

# CRiTERIA project

Short overview and Idiap collaboration



# CRiTERIA Project

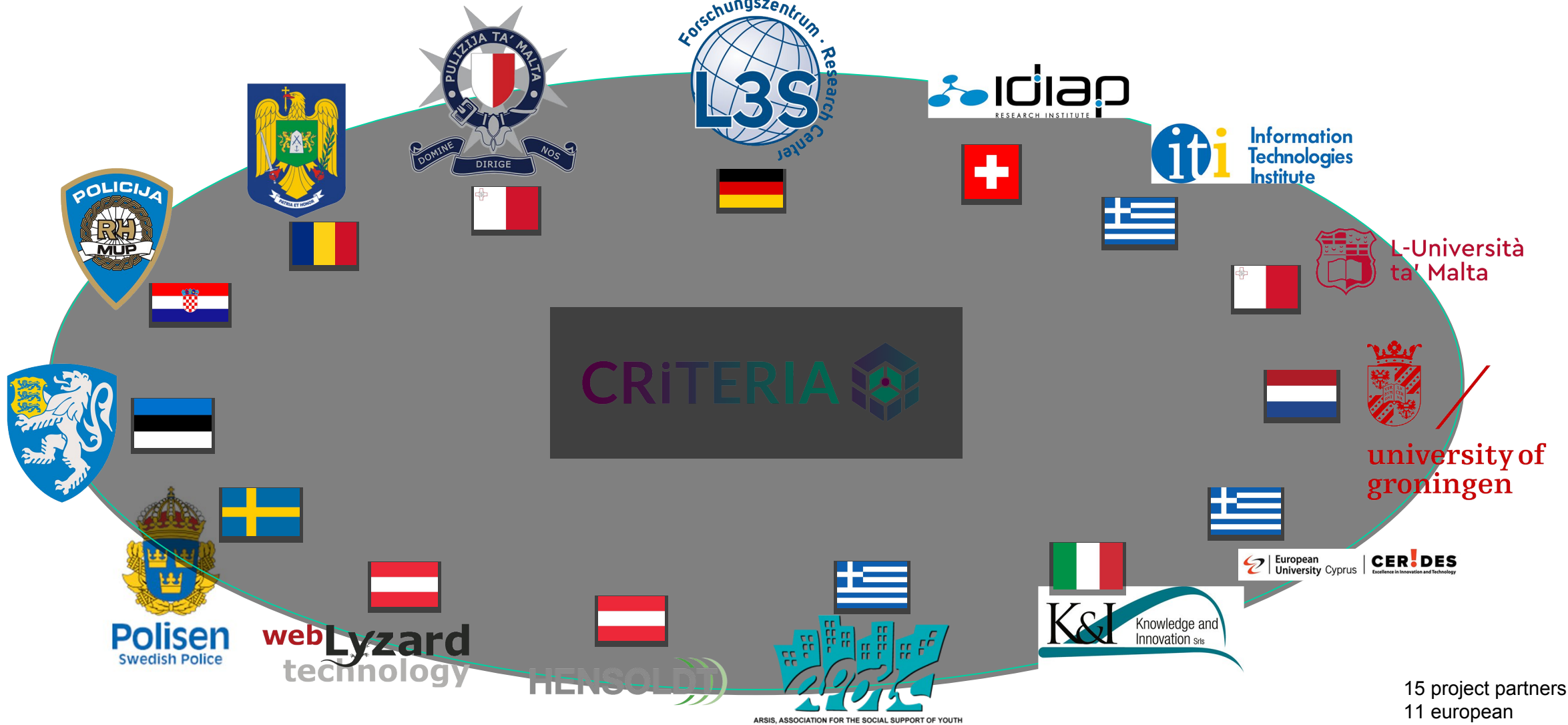
*Comprehensive data-driven **Risk and Threat** assessment methods for the **Early and Reliable Identification, validation and Analysis of migration-related Risks** (CRiTERIA)*



- The main idea behind is to *Expand Existing Risk Analysis Methods and Tools for Border Agencies.*
- A 3-year project that started in September last year.



# CRiTERIA - Interdisciplinary Consortium (Project Partners)



15 project partners  
11 european countries



CRiTERIA - Comprehensive data-driven Risk and Threat Assessment Methods for the Early and Reliable Identification, Validation and Analysis of migration-related risks - has received funding from the European Union's Horizon 2020 research and innovation action program under grant agreement No 101021866.



# CRiTERIA - Goal

- To develop a novel Risk and Vulnerability **Analysis framework for border agencies** that is comprehensive and *human rights sensitive*.
  - Allow a novel **multi-perspective** analysis methodology.
  - Support **multi-source and multi-lingual** analysis technologies and tools.
  - Verifiable, transparent and understandable by humans.



