Dataminr®

AI And Public Data for Humanitarian and Emergency Response

Alex Jaimes, Ph.D. Chief Scientist & SVP of AI
Octavian Udrea, Ph.D. Principal Research Scientist

https://www.lensculture.com/articles/zed-nelson-the-new-nation-makers-of-south-sudan

Jason Patinkin/The Associated Press

Jason Patinkin/The Associated Press

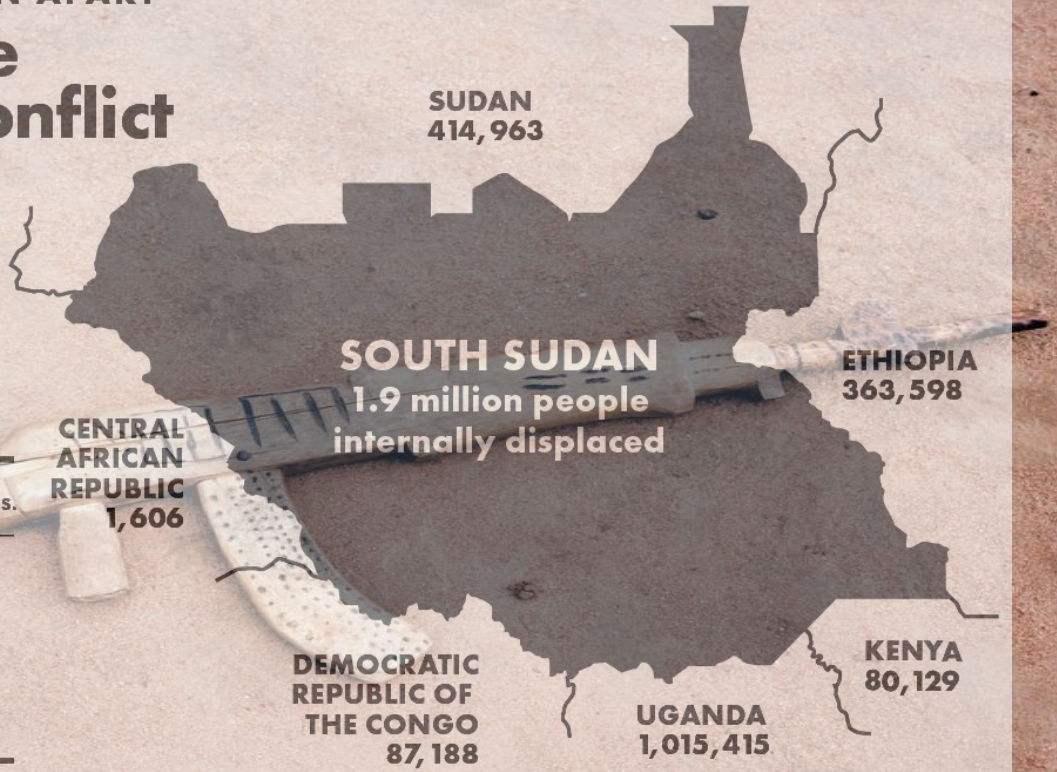
A NEW NATION TORN APART

Families flee to escape conflict

Conflict in the world's newest nation has spread across the country, creating a severe humanitarian crisis. **South Sudan has the third largest refugee crisis,** after Syria and Afghanistan.

3.7 million have fled their homes.

6 out of 10 South Sudanese refugees are children.



As of March 2018

Source: http://data.unhcr.org/SouthSudan/regional.php

MERCY CORPS

https://www.lensculture.com/articles/zed-nelson-the-new-nation-makers-of-south-sudan

A napkin fire...

- Phillip Brickell, a 43 year old ticket collector was notified of a burning tissue at the bottom of an escalator by a commuter. He rode the escalator, found the smoking tissue, extinguished the flames and went back to work. He did not tell anyone about the napkin.
- 15 minutes later, another passenger told another Underground employee about a wisp of smoke as he rode up the station escalator.
- Christopher Hayes, the safety inspector was called. Further reports of smoke and flames were increasingly being reported by commuters and even a policeman.

A napkin fire...

