



THE SENTINEL PROJECT

# Our early warning technologies

Feb 18-19, 2019  
Geneva, CH



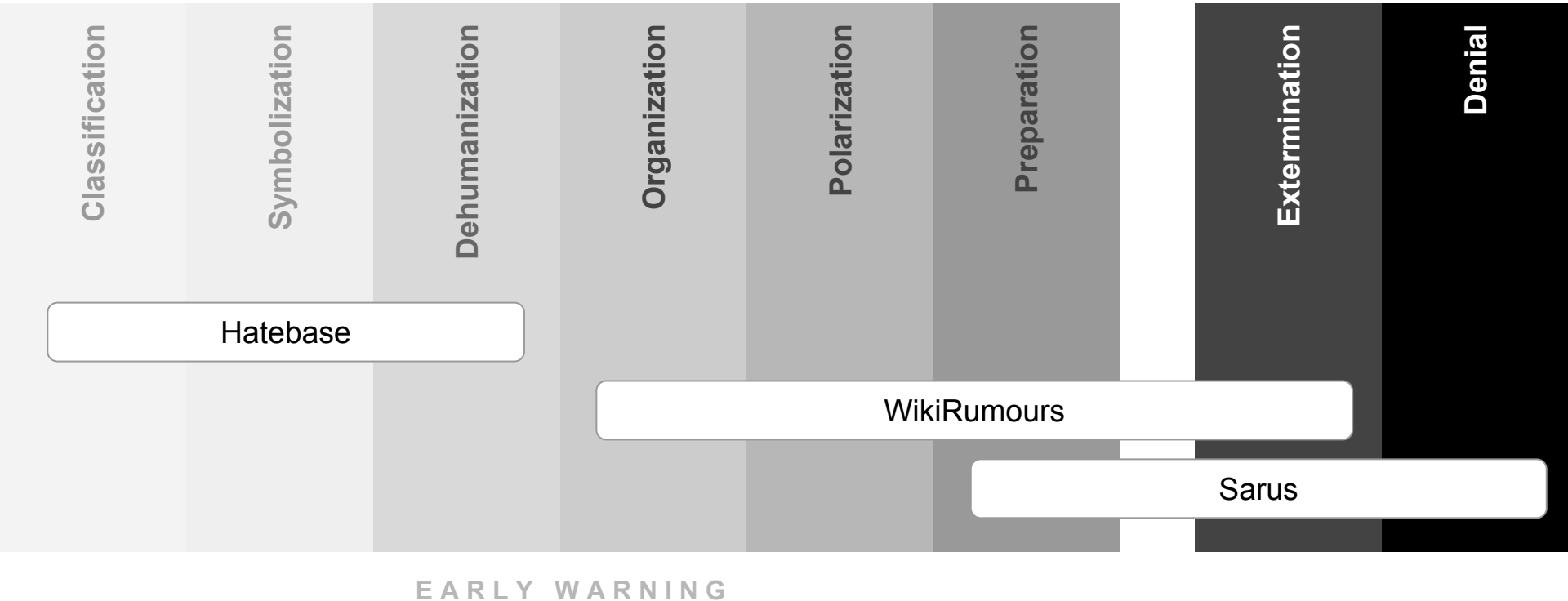
United Nations  
**Human Rights**

OFFICE OF THE HIGH COMMISSIONER FOR HUMAN RIGHTS



**LAW**  
Legal Action Worldwide

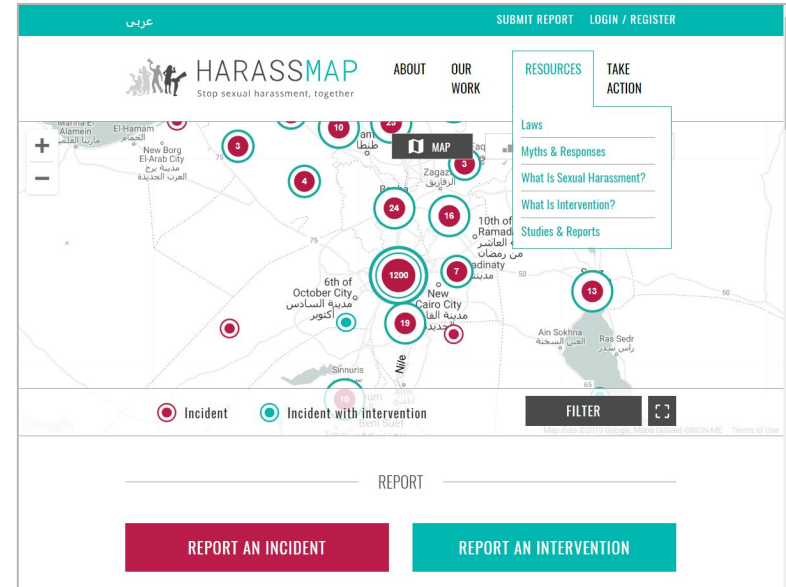
# Technologies of the Sentinel Project, by timeline of intervention



