



# GDELT

## Global Database of Events, Language, and Tone

Kalev Leetaru (Georgetown), Philip Schrodt (Parus Analytical Systems), Patrick Brandt (UT Dallas)

"a realtime social sciences earth observatory cataloging the global world"

### Global Event Database

Physical event database of a quarter-billion georeferenced records of societal-scale behavior in more than 300 categories for all countries 1979-present with daily updates of 100,000 new events per day.

### Global Beliefs Database

Detailed emotional and thematic latent undercurrents of global activity. Captures both context (food vs political protest) and evolving global reaction. Beyond events, captures "global dreams and fears" – the master narratives and themes that undergird global society.

### Global Biographical Database

Detailed biographical entries compiled for all persons, organizations, and social groups appearing in the news, including spatial profiles, relationships with all other groups, beliefs, and surrounding narrative dimensions.

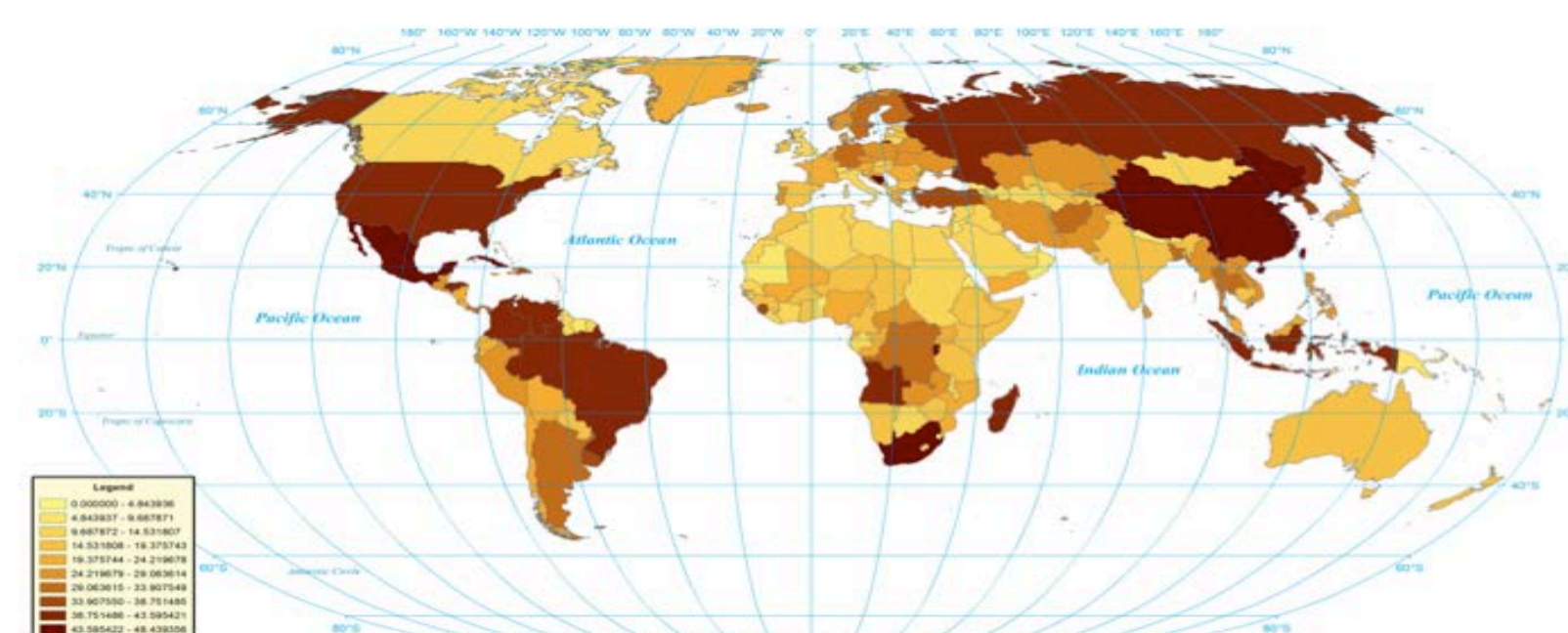
The Global Database of Events, Language, and Tone (GDELT) is an initiative to construct a catalog of human societal-scale behavior and beliefs across all countries of the world over the last two centuries down to the city level globally, to make all of this data freely available for open research, and to provide daily updates to create the first "realtime social sciences earth observatory."

