The Challenge of Collecting and Analyzing Information From Citizens and Social Media in Emergencies: the Crowd4SDG Experience and Tools

Barbara Pernici, Jose Luis Fernandez-Marquez, Carlo Bono, Oguz Mulayin

RCIS 2022, Barcelona   May 19, 2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944

About us

Barbara Pernici
Professor at Politecnico di Milano
Head of the information
systems group
barbara.pernici@polimi.it

Jose Luis Fernandez-Marquez
Senior Lecturer at University of Geneva
Head of the SDG Accelerator
Crowd4SDG Tech. Coordinator
Joseluis.Fernandez@unige.ch

Carlo Bono (PhD Candidate at Politecnico di Milano)
Oguz Mulayim (Postdoc Researcher at IIIA-CSIC)

Crowd4SDG

Outline

- Round table introduction
- Social media and crowdsourcing for disaster response
  - Challenges + potential role of AI for enhancing the tools
  - Tools and role of AI
- Case study Covid, flood Nepal, flood and gender
- Hands-on Part
  - Visual Cit
  - CSProject builder.
- Discussion

UNIVERSITÉ DE GENÈVE · CERN · CSIC Consejo Superior de Investigaciones Científicas · POLITECNICO MILANO 1863 · unitar United Nations Institute for Training and Research · Université de Paris

Crowd4SDG

Social media and crowdsourcing for disaster response

Jose Luis Fernandez-Marquez

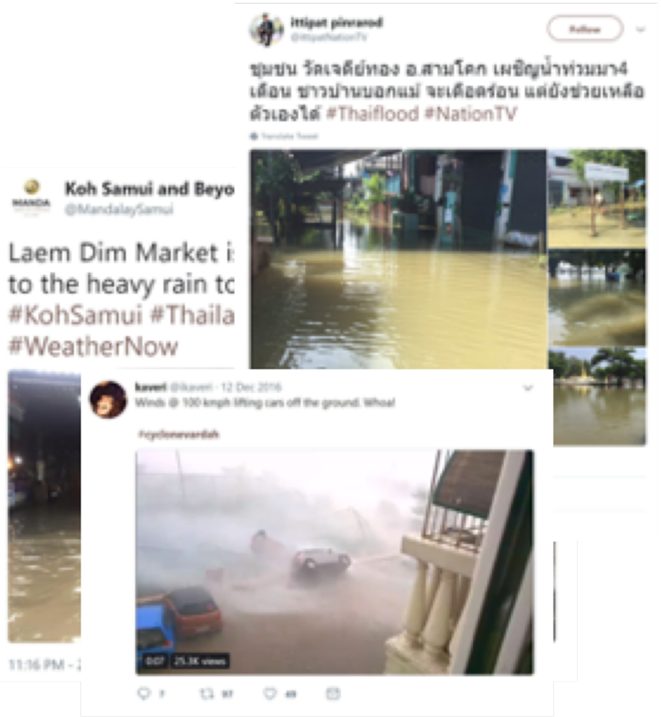
RCIS 2022, Barcelona May 19, 2022

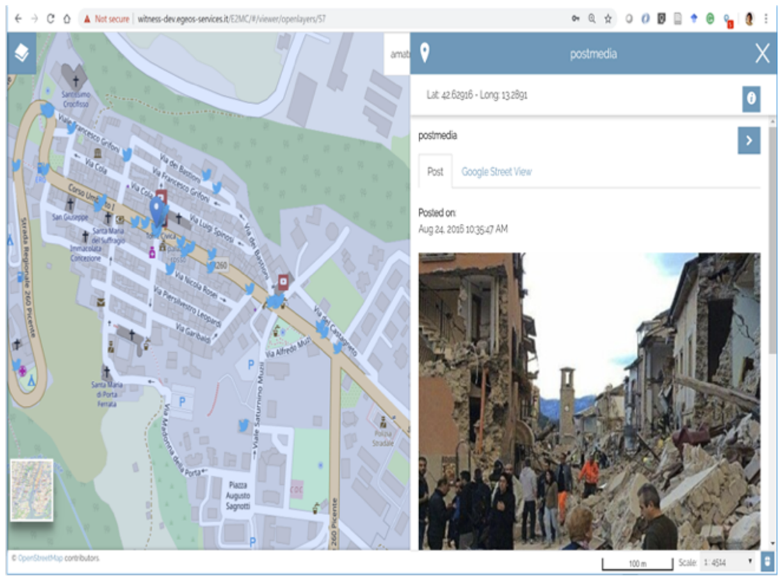
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944

2005
announcement of
Pope Benedict XVI

2013 announcement of
Pope Francis

E2mC: Evolution of Emergency Copernicus services





Crowd4SDG

Crowd4SDG

E2mC: Evolution of Emergency Copernicus services

“Every year natural disasters kill around 90 thousands people and affect close to 160 million people worldwide”
Environmental health in emergencies. WHO

Hurracain Katrina, 2005

Haiti Earthquake, 2010

Chennai floods, 2015

E2mC: Evolution of Emergency Copernicus services

"The first 72 hours after a disaster are crucial; response must begin during that time to save lives" OCHA

Situational Awareness

What is happening and where?

Hurricane Katrina – Satellite Imagine Corporation

Crowd4SDG

E2mC: Evolution of Emergency Copernicus services

Hurricane Michal – Mexico Beach ESRI

Mexico Beach, Florida, last year after Hurricane Michael made landfall.Douglas R. Clifford / Tampa Bay Times via AP

E2mC: Evolution of Emergency Copernicus services

Crowd4SDG

Main Challenges

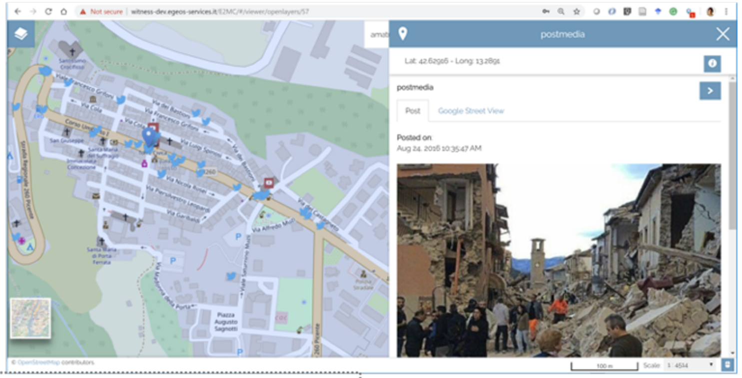
- Process of **large volume** of information
- Assessing the **relevance** of the content and the classification
- **Geolocation** of the content gathered from social media
- **Data quality** and reliability.
- **Timeliness**

