

Who is a (missing) migrant?





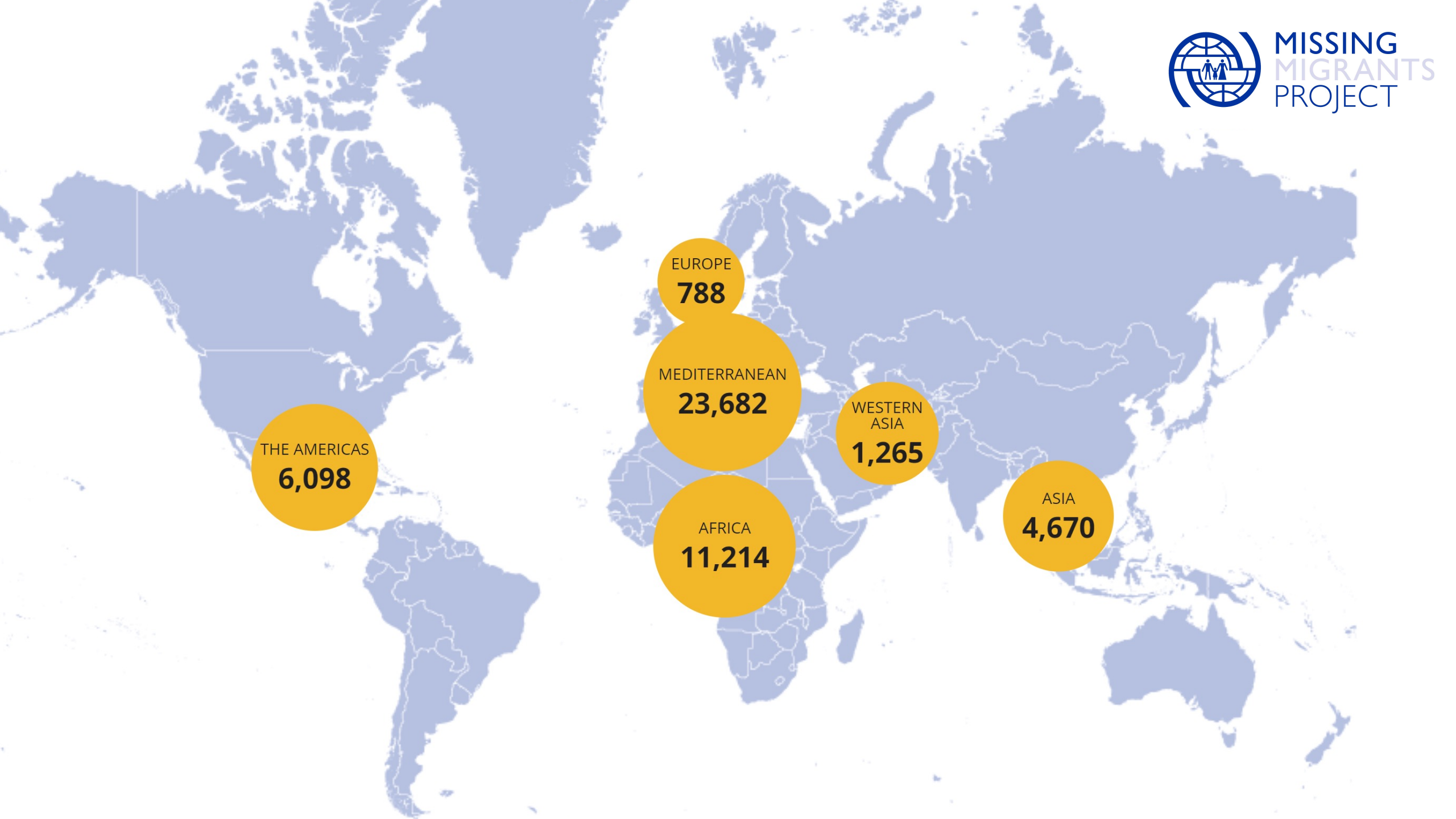






**47,700**

migrant deaths recorded since 2014





# IOM's Missing Migrants Project



Documents deaths and disappearances during migration

**Open-access global database**

SDG indicator 10.7.3

Core team + regional monitors



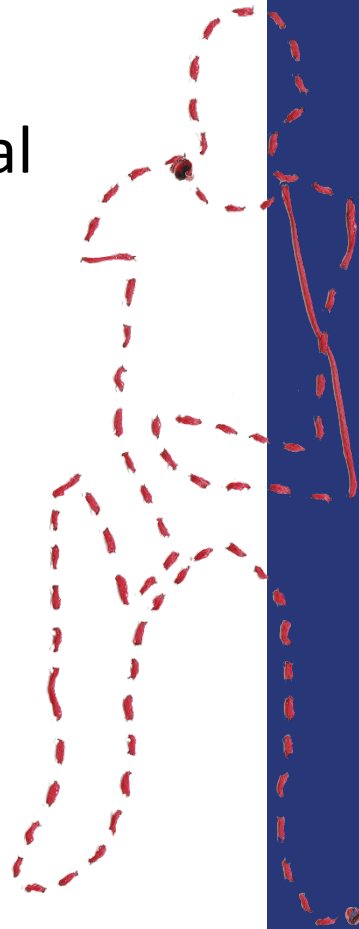
