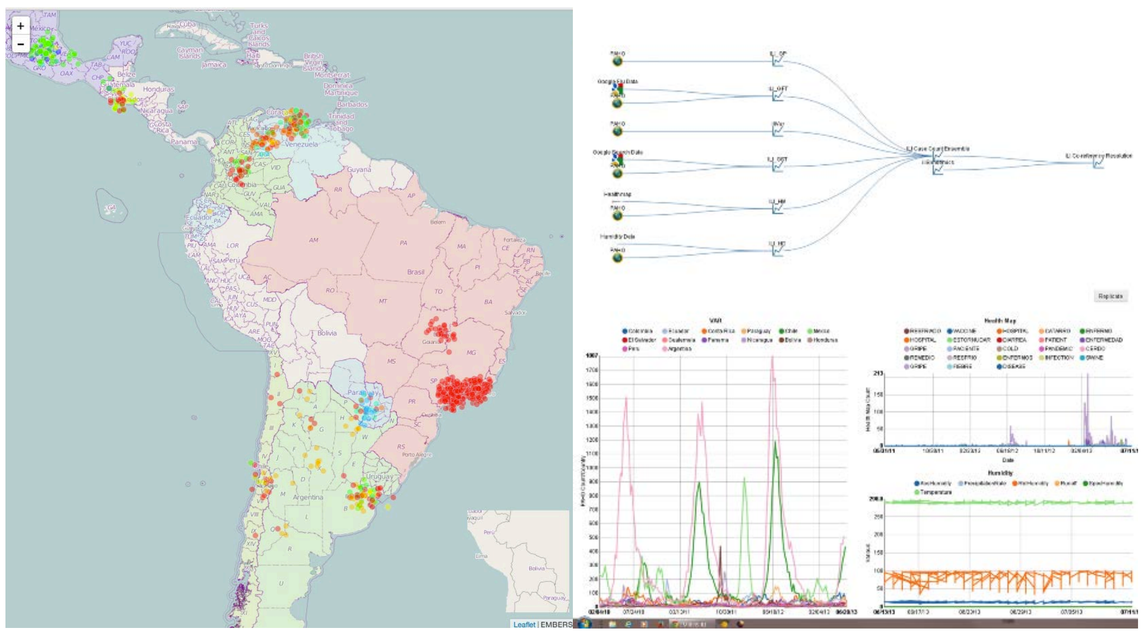
Open Source Indicators (OSI)

Automatically Anticipating and Detecting Significant Societal Events Using Publicly Available Data

Program Manager: Dr. Jason Matheny; E-mail: jason.matheny@iarpa.gov

Goal

Provide early warning of significant societal events through continuous and automated, real-time global monitoring of diverse, publicly available data.



Approach

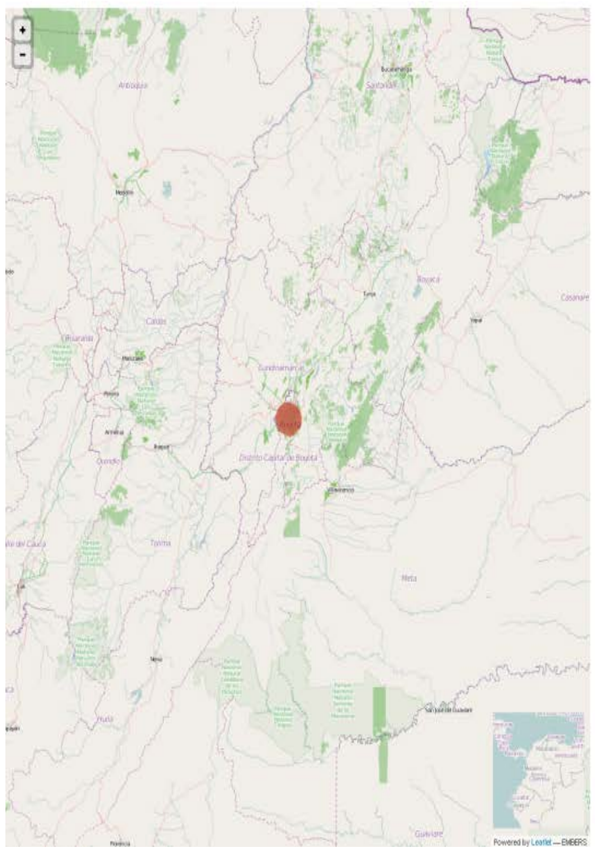
Train automated systems to detect patterns in social media, web search, news feeds, internet traffic, Wikipedia edits, overhead imagery, online reservation systems, financial markets, and other public data that precede events of interest. Events include civil unrest, foreign election results, and disease outbreaks.