Crowd4SDG

Social & crowd platform

Social Media

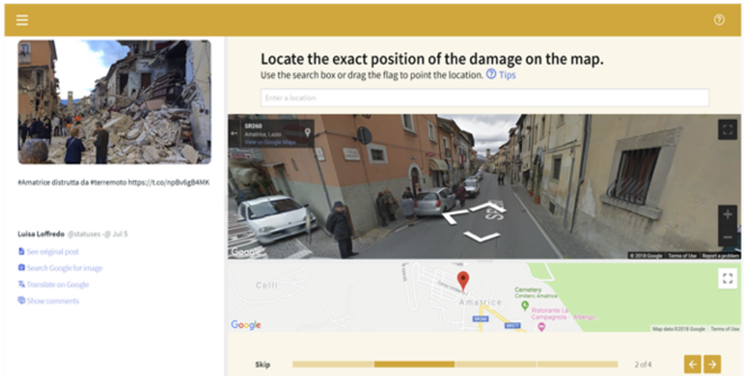
Witness Component



Crowd4EMS

Relevance

Geolocation



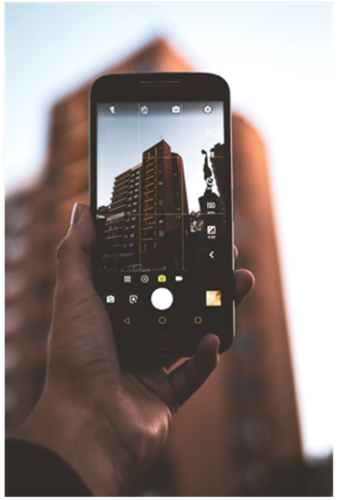
Crowd4SDG

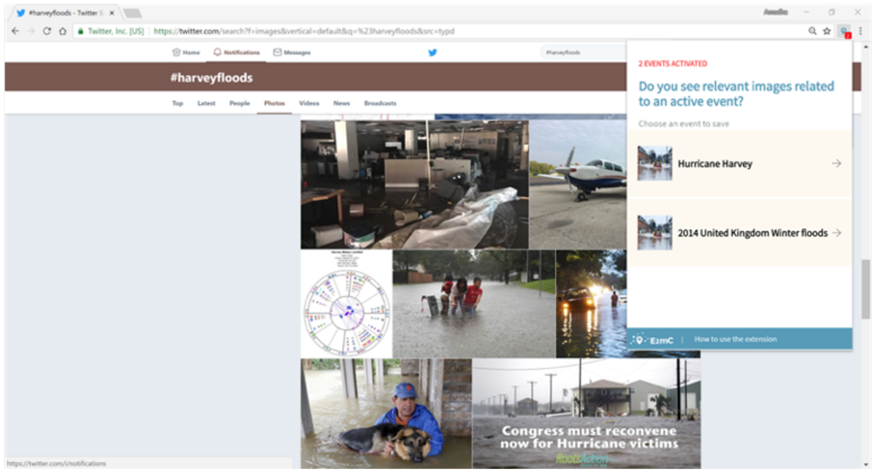
Data Gathering

Crowdsourcing data collection

KoBo toolbox

CROWD4EMS Chrome Extension



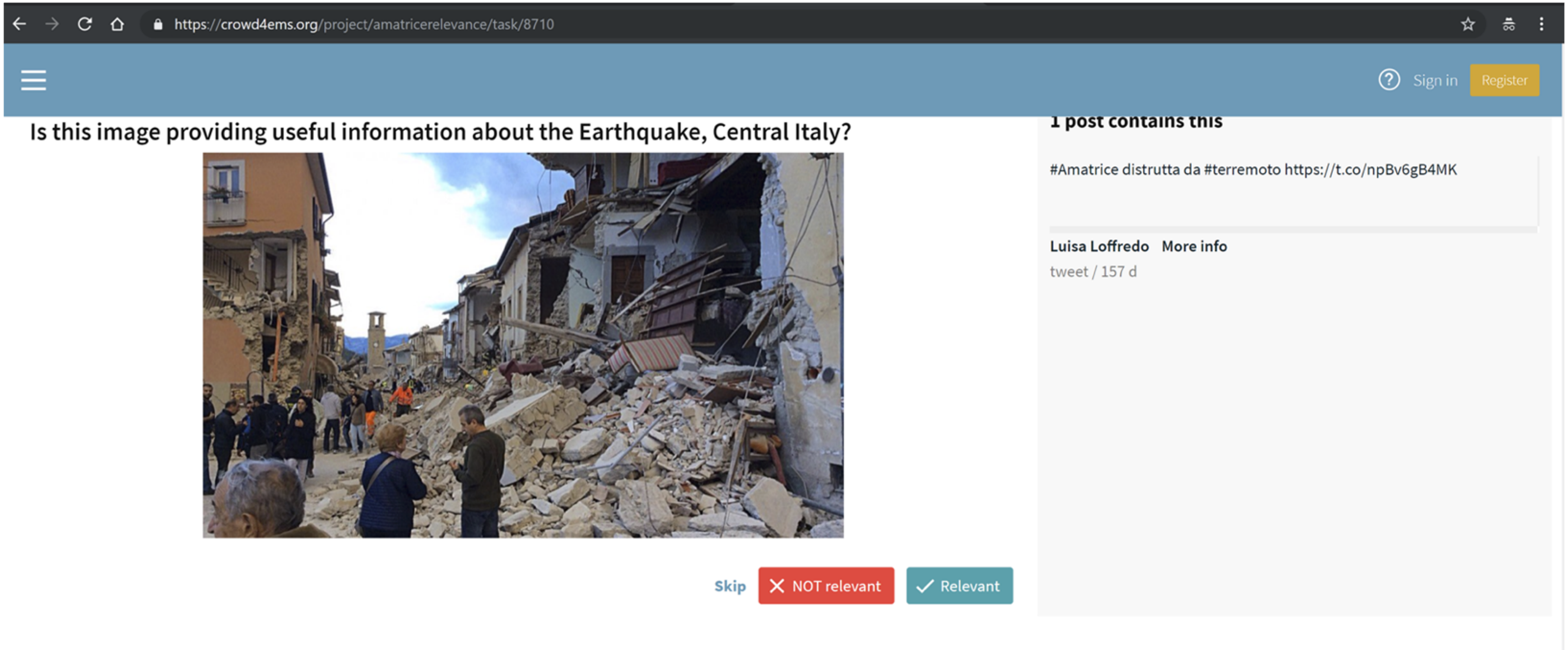


Social Media Data

Social Media

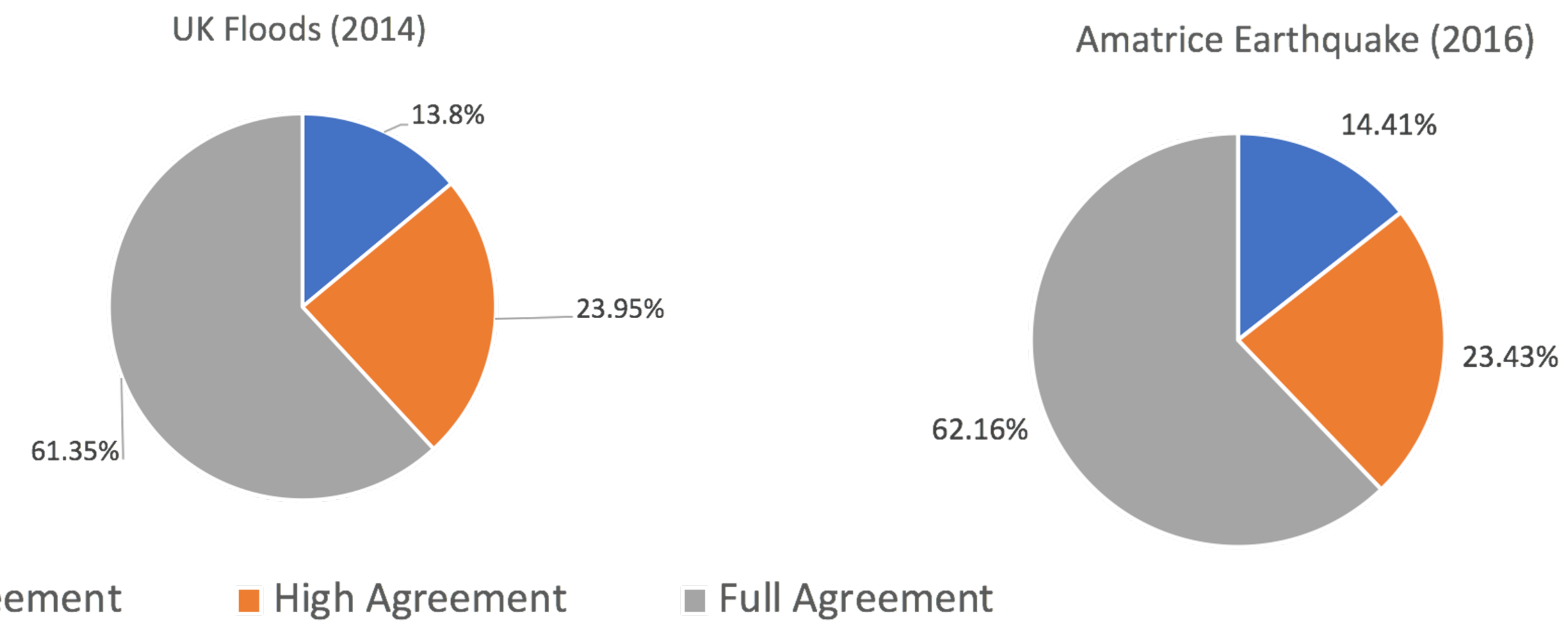
Crowd4SDG

Data Analysis - Relevance



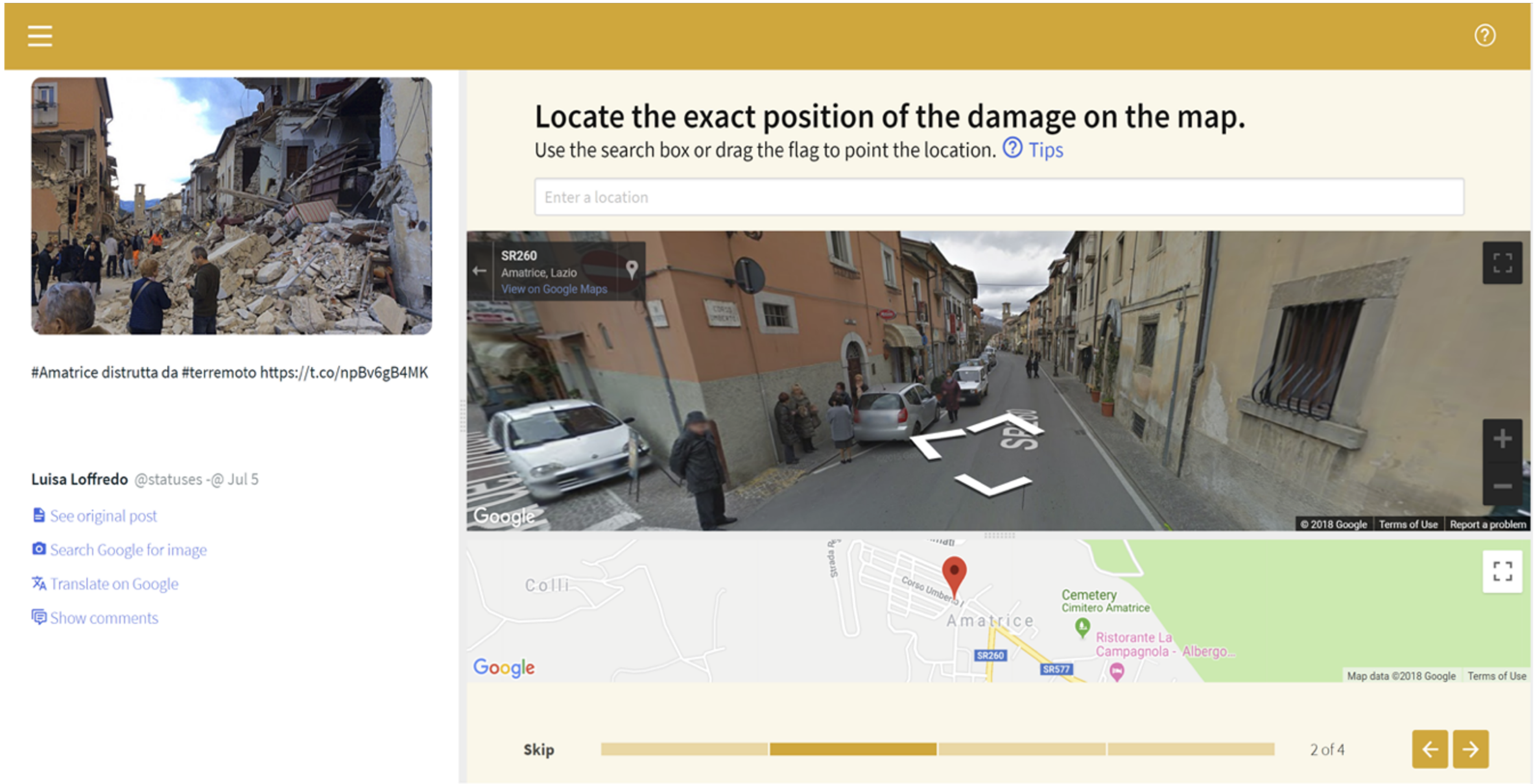
Crowd4SDG

Data Analysis - Relevance



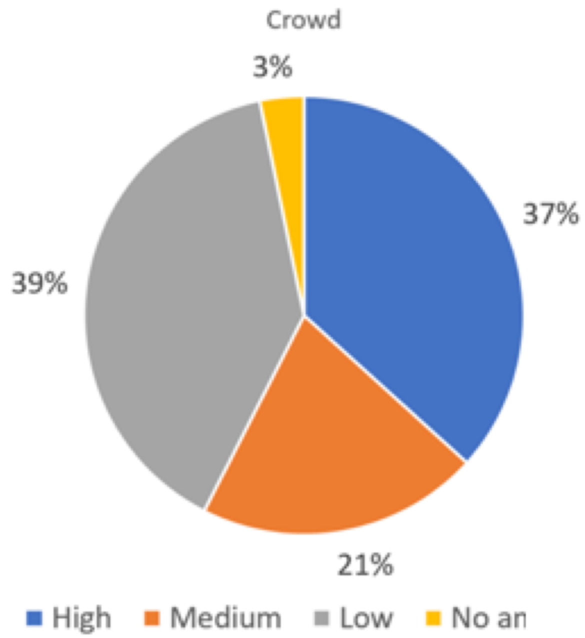
Crowd4SDG

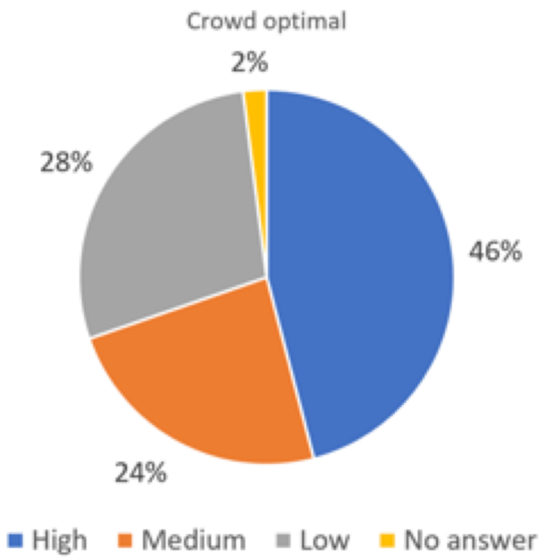
Data Analysis - Geolocation

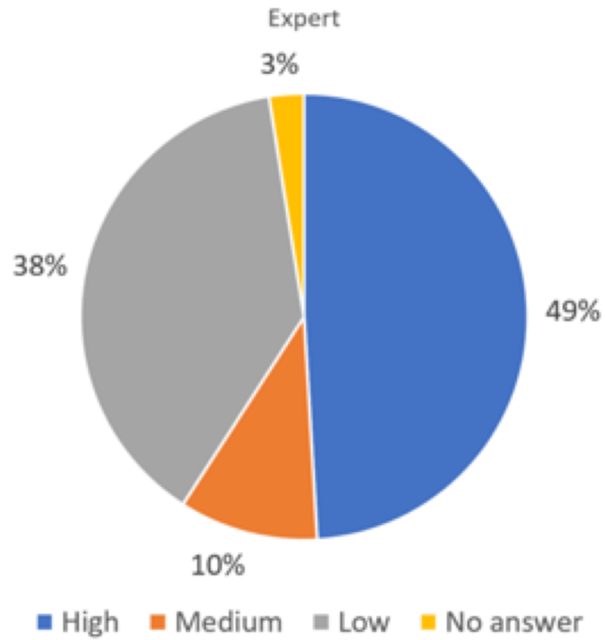


Crowd4**SDG**

Data Analysis - Geolocation

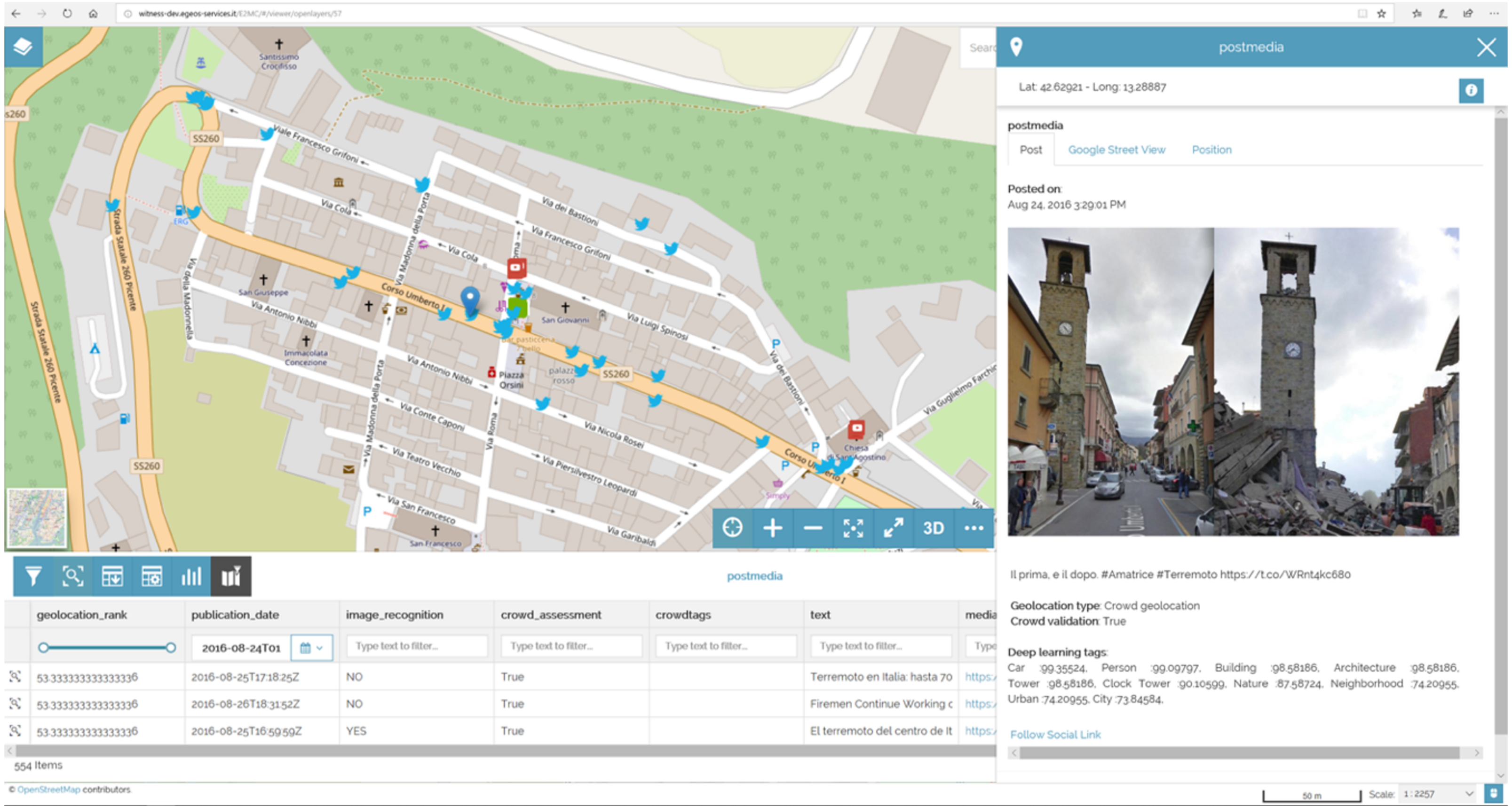






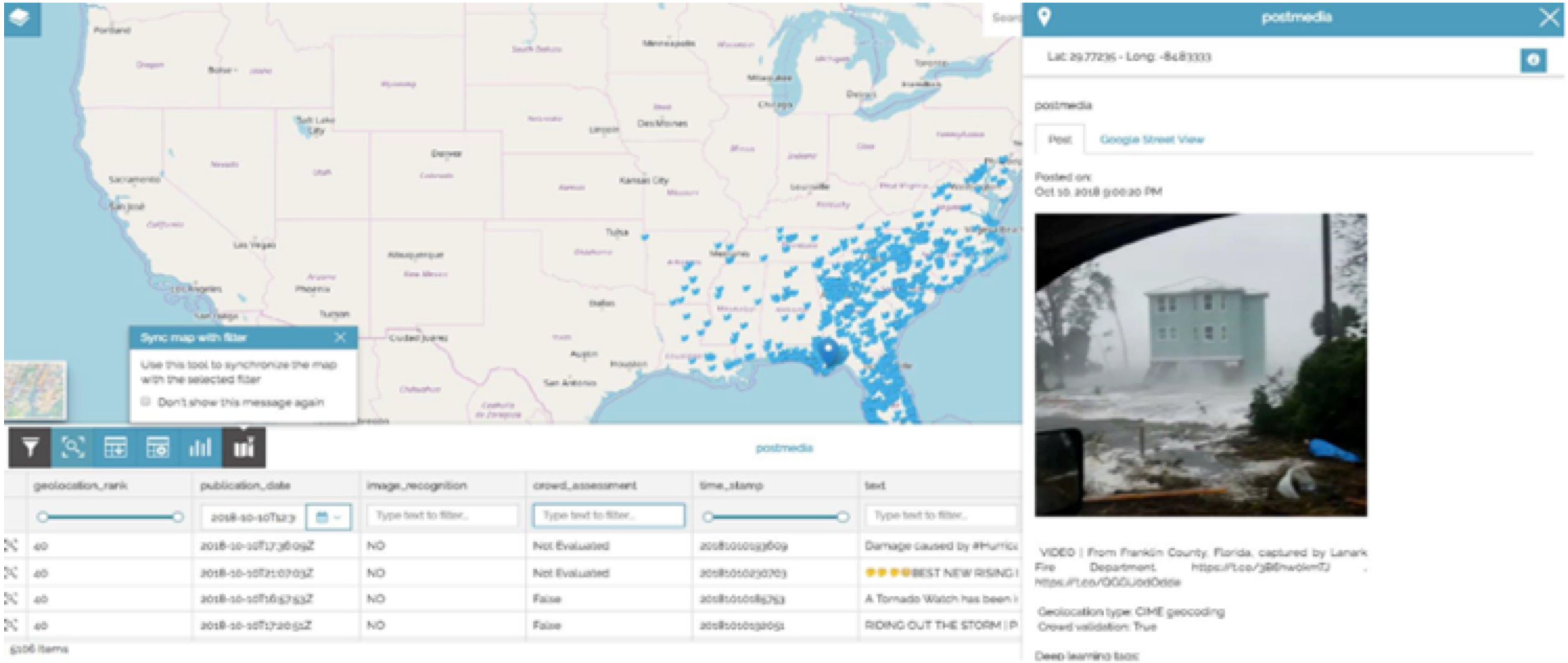
Crowd4SDG

Data Visualization - Witness Component



Crowd4SDG

Data Visualization - Witness Component



Crowd4SDG

Lessons learnt

- Need to improve AI filters to reduce the volume of information before sending that to the crowd
- Geolocation task is still complicated task. It is not s micro task!
- Need to improve the aggregation function to reduce the redundancy needed
- Availability of the crowd
- Engage people beyond data collection and data analysis.

Crowd4SDG

https://Crowd4SDG.eu

Citizen Science Solution Kit

Tool 1: CS Project Builder - Data Analysis including CrowdAnalysis component

Tool 2: CS Logger - Data Collection

Tool 3: Visual Cit - Social media data analyzer

Tool 4: Decidim4CS - Open Governance

https://crowd4sdg.eu/about-2/tools/

Thank you!

www.crowd4sdg.eu

Consensus models

	task_id	user_id	is_fast
0	Audi	John	Yes
1	Audi	Matthew	Yes
2	Audi	Andrew	Yes
3	BMW	John	Yes
4	BMW	Matthew	Yes
5	BMW	Matthew	Yes
6	Seat	John	No
7	Seat	Andrew	No
8	Seat	Matthew	No
9	Maserati	John	Yes
10	Maserati	Andrew	Yes
11	GM	Andrew	Yes
12	GM	Matthew	No
13	GM	John	No
14	Mazda	Andrew	Yes
15	Mazda	John	Yes
16	Mazda	Matthew	Yes
17	Kia	Matthew	Yes
18	Kia	Andrew	Yes
19	Kia	John	No

Crowdsourced annotations

Majority Voting

	Yes	No
Audi	1.0	0.0
BMW	1.0	0.0
GM	0.0	1.0
Kia	1.0	0.0
Maserati	1.0	0.0
Mazda	1.0	0.0
Seat	0.0	1.0

Probabilistic

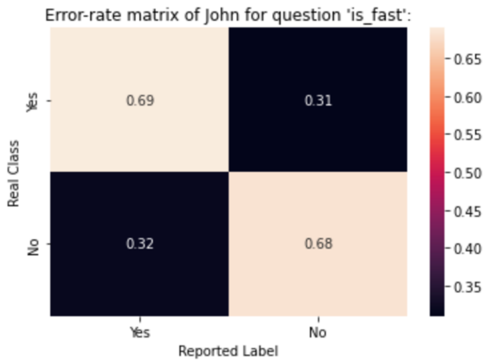
	Yes	No
Audi	1.000000	0.000000
BMW	1.000000	0.000000
GM	0.333333	0.666667
Kia	0.666667	0.333333
Maserati	1.000000	0.000000
Mazda	1.000000	0.000000
Seat	0.000000	1.000000

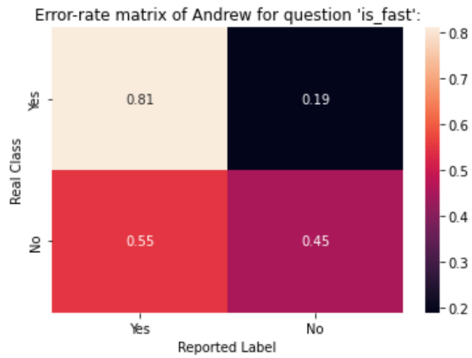
These **do NOT model** individual **annotator** performance.

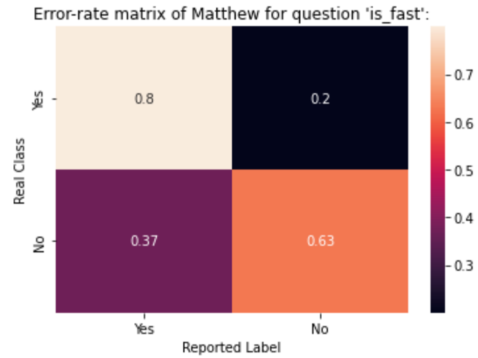
Advanced consensus models - Dawid-Skene

	task_id	user_id	is_fast
0	Audi	John	Yes
1	Audi	Matthew	Yes
2	Audi	Andrew	Yes
3	BMW	John	Yes
4	BMW	Matthew	Yes
5	BMW	Matthew	Yes
6	Seat	John	No
7	Seat	Andrew	No
8	Seat	Matthew	No
9	Maserati	John	Yes
10	Maserati	Andrew	Yes
11	GM	Andrew	Yes
12	GM	Matthew	No
13	GM	John	No
14	Mazda	Andrew	Yes
15	Mazda	John	Yes
16	Mazda	Matthew	Yes
17	Kia	Matthew	Yes
18	Kia	Andrew	Yes
19	Kia	John	No

Annotator error-rates







More reliable consensus is possible thanks to **modelling** the **annotator behavior**.