# Defining 'missing migrant' in MMP

## What is included in MMP?

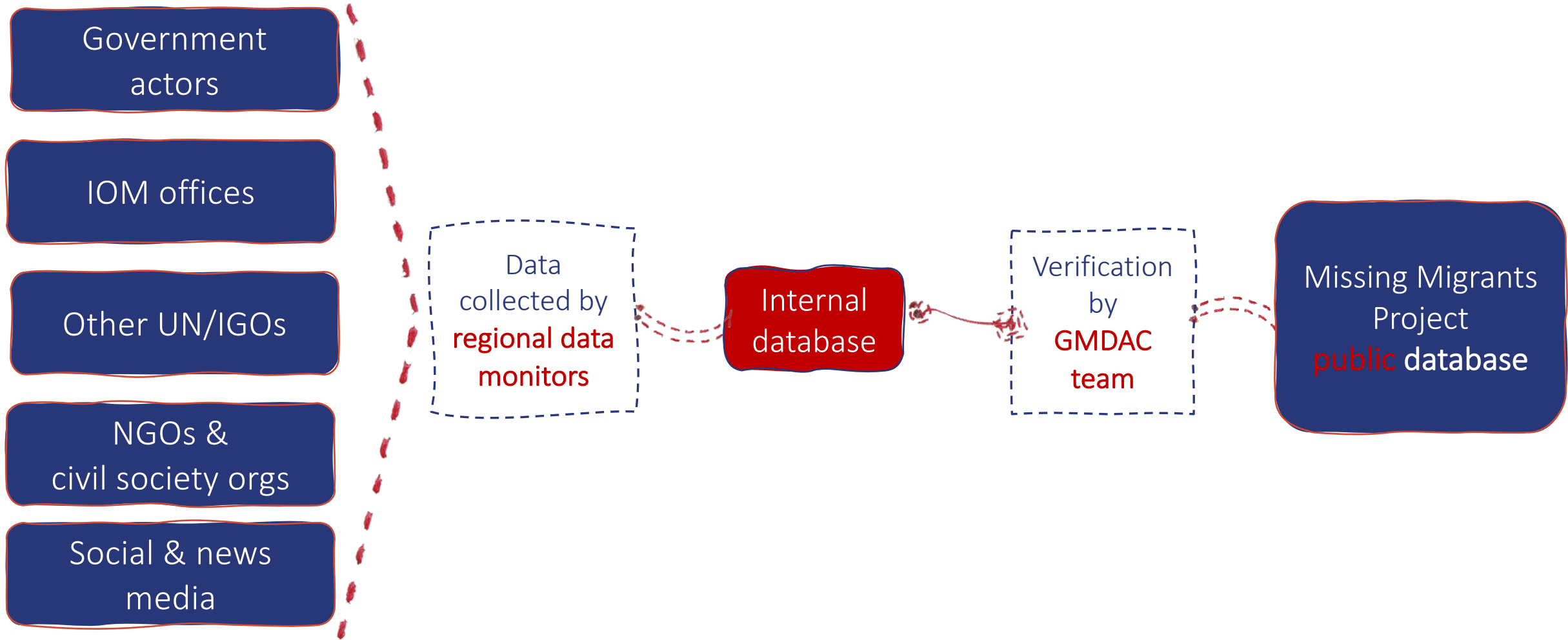
- **Deaths** at the external borders of states, or in the process of migration towards an international destination
- **Disappearances** during migration in which a person is **presumed dead**
- **Rationale:** approach identifies risks linked to irregular migration journeys

