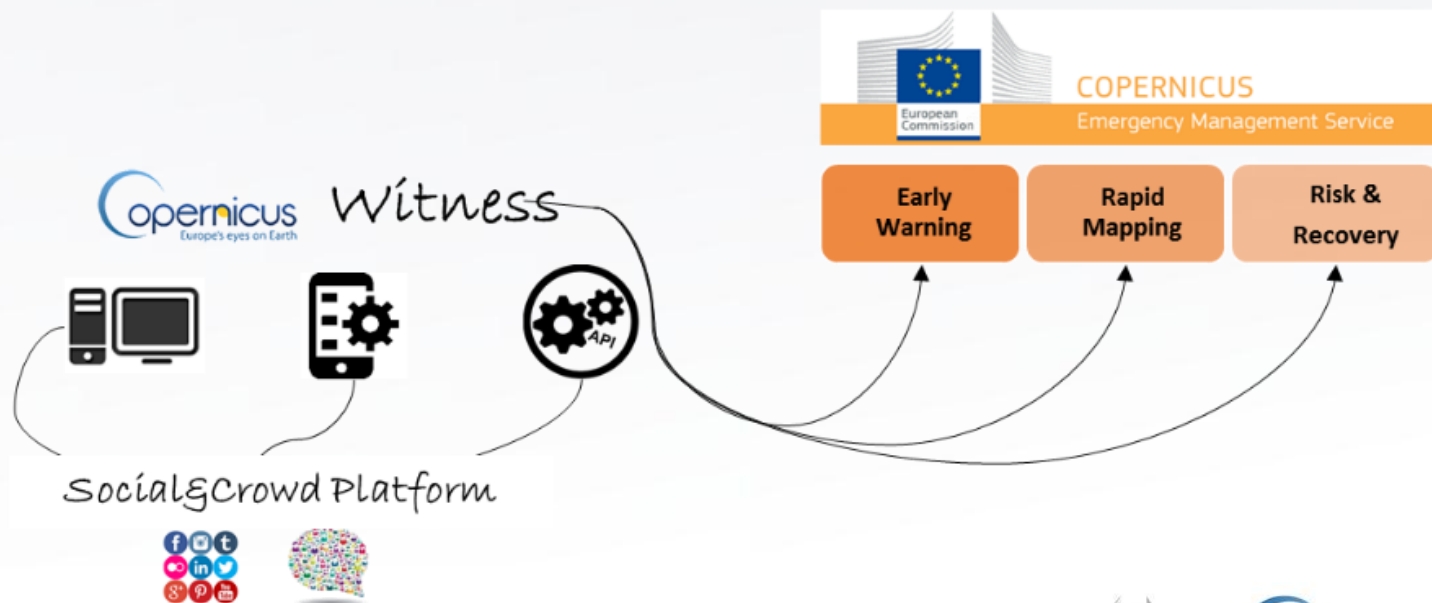
User as «a User»  
User as «a Sensor»



Professional  
Emergency  
Responders



Volunteers,  
Citizens



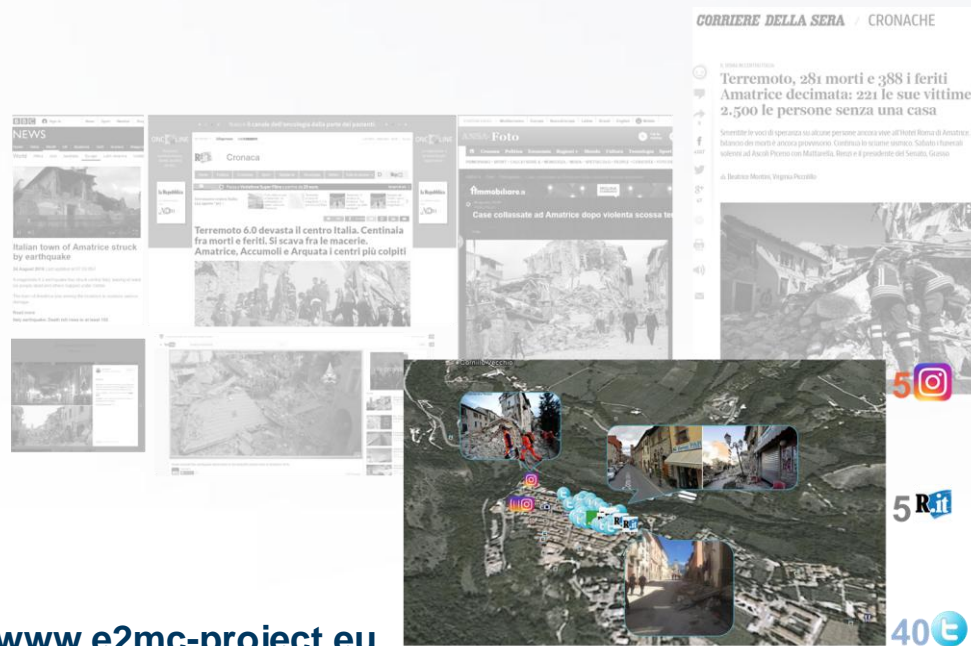




# E<sup>2</sup>mC vision – The Solution



Video features extraction



News and social media crawling, filtering and geolocalization

Today the handling of “unconventional” data (e.g. Twitter, news, ...) is **fully manual**.

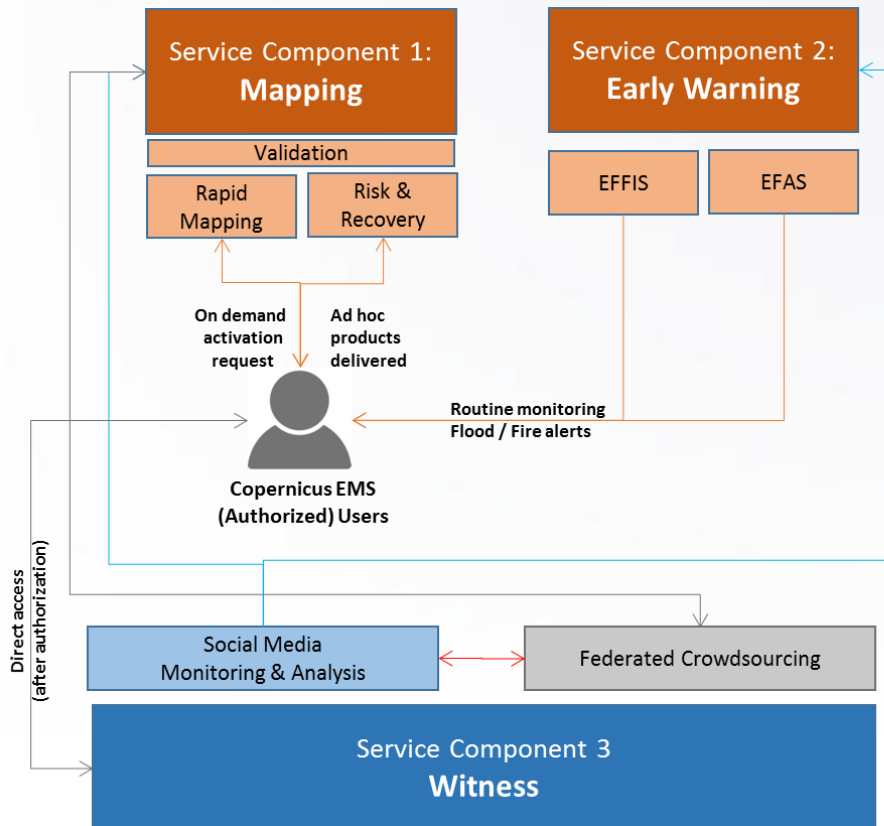
E<sup>2</sup>mC will provide tools to make this process as much as possible **automated**



# E<sup>2</sup>mC Idea – Copernicus EMS evolved version

## A social and crowd platform (S&C)

### COPERNICUS Emergency Management Service (EMS) Evolved version (v2)



Key elements of the Copernicus Witness

- perfectly fitting in the current operational Copernicus EMS.
- a Service Component that serves simultaneously all the different components of the running Copernicus EMS, as it takes into account the needs, requirements and constraints expressed by both the Mapping and the Early Warning
- directly available to Copernicus EMS (Authorized) Users as a **standalone service** ready for being further integrated into specific custom and downstream applications or for being used independently for ad hoc and tailored social media analysis or crowdsourcing campaigns