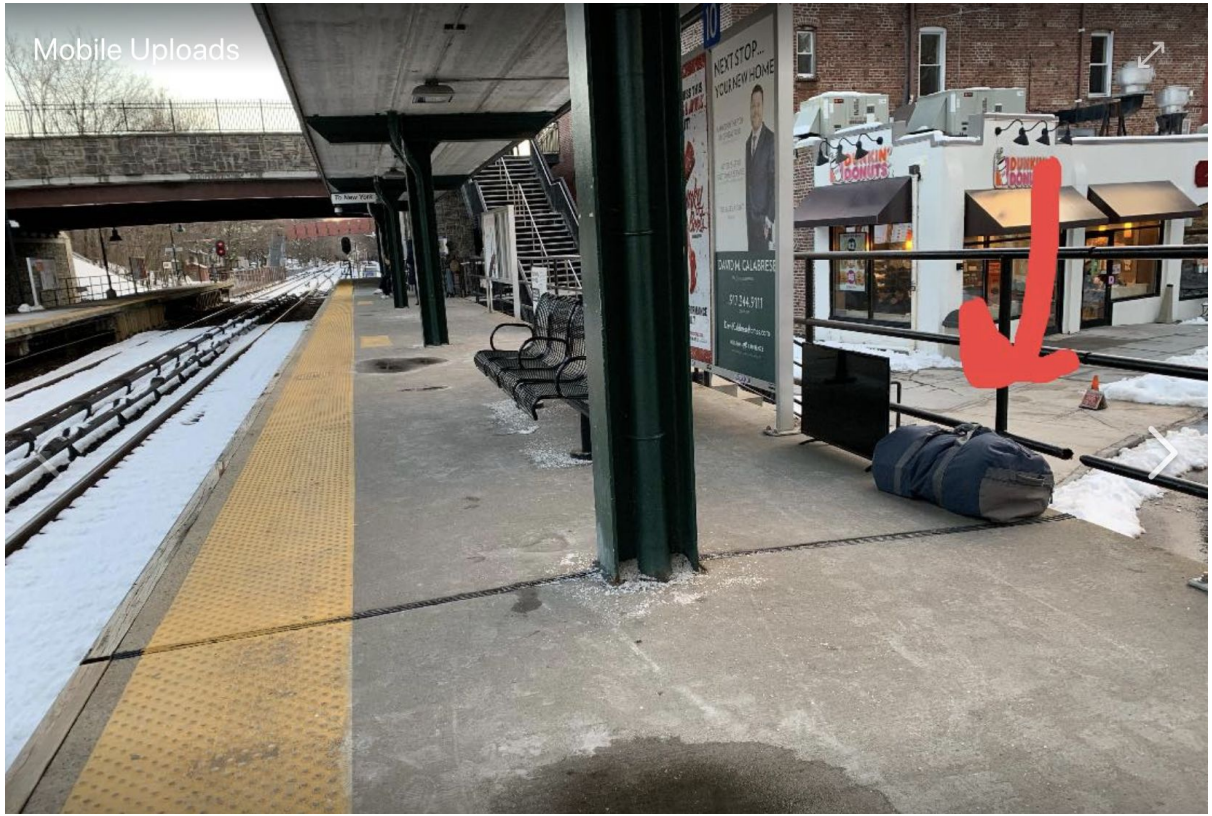
- Hayes did not call the fire brigade. He went into the escalators machine room walking past the sprinkler system, to investigate. The staff had not been trained nor authorized to use it (it was controlled by another department).
- A policeman contacted headquarters and walked up a long staircase outdoors where his radio worked. The fire brigade were notified of a "small fire at King's Cross".
- Fire brigade arrived almost 30 mins after tissue fire.

King’s Cross Fire, 31 casualties (1987)

Kings Cross: What Failed?