## What is excluded from MMP?

- Deaths in immigration **detention**, camp settings, reception centres
- Deaths related to an individual's irregular status while **residing** in a foreign country (e.g. from labour exploitation)
- Deaths related to **internal displacement**
- **Missing persons reports** when there is no presumption of death



# Methodology: 5 main sources of data



# Data gaps and challenges

- **Limited data from official sources:** lack of centralization & disaggregation by migratory status
- **Under-reporting of deaths and disappearances:** migrants or their families may be afraid of reporting deaths or disappearances due to fear of legal consequences
- **Invisibility:** lack of options for safe and legal mobility increases the likelihood that migrants take dangerous, remote routes with the aim of avoiding detection.
- **Remains not recovered and not identified:** bodies lost on remote routes means that identification nearly impossible

## Information gap: impact on families and communities left behind

- **Thousands of family members searching for loved ones** missing during migration with little support

# Input Data

GDELT 2.0

# Filtering

CRISISLEX

MMP master dataset

# Sorting

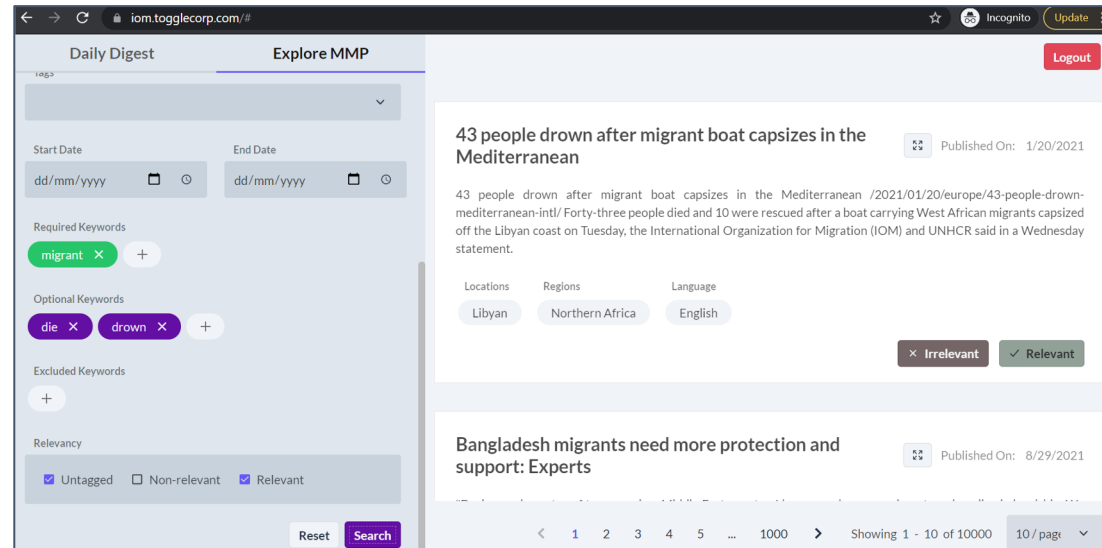
BM25

MMP keywords

GDELT 2.0

MMP master dataset

# Media Monitoring Tool



# Output Data

MMP master dataset

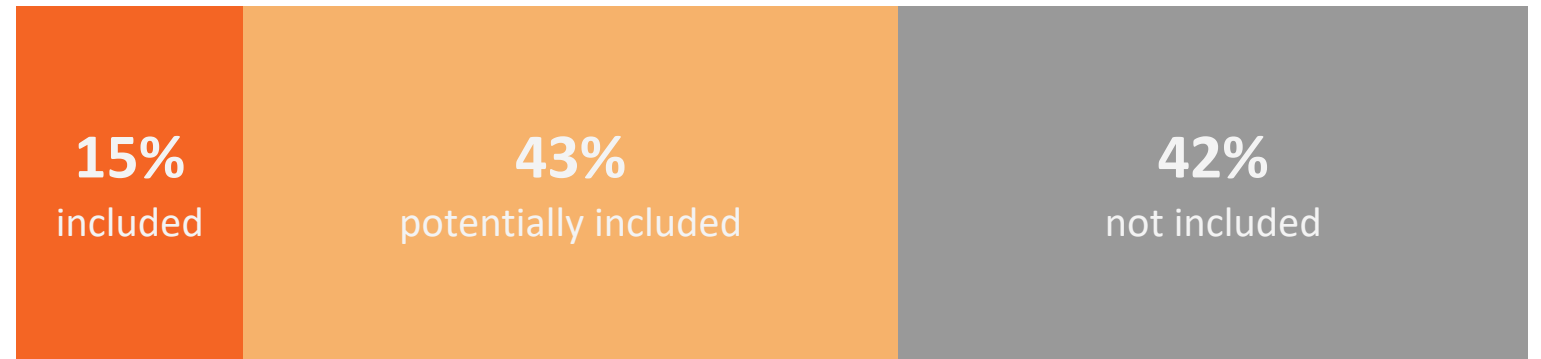
Irrelevant dataset

#english #french #spanish #arabic

# Input Data

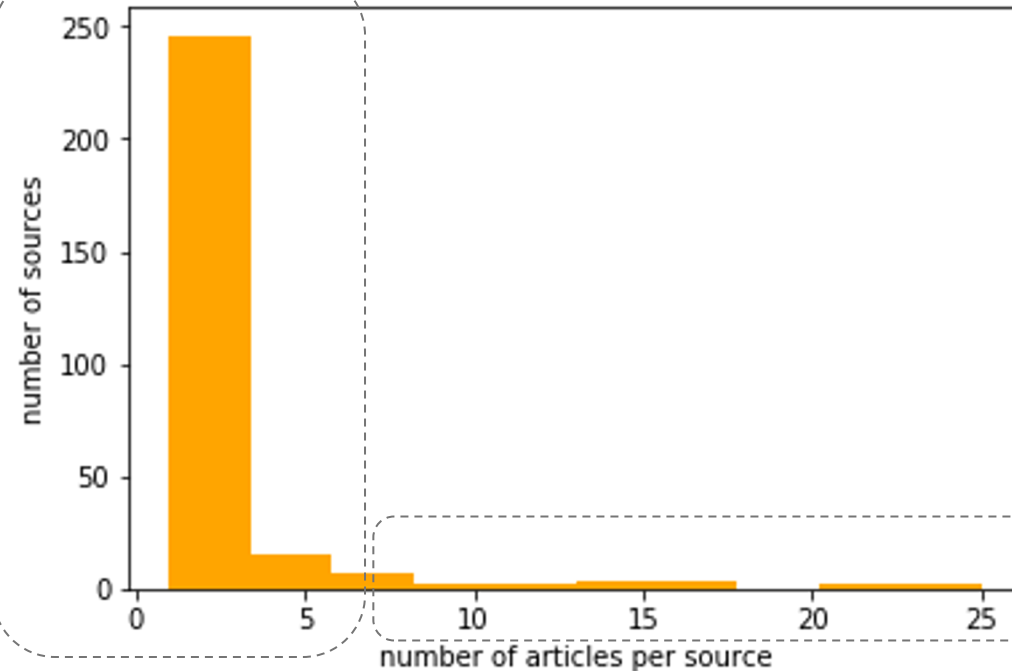
GDELT 2.0