# CRITERIA - Web Dashboard

CRITERIA

[Configuration](#) [Date Range: 10 Nov 2022 – 24 Nov 2022](#) [Data Sources: Multiple / EN](#)

276.5k

[Help](#) [Logout](#)

[Keyword Filter](#) [Data Export](#)

June
July
August
September
October
November

Help Text ▼ To trigger a search, select at least one rectangular marker on the left - global filter, associations, etc. (results will be shown by publication date)

### ASSOCIATIONS

<input type="checkbox"/> world cup	22397
<input type="checkbox"/> ukraine	12589
<input type="checkbox"/> qatar	12436
<input type="checkbox"/> america	34676
<input type="checkbox"/> cup	26665
<input type="checkbox"/> fifa	7124
<input type="checkbox"/> rebounds	7642
<input type="checkbox"/> points	28364
<input type="checkbox"/> russia	13009
<input type="checkbox"/> england	15273
<input type="checkbox"/> u.k.	18728

### METADATA

Sources - Aggregated

<input type="checkbox"/> Sports	66206
<input type="checkbox"/> Other	49831
<input type="checkbox"/> Business & Finance	35564
<input type="checkbox"/> Politics	35094
<input type="checkbox"/> Entertainment & Culture	33767
<input type="checkbox"/> Crime & Law	25562
<input type="checkbox"/> Health	16919
<input type="checkbox"/> Emergency & Accident	6425
<input type="checkbox"/> Education	3406
<input type="checkbox"/> Environment	2330
<input type="checkbox"/> Religion	954

**24 Nov** **BLAG Ghana to construct Cruyff Court in Tema.**  
The Cruyff Foundation, in collaboration with Cargill Ghana Limited and the Tema Metropolitan Assembly (T...  
[graphic.com.gh](#)

**24 Nov** **VIDEO: Prez Akufo-Addo in Qatar for Ghana v Portugal.**  
President Nana Addo Dankwa Akufo-Addo will be in attendance at Stadium 974 today when the Black Stars ...  
[graphic.com.gh](#)

**24 Nov** **Majority has let us down - Minority in Parliament cries out.**  
The Minority caucus in Parliament has said it felt let down and betrayed by the attitude of the Majority group...  
[graphic.com.gh](#)

**24 Nov** **Open herbal unit in every govt hospital - Prof. Samuel Ato Duncan tells Health Ministry.**  
The President of the Ghana Federation of Traditional Medicine Practitioners Associations (GHAFTRAM), Pro...  
[graphic.com.gh](#)

**24 Nov** **United Way Worldwide to unlock youth potential.**  
The President and Chief Executive Officer (CEO) of United Way Worldwide (UWW), Angela F. Williams, has r...  
[graphic.com.gh](#)

**24 Nov** **New low-cost airline Fly Atlantic to operate transatlantic flights from Northern Ireland from 2024.**  
A new low-cost airline is set to offer transatlantic flights from Northern Ireland . The announcement by Fly A...  
[dailymail.co.uk](#)

**24 Nov** **Why some Brazilians won't be wearing their national soccer colors for the World Cup.**  
Fans arrive to the 2014 FIFA World Cup Brazil round of 16 match between Brazil and Chile at Estadio Mineir...  
[npr.org](#)

**24 Nov** **China expands lockdowns as COVID-19 cases hit daily record.**  
Enlarge this image . Residents line up for COVID tests in Beijing, Thursday, Nov. 24, 2022. China is expanding ...  
[npr.org](#)

america arizona australia brazil california  
canada china **CUP** doha egypt england  
episodes facebook fifa florida fox ftx game  
indonesia ireland korea kyiv i.a. layoffs london  
michigan missile moscow nairobi points  
poland qatar rebounds road russia  
shooting state team tennessee texas thanksgiving  
tournament u.k. **ukraine** virginia wales  
washington world yards

### BOOKMARKS

Human Conflict

<input type="checkbox"/> Conflict	28052
<input type="checkbox"/> Migration	10677
<input type="checkbox"/> Geopolitics	2144
<input type="checkbox"/> Human Trafficking	1666