Currently nearly a quarter-billion georeferenced events capture global behavior in more than 300 categories covering 1979 to present with daily updates. Each day GDELT monitors a substantial fraction of the global news media across every corner of the earth and compiles a database of all major societal activity worldwide, all 100% automated and 100% free and open, making the news "computable."

GDELT is designed to help support new theories and descriptive understandings of the behaviors and driving forces of global-scale social systems from the micro-level of the individual through the macro-level of the entire planet.

### Fulltext Geocoding at a Global Scale

An extremely sophisticated contextual fulltext geocoder powers the geographic needs of GDELT, identifying textual references of >8M locations worldwide, disambiguating based on whole-document context, and outputting centroid geographic coordinates: "from text to space". Robust to machine translation and tuned for social media; first paper on the geography of social media.



Global toponym namespace conflicts in GNS+GNIS

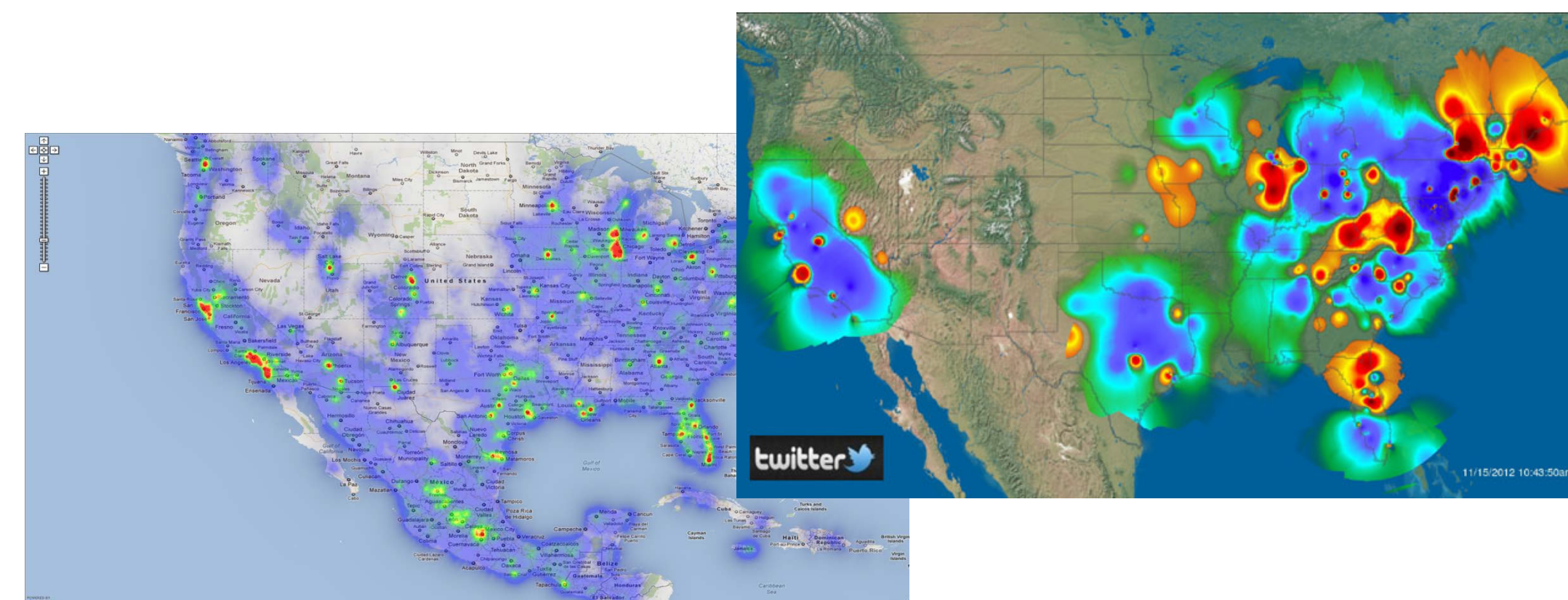
### Accuracy

A number of recent studies have evaluated GDELT's accuracy and found it to match human coding of news media, to align closely with on-the-ground citizen reporting in conflict zones like Syria, to align closely with ICEWS reports, and to exceed the accuracy of ICEWS data when plugged into the ICEWS forecasting models

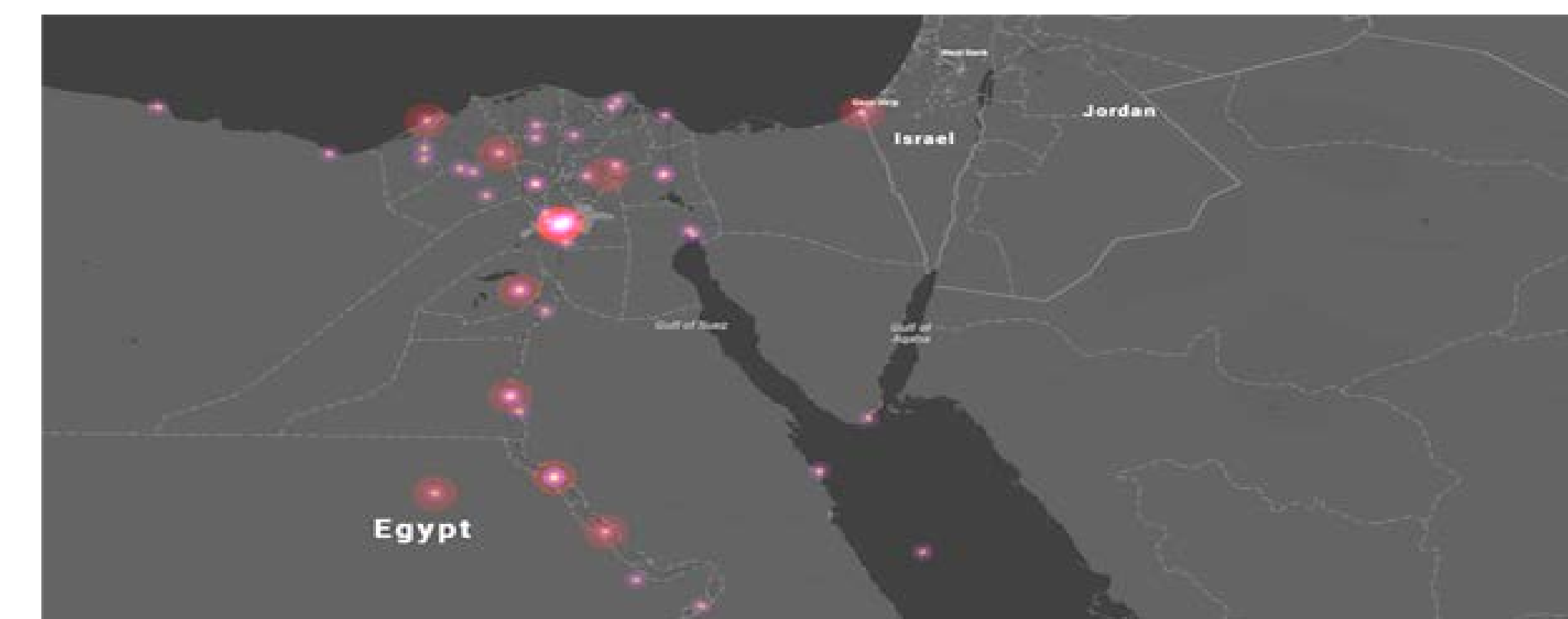