## GDELT Coverage of MMP news articles



**Many sources with few related documents**

e.g.  
china.org.cn  
cibercuba.com



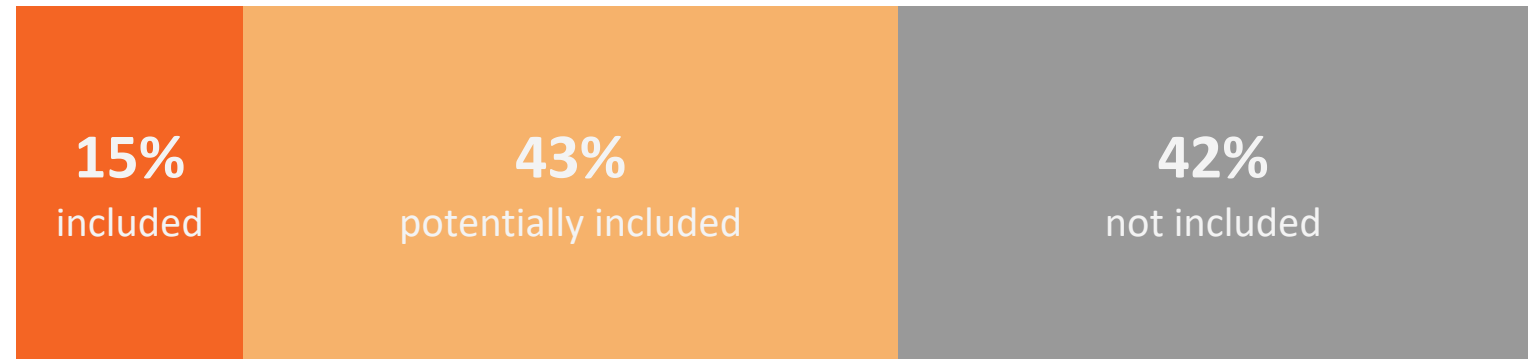
**Few sources with several related documents (up to 30)**

e.g.  
hoytamaulipas.com  
breitbart.com  
debate.com.mx

# Input Data

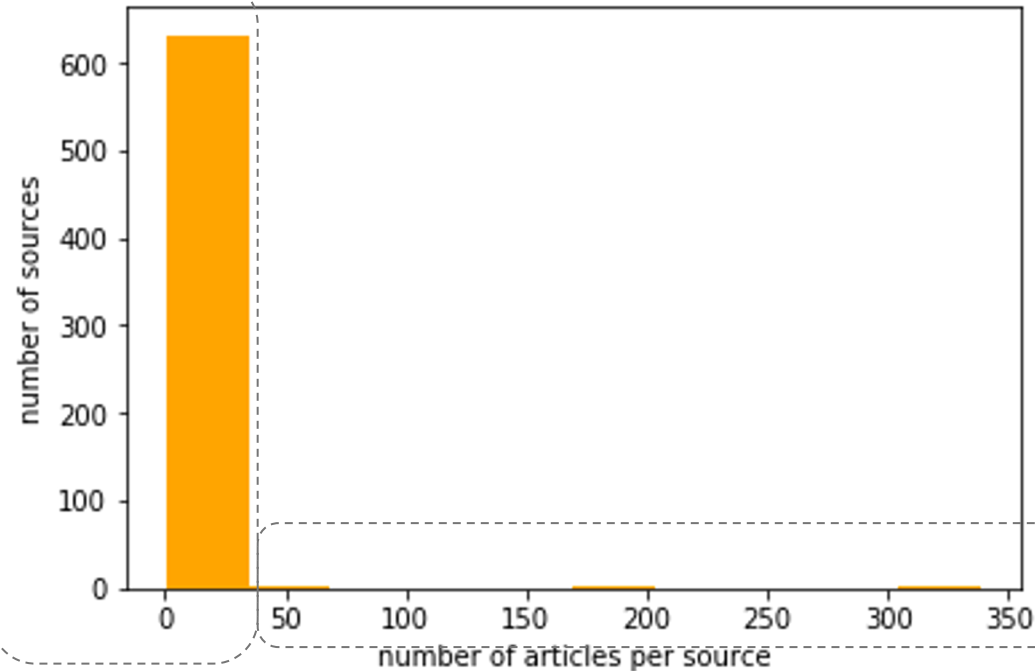
GDELT 2.0

## GDELT Coverage of MMP news articles



**Many sources with few related documents**

e.g.  
g1.globo.com  
galsen221.com



**Few sources with several related documents (up to 340)**

e.g.  
twitter.com  
facebook.com  
hcg.gr

# Filtering

CRISISLEX

MMP master  
dataset

*Set of predefined themes*

VS

*Themes in relevant dataset*



# Sorting

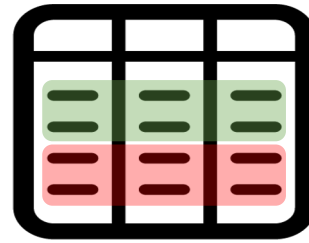
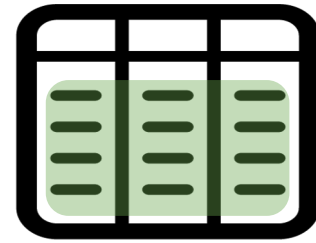
BM25