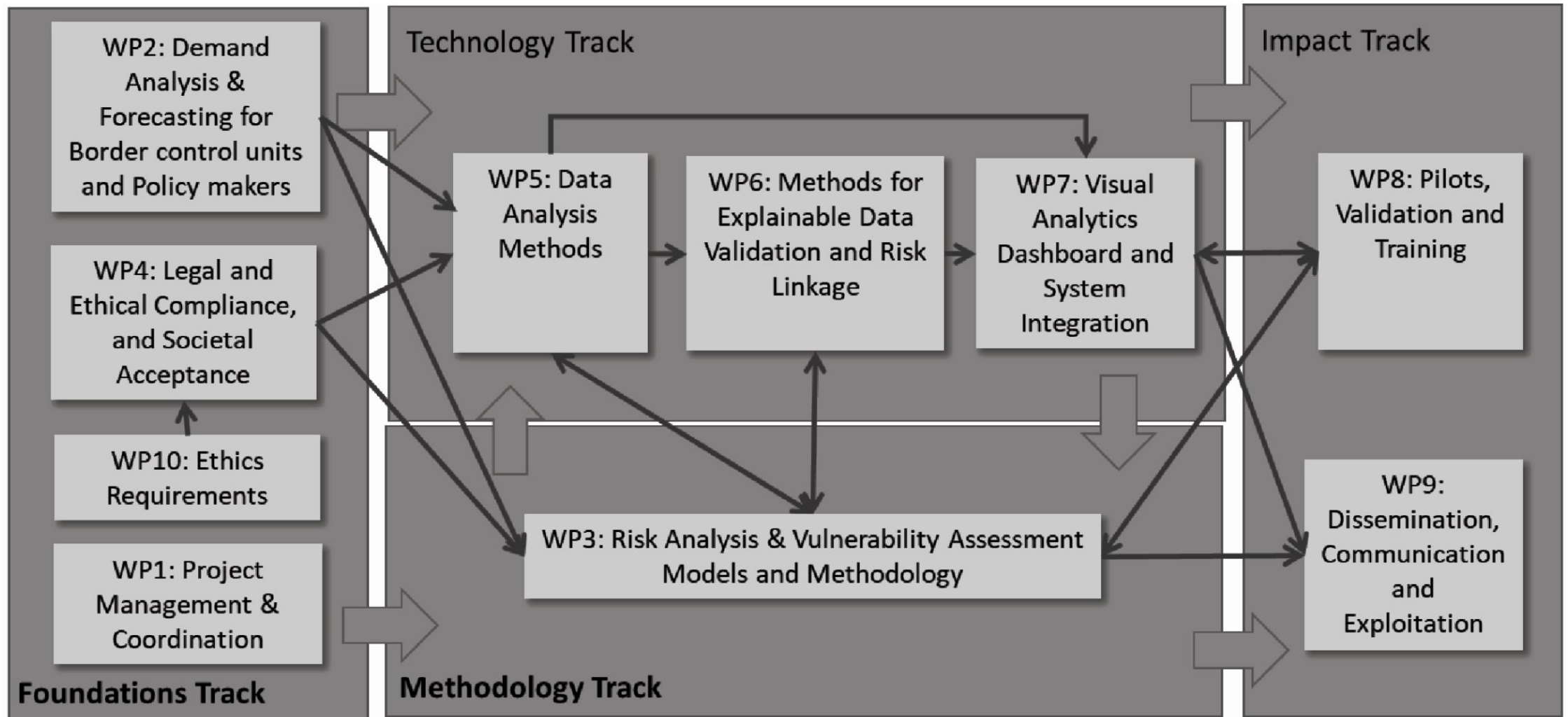
Economic Development

Push/Pull Factors

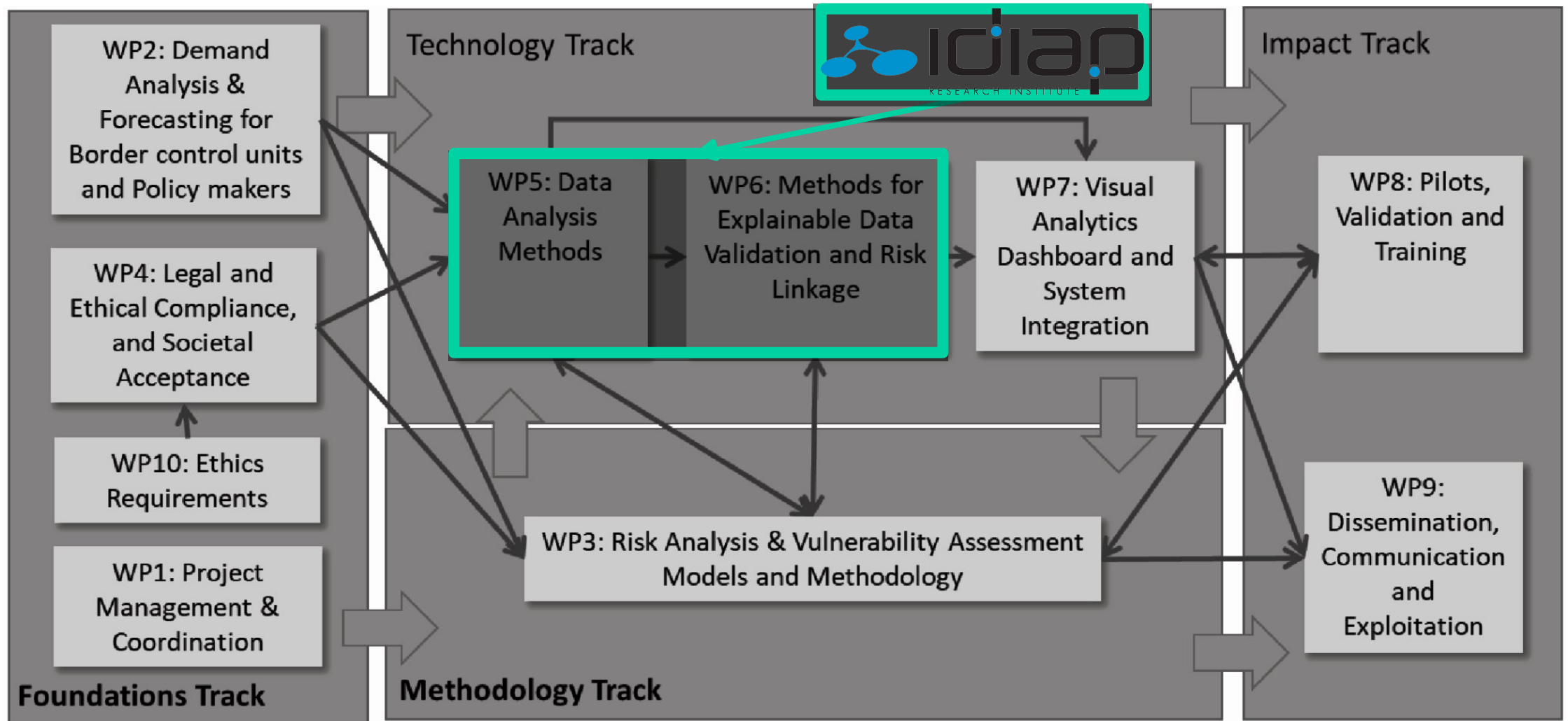
<input type="checkbox"/> Push - Economy	7772
<input type="checkbox"/> Push - Health	7442
<input type="checkbox"/> Pull - Economy	5825
<input type="checkbox"/> Pull - Health	3718
<input type="checkbox"/> Pull - Environment	2287