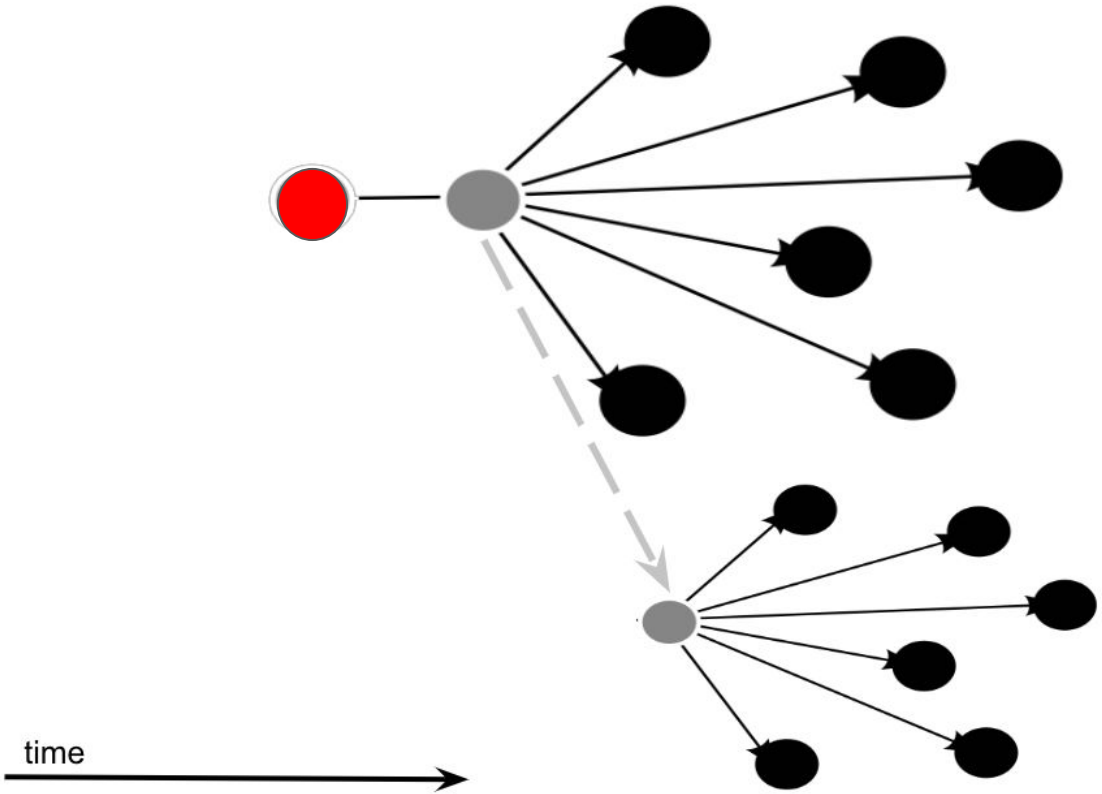
- Various point to point communications
- Little/no inter-departmental communication
- Speed

Fast Forward to 2019

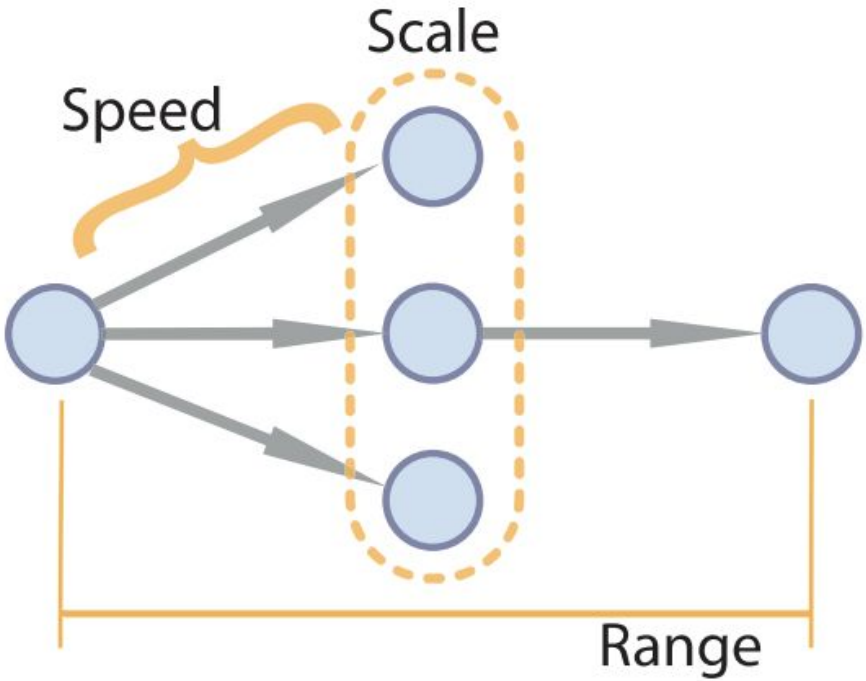


“Found huge duffel bag unattended and called 911. I was transferred to the local station. Dispatcher says it’s not their jurisdiction and directs me to transportation authority. No one showed up during the 10 min I was there waiting to board my train.”

Centralized Model



Broadcast Model (Public data)



https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1468/1896

Social Media in Emergency Response

The convergence of social networks and mobile has thrown the old response playbook out the window. Michael Beckerman, president and CEO of the Internet Association.

During and immediately following Hurricane Sandy, “users sent more than **20 million Sandy-related tweets, despite the loss of cell phone service** during the peak of the storm.” FEMA National Preparedness report (2013).

Twitter is now the first line of communications among government officials when considering event coordination responsibilities that impact 911- asking senior government officials to direct their Twitter followers to check a single account for updates on 911 outages. Karima Holmes, head of the Office of Unified Communications, a 400-person operation serving Washington D.C.