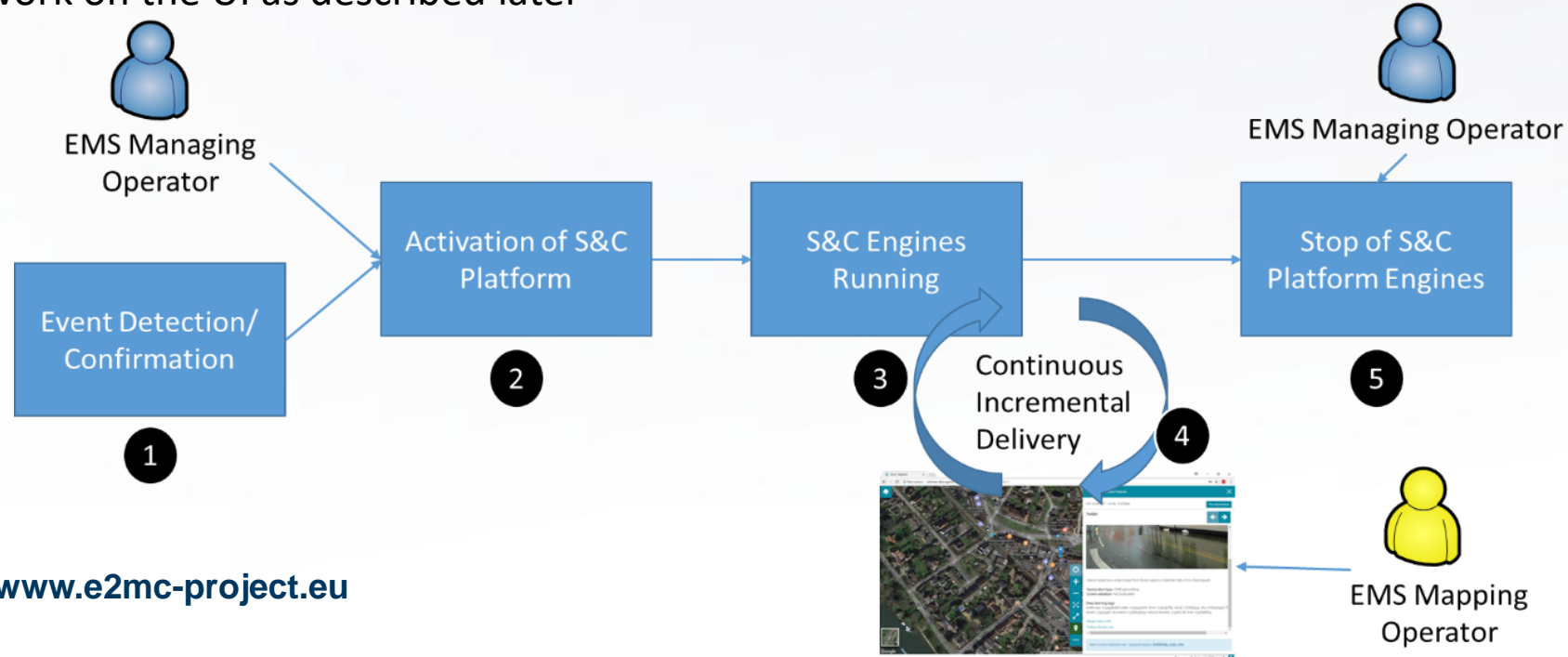




# Activation and management of the S&C platform

This workflow can be described as follows:

1. The Managing Operator or the event detection are in monitoring mode on possible disaster happening globally
2. The platform is activated by the Managing Operator or automatically by the Event Detection
3. The activation triggers the start of several processing engines (e.g. crawlers, geocoding, hot-spot, image filtering)
4. The delivery to the operator is done automatically on a continuous basis and the EMS Mapping operator can visualize and work on the UI as described later





# Operational use of the S&C data workflow

This workflow can be described as follows:

1. The Mapping Operator selects the activation from the Maps menu
2. The Mapping Operator navigates the map.
3. The Mapping Operator can manipulate social data to verify it and, possibly, change or remove it.
4. The Mapping Operator can export the final results to be used in the official enhanced EMS maps







# E<sup>2</sup>mC Idea – Technical Challenges

## Social Media Monitoring & Analysis

- Multilingual (semantics, ontologies), multidisaster, worldwide
- Access to heterogeneous data streams
- Selection of relevant data streams
- Big data problem for systematic monitoring
- Georeferencing strategies
- Identify relevant and independent contents
- Assess quality and reliability

## Federated Crowdsourcing:

- Heterogeneous platforms, with different triggering mechanisms and organizational models
- Data exchange, interoperability
- Exploit crowdsourcing in a SLA ruled environment with specific time constraints
- Crowd building
- Thematic geospatial and crisis oriented vs general purpose platforms
- Exploit crowdsourcing also for enriching social media analysis?

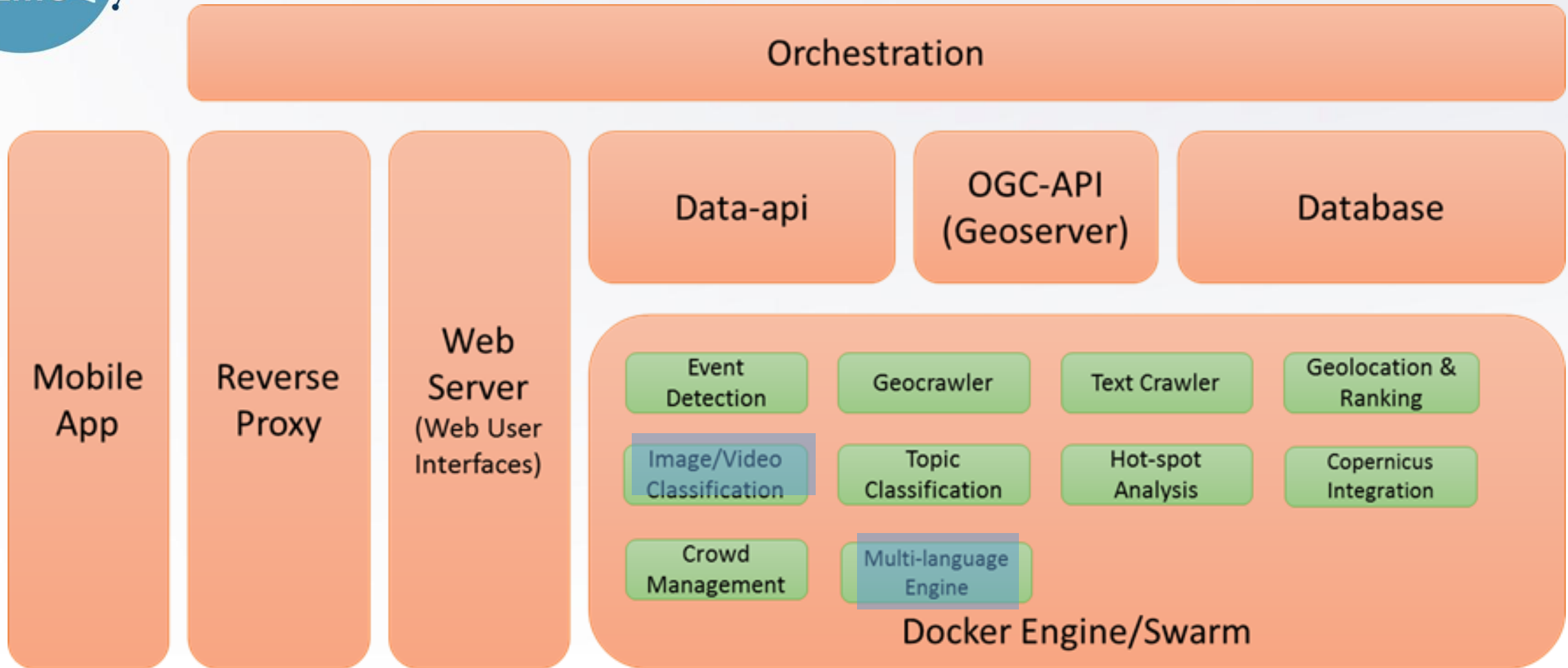


ON THE MOVE





# Prototype S&C Platform Architecture: data-centric and microservices





# Geolocation and ranking component (CIME)

## 2014 South England case study

Location  
Hypotheses

Support  
Based on  
compatibility and  
distance

**Ben von Raumbaboon** @spacebaboon Segui

Flood levels up overnight in **Datchet**. Need wellies for the pavement past **Spices**. **High St** and **Queens Rd** blocked off.