Consensus

	Yes	No
Audi	0.923616	0.076384
BMW	0.946553	0.053447
GM	0.271625	0.728375
Kia	0.716913	0.283087
Maserati	0.848372	0.151628
Mazda	0.923616	0.076384
Seat	0.096134	0.903866

Dawid, A. P., & Skene, A. M. (1979). Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm. *Applied Statistics*, *28*(1), 20. https://doi.org/10.2307/2346806

Crowd4SDG

Case studies

Barbara Pernici

Politecnico di Milano

May 19, 2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944

Motivation

- Support emergency response with evidence-based content from social media (focus on Twitter)
- Driver: SDG 13, Climate action
  - Towards improving indicator 13.1.1 Number of deaths, missing persons and directly affected persons attributed to disasters per 100,000 population

Crowd4SDG case studies

- COVID-19 behavioral impact: **Face masks (VisualCit)**
- **Flood and gender** SDG13 + SDG5
- **TriggerCit: Timeliness** of the delivery in evaluating the **onset** of flood events (Nepal and Thailand)

Building indicators from social media

Social Media provides a trove of information that, if aggregated and analysed appropriately can provide important **statistical indicators to policy makers**.

From images posted to social media

We investigate whether it is possible to obtain such data by aggregating information from **images posted to social media**

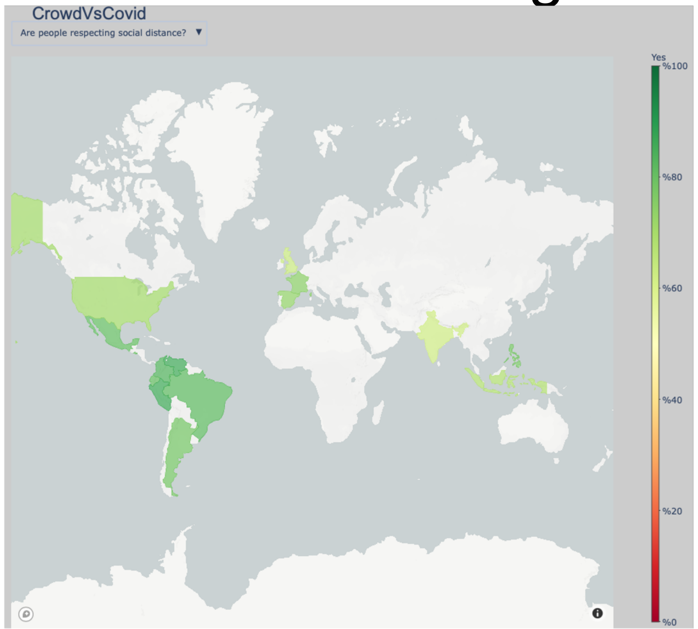
Data collection process on Twitter

1. **Focus** on a specific **goal or situation**
1. Twitter sensing with **language-specific** small word dictionaries
1. Selection of **potentially relevant** image contents with VisualCit
1. **Text analysis** of outputs for dictionary refinement
1. **Geolocation** of posts with CIME
1. **Crowdsourcing** to extract relevant information
1. **Visualize results**

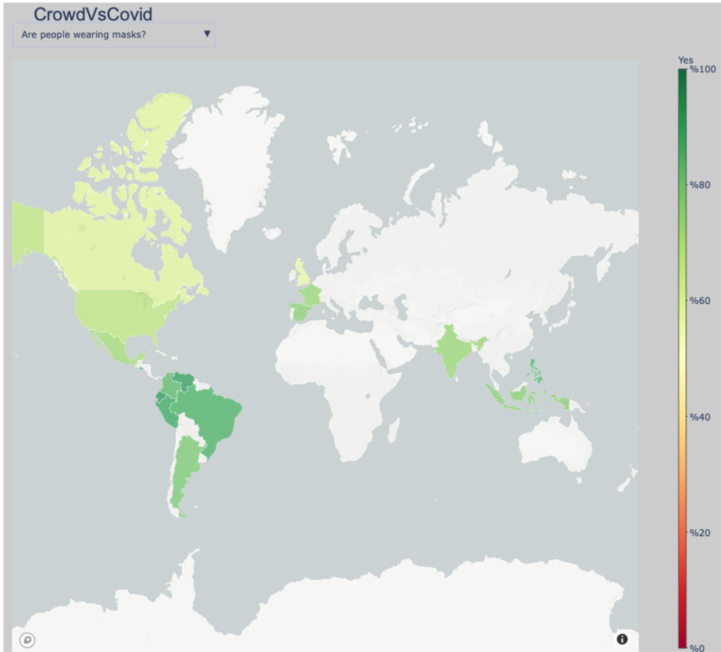
Case studies

COVID-19 - From Tweets to a Face mask dataset and indicators

- Social distancing



- Face mask usage

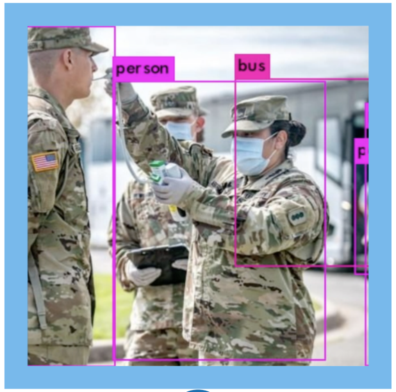


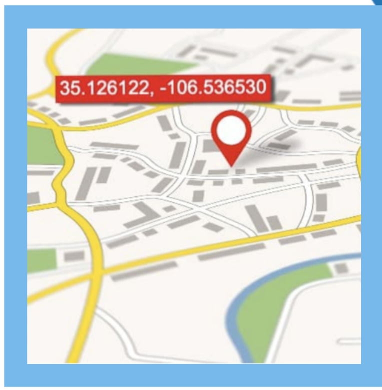
Examples end of August, 2020

V. Negri, et al., Image-based Social Sensing: Combining AI and the Crowd to Mine Policy-Adherence Indicators from Twitter, ICSE, Track Software Engineering in Society, May 2021

Building indicators from social media

- **Collecting images from social media**





Twitter/social crawling

Keyword selection
(with Citizen Scientists)

Duplicate URL & Similar Image Filters

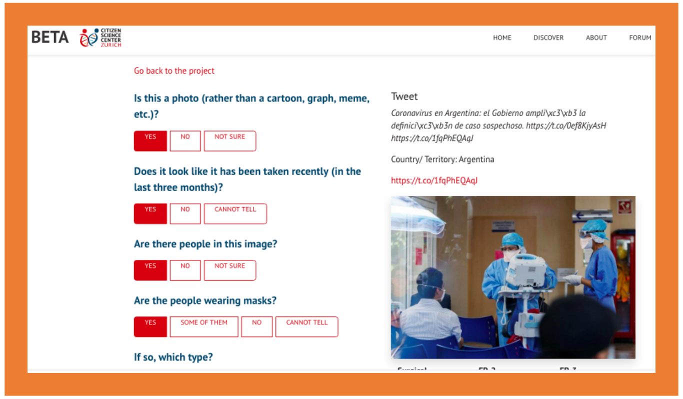
Semantic Image Filters

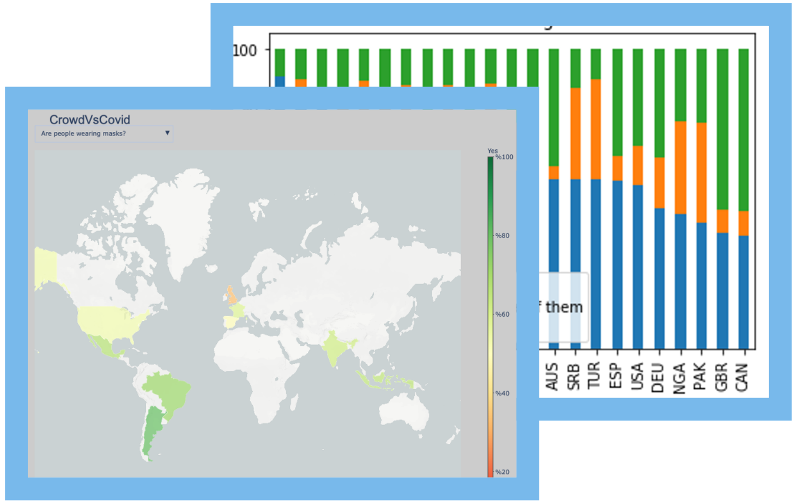
Photo/No photo
Not Safe for Work
Public places
People

Geocoding

Building indicators from social media

- **Deriving indicators**





Crowdsourcing

Project Builder
Asking Citizen Scientists

Aggregate Statistics and ML

Quality evaluation
Building indicators
Building new classifiers

Visualisation

https://pernici.faculty.polimi.it/crowd4sdgpolimi/

Evaluating the impact of floods on gender equality

- Twitter crawling
- Flood classifier (fine-tuning Xception pretrained with ImageNet)

Is this a flood image?

YES NO

Are there people in it?

YES NO

Are there more women than men?

YES NO CAN'T TELL

Do they seem in danger?

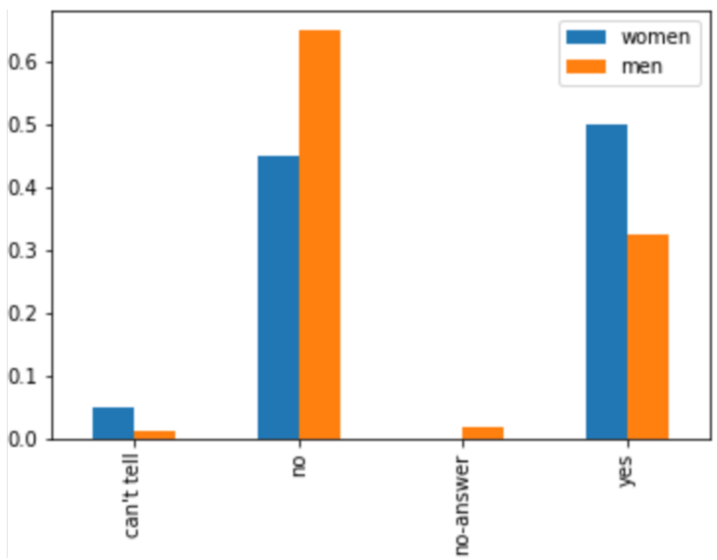
YES NO CAN'T TELL

Are there shelters nearby? (Any place giving

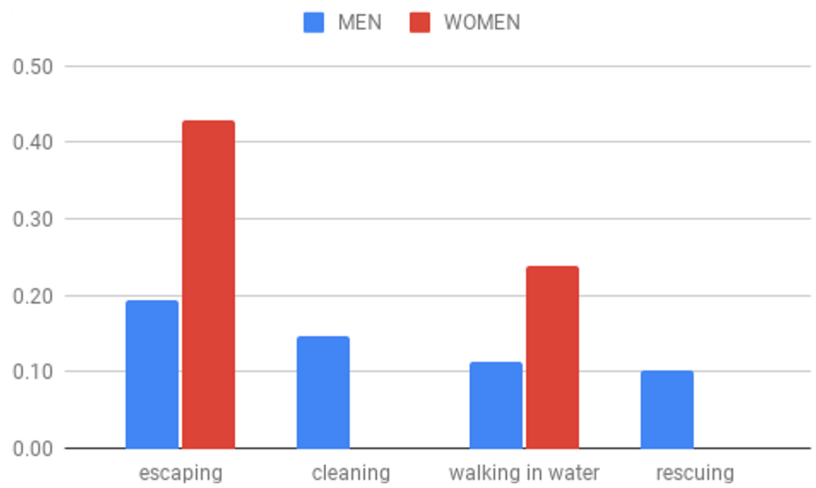
The following link to the tweet might help in answering to the questions :
http://twitter.com/anuuser/status/1391177986897022981

Olimpia Rivera, Juan Felipe Calderon, Paul Planchon, Barbara Pernici, Evaluating the impact of floods on gender equality from social media evidence, 2nd International Research Workshop on Women IS, and Grand Challenges, Dec. 2021

Crowd4**SDG**



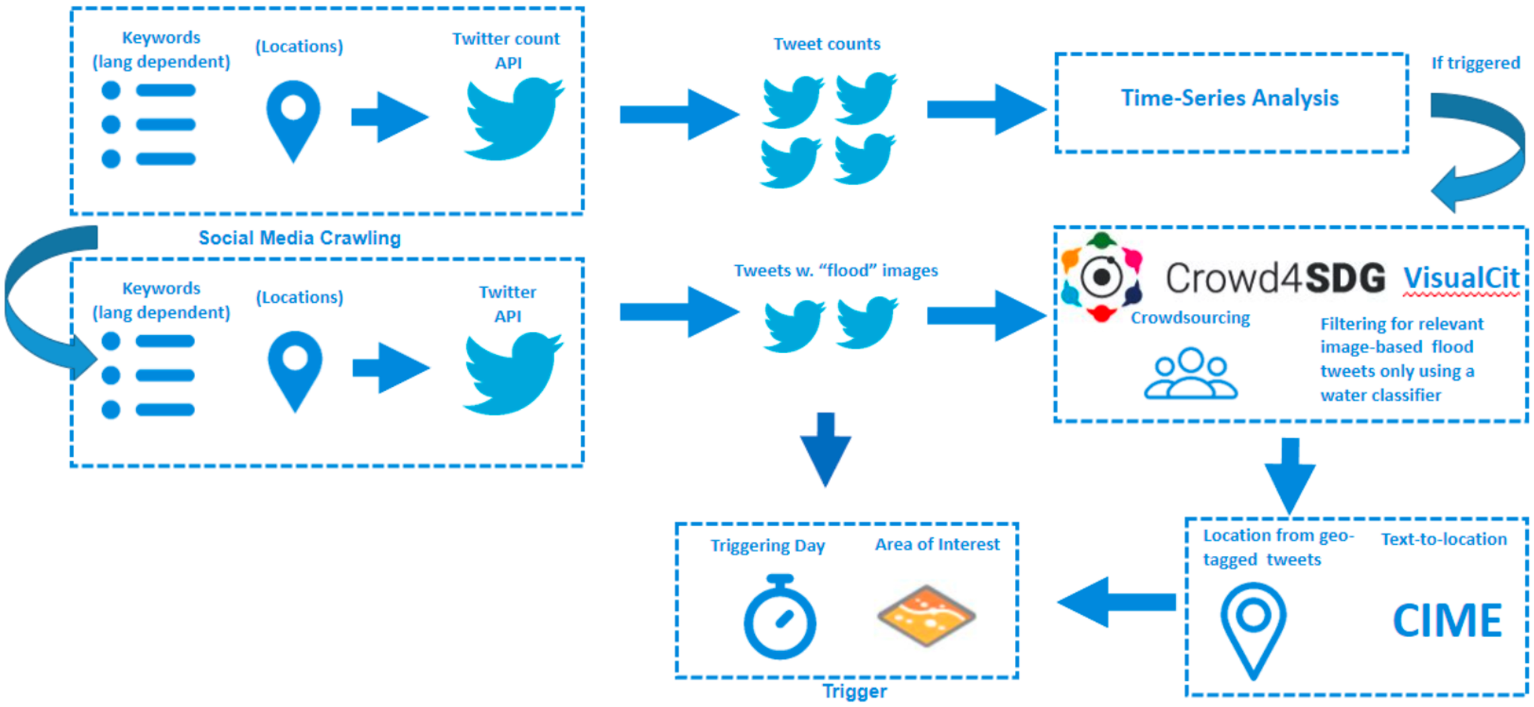
Do they look in danger?



Most popular topics by gender

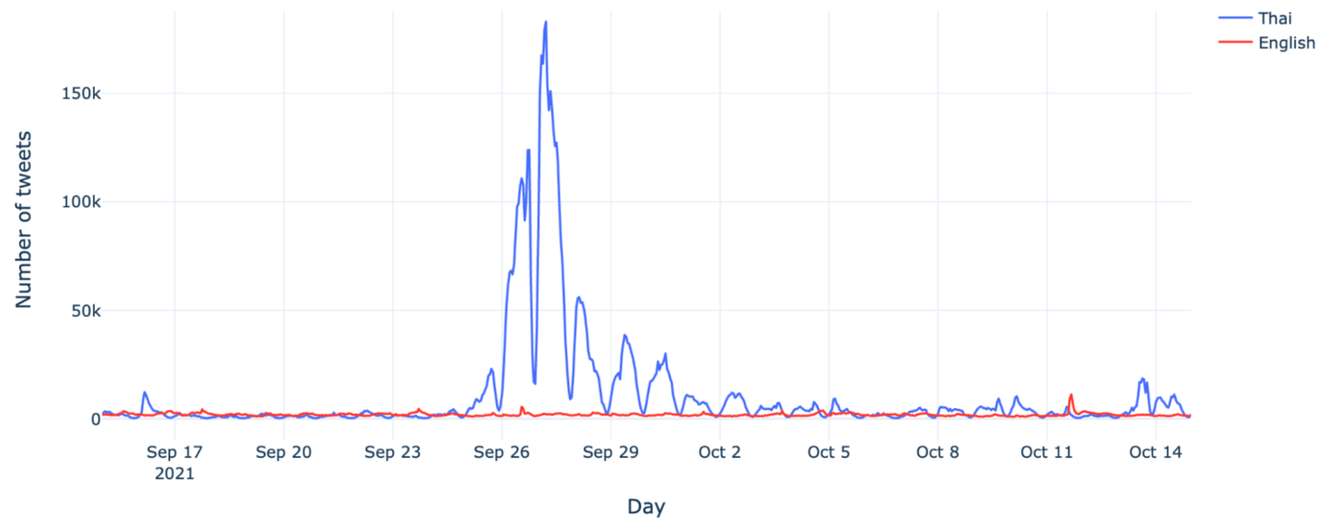
Olimpia Rivera, Juan Felipe Calderon, Paul Planchon, Barbara Pernici, Evaluating the impact of floods on gender equality from social media evidence, 2nd International Research Workshop on Women IS, and Grand Challenges, Dec. 2021