Each disaster sparks its own complex web of fast-paced information exchange, it can both improve disaster response and allow affected populations to take control of their situation as well as feel empowered. Mark Keim, associate director for science in the Office of Environmental Health Emergencies at the U.S. CDC.

https://www.ustelecom.org/emergency-911-service-and-the-role-of-social-media/

Types of Emergencies

- Chemical
- Drought
- Earthquake
- Fire
- Flu (COVID)
- Food Safety
- Heat Wave
- Highway Safety
- Hurricane
- Landslide
- Nuclear explosion
- Poisoning
- Power outage
- Terrorism
- Thunderstorm
- Tornado
- Tsunami
- Volcano
- Water Safety
- Wildfire
- Winter Storm

https://www.redcross.org/get-help/how-to-prepare-for-emergencies/types-of-emergencies.html

Types of Emergencies

- Smaller scale
  - Traffic accidents
  - Shootings
  - Gas explosions
  - Manhole explosions
  - Fallen trees
  - ...

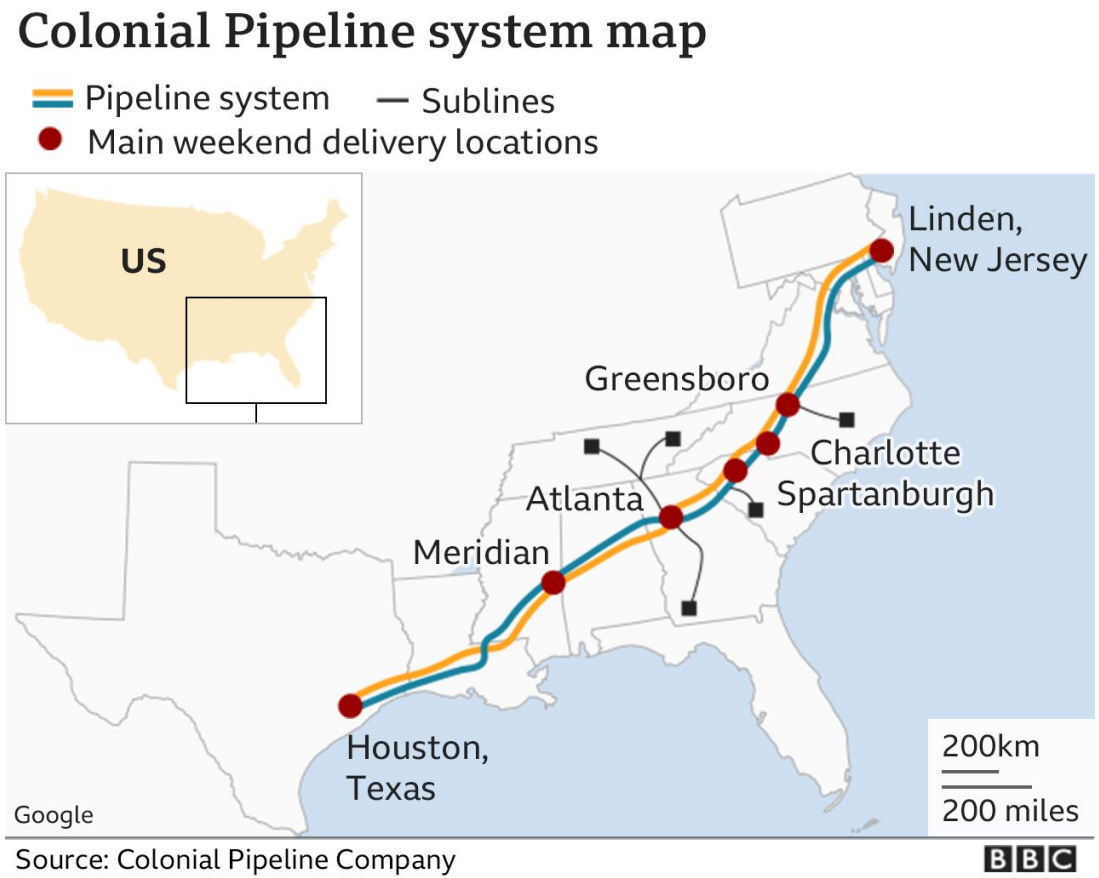
Cyber

https://pixabay.com/illustrations/cyber-attack-virus-protection-1654709/

Cyber

https://en.wikipedia.org/wiki/Colonial_Pipeline_cyberattack

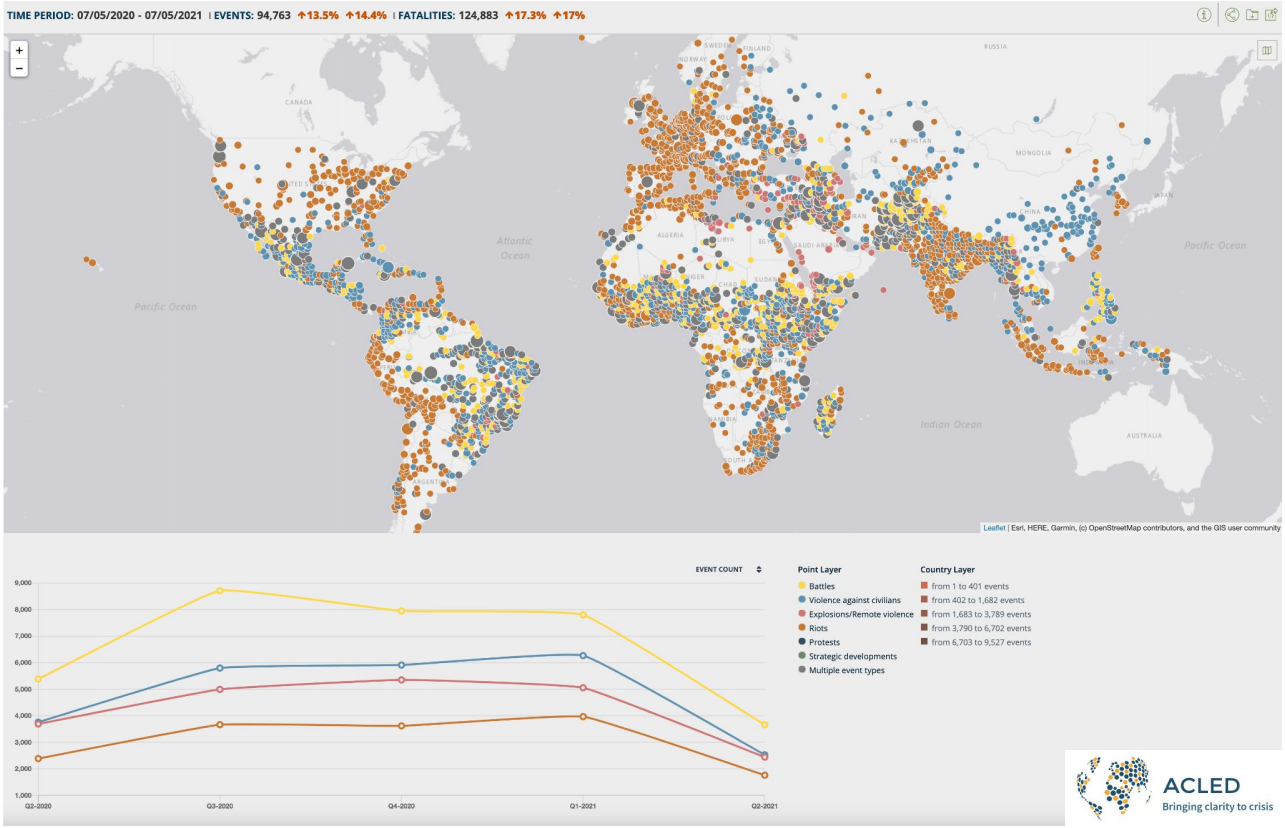
Cyber


Colonial Pipeline system map
Pipeline system
Sublines
Main weekend delivery locations
US
Linden, New Jersey
Greensboro
Charlotte
Spartanburgh
Atlanta
Meridian
Houston, Texas
200km
200 miles
Google
Source: Colonial Pipeline Company
BBC

Cyber

- Ransomware
- Expired certificates
- DOS attacks
- Data leaks
- Phishing
- Malware ….

Important Events



https://onthemap.ces.census.gov/em/

https://acleddata.com/dashboard/#/dashboard

“Critical Events”