00:28 - 11 feb 2014

2 Retweet 1 Mi piace



id: 433155447894581248  
 publication\_date: 2014-02-11 08:28:01  
 Location: **Queens Road, Datchet CP, Datchet, Windsor and Maidenhead, South East, England @ (51.484065700000002, -0.58114310000000002)**  
 Flood levels up overnight in Datchet. Need wellies for the pavement past Spices. High St and Queens Rd blocked off. <http://t.co/7Rs3D9GKRe>



# Geolocation and ranking component Innovation vs. state of the art

With respect to the state of the art, the proposed approach:

- **Exploits OpenStreetMap** (as opposed to the more common use of GeoNames) as gazetteer, handling many more candidate locations and, thus, providing at the same time greater precision
- **Exploits the social networks** associated with social media to increase precision and recall
- Aims at **extracting the locations of the posts' content**, as opposed to the locations of the posting users
- Multilingual: NER with Stanford core multilingual and Rosette

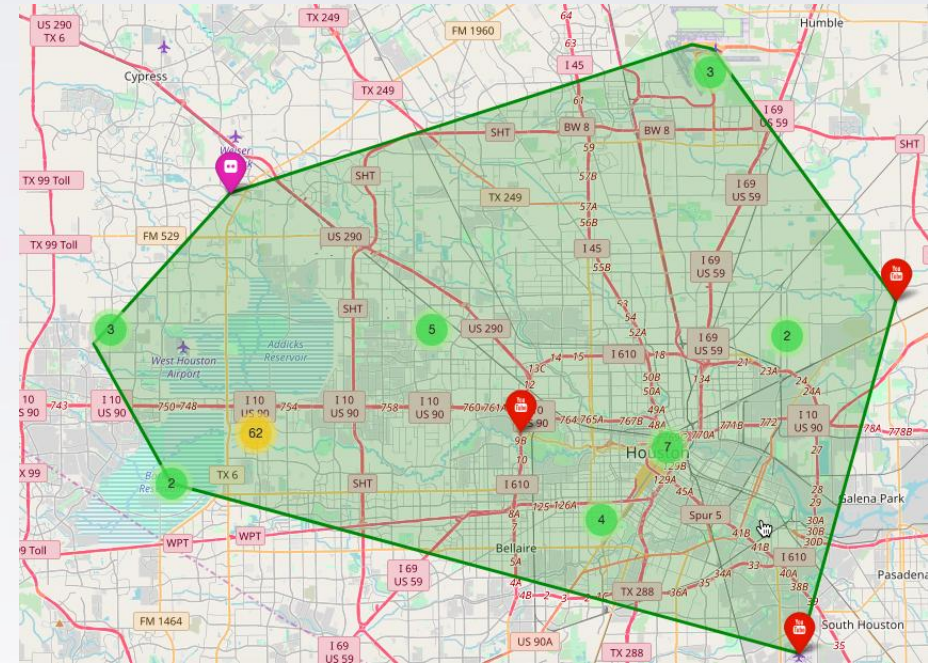
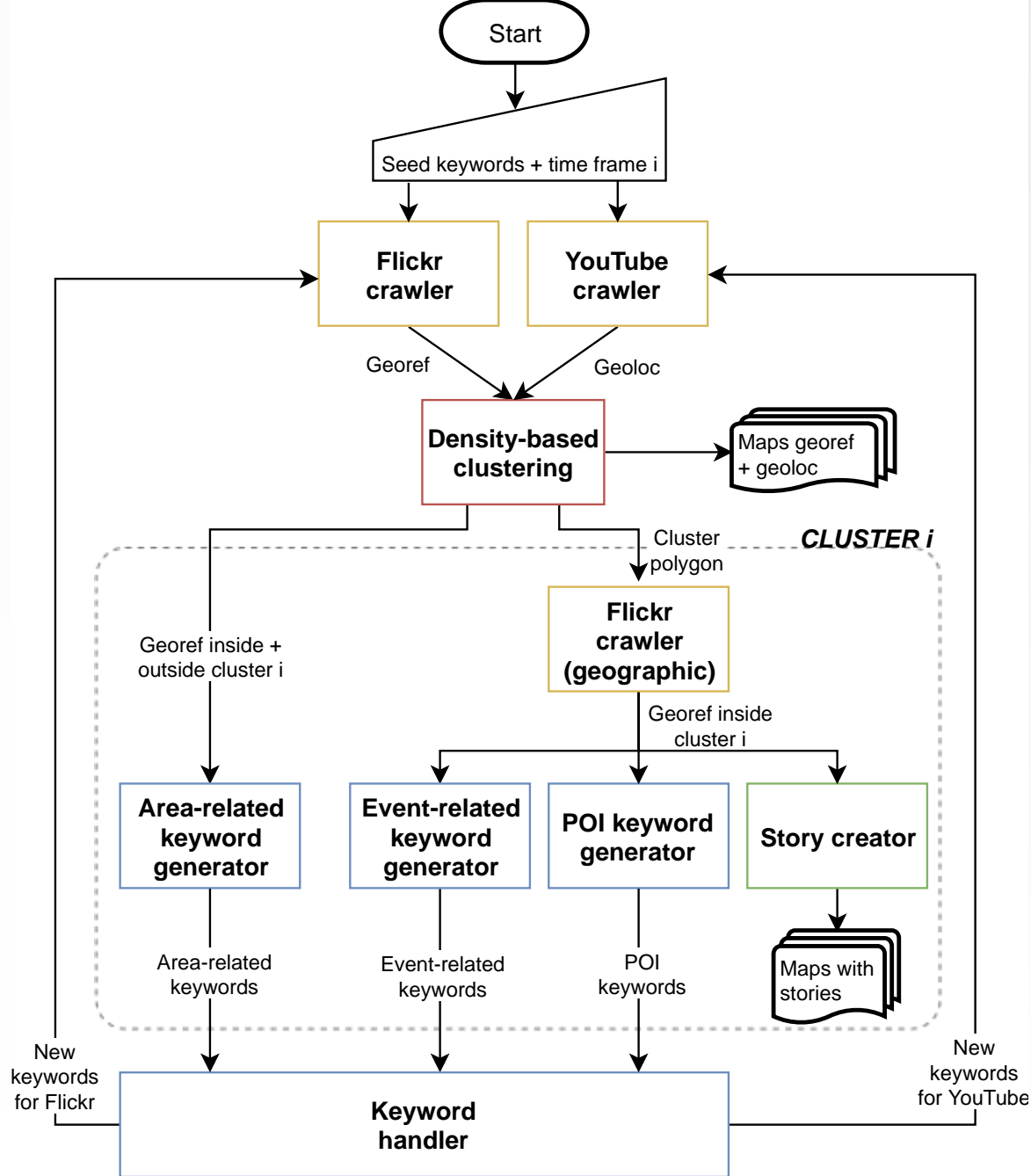




# Iterative multi-crawling for Flickr and YouTube

- Goal: improve the recall of flickr images and YouTube videos
- How: extracting new search keywords from clustering flickr posts – language independent as it uses tags.

Work done in collaboration with A. Autelitano, G. Scalia, B. Pernici



First keyword: flood

New keywords:

"houston flood", "tropical storm harvey",  
 "tropical storm flood", "street flooding",  
 "allen parkway", "buffalo bayou flood",  
 "tsharvey2017", "houston2017", "houston",  
 "harvey", "flooding", "hurricane harvey", "floods", "flooded"

Work done in collaboration with A. Autelitano, G. Scalia, B. Pernici



# Creating Flickr histories

A screenshot of a web application interface for creating Flickr histories. The main window displays a large photo of a wet road on Eldridge Parkway. Below the photo, it shows the date taken (2017-08-28 13:25:12) and the URL. To the right, a sidebar shows a list of other photos taken at the same location, each with its own date and URL. The interface includes a map at the bottom showing the location of Eldridge Parkway.

**POI name:** Eldridge Parkway  
**POI types:** ['highway=secondary']  
**longitude:** -95.625503  
**latitude:** 29.760872

**date\_taken:** 2017-08-28 13:25:12  
**url:** <https://www.flickr.com/photos/zeoeyann/36071274793/>

**POI name:** Eldridge Parkway  
**POI types:** ['highway=secondary']  
**longitude:** -95.625503  
**latitude:** 29.760872