- **New GDELT-TABARI engine.** Specially configured and enhanced version of TABARI with new dictionaries running inside of massive pipelined architecture that rewrites each document to tune it for TABARI, resulting in dramatic increase in both recognized events and captured detail, especially in complex and fluid situations with incomplete information streams.
- **Nearly as accurate as human coding.** Recent paper found "hand coding full news stories does not lead to significant improvements in the accuracy or depth of actor information compared with machine coding by TABARI".
- **Closely aligned with on-the-ground citizen reporting.** Recent paper "identif[ied] strong correlation between GDELT violent events and Syria Tracker death tolls".
- **Only event stream to achieve 80% accuracy on DARPA ICEWS forecasting acceptance tests.** During the original DARPA ICEWS acceptance tests that independently measured the accuracy of each of the major event coding systems, only the TABARI engine achieved 80% forecasting accuracy. Systems by the other ICEWS vendors were not competitive and were unable to achieve the DARPA minimal acceptance accuracy levels. GDELT uses the new GDELT-TABARI engine which dramatically improves on the accuracy of the original TABARI engine that passed DARPA ICEWS acceptance tests.
- **Highly correlated with ICEWS.** Emerging comparisons are finding GDELT is highly correlated with ICEWS in all key AOIs tested, with GDELT having vastly higher spatial resolution and significantly greater capture of fluid situations.
- **Outperforms ICEWS data on ICEWS forecasting models.** Recent paper found that "GDELT data performs as well or better than the data in the original ICEWS" and "GDELT does a better job than the available ICEWS of predicting the five events of interest that were defined by the ICEWS project".