TriggerCit: Main phases



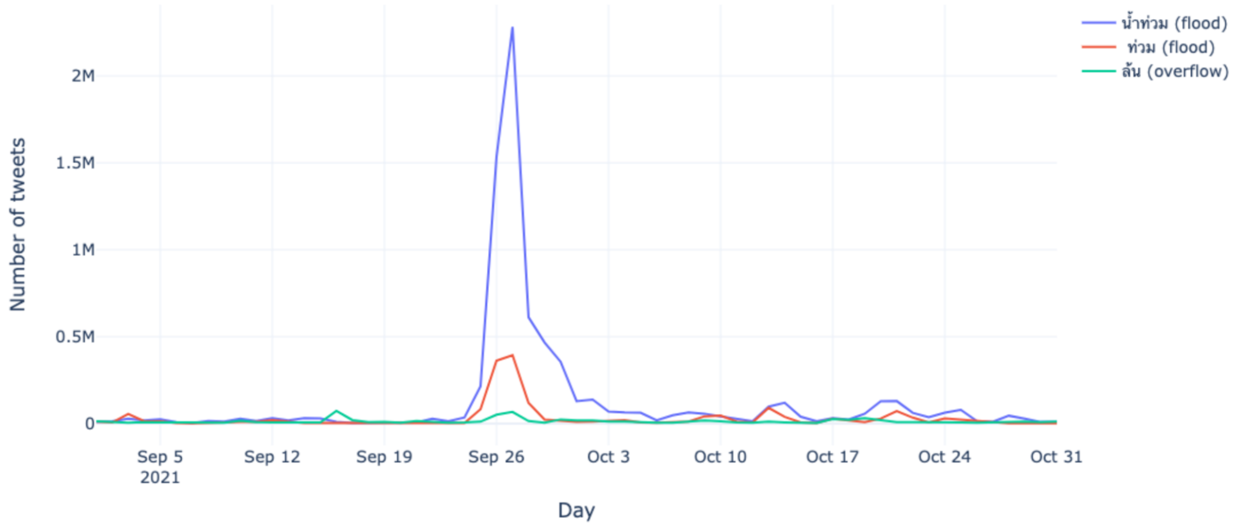
Multilingual context and event detection

We focused on very small, no-cost dictionaries.
Language-based effect of events on social media volumes



Experimental results – Thailand case study

The event onset is clearly identified



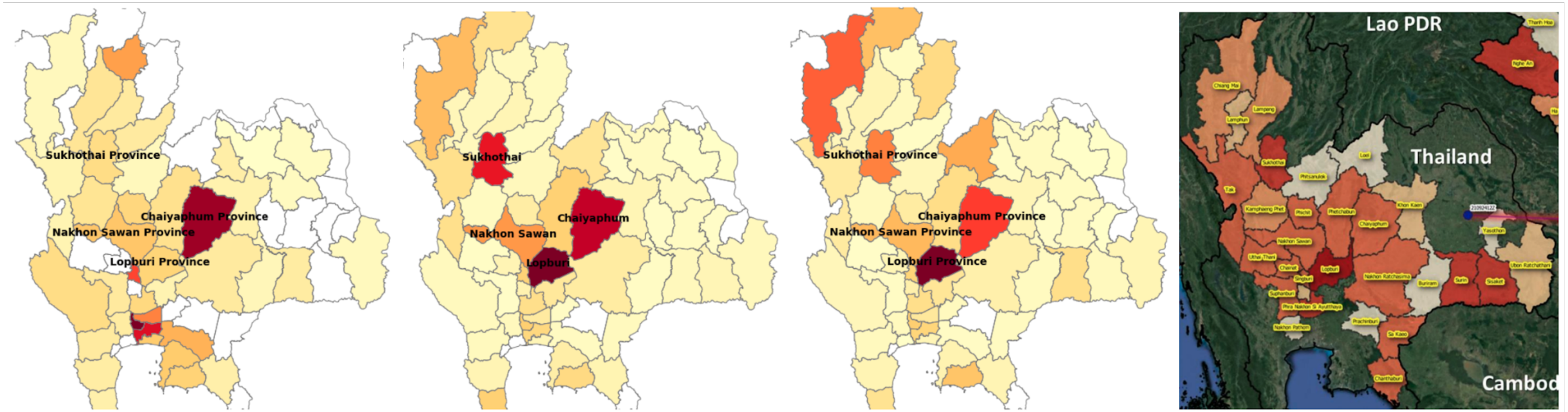
Tweet counts for seed dictionary entries

Data	Count
All tweets, September 26-27	4’145’447
No retweets	66’868
Containing images	6’292
Native Twitter locations	227
Overall images	8’774
Passed VisualCit filters	3’056
Places geolocated by CIME	1’671

Data volumes through processing steps

Crowd4**SDG**

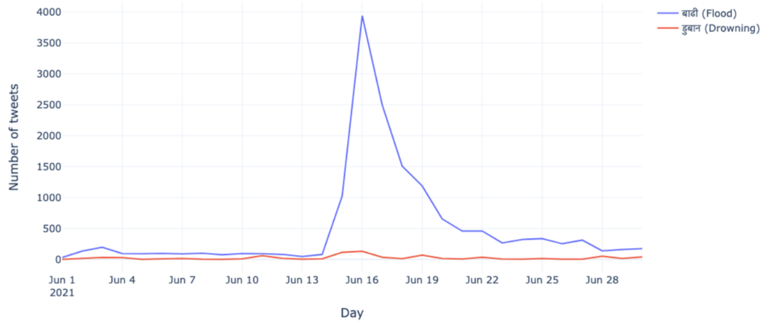
Experimental results – Thailand case study



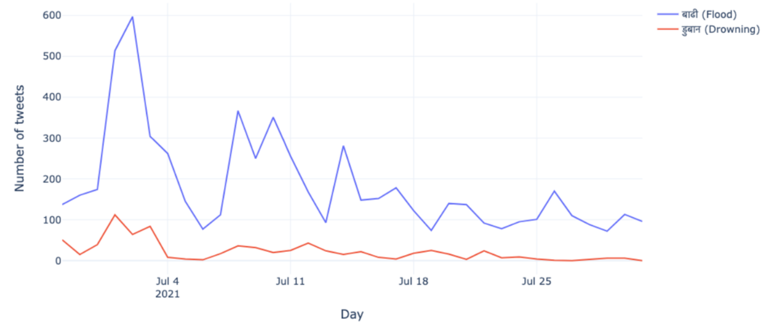
(a) Native Twitter geolocation (b) VisualCit geolocation (c) VisualCit geolocation with extended tweet set (d) Affected people

Geolocations / inhabitants ratio by region (a) Twitter native geolocations, (b) Twitter native + CIME geolocated, (c) Twitter native + CIME geolocated from extended dataset (with images + promising text-only tweets), and
(d) Number of affected persons by region at September, 28th (source: ReliefWeb)

Experimental results – Nepal case study



(a) Nepali entries in June



(b) Nepali entries in July

Tweet counts for seed dictionary entries

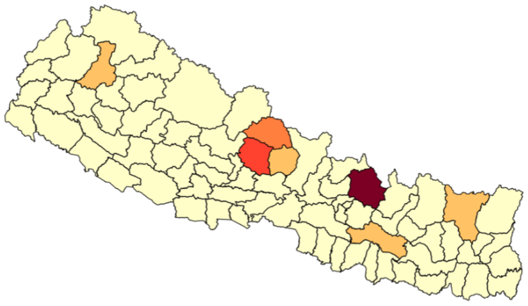
Data	June 16-17	July 1-2
All tweets	6’639	1’225
No retweets	2’807	594
Containing images	261	63
Native Twitter locations	8	10
Overall images	391	80
Passed through VisualCit	218	55
Places geolocated with CIME	51	10

Data volumes through processing steps

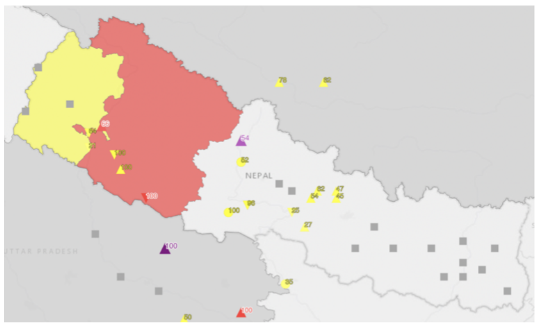
The onset of an announcer sub-event is clearly identified

Experimental results – Nepal case study

Comparing June 16-17 and July 1-2 events with other sources

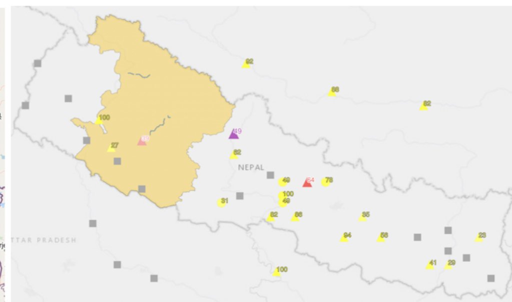


(a) Nepali tweets geolocated with CIME - June 16-17, 2021

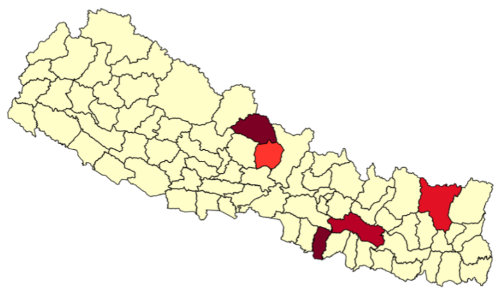


(b) GloFAS reporting points and Rapid Impact Assessment map - June 17, 2021

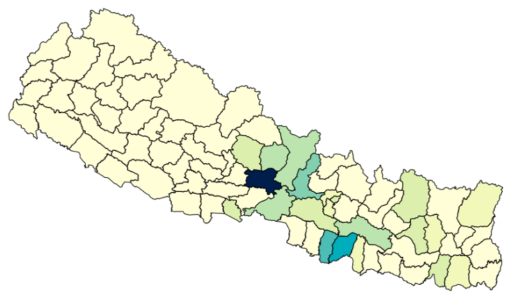
(a) UNOSAT Flood AI Monitoring Dashboard - July 1-2, 2021



(b) GloFAS reporting points and Rapid Impact Assessment map - July 2, 2021



(c) Geolocated Nepali tweets by district - July 1-2, 2021



(d) Nepal Disaster Risk Reduction portal incident count by district - July 1-2, 2021

Experimental results – Timeliness

Time of the events as reported from different sources

	Dates / Countries	**Thailand**	**Nepal**
Activation-based	UNOSAT activation	28/09/2021	30/06/2021
	Copernicus EMS activation	None	None
Manual	GDACS Disaster Alerts	27/09/2021 (Green Alert)	28/06/2021 (Green Alert)
Forecast	GloFAS	24/09/2021	28/06/2021
Manual	FloodList reported news	27/09/2021	04/07/2021
	TriggerCit	26/09/2021	02/07/2021

What we are currently working on

- Use **validated** event data as a ground truth, to benchmark and enhance the architecture components, evaluating uncertainty in space and time
- Explore the robustness of a supervised approach for automated event **triggering**
- Additional for data layers (additional systems, sensors, models, media) for an integrated **multi-modal** approach

References

- Carlo Bono, Barbara Pernici, Jose Luis Fernandez-Marquez, Amudha Ravi Shankar, Mehmet Oğuz Mülâyim, Edoardo Nemni, TriggerCit: Early Flood Alerting using Twitter and Geolocation - a comparison with alternative sources, 2022, accepted for presentation at ISCRAM 2022,May 2022 preprint: https://doi.org/10.48550/arXiv.2202.12014
- Carlo Bono, Barbara Pernici. (2022). Twitter dataset of flood-related images for September 2021, Thailand and June/July 2021, Nepal floods [Data set]. Zenodo. https://doi.org/10.5281/zenodo.6327151
- V. Negri, D. Scuratti, S. Agresti, D. Rooein, G. Scalia, J. L. Fernandez-Marquez, A. Ravi Shankar, M. Carman and B. Pernici, Image-based Social Sensing: Combining AI and the Crowd to Mine Policy-Adherence Indicators from Twitter, ICSE, Track Software Engineering in Society, May 2021 link to dataset and paper
- G. Scalia, C. Francalanci, B. Pernici, CIME: Context-aware geolocation of emergency-related posts, GeoInformatica, 26 (1), 125-157 https://link.springer.com/article/10.1007/s10707-021-00446-x

Thank you!

www.crowd4sdg.eu

Crowd4**SDG**

Social media analysis with VisualCit

crawling, extracting information from social media posts

Barbara Pernici, Carlo Bono, Politecnico di Milano

barbara.pernici@polimi.it

RCIS Tutorial, May 19, 2022

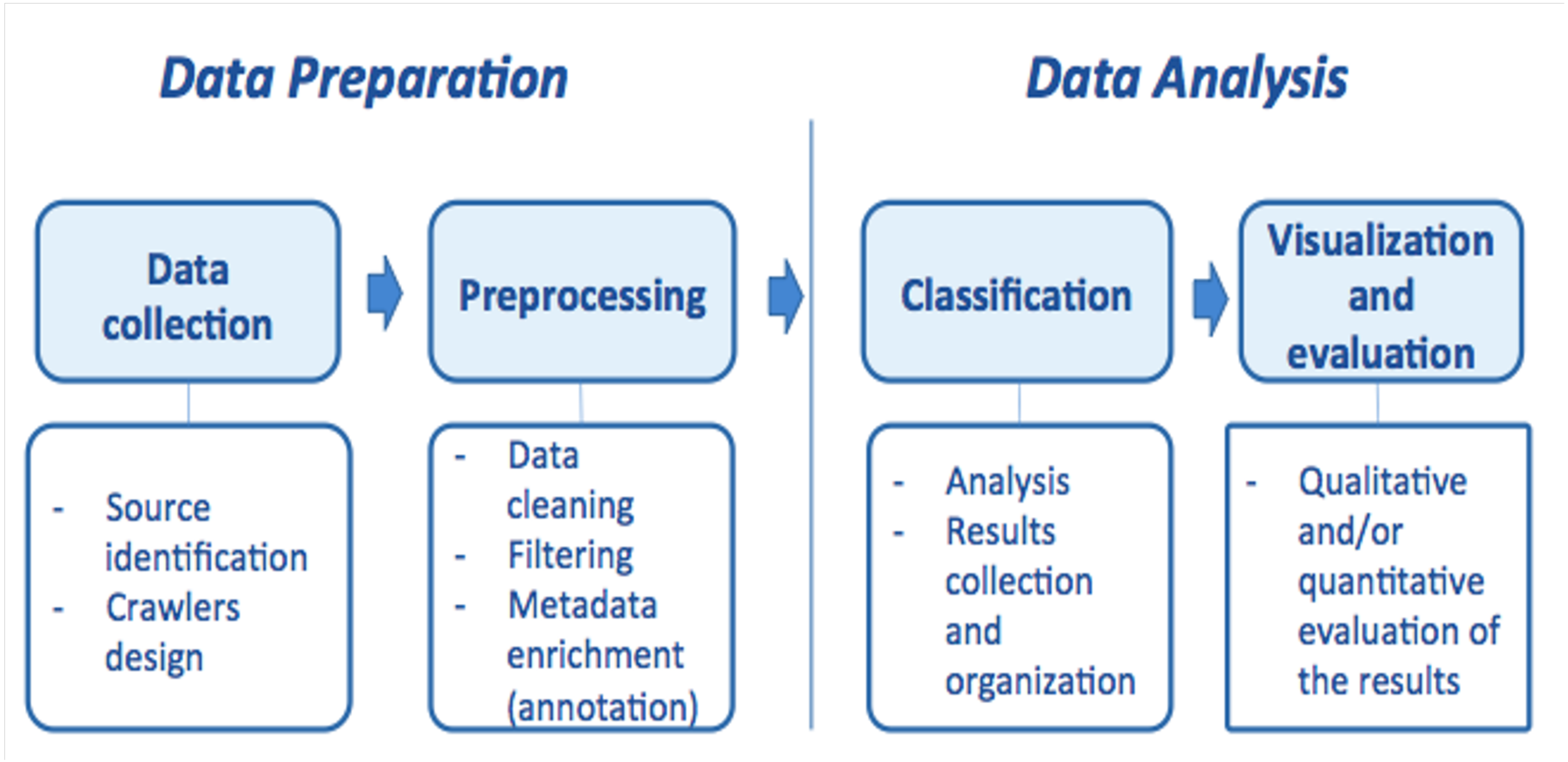
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944

Crowd4SDG

Agenda

- Extracting information from Twitter for emergencies (focus on floods)
- Analysis Pipelines
- VisualCit
- Hands on

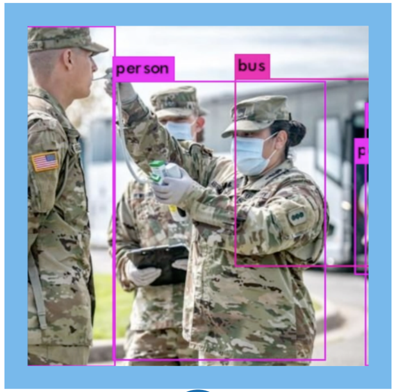
Crowd4SDG: VisualCit pipeline

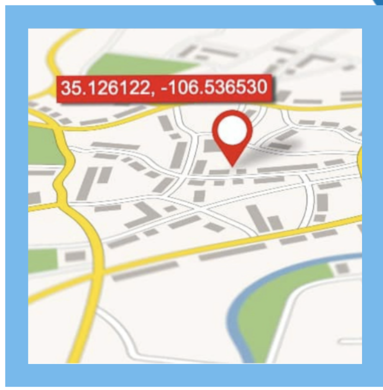


C. Cappiello et al., Modeling Adaptive Pipeline for Crowd Enhanced Processes, ER 2021, Oct. 2021

VisualCit Pipeline
to build indicators from social media

- **Data Preparation**
- **Collecting images from social media**





Twitter/social crawling

Keyword selection
(by Citizen Scientists)