**date\_taken:** 2017-08-28 13:25:12  
**url:** <https://www.flickr.com/photos/zeoeyann/36071274793/>

**date\_taken:** 2017-08-28 13:25:12  
**url:** <https://www.flickr.com/photos/zeoeyann/36071274793/>

**date\_taken:** 2017-08-28 13:25:12  
**url:** <https://www.flickr.com/photos/zeoeyann/36071274793/>

Work done in collaboration with A. Autelitano, G. Scalia, B. Pernici





# Geocrawler, Topic Extraction, Spatial Hotspots

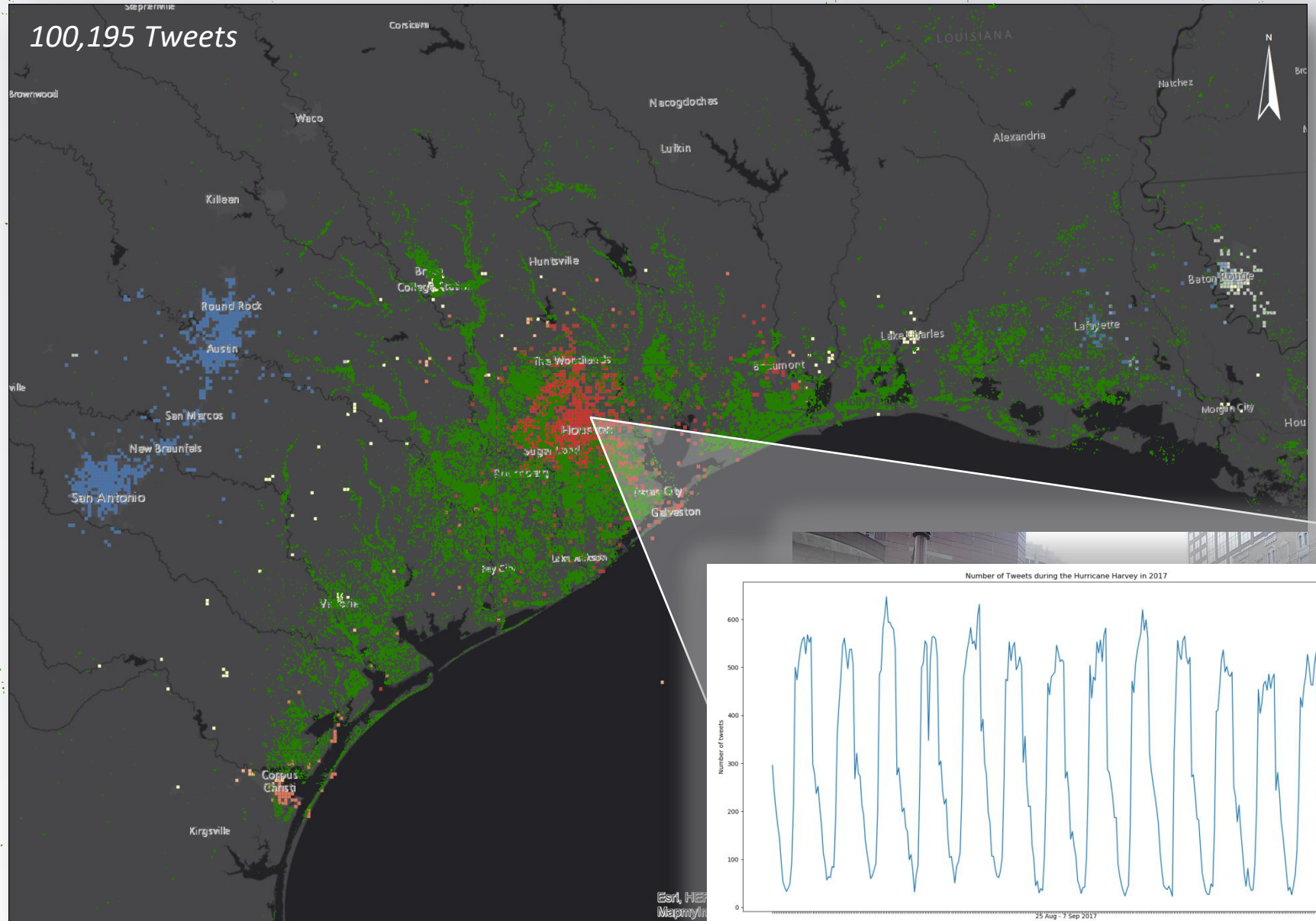
- **Geocrawler** up and running (*Twitter, Flickr, Youtube*)
- **Topic extraction** working satisfactorily (*proven for 3 different use cases*)
- **Spatial hotspot analysis** delivers useful results (*affected vs. non-affected areas*)
- **Accuracy assessment** shows convincing results (*statistical validation to compare the E2mC Witness with current EMS outputs*)



# Hurricane Harvey

- 6,706 tweets
- Houston: hot spot
- San Antonio and Austin: cold spots
- Coastal regions: hot spots
- Large number of **weather station's bots** partly creating hot spots

Hurricane Harvey 2017

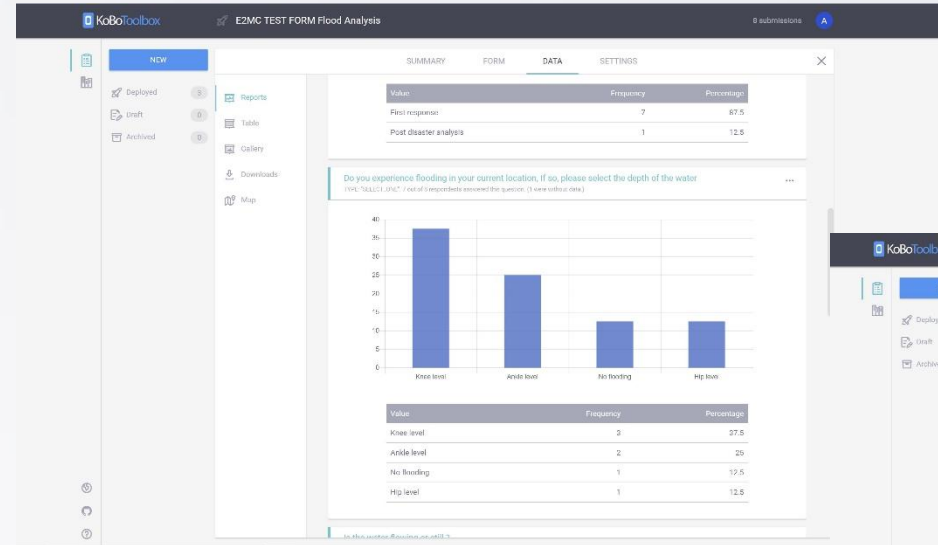
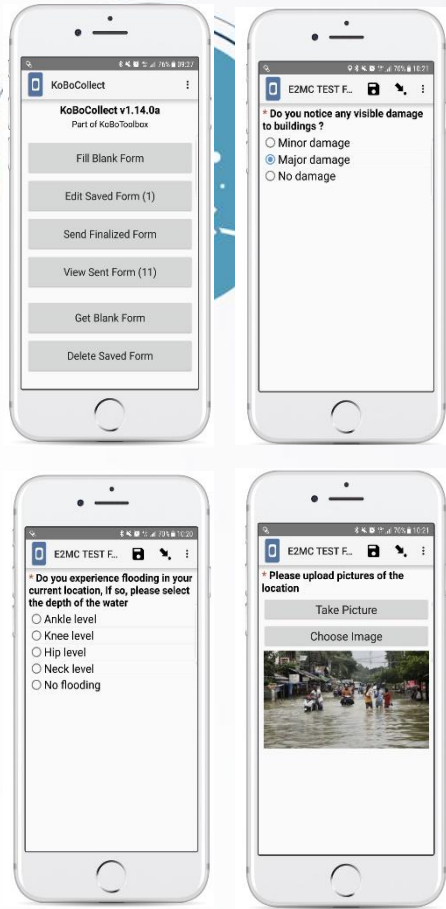




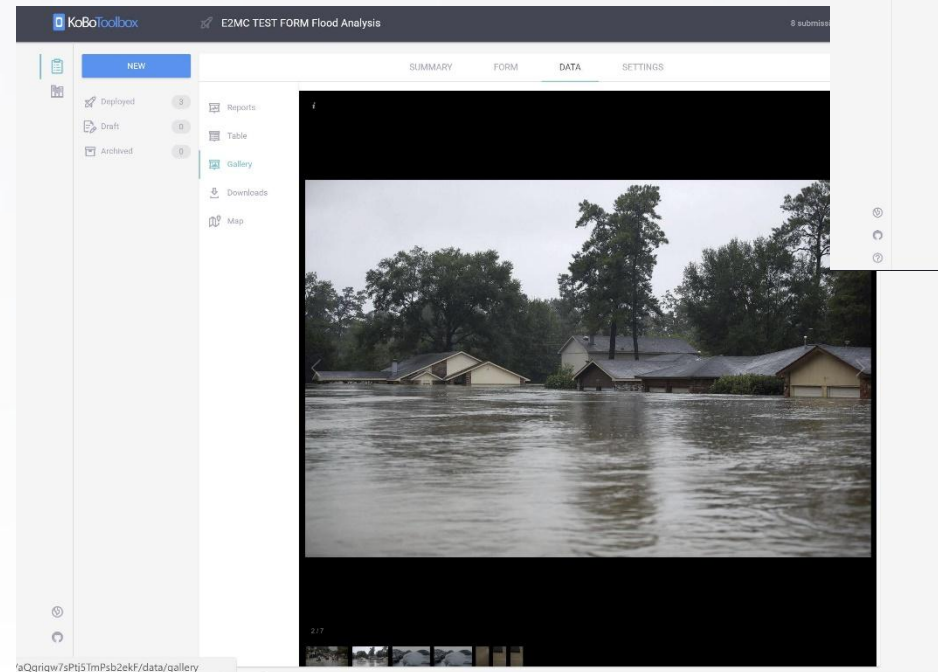
# Mobile Phone Collector App : KOBO COLLECT

- **Goal:** To Allow local communities (experts and/or volunteers) to contribute media information (pictures, video) along with the forms attached
- **Platforms under consideration:** Epicollect, KoboCollect, ODK, ASIGN
- **Features:**
  - Easy creation of missions without programming skills (e.g. report damaged buildings)
  - Easy to adapt to the need of each crisis
  - Easy collection of data
  - Easy integration on the E2mC project





Validation status	start	end	Please enter...	Please select...	Do you experi...	Is the water l...	Do you notice...
Open	May 3, 2018 9...	May 3, 2018 1...	User11	First response	Knee level	Flooding	Major damage
Open	May 3, 2018 1...	May 3, 2018 1...	C	First response	Hip level	Howing	No damage
Open	May 3, 2018 1...	May 3, 2018 1...	B	First response	Ankle level	Still	Major damage
Open	May 3, 2018 1...	May 3, 2018 1...	A	First response	No flooding		
Open	May 3, 2018 1...	May 3, 2018 1...	User34	First response	Knee level	Still	Major damage
Open	May 2, 2018 1...	May 3, 2018 1...	User32	Post disaster ...			
Open	May 2, 2018 1...	May 3, 2018 1...	User32	First response	Knee level	Flooding	Major damage
Open	May 2, 2018 1...	May 2, 2018 1...	User17	First response	Ankle level	Still	No damage