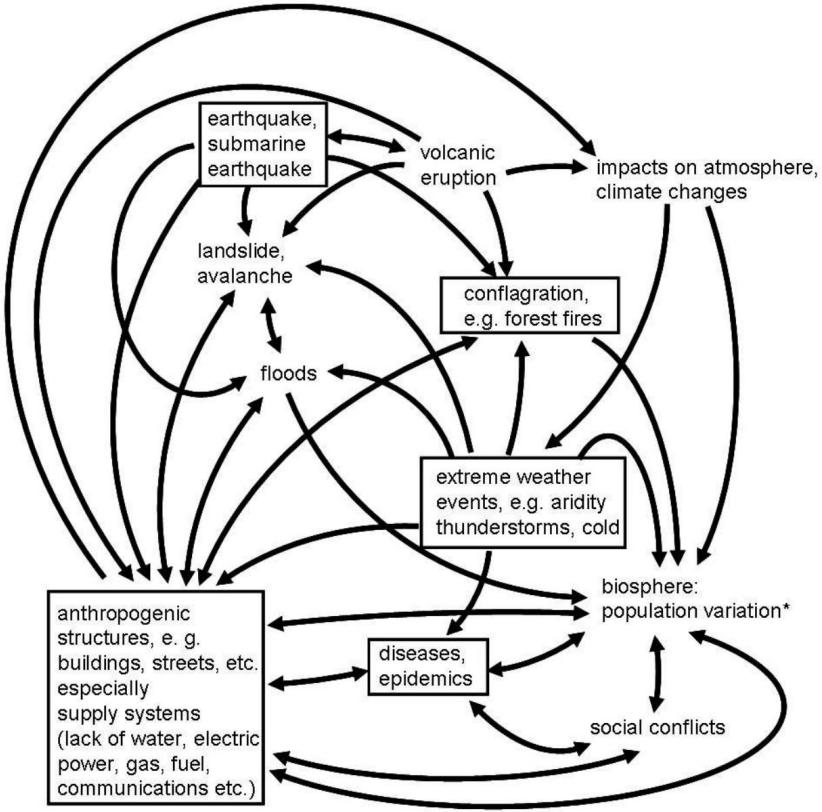
- Vocabulary
- Geographic scope & scale
- Time scope & scale
- Strength of signals
- Importance- what’s/who’s affected? Where? Why?

Why are Emergencies So Complicated?

- Unexpected events, speed
- Fragmented information
- Geographic spread
- Efficient deployment of resources
- Logistics

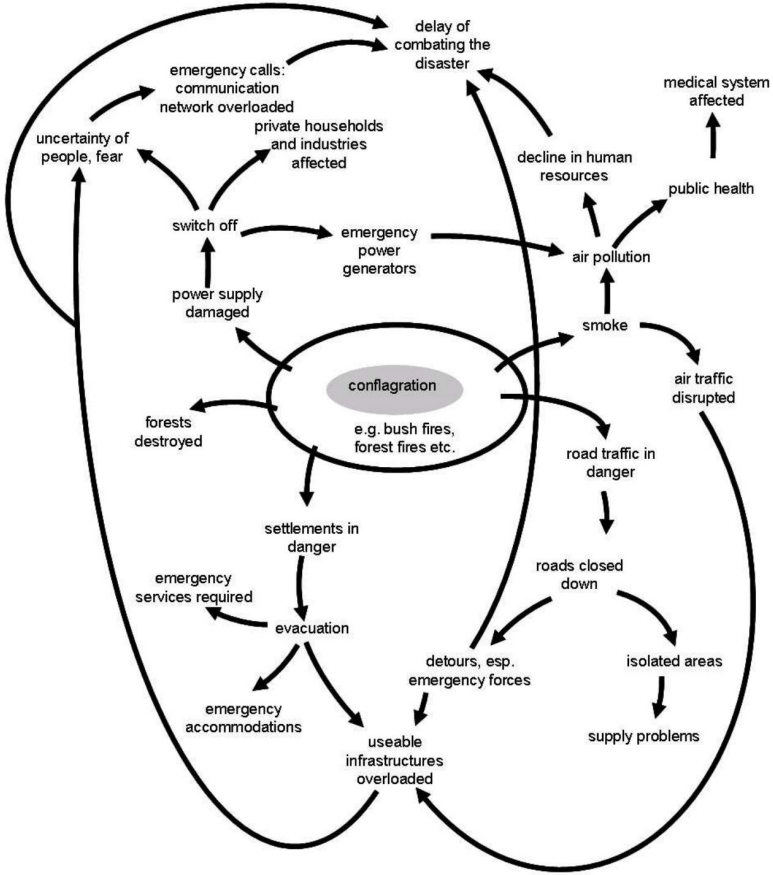
Causality (“after-shocks”)

Causality Networks



Helbing et. al. 2006

Causality Networks



Helbing et. al. 2006

Applications

Realtime public data (text, images, video, audio, sensors)

All in different formats, many languages, worldwide coverage
from hundreds of thousands of data sources.

AI Platform

For Businesses

Identify and respond to emerging risks across your enterprise.