Duplicate URL & Similar Image Filters

Semantic Image Filters

Photo/No photo
Not Safe for Work
Public places
People

Geocoding

CIME algorithm

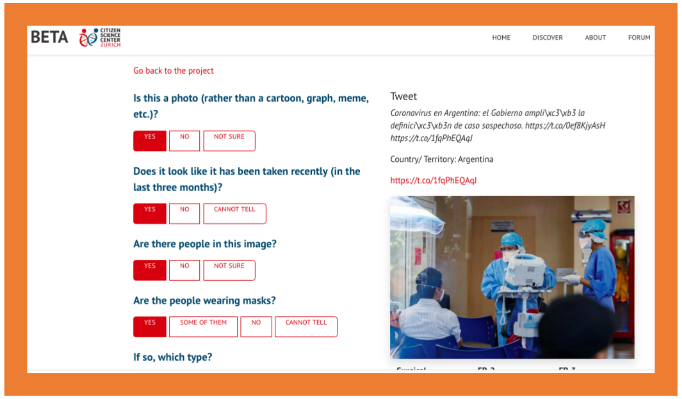
Pipeline to build indicators from social media

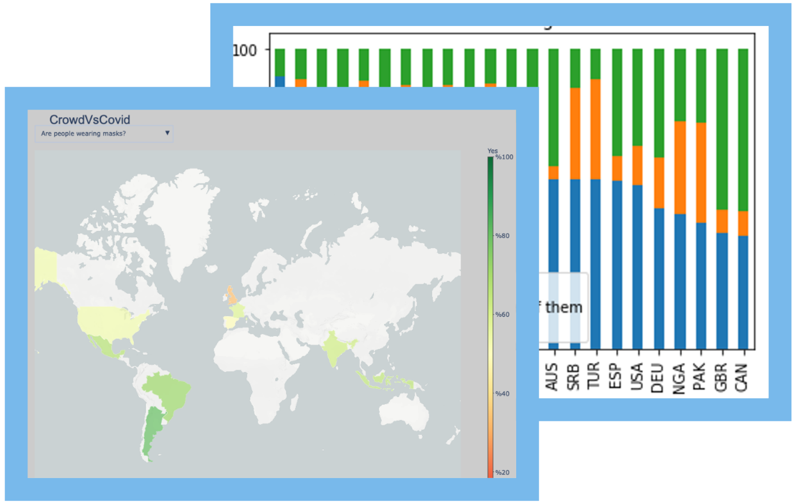
Crowd4SDG

POLITECNICO MILANO 1863

VisualCit

- **Data analysis - Deriving indicators**





Crowdsourcing

Project Builder
Asking Citizen Scientists

Aggregate Statistics and ML

Quality evaluation
Building indicators
Building new classifiers

Visualisation

Thematic maps
Benchmarks

https://pernici.faculty.polimi.it/crowd4sdgpolimi/

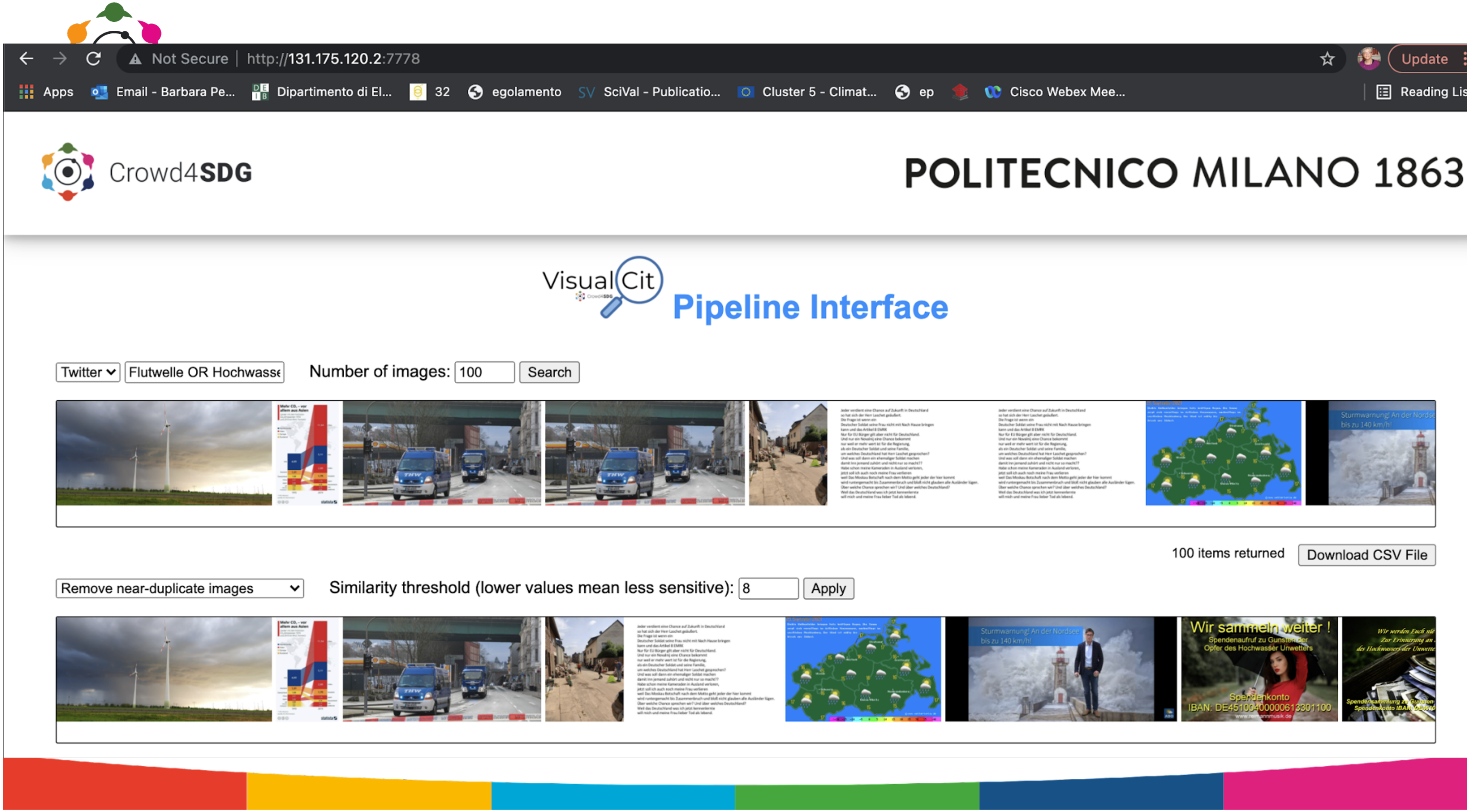
Building a project with VisualCit

0) Define your goals

1) Select images with VisualCit
   - Keywords
   - Filters

- Try it:
  - http://visualcit.polimi.it:7778/

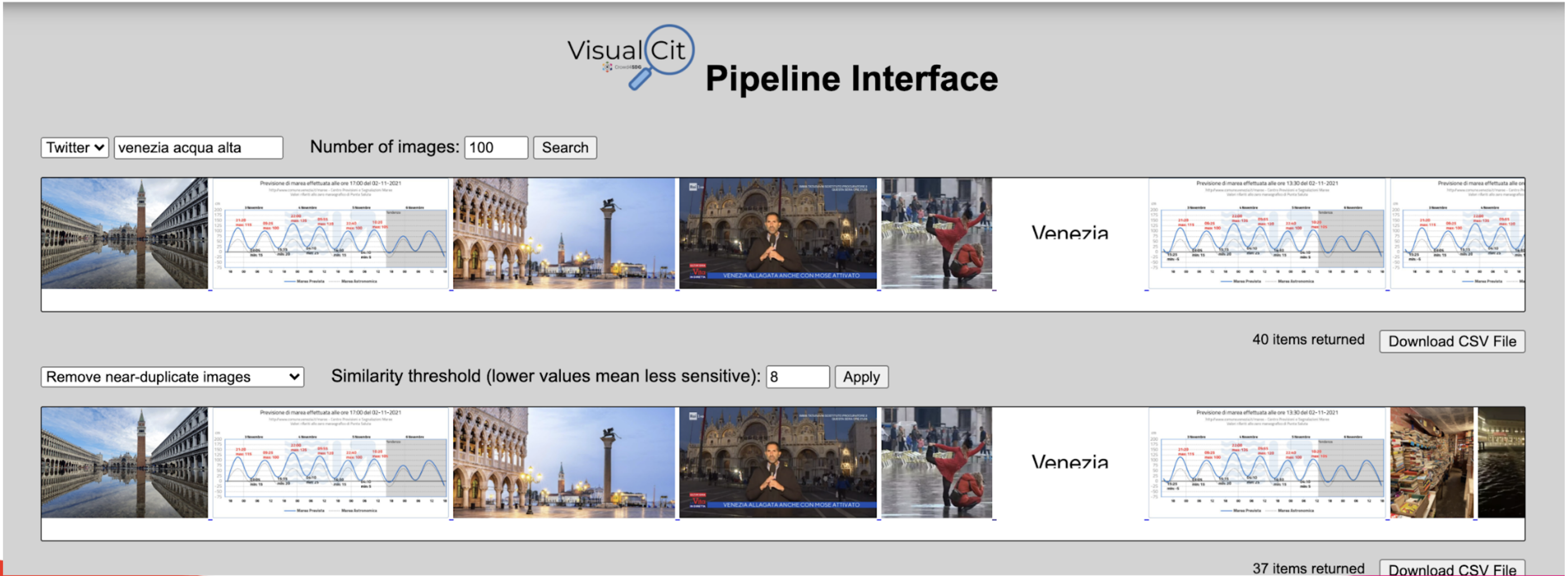
2) Use the collected information
   - Set up a crowdsourcing initiative (design questions, identify crowd)
   - Compute indicators


Not Secure
http://131.175.120.2:7778
Update
Apps
Email - Barbara Pe...
Dipartimento di El...
32
egolamento
SciVal - Publicatio...
Cluster 5 - Climat...
ep
Cisco Webex Mee...
Reading Lis
Crowd4SDG
POLITECNICO MILANO 1863
VisualCit
Pipeline Interface
Twitter
Flutwelle OR Hochwasse
Number of images:
100
Search
Sturmwarnung! An der Nordse
bis zu 140 km/h!
100 items returned
Download CSV File
Remove near-duplicate images
Similarity threshold (lower values mean less sensitive):
8
Apply
Sturmwarnung! An der Nordsee
bis zu 140 km/h!
Wir sammeln weiter !
Spendenkonto

Crowd4SDG

Crowd4SDG

POLITECNICO MILANO 1863

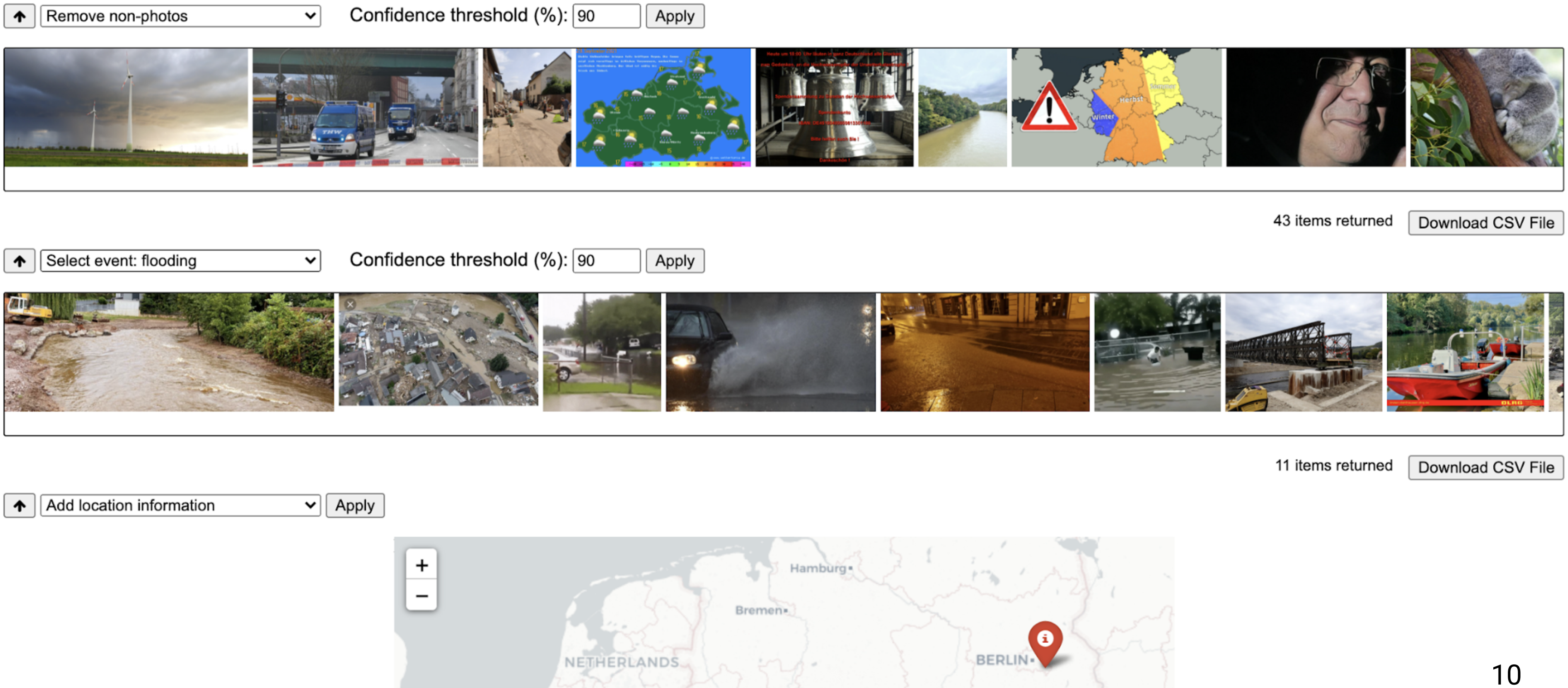


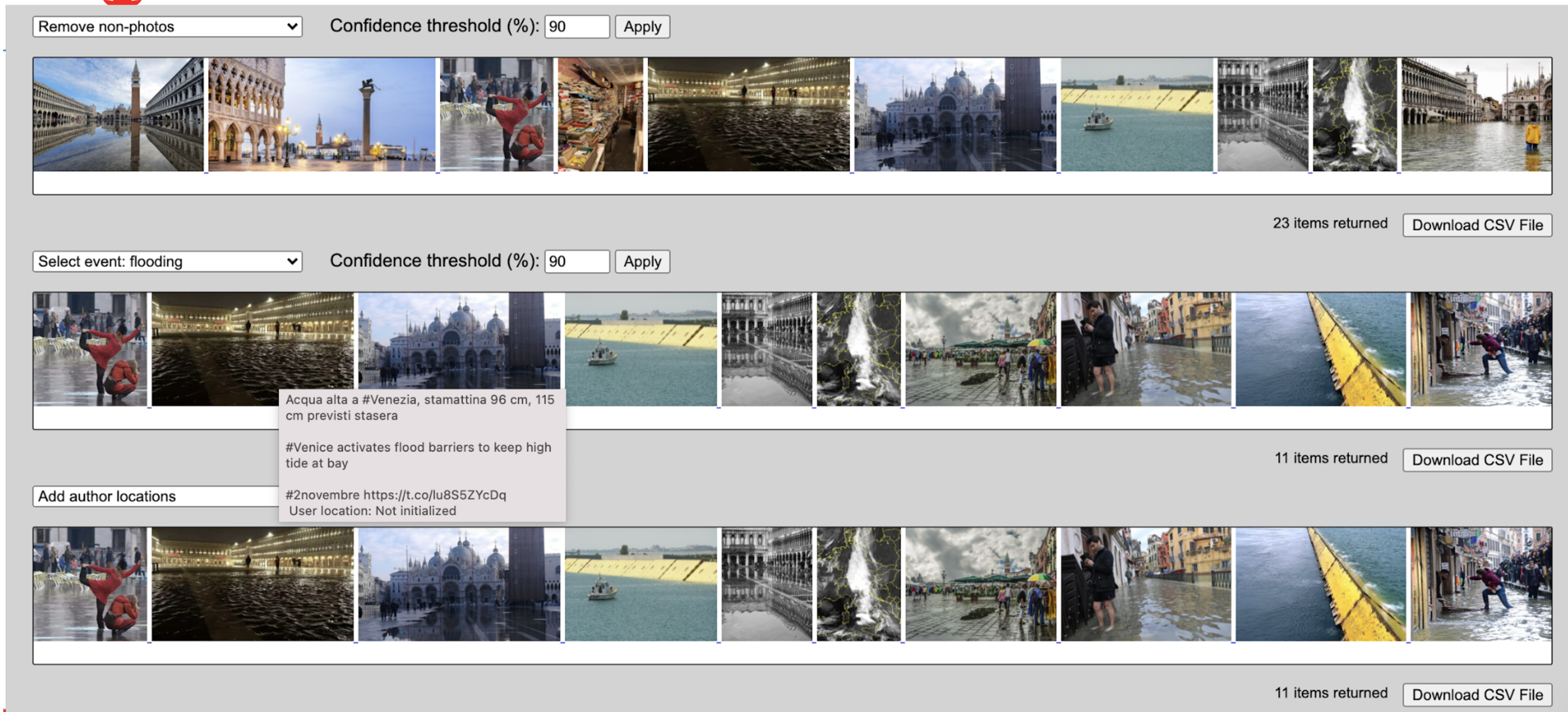
E2mC Keywords for floods in English and Italian

- en= inundate,inundation,overflow,riverlevel,waterflow,swamped,mudslip,engulf,drainage,flooded,landslide,flooding,riverflow,immerse,torrential,flashflooding,mudflow,floodwater,rainfall,hailstorm,cloudburst,mudslip,precipitation,deluge,deluged,engulfed,flood,landslip,mudslide,submerged,swamp,torrent,waterlevel,tsunami,hail