# CRiTERIA - Work Packages



# CRiTERIA - Work Packages



CRiTERIA - Comprehensive data-driven Risk and Threat Assessment Methods for the Early and Reliable Identification, Validation and Analysis of migration-related risks - has received funding from the European Union's Horizon 2020 research and innovation action program under grant agreement No 101021866.



# Idiap contributions to CRiTERIA (Overview)



- Fact-Checking system (WP6)<sup>1</sup>.
- News sources reliability estimation model (WP6).
  - Pipeline to download, process and index huge snapshots of the WWW containing news articles (WP5).
- Cause-effect relationship identification and extraction on news articles (WP6)<sup>2,3</sup>.

[1] [Claim-Dissector: An Interpretable Fact-Checking System with Joint Re-ranking and Veracity Prediction](#) (Fajcik et al)

[2] [Efficient Causal Relation Identification Through a Prompt-based Few-shot Approach](#) (Burdisso et al)

[3] [Extracting Cause-Effect-Signal Triplets via Pre-trained Autoregressive Language Model](#) (Fajcik et al)







**Thanks!**  
*Questions?*

More info:

[www.roject-criteria.eu](http://www.roject-criteria.eu)



CRiTERIA - Comprehensive data-driven Risk and Threat Assessment Methods for the Early and Reliable Identification, Validation and Analysis of migration-related risks - has received funding from the European Union's Horizon 2020 research and innovation action program under grant agreement No 101021866.



# Idiap Research Institute

## Contributions to CRiTERIA

(as of Nov. 2022)



Sergio  
Burdisso



Martin  
Fajčík

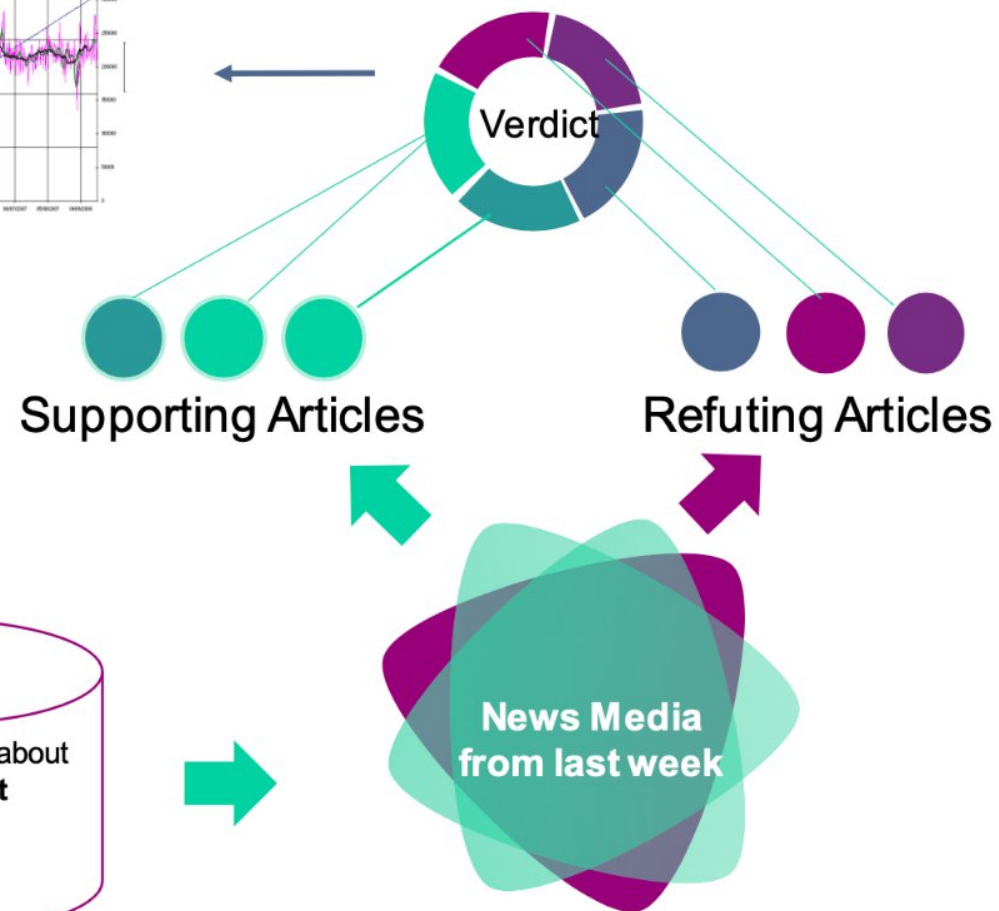
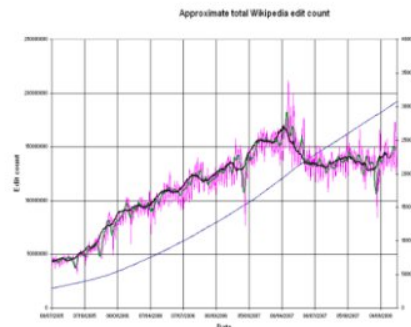