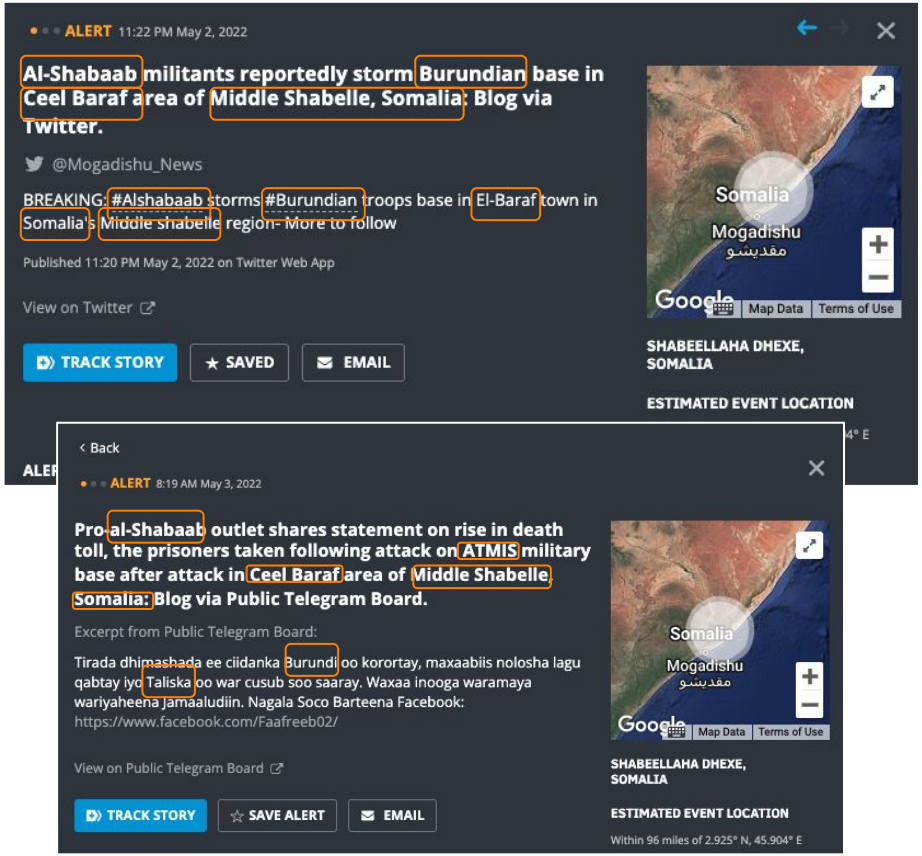
For the Public Sector

Enabling the fastest real-time response.

For Newsrooms

Gain the earliest edge in discovering stories that matter to your audience.

Knowledge for Real-Time Event Detection at Internet Scale



Our knowledge platform is used for:

- **Event detection** - what is alertable?
- **Alert generation** - what should be communicated about an event?
- **Personalization/recommendation** - who is interested in an alert?

We value:

- **Timeliness** - alert as early as possible
- **Accuracy** - send relevant alerts
- **Clarity** - information should be accessible/actionable

Entity Collection for Event Detection

- Large number of sources: public KGs, subscription services for public data, automated extraction from social media
- Semi-automated / Human-In-The-Loop collection process

- **Timeliness:** Cost/benefit analysis of entities should avoid noise

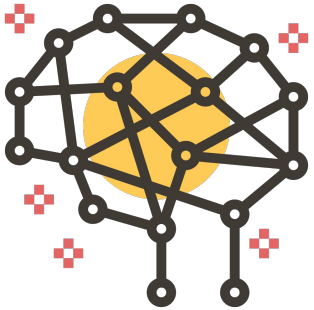
- **Accuracy:** Expert and automated validation

- **Clarity**: Multiple sources + expert curation (e.g., for editorial styles)

Event Detection: Recognizing Entities

Name Variants (strings, regexes)

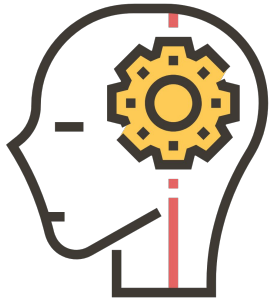
Timeliness: allows fast deployment of new/emerging entities



AI Models

Named Entity Recognition (DL, multilingual)
+ Entity Linking (ML)
End-to-end Entity Linking (DL)

Accuracy: Recognizes entities beyond their known aliases, resolves (links) them based on context



Expert rules

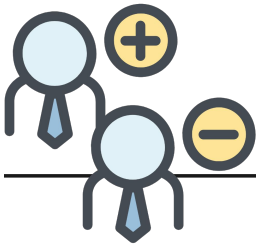
Clarity: Provide formatting and style rules to make information actionable