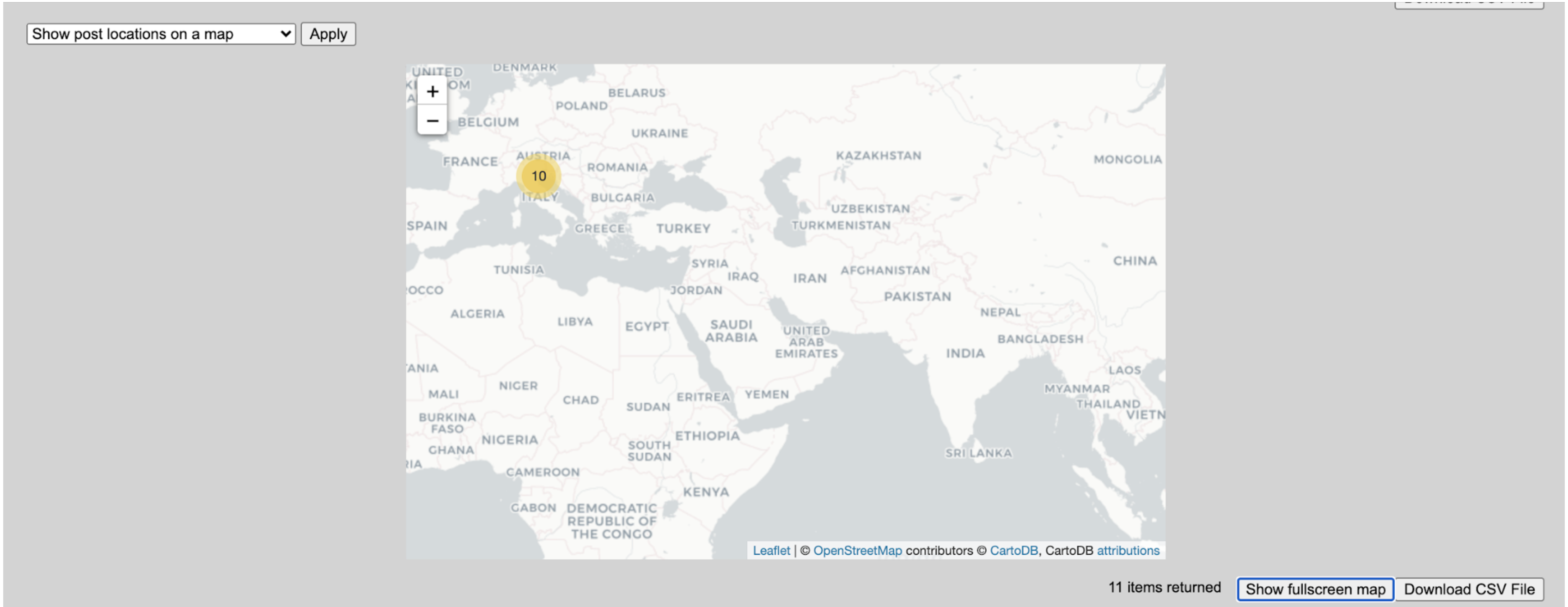
- it= alluvione,diga,inondazione,inondazioni,dell'alluvione,torrente,pioggia,piovosità,straripamento,nubifragio,marea,allagamenti,erosione,frana,sommersi,sommersa,bonifica,aumentare,inondato,sommerso,eccezionale,deflusso,precipitazione,diluvio,argine,annegato,salire,terrapieno,allagamento,straripare,palude,drenaggio,valanga,grandine,scolo,lungofiume

VisualCit filters




Remove non-photos
Confidence threshold (%): 90
Apply
23 items returned
Download CSV File
Select event: flooding
Confidence threshold (%): 90
Apply
Acqua alta a #Venezia, stamattina 96 cm, 115 cm previsti stasera
#Venice activates flood barriers to keep high tide at bay
#2novembre https://t.co/lu8S5ZYcDq
User location: Not initialized
11 items returned
Download CSV File
Add author locations
11 items returned
Download CSV File

Geolocate


Show post locations on a map
Apply
UNITED KINGDOM
DENMARK
BELARUS
POLAND
BELGIUM
UKRAINE
FRANCE
AUSTRIA
ROMANIA
10
ITALY
BULGARIA
KAZAKHSTAN
MONGOLIA
UZBEKISTAN
TURKMENISTAN
SPAIN
GREECE
TURKEY
SYRIA
IRAQ
IRAN
AFGHANISTAN
CHINA
TUNISIA
JORDAN
PAKISTAN
ALGERIA
LIBYA
EGYPT
SAUDI ARABIA
NEPAL
UNITED ARAB EMIRATES
BANGLADESH
INDIA
LAOS
MYANMAR
THAILAND
MALI
NIGER
CHAD
SUDAN
ERITREA
YEMEN
BURKINA FASO
NIGERIA
ETHIOPIA
GHANA
SOUTH SUDAN
SRI LANKA
CAMEROON
KENYA
GABON
DEMOCRATIC REPUBLIC OF THE CONGO
Leaflet | © OpenStreetMap contributors © CartoDB, CartoDB attributions
11 items returned
Show fullscreen map
Download CSV File

↑ Add location information ⌄ Apply

11 items returned | Show fullscreen map | Download CSV File

-Select- ⌄ Apply

Reset Page

Get configuration for this pipeline

VisualCit

Interactive interface: visualization in browser

http://visualcit.polimi.it:7778/

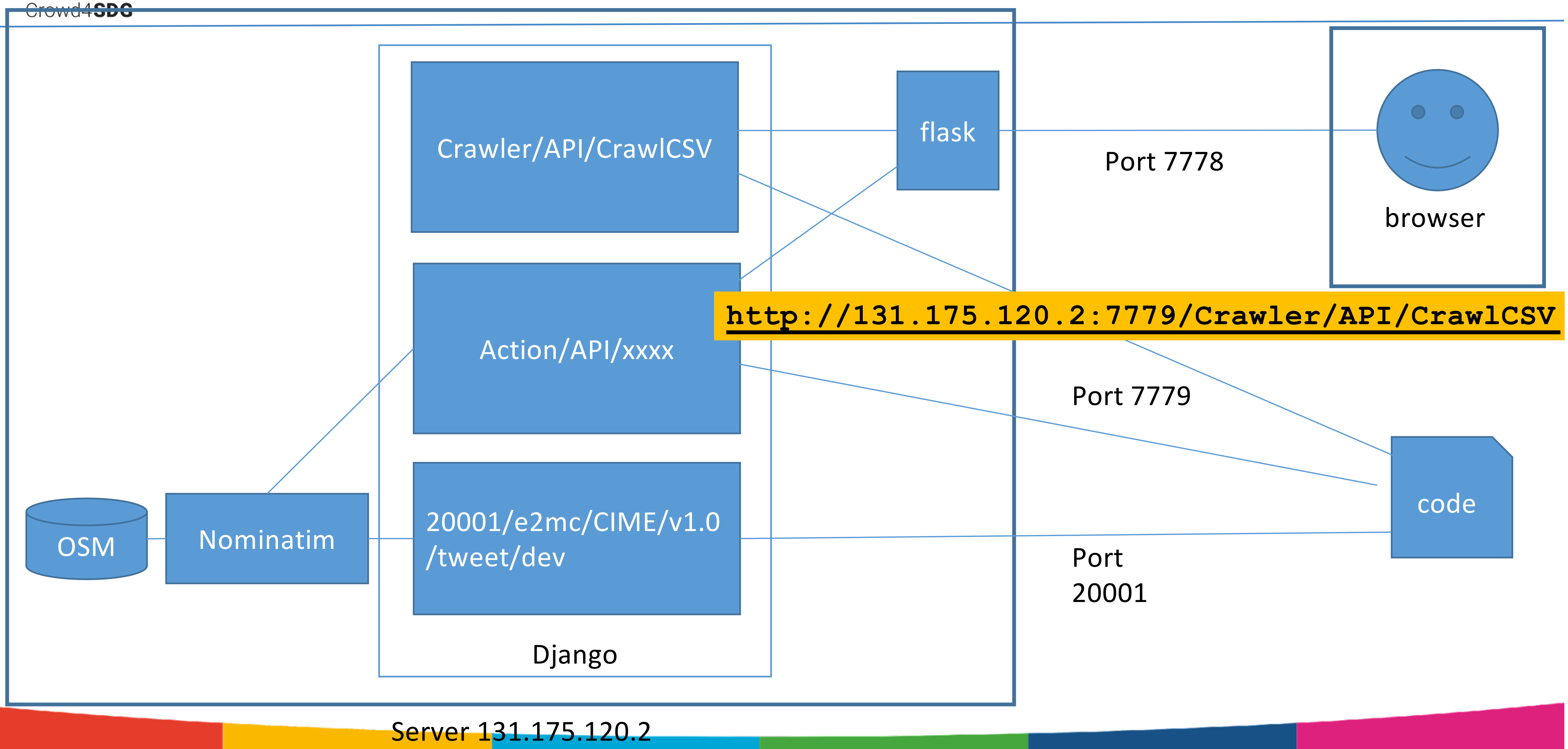
Web service: can be invoked inside a Python programme

See link for details (discussed later)

Some **test data** can be found in

Link to folder

Using VisualCit a service



VisualCit endpoints

Endpoint visualization

See services with Command + Option + I on MacOS
Or Command F11 on Windows
then Network

Documentation of available endpoints for VisualCit web services link

In VisualCit interactive service
Download configuration (Get pipeline configuration)

Crowd4SDG

- Carlo Bono, Barbara Pernici, Jose Luis Fernandez-Marquez, Amudha Ravi Shankar, Mehmet Oğuz Mülâyim, Edoardo Nemni, TriggerCit: Early Flood Alerting using Twitter and Geolocation - a comparison with alternative sources, 2022, accepted for presentation at ISCRAM 2022, May 2022 preprint: https://doi.org/10.48550/arXiv.2202.12014
- Carlo Bono, Barbara Pernici. (2022). Twitter dataset of flood-related images for September 2021, Thailand and June/July 2021, Nepal floods [Data set]. Zenodo. https://doi.org/10.5281/zenodo.6327151
- V. Negri, D. Scuratti, S. Agresti, D. Rooein, G. Scalia, J. L. Fernandez-Marquez, A. Ravi Shankar, M. Carman and B. Pernici, Image-based Social Sensing: Combining AI and the Crowd to Mine Policy-Adherence Indicators from Twitter, ICSE, Track Software Engineering in Society, May 2021 link to dataset and paper
- G. Scalia, C. Francalanci, B. Pernici, CIME: Context-aware geolocation of emergency-related posts, GeoInformatica, 26 (1), 125-157 https://link.springer.com/article/10.1007/s10707-021-00446-x

Thank you!

www.crowd4sdg.eu

Details on classifiers

Meme: based on VGG16 (custom)

Flood: based onXception (custom)

Object: YOLOv5 (also DETR di Facebook, only available in backend)

Scene: PlacesCNN (trained with Places365, based on VGG16)

NSFW: based on MobileNetV2 (not customized)

Crowd4SDG

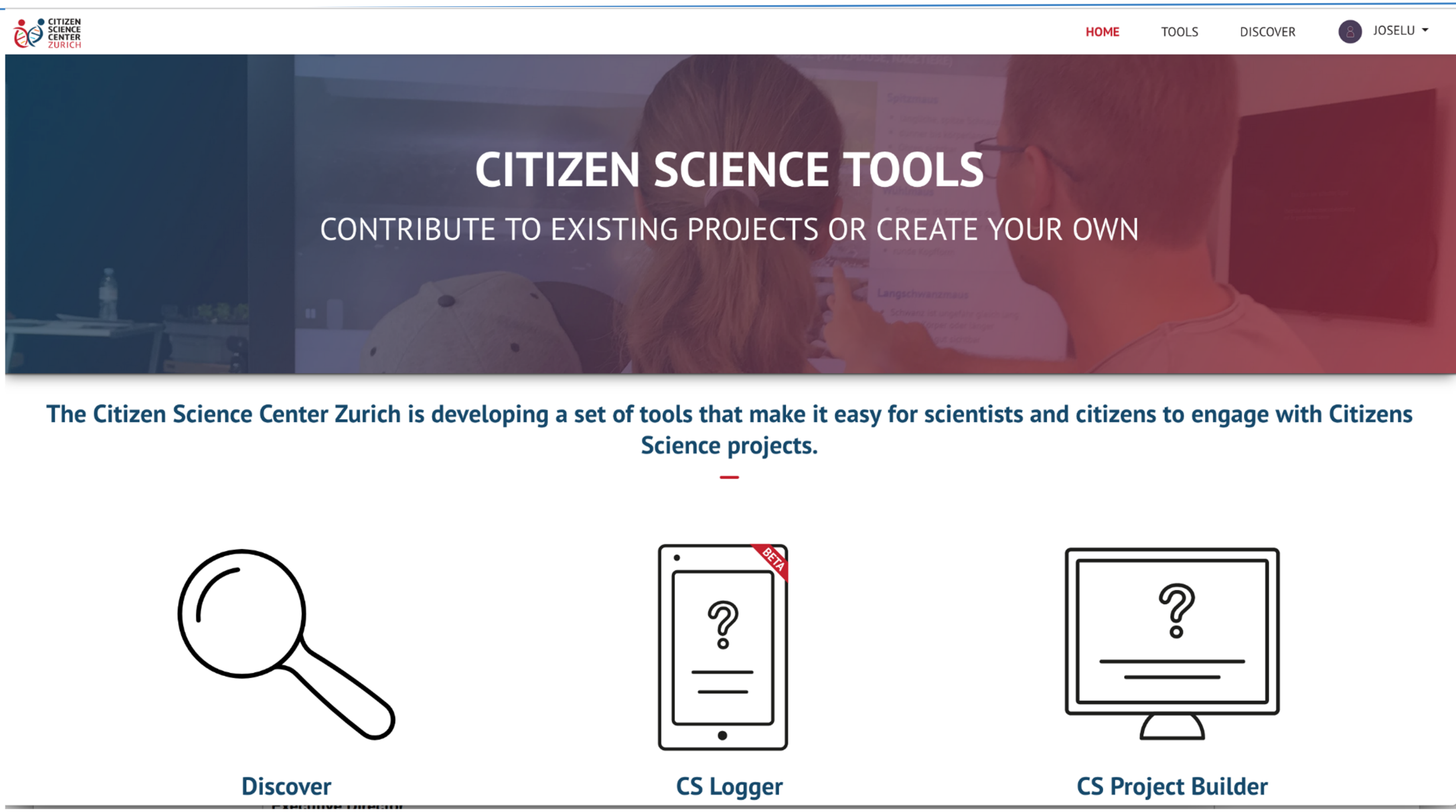
Citizen Science Project Builder

Jose Luis Fernandez-Marquez

RCIS 2022, Barcelona May 19, 2022

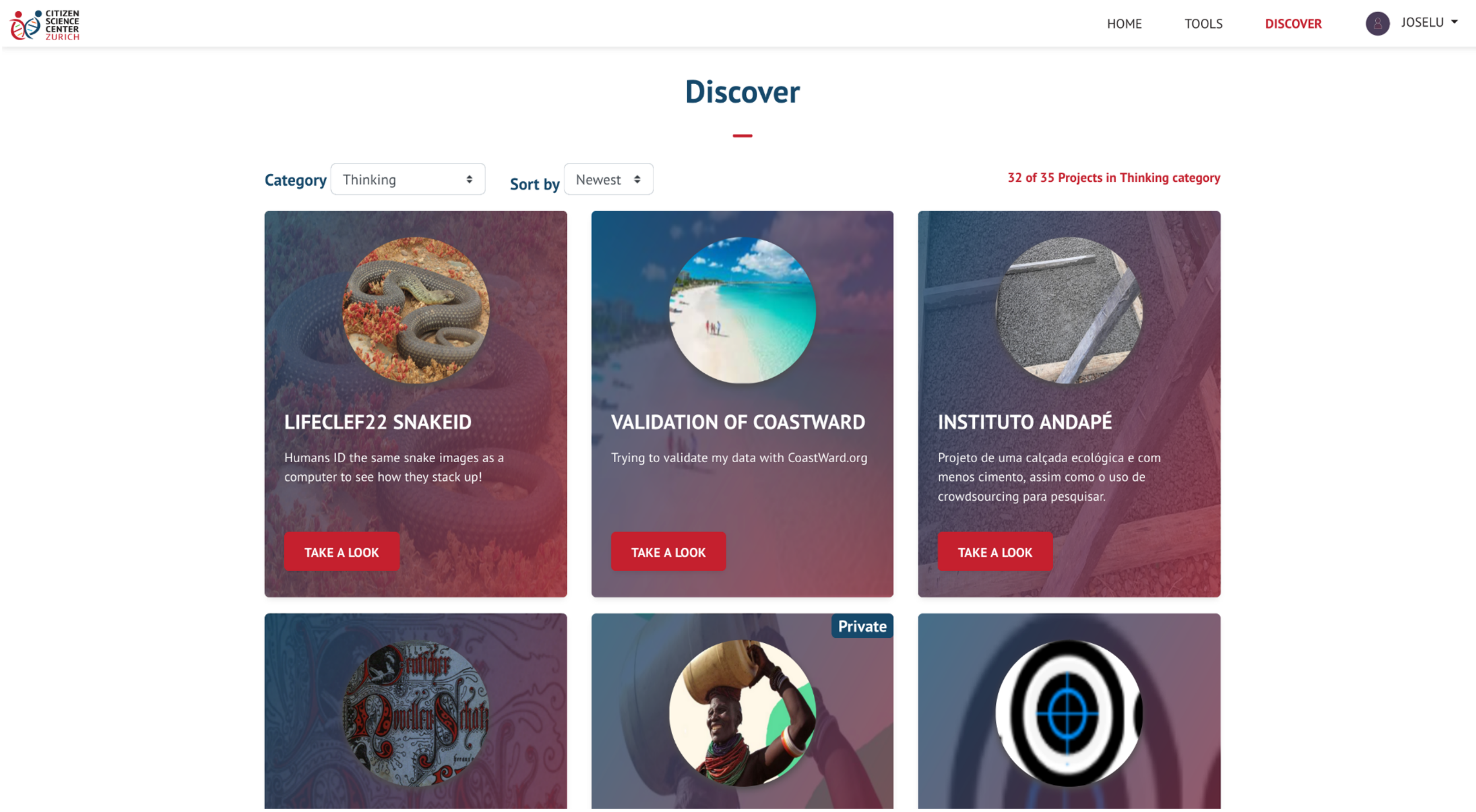
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944

Crowd4SDG


CITIZEN SCIENCE CENTER ZURICH
HOME
TOOLS
DISCOVER
JOSELU
CITIZEN SCIENCE TOOLS
CONTRIBUTE TO EXISTING PROJECTS OR CREATE YOUR OWN
The Citizen Science Center Zurich is developing a set of tools that make it easy for scientists and citizens to engage with Citizens Science projects.
BETA
Discover
CS Logger
CS Project Builder

Crowd4SDG

Discovering new projects



Contribute to an existing project

CITIZEN SCIENCE CENTER ZURICH

HOME TOOLS DISCOVER JOSELU

Go back to the project

What Snake is This?