Petr  
Motlíček



# Fact Checking in Risk Assessment

- For a **specific location**
  - Prepare a **set of automatically analyzed claims** for a given risk indicator.
- Analyze their occurrence/veracity **over time**.



- Layoffs in Petersburg cause a wave of unemployment.
- A factory was closed in Pskov oblast.
- Companies pulling back from Novgorod.

A set of claims about **Unemployment push factor** in border area



# Fact Checking System

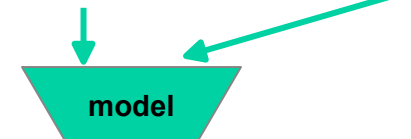
- A novel SOTA system trained on **standard fact-checking dataset** FEVER (Throne et al., 2018)
- **Provides explanations** on document/paragraph/sentence/token level.
- Reasoning over whole **Wikipedia**.
- Soon to be published: Martin Fajčík, Petr Motlíček, and Pavel Smrž. "[Claim-Dissector: An Interpretable Fact-Checking System with Joint Re-ranking and Veracity Prediction.](#)" *arXiv preprint arXiv:2207.14116* (2022).
- **TODO:** Adapt Fact-Checking system to the **news domain data**.
  - Can we trust in the retrieved articles? (next slide)

Wikipedia



## Claim

"American Sniper (book) is about a sniper with 170 officially confirmed kills."



**American Sniper\_RS(0.97)\_PS[0.91]**

-- American Sniper is a 2014 American biographical war drama film directed by Clint Eastwood and written by Jason Hall

**American Sniper\_RS(0.95)\_PS[0.96]**

-- The film follows the life of Kyle, who became the deadliest marksman in U.S. military history with 255 kills from four tours in the Iraq War, 160 of which were officially confirmed by the Department of Defense

**American Sniper (book)\_RS(0.93)\_PS[0.79]**

-- American Sniper : The Autobiography of the Most Lethal Sniper in U.S. Military History is an autobiography by United States Navy SEAL Chris Kyle, written with Scott McEwen and Jim DeFelice

**American Sniper (book)\_RS(0.92)\_PS[1.00]**

-- With 255 kills, 160 of them officially confirmed by the Pentagon, Kyle is the deadliest marksman in U.S. military history

**American Sniper\_RS(0.44)\_PS[0.21]**

-- The film became a

**American Sniper (book)\_RS(0.42)\_PS[0.31]**

-- Its film adaptation directed by Clint Eastwood and featuring Bradley Cooper as Kyle was released in 2014

↓  
**Refute (P=0.9)**

James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018. FEVER: a Large-scale Dataset for Fact Extraction and Verification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 809–819, New Orleans, Louisiana. Association for Computational Linguistics.



CRITERIA - Comprehensive data-driven Risk and Threat Assessment Methods for the Early and Reliable Identification, Validation and Analysis of migration-related risks - has received funding from the European Union's Horizon 2020 research and innovation action program under grant agreement No 101021866.