## Mobile Collector App

- Can share pictures of affected areas (geo-tagged)
  - Can fill assessment forms related to affected areas
  - Can send a geo-located text
- E2mC I [www.e2mc-project.eu](http://www.e2mc-project.eu)**

<https://kobo.humanitarianresponse.info>

## Collector DASHBOARD

- Gallery of images uploaded
- Summary of the assessment forms
- Geo-tagged images on a map



# Image analysis



Image Library



TensorFlow



Flood / No Flood Response

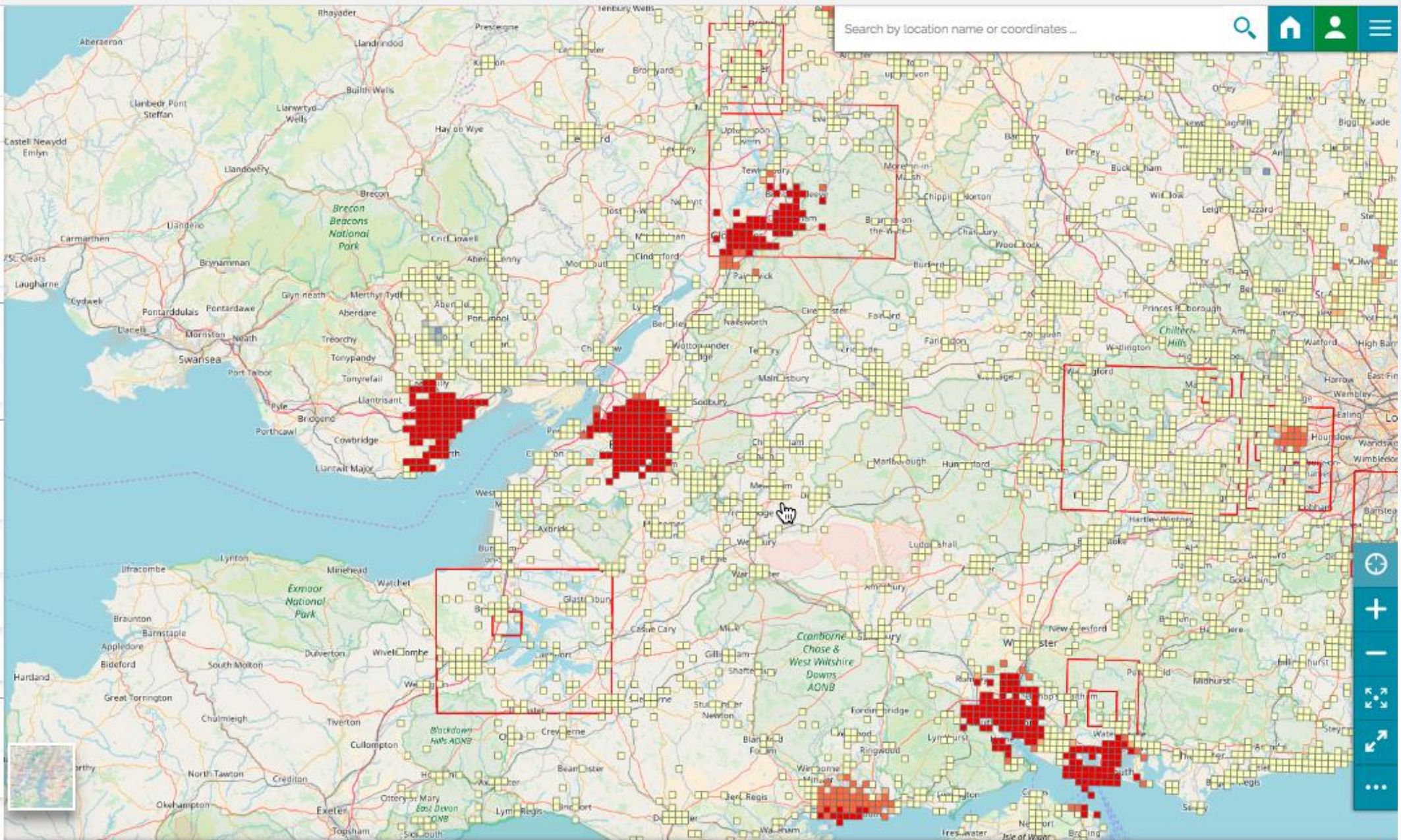
# Demo Prototype



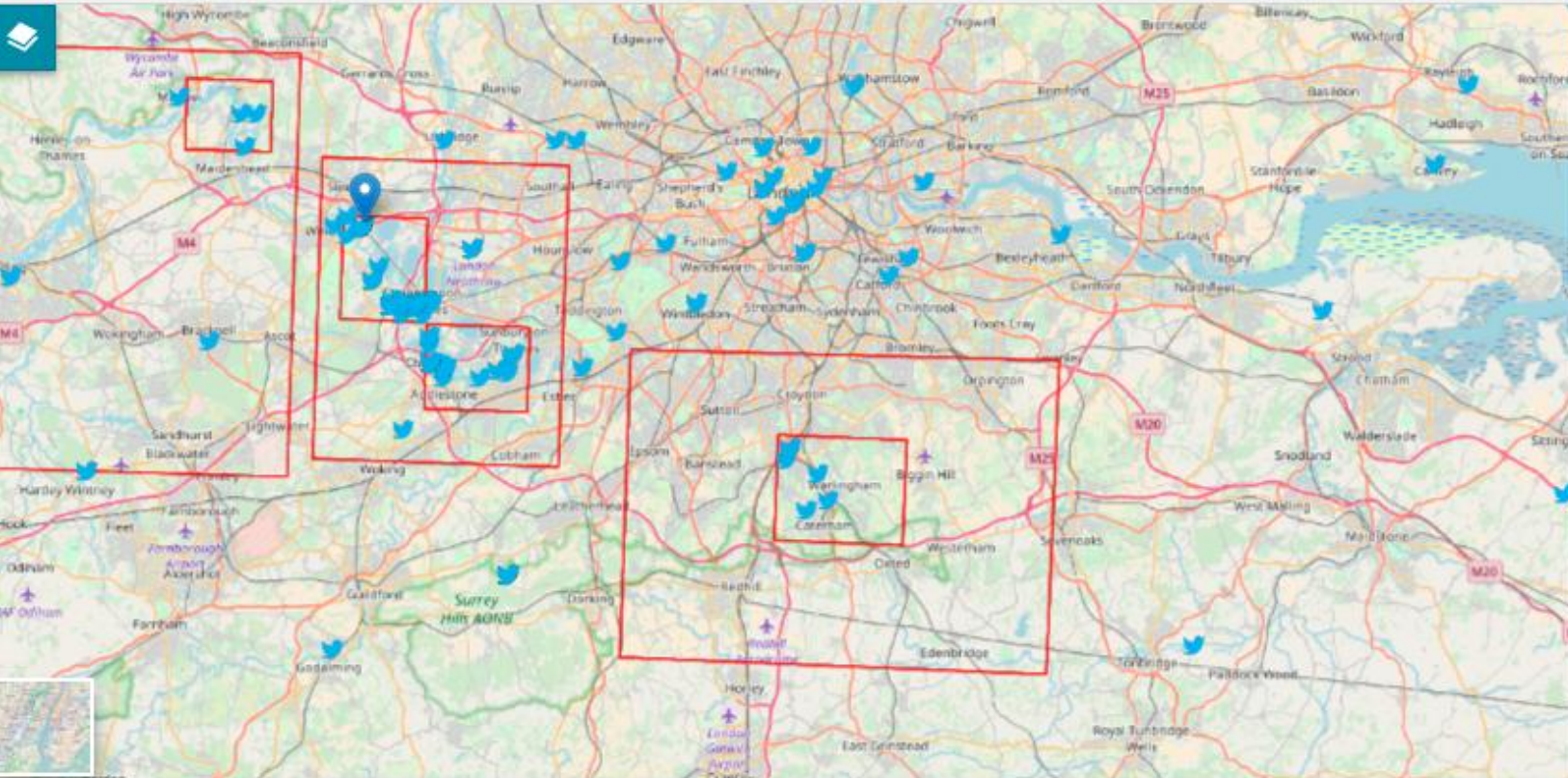
- Demonstrated at demo stand



- Filter layers
- Add Layer
- Social
  - Twitter
- GDELT
  - gdelt\_uk\_flood
- HotSpotAnalysis
  - Twitter
  - Youtube
  - Flickr
  - UK\_HotSpotAnalysis
- Copernicus
  - EMSR06g\_area\_of\_i...
  - EMSR06g\_crisis\_inf...