MMP keywords

GDELT 2.0

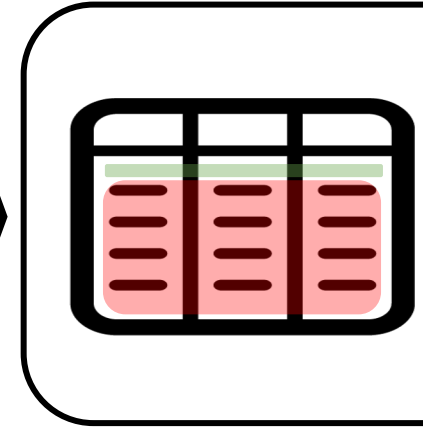
MMP master dataset

## Distribution of data



853 relevant  
(29%)

2050 irrelevant  
(71%)



x77  
datasets

11 relevant  
(0.5%)

2050 irrelevant  
(99.5%)

## Information Retrieval Task

- *Elastic Search*
  - *BM25 similarity metric*
  - *Fine-tuning of hyperparameters according to MMP dataset (grid search)*

## Evaluation Framework

- *Recall@n*

# Sorting

BM25

MMP keywords

GDEL2.0

MMP master dataset

R@11 = 0.36

k1 = 0

b = 0

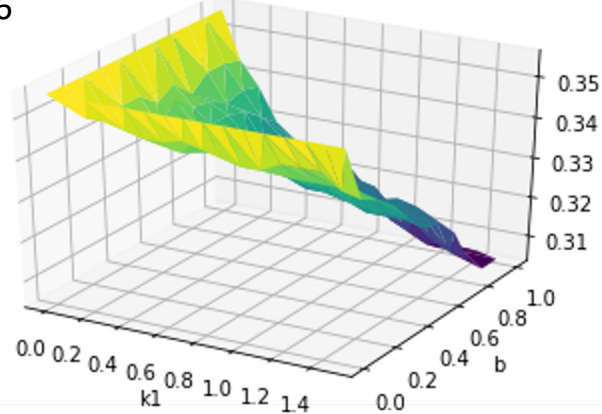
k1 = 0

b = 1

k1 = 1.5

b = 0

Recall@11 for English Language

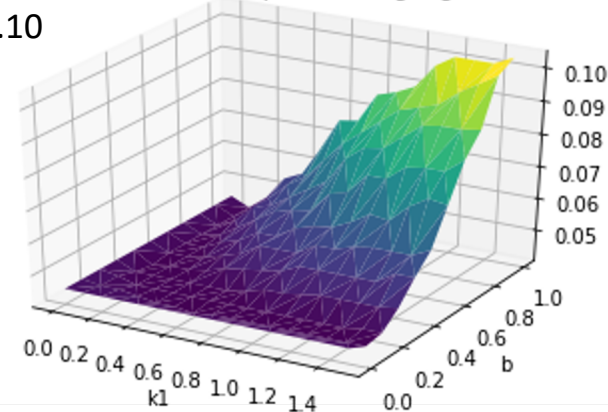


R@4 = 0.10

k1 = 1.5

b = 1

Recall@4 for Spanish Language

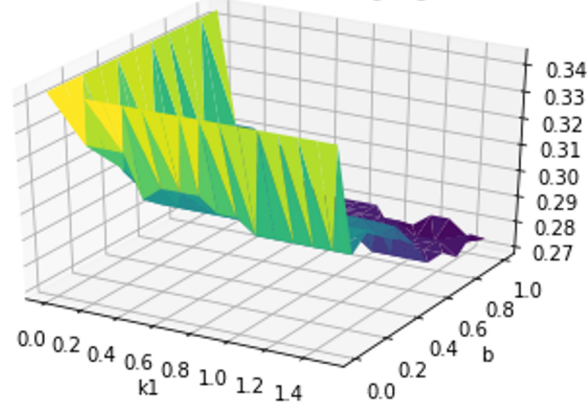


R@2 = 0.34

k1 = 0

b = 0

Recall@2 for French Language



R@1 = 0.01

k1 = 0

b = 0

Recall@1 for Arabic Language