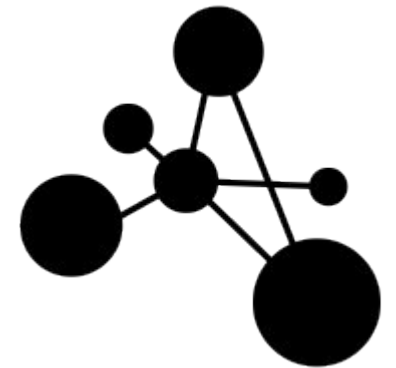


# News Sources Reliability Estimation

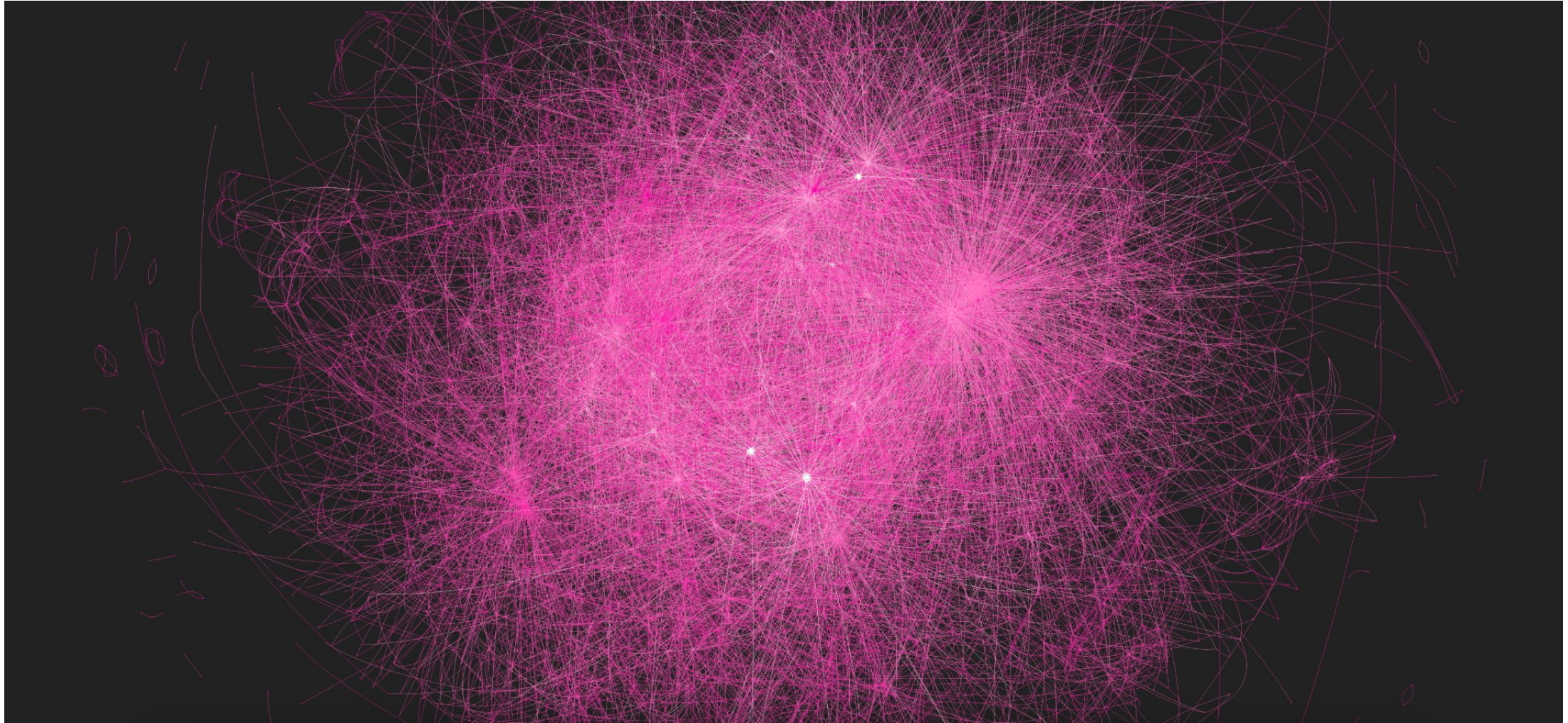
- Given a list of reliable and unreliable news sources, estimate the reliability of all other news sources based on how they're related to each other.
  - But... how news sources are related to each other in real life?

## Our contributions:

- Created a pipeline to download, process and index huge snapshots of the WWW containing news articles ([CC-News](#) snapshots).
- Automatic generation of network graph with (weighted) links among news sources for each language.
- Designed a reinforcement learning model to learn the reliability value of each source.
  - Accuracy above 85% using Wikipedia's "[perennial sources](#)" list as ground truth (after community consensus).