**Feature Info** ✕

Lat: 51.48651 - Long: -0.57678 [More Info](#)

Twitter ← →

Posted on:  
Feb 11, 2014 2:59:59 PM



Map navigation icons: Filter, Search, Layers, Full Screen, Print

Twitter

geolocation_rank	image_recognition	crowd_assessment	time_stamp	text	mediart
>50	yes	Relevant	Type number or expressi	Type text to filter..	Type text to filter..
100	Yes	Relevant	20140210081553	Bristol Road (B29 6LX) Birmir	http://pbs.twimg.com/medi
100	Yes	Relevant	20140210091257	@micky_norcross @Hi, can y	http://pbs.twimg.com/medi
100	Yes	Relevant	20140210092458	@chelseafc Hi, can you re-tw	http://pbs.twimg.com/medi
100	Yes	Relevant	20140210092555	@GazGShore Hi, can you re-l	http://pbs.twimg.com/medi
100	Yes	Relevant	20140210094111	@kirk_official Hi, can you re-t	http://pbs.twimg.com/medi

Manor Hotel now under threat from flood waters in Datchet. <http://t.co/8ALrz1uqph>

Geolocation type: CIME geocoding  
Crowd validation: Not Evaluated

Deep learning tags:  
waterway: 0.94598466, water: 0.93432206, town: 0.9149769, canal: 0.6798334, city: 0.60974437, flood street: 0.5913547, recreation: 0.56625634, natural disaster: 0.5261718, tree: 0.51658815.

[Street View Link](#)

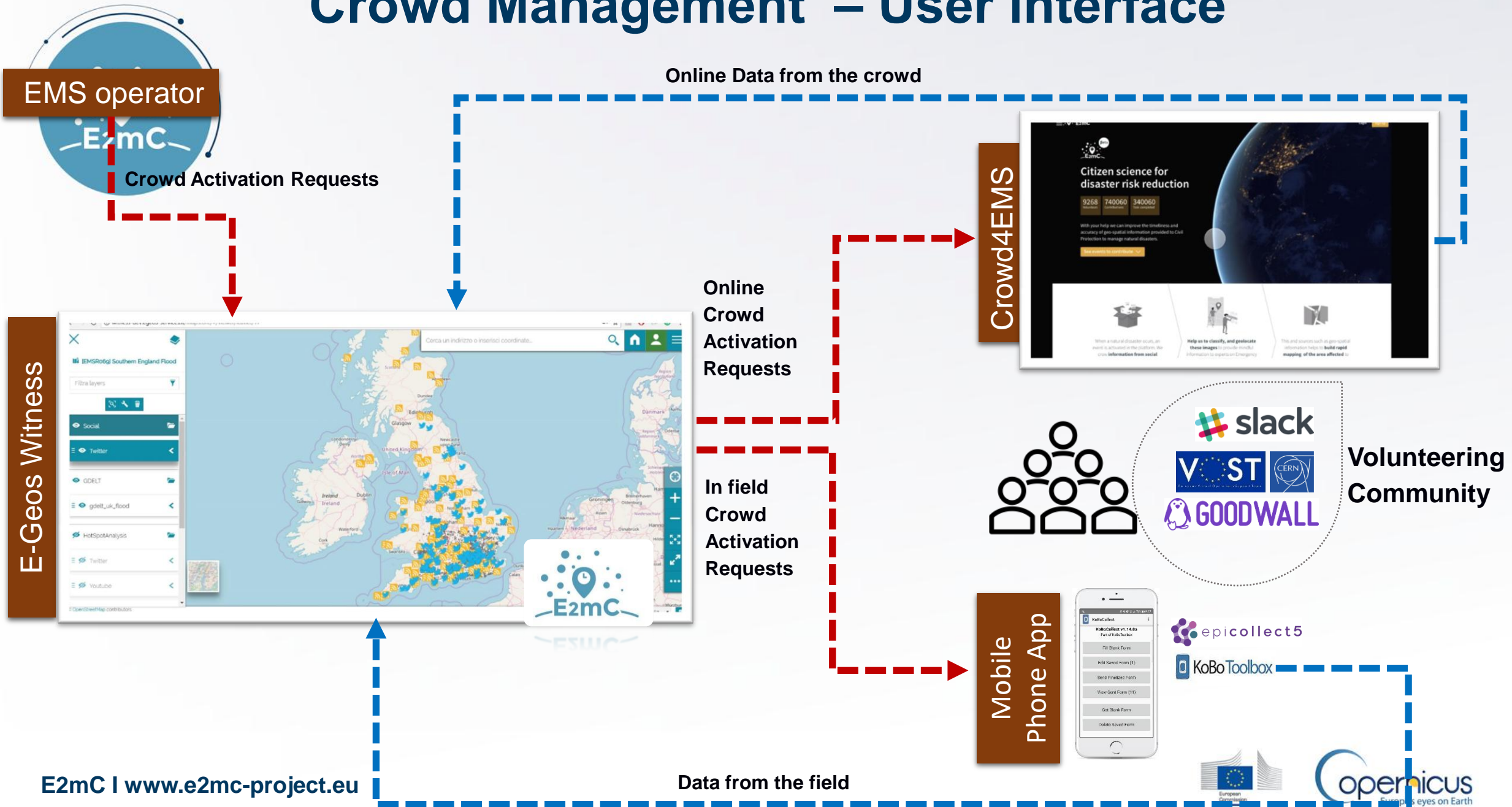
There are no features for the following layers: EMSR06g\_crisis\_information\_poly\_view