This photo was taken in: **Lusaka - Zambia**

This photo was taken by: **iNaturalist**

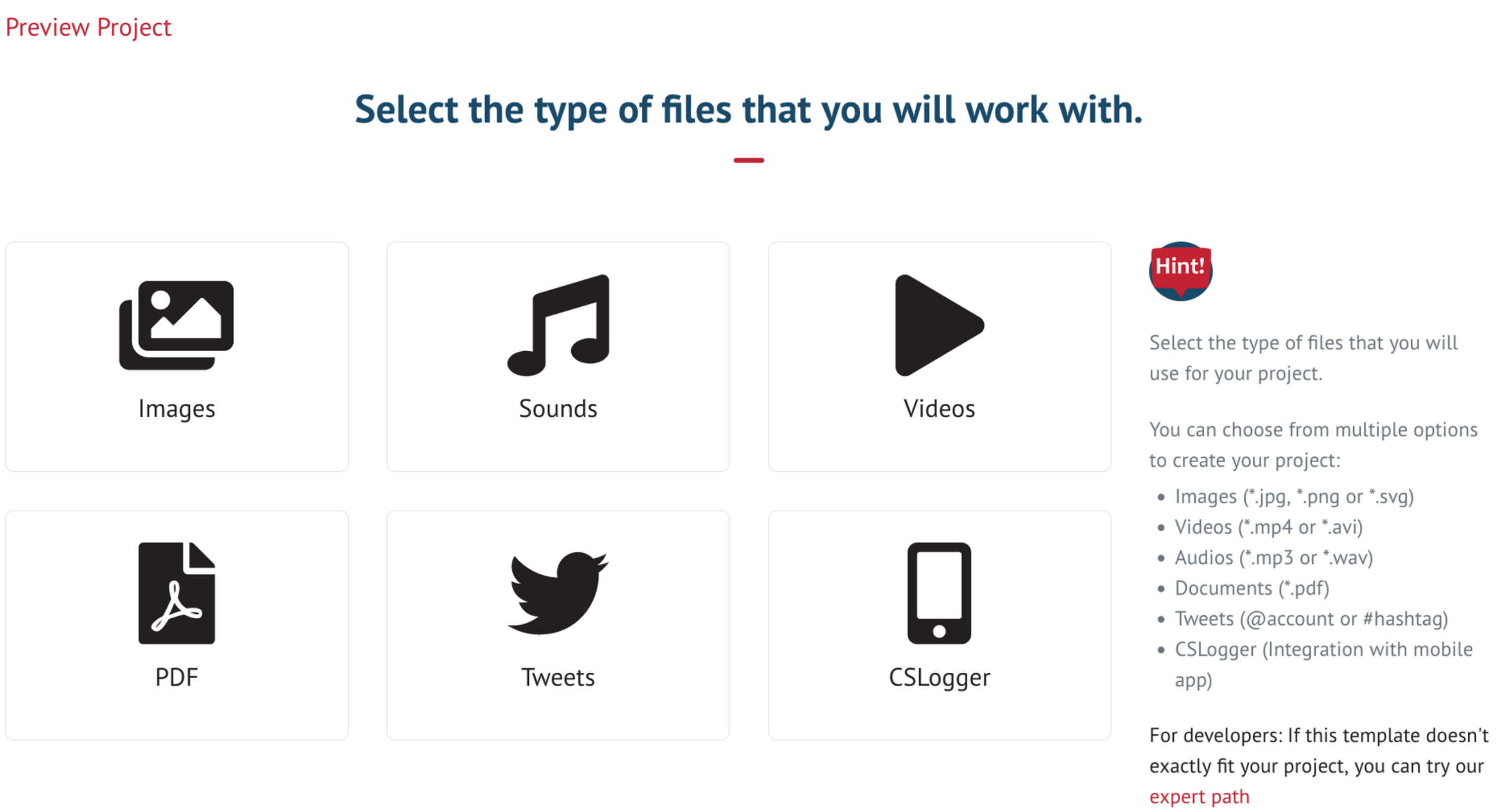
Type at least 2 letters for Family, Genus, Binomial or Common Name

More info

SUBMIT SKIP

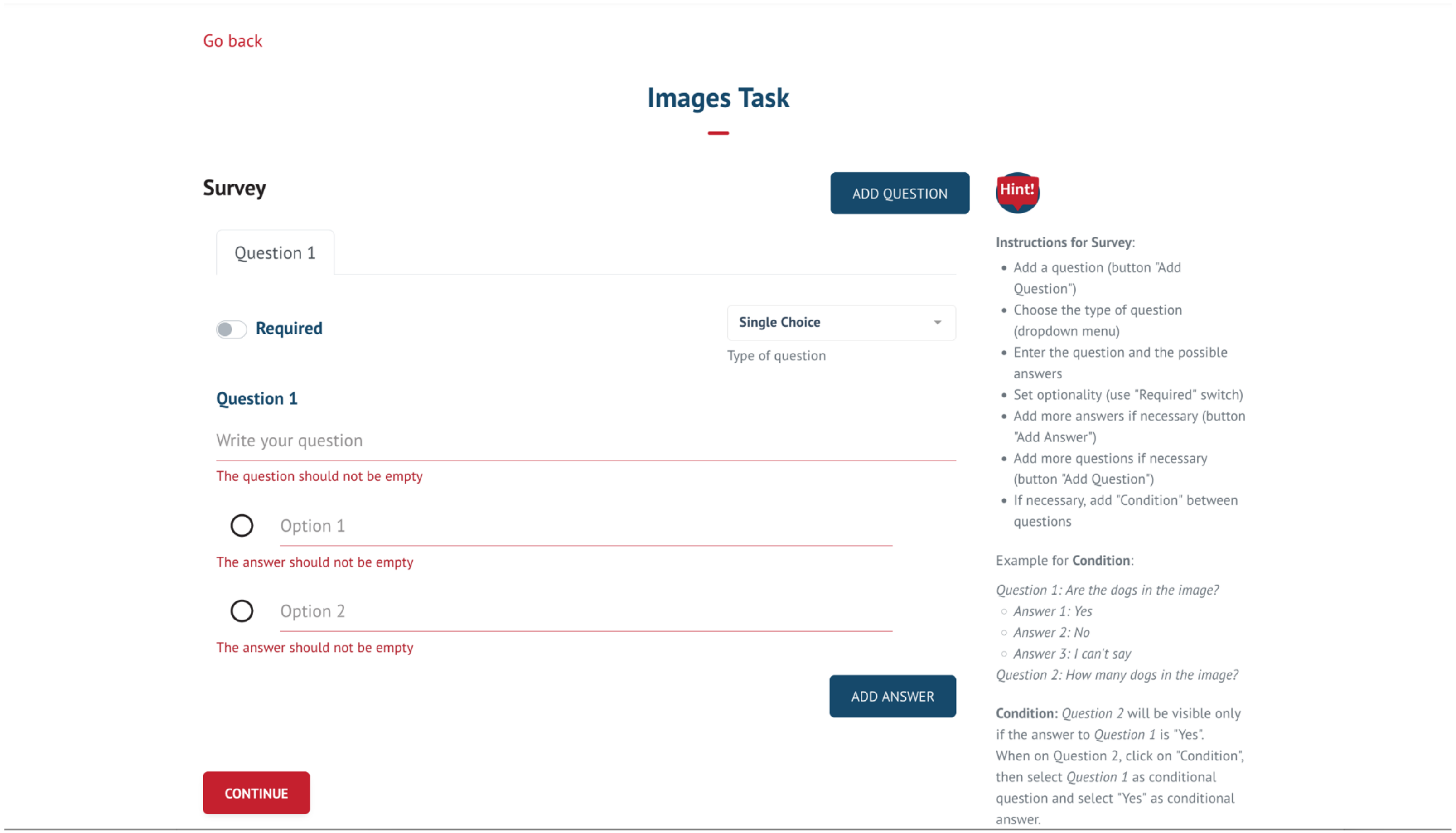
You have completed: 0 tasks out of 150

Create a new project



Create a new project

Create a new project



Crowd4**SDG**

Create a new project - Import tasks

Select your files

Amazon S3 bucket

Dropbox

Flickr

csv File

Hint!

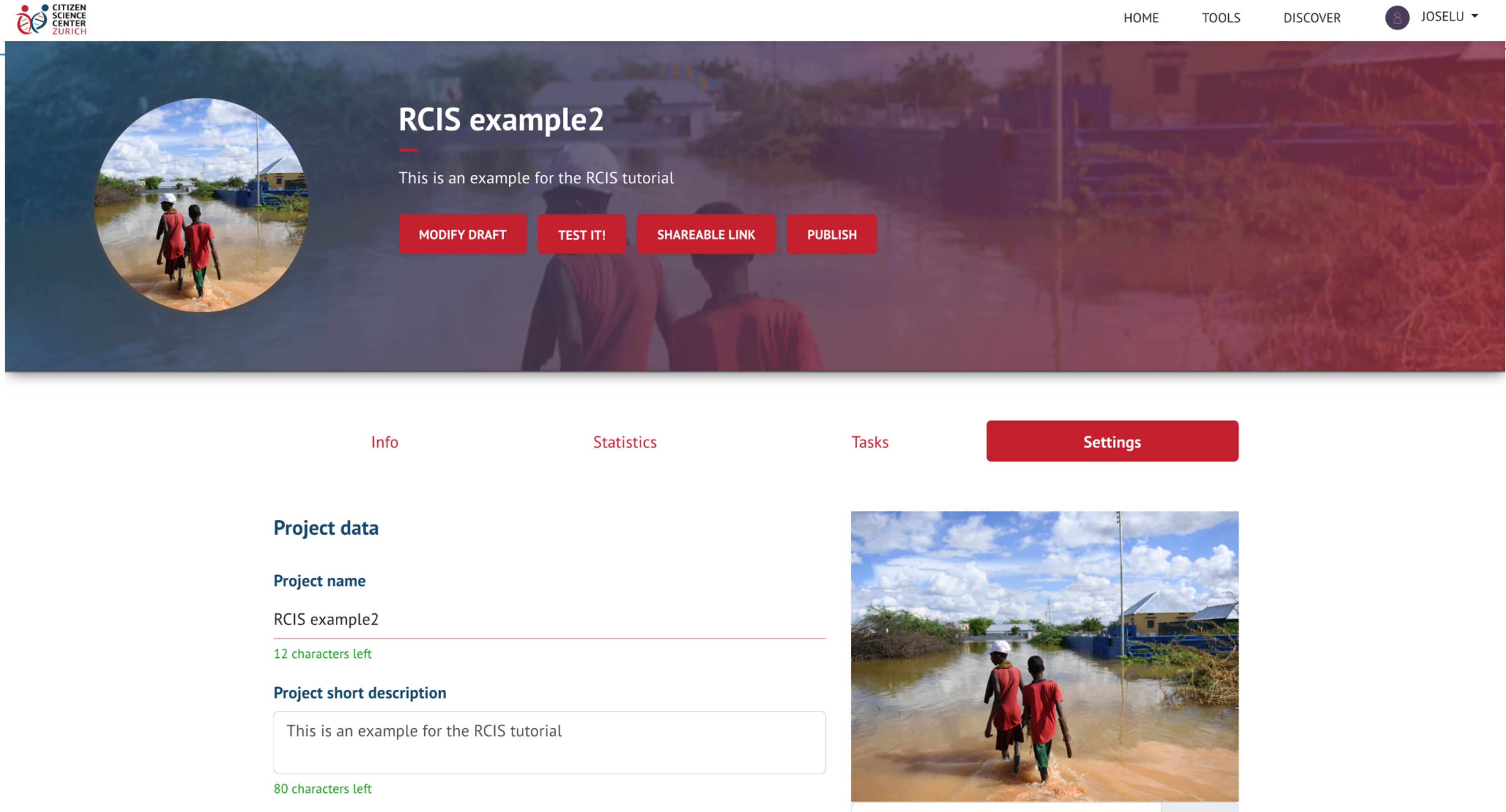
Select the source of your data.

Your data should be available via url, in particular:

- File hosted on cloud services such as Dropbox, Flickr or Amazon S3 bucket.

Make sure you are authorized to use the data, ie. use free licensed files (Creative Commons or alike), your own files, or copyright files that you are authorized to use.

Our project is created!



Contributing to the project

Go back to the editor

Do you see a flooded area in this picture?

- ( ) Yes
- (•) No
- ( ) I am not sure

SUBMIT SKIP

You have completed: 5 tasks out of 24

Managing your project

Info Statistics **Tasks** Settings

Import Tasks

Import new tasks (files)

IMPORT

Export Tasks

Export tasks to JSON, CSV, or a CKAN server

EXPORT

Task Presenter

Edit the task presenter (only for experts)

EDITOR

Settings

Configure the task scheduler, redundancy, etc.

SETTINGS

Browse

Check the progress on tasks.

BROWSE

Task settings

Task settings

Task scheduler

Change how the tasks are delivered to volunteers

SCHEDULER

Task priority

Change the priority of the tasks

PRIORITY

Delete Tasks

Delete all the tasks and associated answers

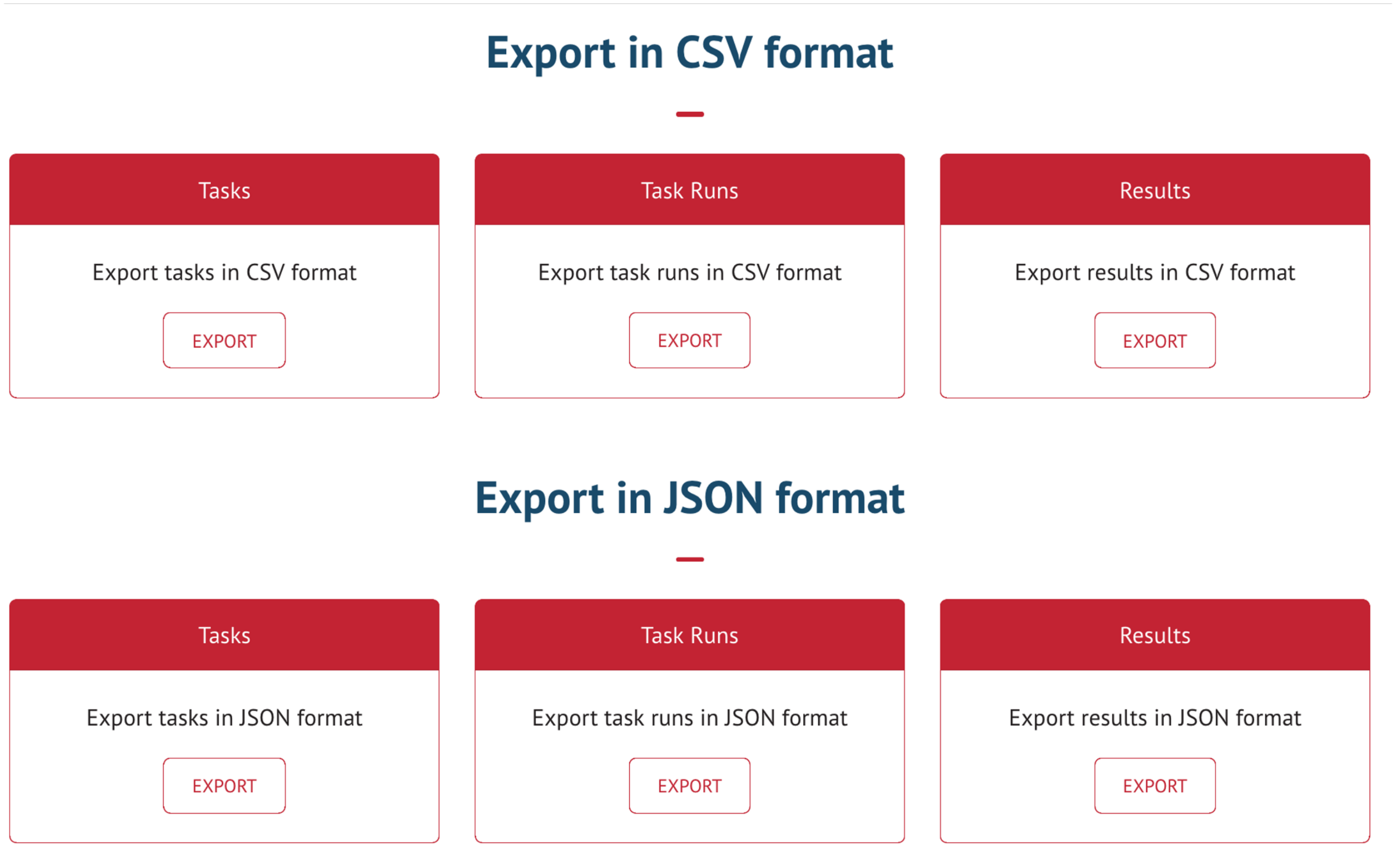
DELETE

Tasks Redundancy

Change the redundancy level for the tasks

REDUNDANCY

Export your data



What's next?