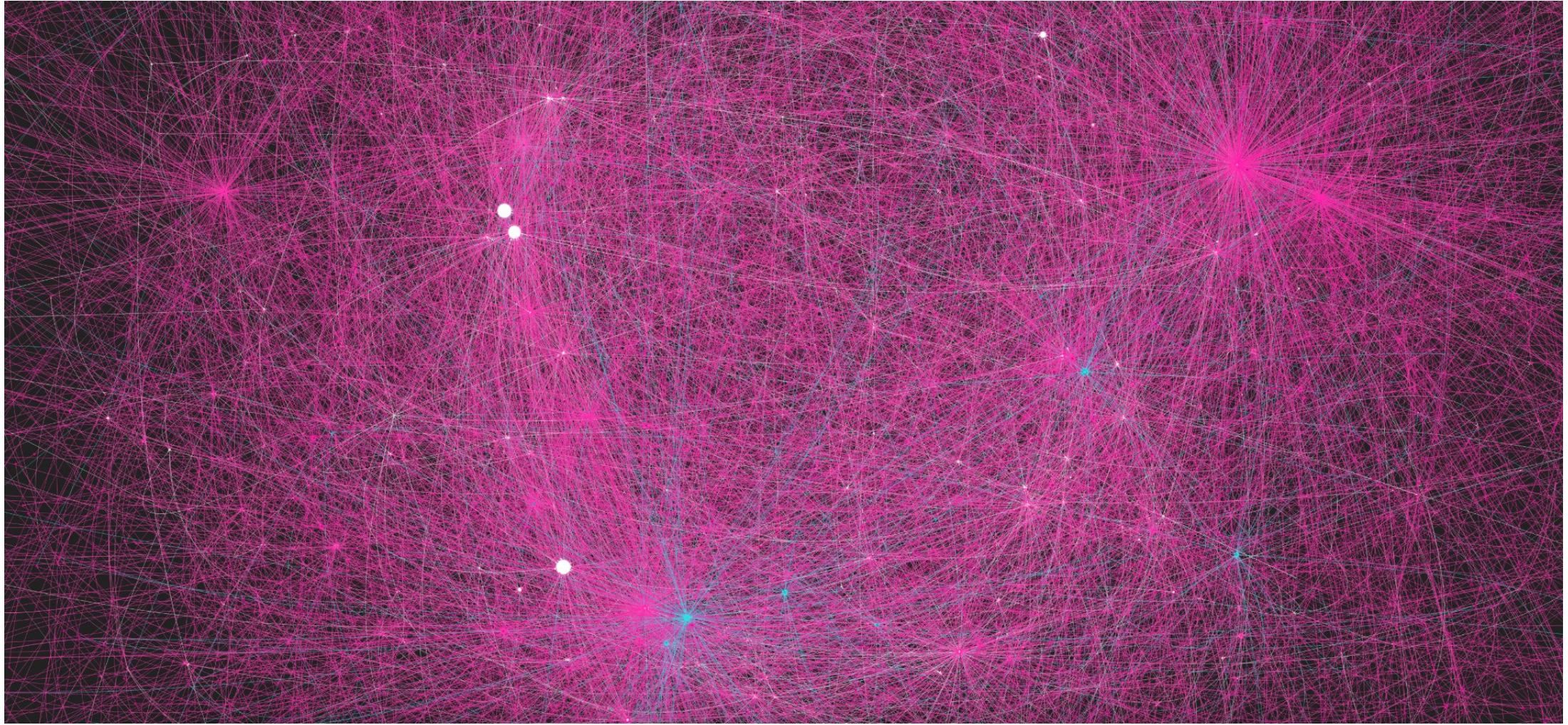
# News Graph (English)



CRiTERIA - Comprehensive data-driven Risk and Threat Assessment Methods for the Early and Reliable Identification, Validation and Analysis of migration-related risks - has received funding from the European Union's Horizon 2020 research and innovation action program under grant agreement No 101021866.