### A Database of Global Beliefs and Global Biographies

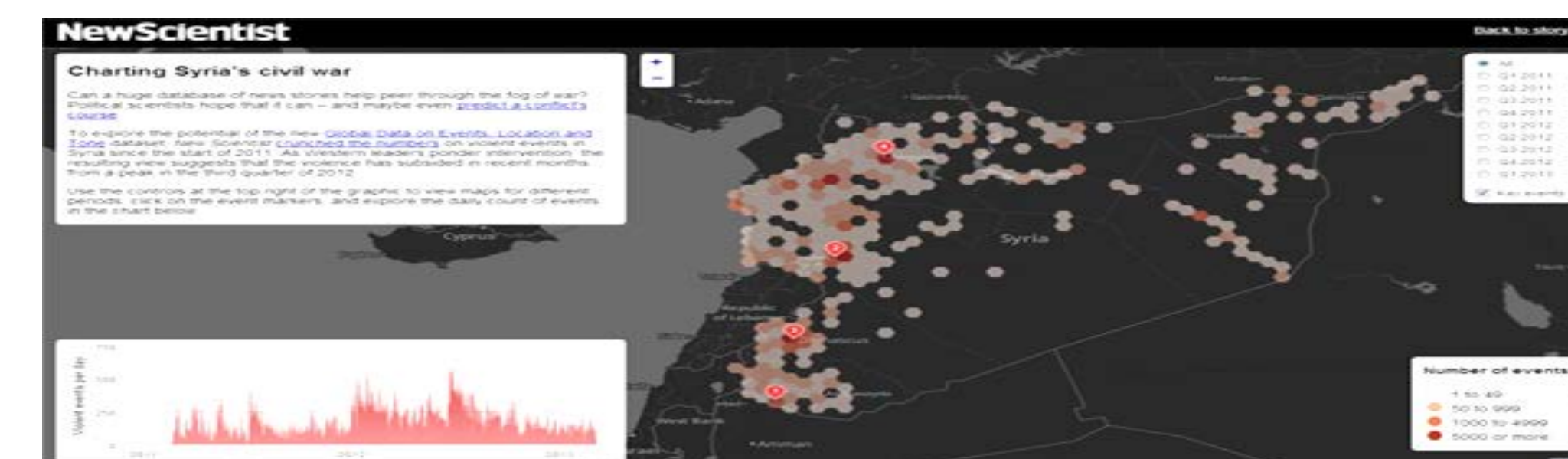
GDELT georeferences global emotions and thematic narratives and places them into a dynamic spatial network. Contextualizes events and captures realtime intensity of master narratives across the world, along with emergent societal-scale "dreams and fears". Produces spatial profiles of leaders and organizations over time. First realtime global-scale spatially-aware latent narrative catalog, can be integrated for realtime assessment of local populations.