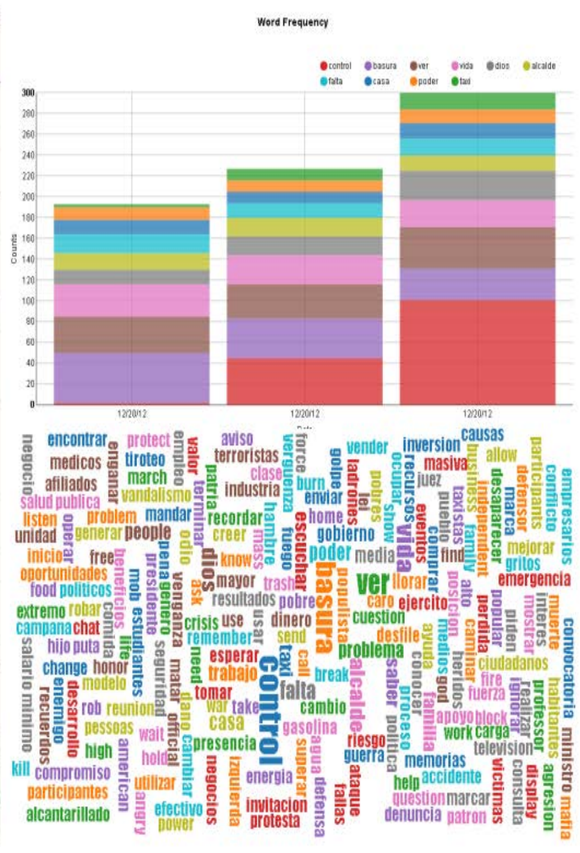
Program schedule: April 2012 - April 2015.

Evaluation

Methods are evaluated against thousands of real-world events in Latin America, the Middle East, and North Africa. Forecasts are scored for accuracy, lead-time, precision, and recall, compared to the earliest news mention.

OSI has achieved an average 8-day lead-time for warnings of civil unrest events, 6-day lead-time for rare disease events, 85% accuracy for political elections, 70% accuracy for influenza cases, and accurate geolocation of social media to within 6km of actual locations.





Event Typology

01 – Civil Unrest	011 – Employment & Wages	0111 – Non-violent Civil Unrest
		0112 – Violent Civil Unrest
	012 – Housing	0121 – Non-violent Civil Unrest
		0122 – Violent Civil Unrest
	013 – Energy & Resources	0131 – Non-violent Civil Unrest
		0132 – Violent Civil Unrest
	014 – Other Economic Policies	0141 – Non-violent Civil Unrest
		0142 – Violent Civil Unrest
	015 – Other Government Policies	0151 – Non-violent Civil Unrest
		0152 – Violent Civil Unrest
	016 – Other	0161 – Non-violent Civil Unrest
		0162 – Violent Civil Unrest
	017 - Unspecified	0171 – Non-violent Civil Unrest
		0172 – Violent Civil Unrest
02 – Vote	021 – Election	0211 – President/Prime Minister
		0212 – Governor
		0213 – Mayor
	022 – Referendum	0221 – "Yes"
		0222 – "No"
03 – Infectious Human Illness	031 – Rare Diseases	0311 – Bolivian Hemorrhagic Fever (Machupo)
		0312 – Cholera
		0313 – Hantavirus
		0314 – Yellow Fever
	032 – Pandemic	
	033 – Influenza Like Illness (ILI)	
04 – Economy	041 – Stock Index	0411 – Stock Index Increases
		0412 – Stock Index Decreases
	042 – Currency Exchange	0421 – Currency Exchange Increases
		0422 – Currency Exchange Decreases

Potential Impact

Significantly improved indications and warning capability for events not typically covered by traditional analysis.

Notable media attention:

- **The Wall Street Journal – "**U.S. Intelligence Community Explores More Rigorous Ways to Forecast Events," 9/5/14
- **USA Today** – "Data analysis allows researchers to predict disease outbreaks," 2/13/14
- **The Washington Post** – "The intelligence community gets social," 9/19/11.

Notable journal articles:

- **Science** – "The parable of google flu traps in big data analysis," Lazer et al., 3/14/14
- **PLOS ONE** – "The Twitter of Babel: Mapping world languages through microblogging platforms," Mocanu et al., 4/18/13

Research Teams

Harvard Medical School
San Diego State University
BASIS Technology
The University of Utah
Carnegie Mellon
Raytheon BBN Technologies
MITRE