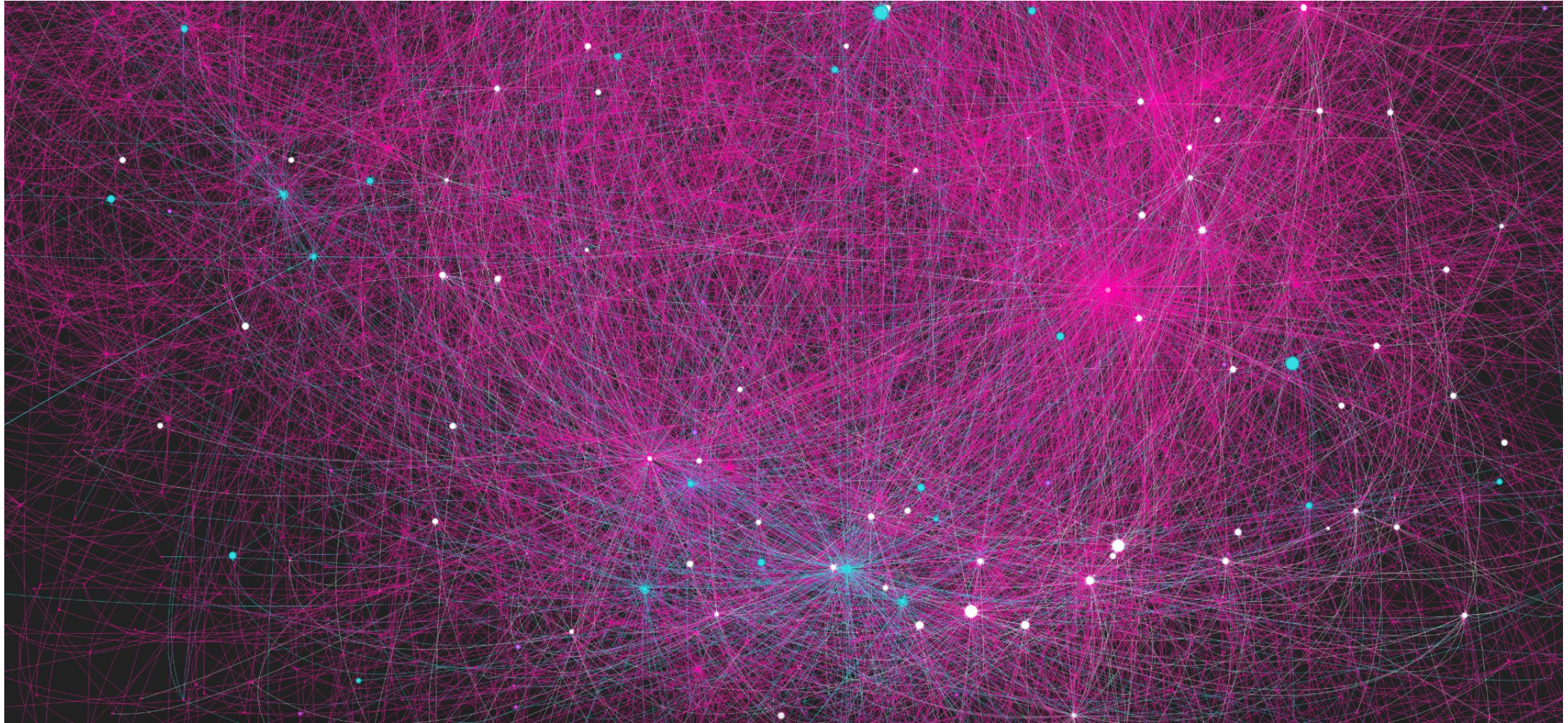
# News Graph + Ground Truth Seeds



CRiTERIA - Comprehensive data-driven Risk and Threat Assessment Methods for the Early and Reliable Identification, Validation and Analysis of migration-related risks - has received funding from the European Union's Horizon 2020 research and innovation action program under grant agreement No 101021866.



# News Graph + Learned Reliability Values



CRiTERIA - Comprehensive data-driven Risk and Threat Assessment Methods for the Early and Reliable Identification, Validation and Analysis of migration-related risks - has received funding from the European Union's Horizon 2020 research and innovation action program under grant agreement No 101021866.





# Reliability Estimation Ranking Example

## Top-15 most reliable sources:

1. [bbc.co.uk](http://bbc.co.uk) (17.2920)
2. [nytimes.com](http://nytimes.com) (16.2787)
3. [theguardian.com](http://theguardian.com) (15.4281)
4. [washingtonpost.com](http://washingtonpost.com) (10.2857)
5. [huffpost.com](http://huffpost.com) (9.7438)
6. [latimes.com](http://latimes.com) (9.3507)
7. [reuters.com](http://reuters.com) (8.0091)
8. [usatoday.com](http://usatoday.com) (7.9496)
9. [bloomberg.com](http://bloomberg.com) (7.8193)
10. [cnn.com](http://cnn.com) (7.8148)
11. [foxbusiness.com](http://foxbusiness.com) (7.5828)
12. [telegraph.co.uk](http://telegraph.co.uk) (7.5106)
13. [forbes.com](http://forbes.com) (7.4331)
14. [npr.org](http://npr.org) (7.4250)
15. [thehill.com](http://thehill.com) (7.2795)

## Top-15 least reliable sources:

1. [thepointsguy.com](http://thepointsguy.com) (-13.1981)
2. [swarajyamag.com](http://swarajyamag.com) (-12.6034)
3. [infowars.com](http://infowars.com) (-12.5987)
4. [veteranstoday.com](http://veteranstoday.com) (-12.4937)
5. [breitbart.com](http://breitbart.com) (-12.3822)
6. [globalresearch.ca](http://globalresearch.ca) (-11.8422)
7. [newswars.com](http://newswars.com) (-11.7647)
8. [oann.com](http://oann.com) (-6.6659)
9. [projectveritas.com](http://projectveritas.com) (-6.6637)
10. [vgchartz.com](http://vgchartz.com) (-6.6618)
11. [bild.de](http://bild.de) (-6.6575)
12. [dailysabah.com](http://dailysabah.com) (-6.6462)
13. [telesurenglish.net](http://telesurenglish.net) (-6.6385)
14. [presstv.com](http://presstv.com) (-6.6230)
15. [tass.com](http://tass.com) (-6.6217)



# Cause-effect relation identification in News...

EMNLP's CASE workshop challenge: "Event Causality Identification with Causal News Corpus".

- Efficient Causal Relation Identification Through a Prompt-based Few-shot Approach (Burdisso et al).
- Extracting Cause-Effect-Signal Triplets via Pre-trained Autoregressive Language Model (Fajcik et al).