# ONGOING WORK

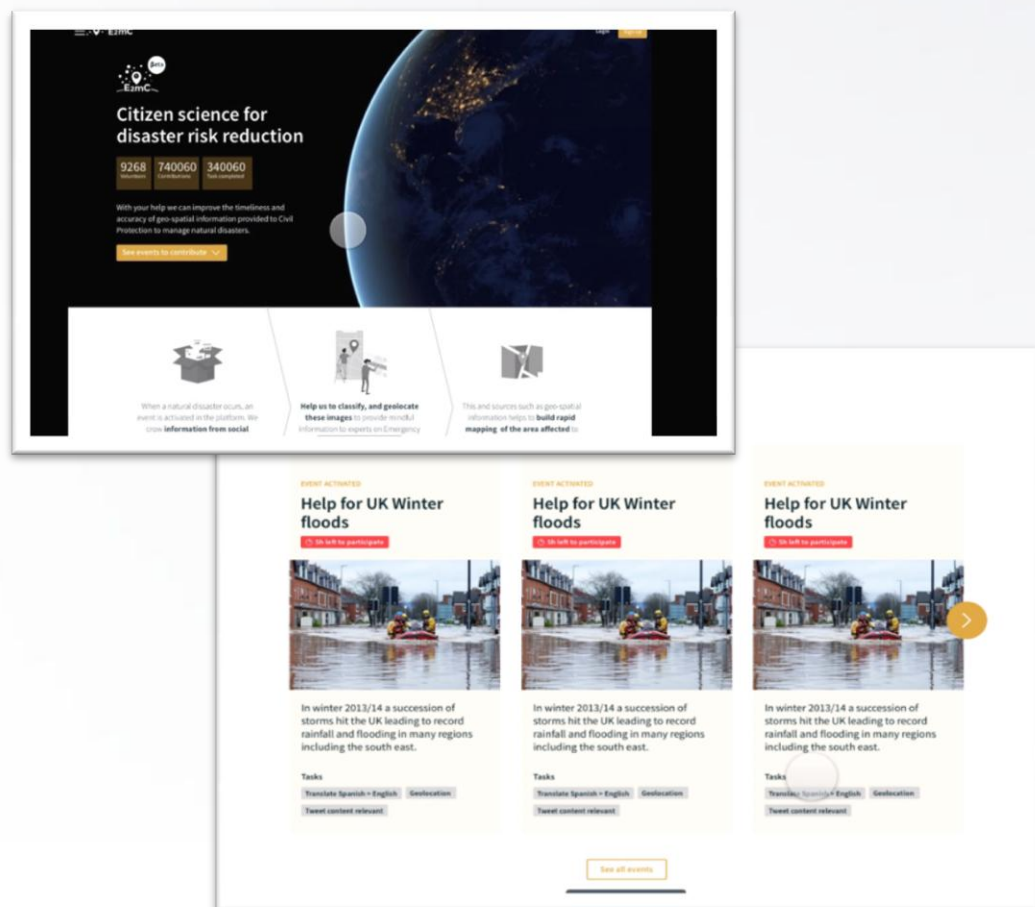


# Crowd Management – User interface



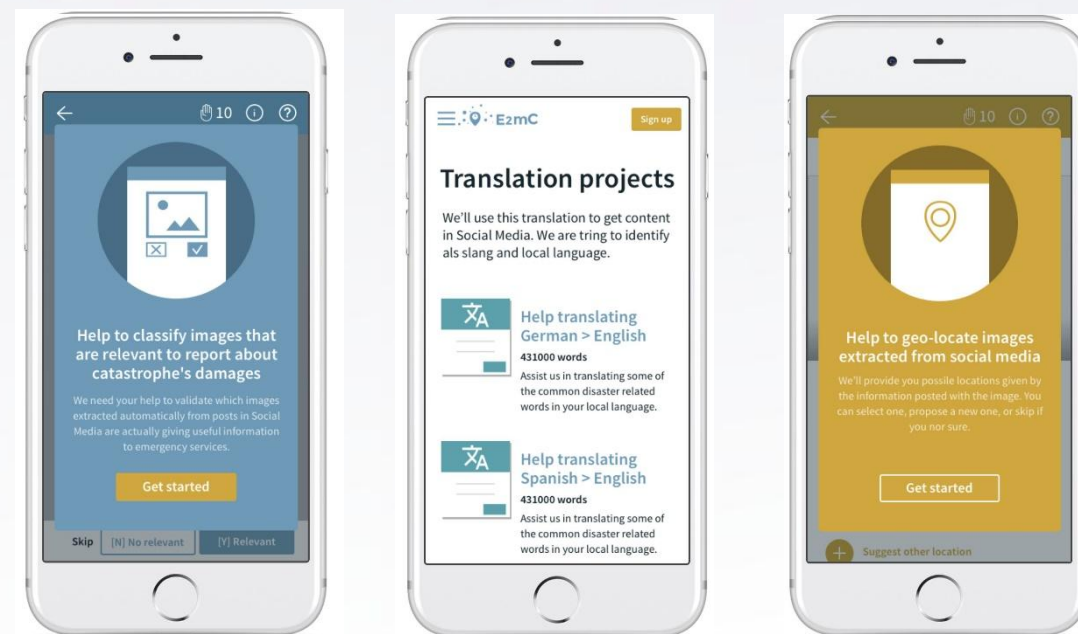


# User Interface (Crowd4EMS)



[www.e2mc.pybossa.com](http://www.e2mc.pybossa.com)

the crowd4EMS platform for activating *crowd-sourcing*

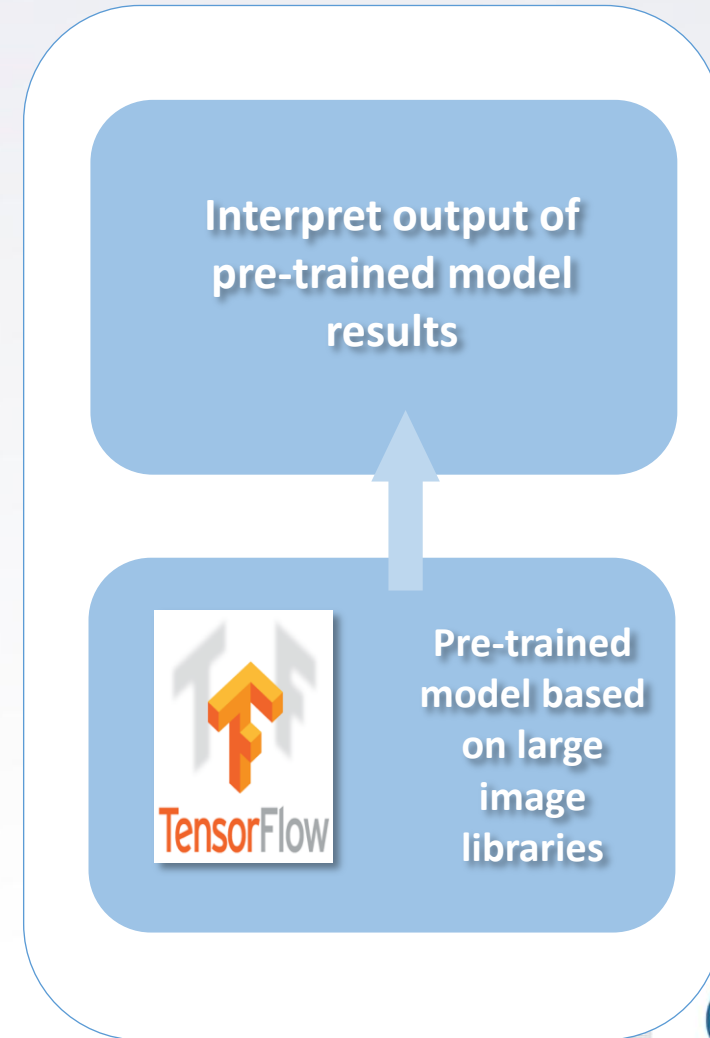
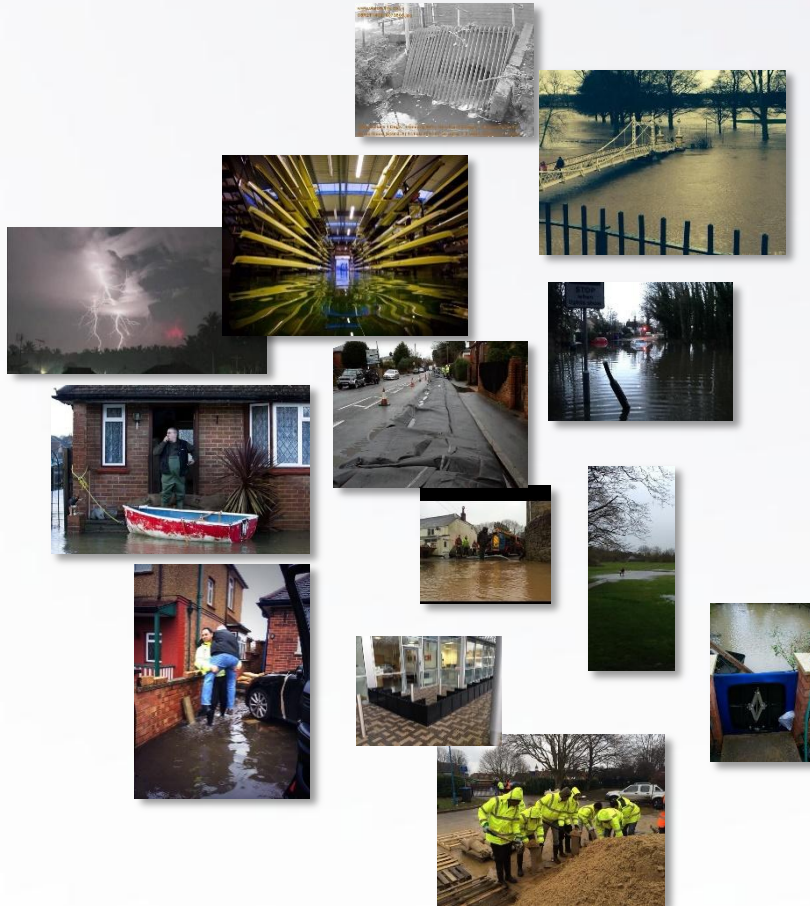


tasks that are available for the volunteers to contribute



# Recognition process

## Interpreting deep learning (ImageNet) classes

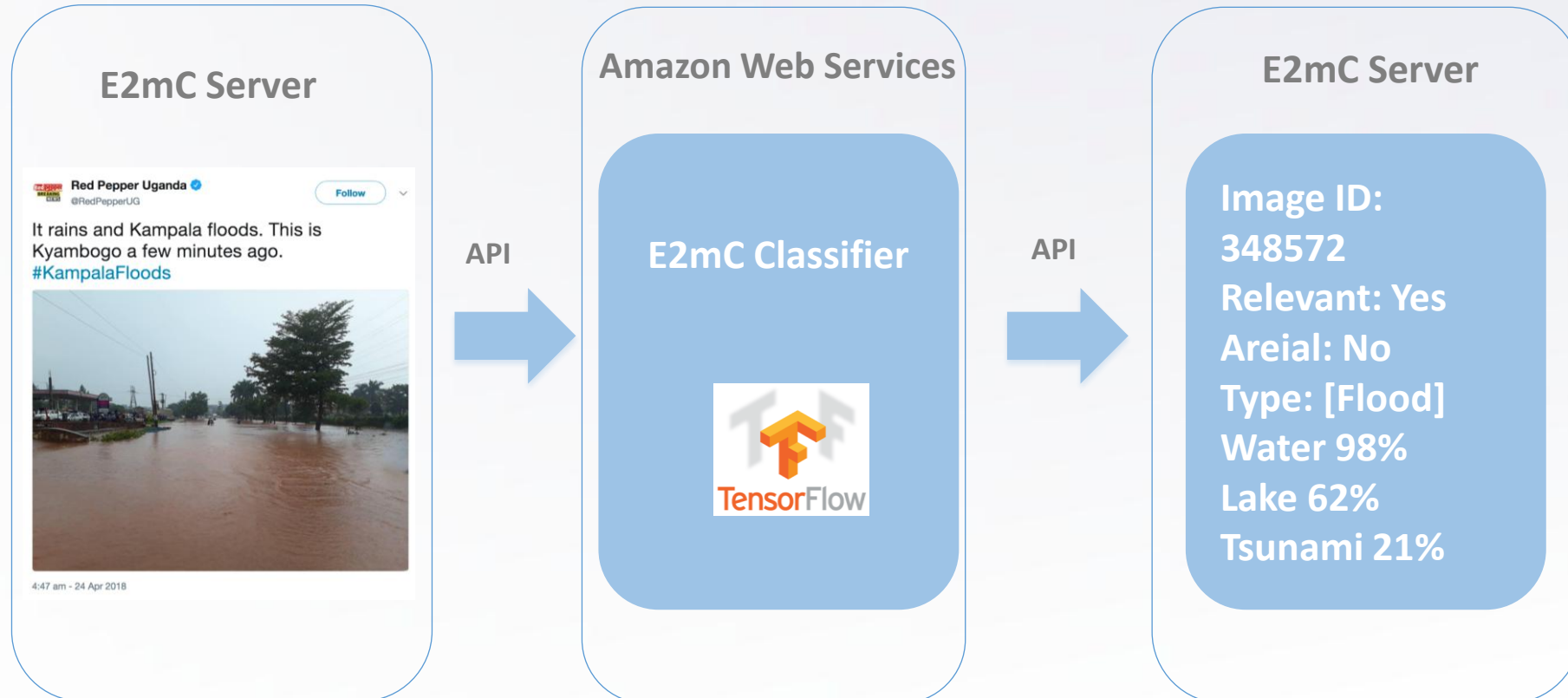






# Recognition process

## Image classification



**NB: Videos are handled in the same way. The video is split into individual sequences. The sequences are interpreted like photos**