Knowledge Graph

Emerging Entities and Entity/Relationship Extraction

- Emergency/humanitarian response has special challenges around emerging entities
  - Retrospective analysis over CrisisNLP dataset shows **54%** of entities exist in Wikipedia pre-event, and **76%** post-event
  - IE / Deep Retrieval approaches for finding emerging entities based on mention co-occurrence
- **Timeliness, Accuracy**: Fast iteration between IE/DL pipelines and expert feedback is critical

AI entity/relationship extraction

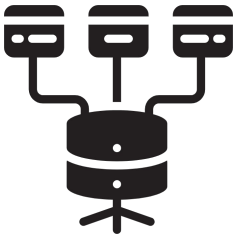


Expert feedback

Knowledge Graph

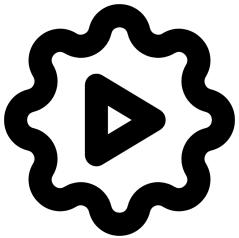
Entity Management and Quality Assurance

- Competing requirements: support expert-curated inputs/edits (**clarity**), collection/imports from external sources (**timeliness**) and multi-source validation/QA (**accuracy**)



Graph Data Model for the complex life-cycle of entities and associated data (QA, sources, validations, feedback)

Quality Assurance process similar to that of code - entity-based models are validated via tests, data replays and live deployment in isolated environments

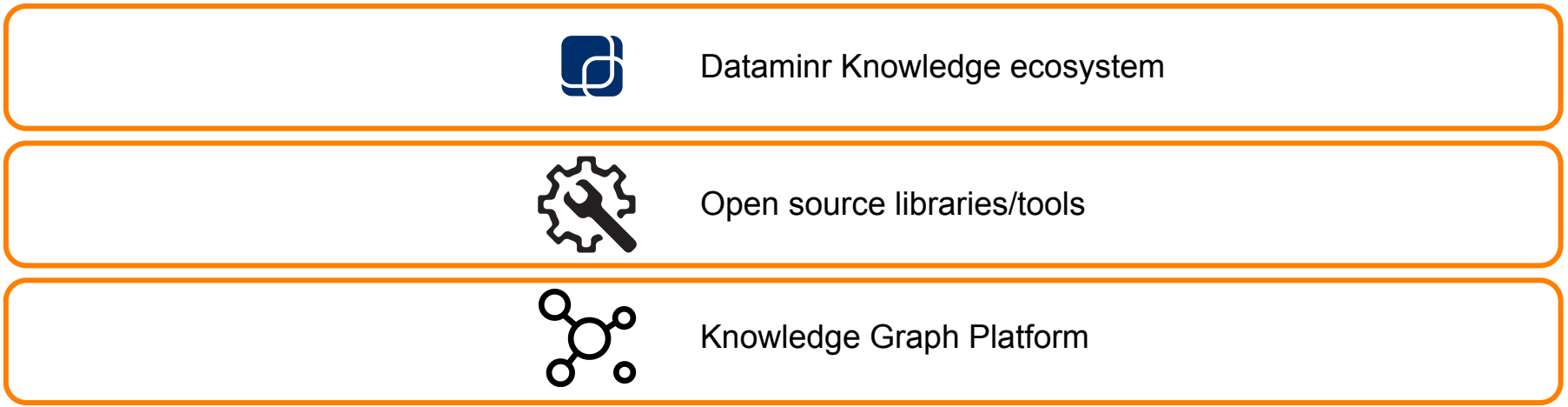


Automated enrichment/ validation processes - e.g., fetching additional information about new entities from known sources and prepare for review.

Standardized annotation and labeling practices led by HCI experts.

Entity Management Tooling

- **Accuracy, Clarity:** In-house entity management tooling on top of KG infrastructure to support entity collection, QA and AI requirements
  - An open-source ecosystem around basic Knowledge Graph tools allows us to reuse rather than re-invent



Change Detection

- **Accuracy, clarity:** Maintain accurate relationships and entity data is challenging for many entities.
  - Few planned changes: e.g., elections
  - Many unplanned changes worldwide, in multiple languages: corporate & public roles, new organizations, name changes

Being proactive vs. reactive: automated enrichment, validation and extraction services that find new evidence.

Versioning: maintaining versions of entities/relationships in the graph, to support retrospective search and analysis

Summary of lessons learned

- Your requirements (e.g., **Timeliness, Accuracy, Clarity**) will dictate how you should use Knowledge Graphs.
- AI platforms at Internet scale requires streamlined processes. Knowledge Graphs will make things easier.
- Cultivate a *graph* mindset in the organization - show how connected entities support your requirements.
- Aggressively reuse QA practices from the Software Engineering community.

Thank you!