Heatmap of spatial footprint of Mexican drug cartels based on global media; Realtime heatmap of emotion on Twitter



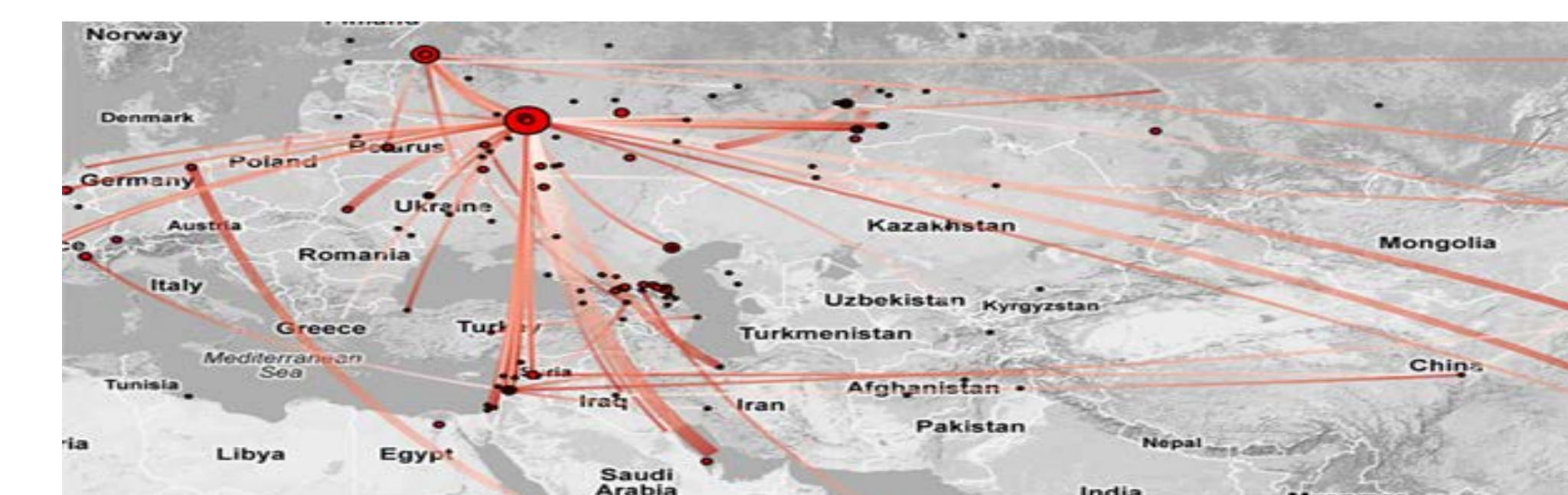
Dashboard of Egyptian unrest August 9-17th from global news (John Bieler)



Dashboard of Syrian civil war 2011-2013 by New Scientist



Animated map of worldwide protests 1979-present (John Bieler)



Spatial network visualization of Russian interactions (Rolf Fredheim)

We would like to specifically acknowledge the following organizations in making this research possible: Yahoo! Fellowship, Georgetown University, BBC Monitoring, Reed Elsevier's LexisNexis Group, Google and Google News, the School of Economic, Political and Policy Sciences at the University of Texas, Dallas, NSF grant SES-1004414, SES-092105, US Army CERL, and a Fulbright-Hays Research Fellowship for work at the Peace Research Institute.