Consensus models

	task_id	user_id	is_fast
0	Audi	John	Yes
1	Audi	Matthew	Yes
2	Audi	Andrew	Yes
3	BMW	John	Yes
4	BMW	Matthew	Yes
5	BMW	Matthew	Yes
6	Seat	John	No
7	Seat	Andrew	No
8	Seat	Matthew	No
9	Maserati	John	Yes
10	Maserati	Andrew	Yes
11	GM	Andrew	Yes
12	GM	Matthew	No
13	GM	John	No
14	Mazda	Andrew	Yes
15	Mazda	John	Yes
16	Mazda	Matthew	Yes
17	Kia	Matthew	Yes
18	Kia	Andrew	Yes
19	Kia	John	No

Crowdsourced annotations

Majority Voting

	Yes	No
Audi	1.0	0.0
BMW	1.0	0.0
GM	0.0	1.0
Kia	1.0	0.0
Maserati	1.0	0.0
Mazda	1.0	0.0
Seat	0.0	1.0

Probabilistic

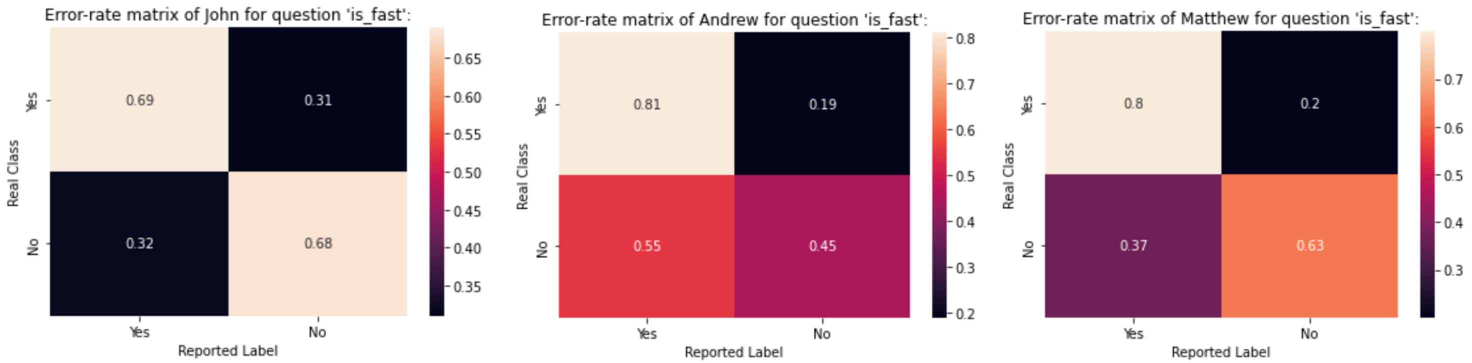
	Yes	No
Audi	1.000000	0.000000
BMW	1.000000	0.000000
GM	0.333333	0.666667
Kia	0.666667	0.333333
Maserati	1.000000	0.000000
Mazda	1.000000	0.000000
Seat	0.000000	1.000000

These **do NOT model** individual **annotator** performance.

Advanced consensus models - Dawid-Skene

	task_id	user_id	is_fast
0	Audi	John	Yes
1	Audi	Matthew	Yes
2	Audi	Andrew	Yes
3	BMW	John	Yes
4	BMW	Matthew	Yes
5	BMW	Matthew	Yes
6	Seat	John	No
7	Seat	Andrew	No
8	Seat	Matthew	No
9	Maserati	John	Yes
10	Maserati	Andrew	Yes
11	GM	Andrew	Yes
12	GM	Matthew	No
13	GM	John	No
14	Mazda	Andrew	Yes
15	Mazda	John	Yes
16	Mazda	Matthew	Yes
17	Kia	Matthew	Yes
18	Kia	Andrew	Yes
19	Kia	John	No

Annotator error-rates

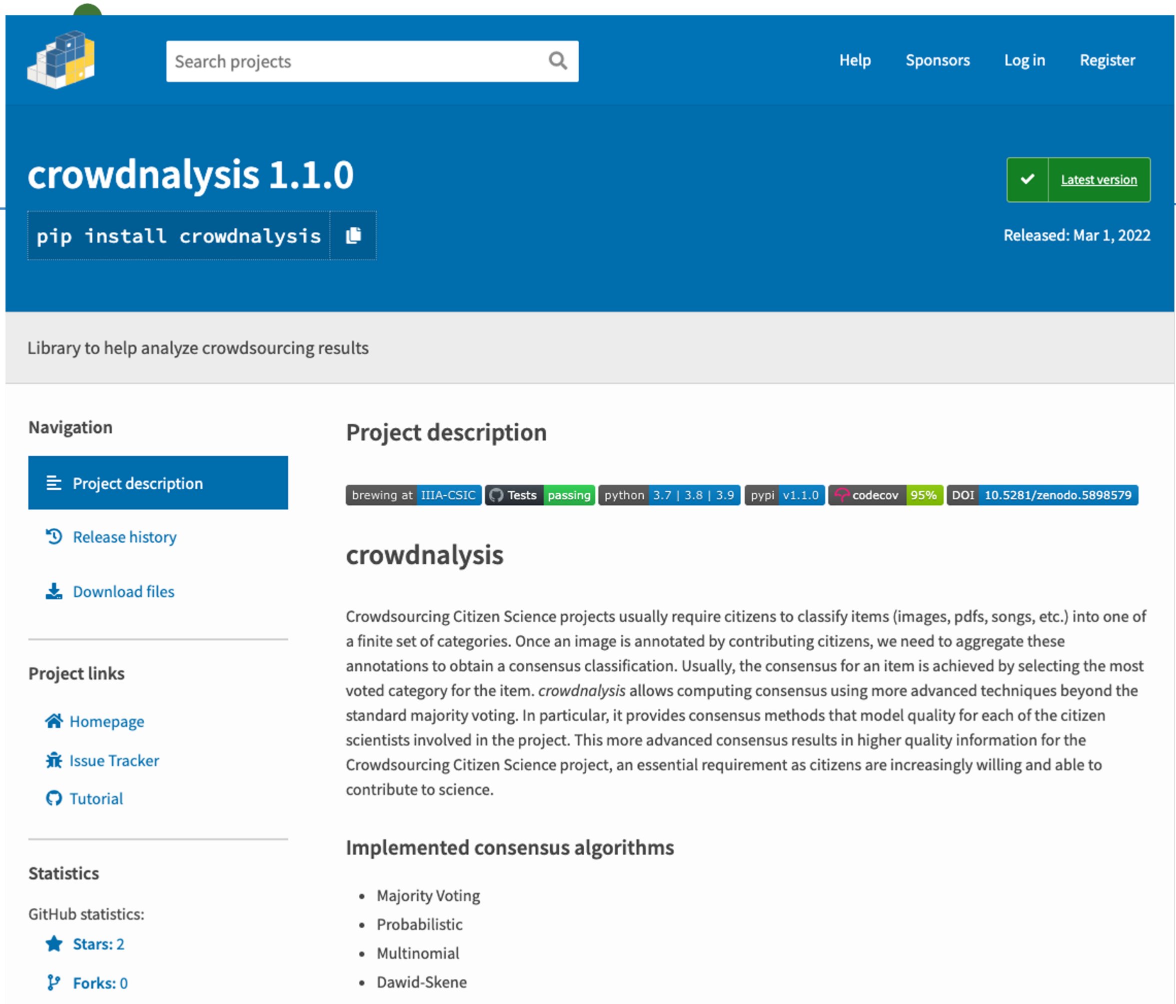


More reliable consensus is possible thanks to **modelling** the **annotator behavior**.

Consensus

	Yes	No
Audi	0.923616	0.076384
BMW	0.946553	0.053447
GM	0.271625	0.728375
Kia	0.716913	0.283087
Maserati	0.848372	0.151628
Mazda	0.923616	0.076384
Seat	0.096134	0.903866

Dawid, A. P., & Skene, A. M. (1979). Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm. *Applied Statistics*, *28*(1), 20. https://doi.org/10.2307/2346806



crowdnalysis

analyzing crowdsourced data

- Import **annotation data** with preprocessing
- Calculate **inter-rater reliability**
- **Model** annotators
- Compute the **consensus** on annotations
- Conduct **prospective analysis** (e.g., '*accuracy vs. number of annotations*')
- **Visualize** consensus & annotator **error-rates**
- and more ...

python™ Package Index

https://pypi.org/project/crowdnalysis/

👉 [Google Colab Notebook] for a hands-on experience

Thank you!

www.crowd4sdg.eu

UNIVERSITÉ DE GENÈVE · CERN · CSIC Consejo Superior de Investigaciones Científicas · POLITECNICO MILANO 1863 · unitar United Nations Institute for Training and Research · Université de Paris

Crowd4SDG

crowdnalysis: *Analyzing crowdsourced data*

Oguz Mulayim

oguz@iiia.csic.es

Artificial Intelligence Research Institute (IIIA), CSIC

RCIS Tutorial, May 19, 2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 872944

Crowd4SDG

Consensus models

	task_id	user_id	is_fast
0	Audi	John	Yes
1	Audi	Matthew	Yes
2	Audi	Andrew	Yes
3	BMW	John	Yes
4	BMW	Matthew	Yes
5	BMW	Matthew	Yes
6	Seat	John	No
7	Seat	Andrew	No
8	Seat	Matthew	No
9	Maserati	John	Yes
10	Maserati	Andrew	Yes
11	GM	Andrew	Yes
12	GM	Matthew	No
13	GM	John	No
14	Mazda	Andrew	Yes
15	Mazda	John	Yes
16	Mazda	Matthew	Yes
17	Kia	Matthew	Yes
18	Kia	Andrew	Yes
19	Kia	John	No

Crowdsourced annotations

Majority Voting

	Yes	No
Audi	1.0	0.0
BMW	1.0	0.0
GM	0.0	1.0
Kia	1.0	0.0
Maserati	1.0	0.0
Mazda	1.0	0.0
Seat	0.0	1.0

Probabilistic

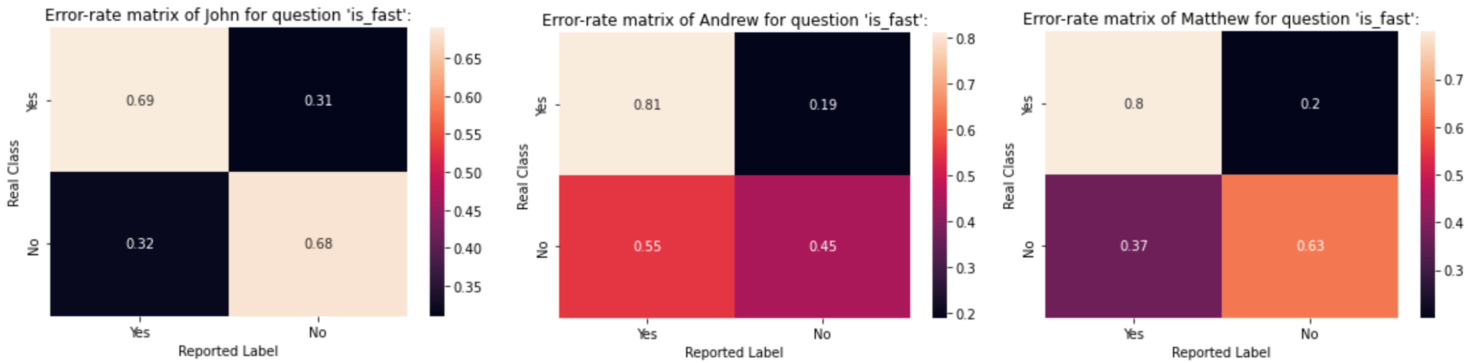
	Yes	No
Audi	1.000000	0.000000
BMW	1.000000	0.000000
GM	0.333333	0.666667
Kia	0.666667	0.333333
Maserati	1.000000	0.000000
Mazda	1.000000	0.000000
Seat	0.000000	1.000000

These **do NOT model** individual **annotator** performance.

Advanced consensus models - Dawid-Skene

	task_id	user_id	is_fast
0	Audi	John	Yes
1	Audi	Matthew	Yes
2	Audi	Andrew	Yes
3	BMW	John	Yes
4	BMW	Matthew	Yes
5	BMW	Matthew	Yes
6	Seat	John	No
7	Seat	Andrew	No
8	Seat	Matthew	No
9	Maserati	John	Yes
10	Maserati	Andrew	Yes
11	GM	Andrew	Yes
12	GM	Matthew	No
13	GM	John	No
14	Mazda	Andrew	Yes
15	Mazda	John	Yes
16	Mazda	Matthew	Yes
17	Kia	Matthew	Yes
18	Kia	Andrew	Yes
19	Kia	John	No

Annotator error-rates

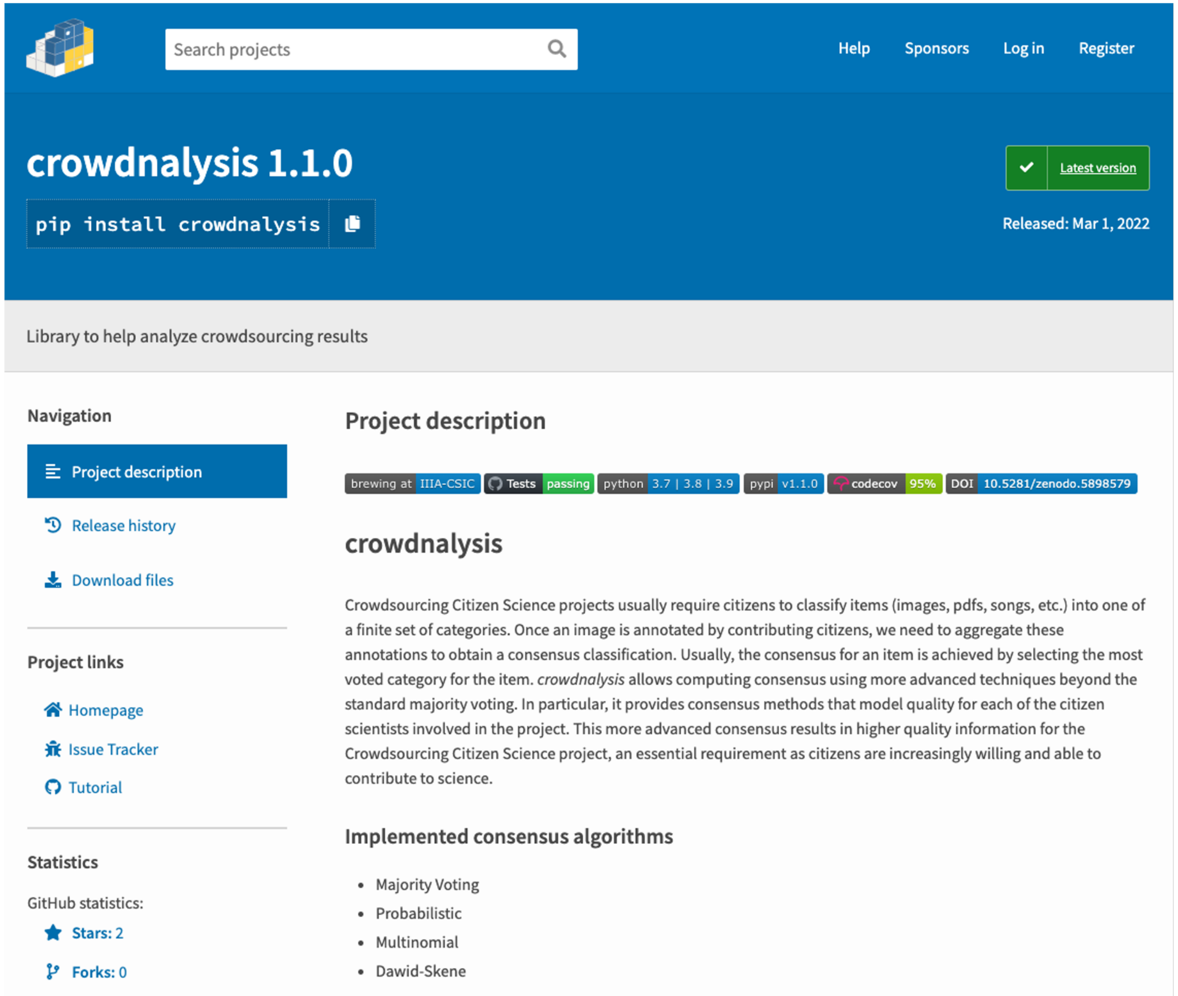


Consensus

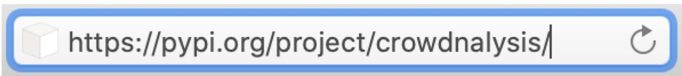
	Yes	No
Audi	0.923616	0.076384
BMW	0.946553	0.053447
GM	0.271625	0.728375
Kia	0.716913	0.283087
Maserati	0.848372	0.151628
Mazda	0.923616	0.076384
Seat	0.096134	0.903866

More reliable consensus is possible thanks to **modelling** the **annotator behavior**.

Dawid, A. P., & Skene, A. M. (1979). Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm. *Applied Statistics*, *28*(1), 20. https://doi.org/10.2307/2346806



python Package Index



crowdnalysis

analyzing crowdsourced data

- Import **annotation data** with preprocessing
- Calculate **inter-rater reliability**
- **Model** annotators
- Compute the **consensus** on annotations
- Conduct **prospective analysis** (e.g., '*accuracy vs. number of annotations*')
- **Visualize** consensus & annotator **error-rates**
- and more ...

👉 **Google Colab Notebook** for a hands-on experience

Thank you!

www.crowd4sdg.eu