• Mapping complete English text of Wikipedia: 80M locations and 40M dates via fulltext geocoding
• First large-scale examination of the geography of social media: Global Twitter Heartbeat
• Tracing spread of ideas through space over millions of books
• Spatial visualization of millions of declassified State Dept cables
• Compiling the world’s constitutions in digital form
• First large-scale study of how social media is used in conflict
• Mapping half a million hours of American television news (2.7 billion words of closed captioning)
• First live emotional “leaderboard” for television (NBC/SyFy)
• All the images of 600 million pages of books from 1,000 libraries and dating back 500 years
Datasets

- **NEWS**: Worldwide local news coverage in 100 languages (65 live translated) – online news preserved via Internet Archive
- **TELEVISION**: Collaboration with the Internet Archive to process more than 100 television stations across the US, updating daily
- **ACADEMIC LITERATURE**: 21 billion words covering 70 years (JSTOR/DTIC/CORE/CITeseer/IA)
- **BOOKS**: Collaboration with Internet Archive and HathiTrust to process 3.5 million books 1800-2015
- **HUMAN RIGHTS**: Half century of worldwide human rights reports
- **IMAGERY**: Large fraction of global news imagery processed via deep learning: objects/activities, OCR, logos, facial sentiment, geolocation
BAGHDAD, Iraqi leaders criticized Turkey on Monday for bombing Kurdish militants in northern Iraq with airstrikes that they said had left at least one woman dead.

Event Code: 111
Source: IRQ GOV
Target: TUR

Event Code: 223
Source: TUR
Target: IRQKRD REB
Physical Unrest

Positive/Negative: “Cautiously Optimistic” Trending

Anxiety
US Ebola News Coverage

- March 2014 WHO announcement
- First American infections
- Eric Duncan arrives in Dallas

Number American television news broadcasts per week mentioning "ebola"

- Steady ascent towards more and more positive coverage as “Western medicine miracles to the rescue” theme dominates coverage

Average “tone” of English language media coverage of “ebola”
Carbon Capture & Sequestration

• English coverage of CCS 2010-2015
• 32,000 websites, 250,000 people, 140,000 organizations, 50,000 locations
• Green cluster (center): senior American policymakers
• Green cluster (lower): “cap and trade” politicians
• Red cluster (bottom): American lawmakers on Congressional energy committee or sponsoring energy-related legislation
• Purple cluster (top right): climate skeptics
• Yellow (upper left): Australian politicians
• Pink (upper center): British politicians
• Periphery of all clusters: journalists and financial analysts who feature prominently in coverage or who write much of the coverage – Karolin Schaps (Reuters) and Alex Morales (Bloomberg News London) are attached to British political cluster; Tom Friedman is attached to American political cluster
• Red: Actual Ukraine
• (r=0.49)
Reimagining the Book
Kalev Hannes Leetaru

kalev.leetaru5@gmail.com
http://kalevleetaru.com
http://gdeltproject.org/