# Cooperation on geolocating aerial images with street view

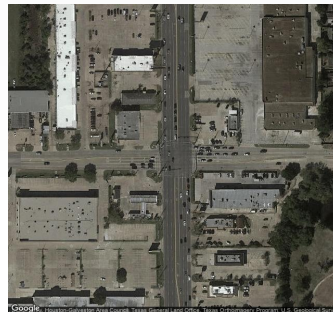
## Observations:

- Aerial images can be more easily geolocated as details (image POIs) and outlines remain the same before and after an event (for example, flooded roads change color, but they do not change shape).
- Aerial images cover a broader area, so their geolocation with street view is computationally feasible even if they are not precisely geolocated by CIME algorithm or native geocoordinates.

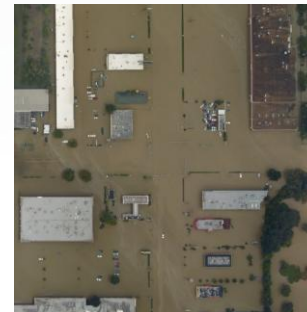
## Goal:

- conduct exploratory research to verify whether and to what extent:
  - Terranea can classify aerial images correctly
  - CIME can provide a geolocation that is "good enough"
  - CIME geolocation can be effectively improved through image comparison with street view images

*Area before flood*



*Area after flood*





# Geocrawler, Topic Extraction, Spatial Hotspots

- Integration of **more social media channels**  
(Facebook, Instagram, Foursquare, ...)
- **Intelligent pre-processing** algorithms for improved analysis results
- Improved **topic extraction** (semi-supervised machine learning)
- **Accuracy assessment**  
(statistical validation for additional use cases and data sources)



# Future work



- Multilanguage analysis
- Image recognition
- Ranking
- Integration of new functionalities
- Validation
- Experimentation on new cases



# Publications

- Clemens Havas, Bernd Resch, Chiara Francalanci, Barbara Pernici, Gabriele Scalia, Jose Luis Fernandez-Marquez, Tim Van Achte, Gunter Zeug, Rosy Mondardini, Domenico Grandoni, Birgit Kirsch, Milan Kalas, Valerio Lorini, Stefan Rüping, E2mC: Improving Emergency Management Service Practice through Social Media and Crowdsourcing Analysis in Near Real Time, Sensors, Dec. 2017
- Jose Luis Fernandez Marquez, Chiara Francalanci, Sharada Mohanty, Rosy Mondardini, Barbara Pernici, Gabriele Scalia, E2mC: Improving Rapid Mapping with Social Network Information, itAIS'17, Springer, 2018
- Barbara Pernici, Chiara Francalanci, Gabriele Scalia, Exploratory spatio-temporal queries in evolving information, MATES VLDB Workshop, Munich, Sept. 2017, Springer 2018
- D. Grandoni, L. De Vendictis, C. Francalanci, B. Pernici, G. Scalia, J.L. Fernandez, R. Mondardini, “The E2mC Project: An Innovative Approach to Combine Social Media and Crowdsourcing for Rapid Mapping”, “Citizen Science” session, Earth Observation Open Science 2017 Conference, ESRIN, Frascati, 25-28 September 2017
- Chiara Francalanci, Paolo Guglielmino, Matteo Montalcini, Gabriele Scalia, Barbara Pernici: IMEXT: A method and system to extract geolocated images from Tweets – Analysis of a case study. RCIS 2017, Brighton, UK, 382-390, IEEE.



- Barbara Pernici, Chiara Francalanci, Gabriele Scalia, Marco Corsi, Domenico Grandoni and Mariano A. Biscardi, Geolocating social media posts for emergency mapping, demo paper, The 5th International Workshop on Social Web for Disaster Management (SWDM'18), Los Angeles, Feb. 2018
- Chiara Francalanci, Barbara Pernici, Gabriele Scalia, Gunter Zeug, Talking about places: Considering context in geolocation of images extracted from tweets, Short paper, GI-Forum, Salzburg, July 2018





## Partners

**e-geos**  
AN ASI / TELESPAZIO COMPANY

**UNIVERSITÄT  
SALZBURG**

**terranea**

**Université**  
**de Strasbourg**

  
**POLITECNICO  
MILANO 1863**

**PSC Europe**  
Public Safety Communications Europe

  
**SIRS**

  
**Fraunhofer  
FOKUS**



CAMPUS VESTA



**GAFAG**

  
**UNIVERSITÉ  
DE GENÈVE**



# Thank you

[WWW.E2mC-PROJECT.EU](http://WWW.E2mC-PROJECT.EU)



*The E2mC project is co-funded by the European Commission /  
H2020 Programme Grant Agreement No.730082".*





# System components

- The E2mC visualizer is a web application
- Geographic web server GeoServer 2.9.4, using a WMS OGC standard service for layering.
- Database: PostgreSQL 9.6 with the spatial extension PostGIS
- Google Cloud Vision
- Gazetteer: OpenStreetMap with Nominatim
- NER: Stanford CoreNLP / Rosette





## ADDITIONAL SLIDES

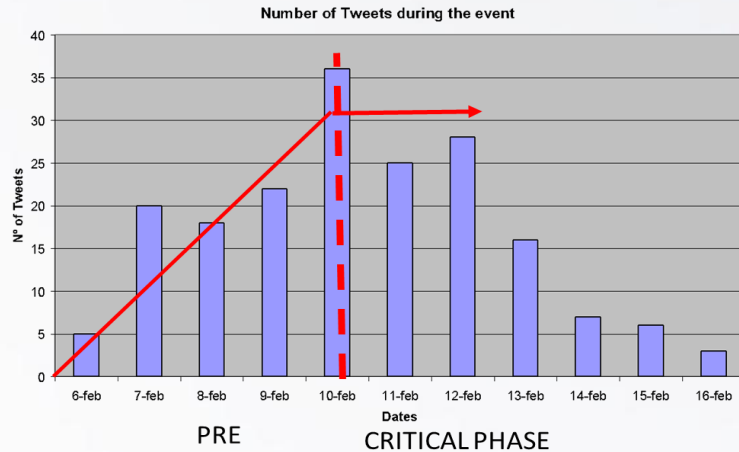


# Geolocation and ranking component sample results on UK flood 2014 (Feb. 10-15, 2018)

- 108,757 tweets (3,333 georeferenced tweets)
- **310** georeferenced images
- **695** geolocated images (local + global)
  - Further improved to **806** considering also links from Tweets to other social media
  - Additional 1016 images from Flickr, 21 from Instagram (but steep growth rate after 2014), 39 videos from Youtube
- The geolocated images from Twitter have been manually analyzed to evaluate them
  - Considering only the mapped areas of Copernicus, **79%** are relevant (55% relevant for not mapped areas)
  - In comparison, 66% of georeferenced images are relevant (36% relevant for not mapped areas)
- There exists a correlation among images in mapped areas and relevance



# E<sup>2</sup>mC State of the art analysis Flood example (UK 2014)



## Key findings

**Social Media sources.** Among the more diffuse Social Media (e.g. Twitter, Facebook, Instagram, YouTube, Pinterest, etc.) Twitter has resulted the most relevant channel to achieve information, not only as primary source provided by the users, but also as indirect way to access to other social data content,

**Type of information: text, photo, video.** Data containing panoramic videos or photo of the affected areas are more relevant with respect to message containing only textual information,

**User: private, public, institutional.** The most relevant information about the crisis in terms of infrastructures damages, flood areas, etc. are provided in largest part from public entities or institutions,

**Geolocation.** Less than 3% of analysed Tweets have been geotagged and, in most cases, the position of the Tweet was located outside the crisis event area.

**Information redundancy.** Information redundancy is another factor to get and filter, in a rapid way, only those are reliable and relevant for the event.





# E<sup>2</sup>mC – High level functions