# The Sentinel Project also assists regional partners and other NGOs with various technology projects

HarassMap is a platform which:

1. Provides **relevant, location-based information** to assist victims of sexual harassment
2. Tracks incidents and interventions to help better reveal the breadth of the problem and **drive changes in policy**



The Sentinel Project invests in technology which is:

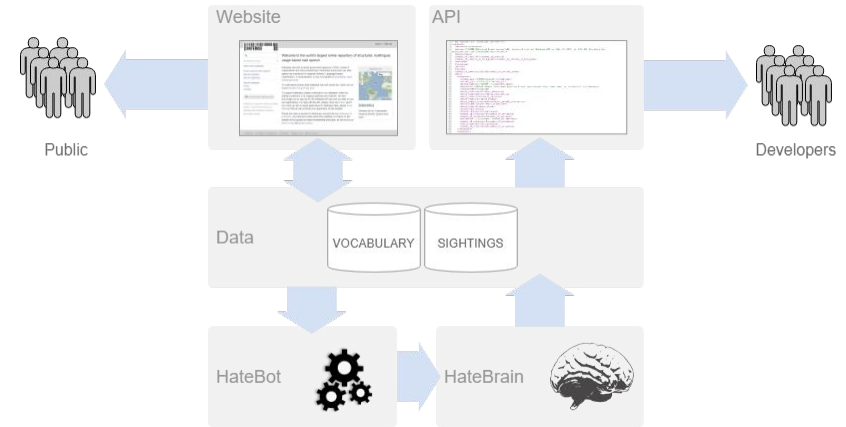
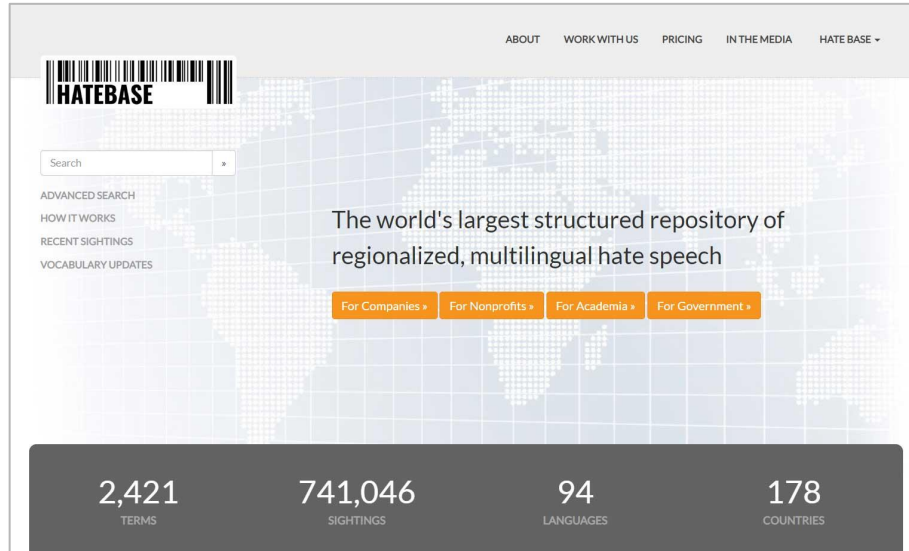
1. Practical and based on real-world requirements
2. Implementable with limited funds / resources
3. Inherently measurable
4. Politically and ideologically neutral
5. Security- and privacy-oriented
6. Open source / open data
7. Voluminously accessible through open APIs



# Hatebase



# Hatebase is a technology platform for monitoring and analyzing **multilingual** and **regionalized** hate speech



Hatebase is built around a natural language processing (NLP) engine called **HateBrain**, which...

- Recognizes hate speech terms, even if obfuscated (e.g. leetspeak)
- Eliminates homonyms using rudimentary language detection
- Recognizes clinical (non-hateful) contexts
- Assesses the probability of hateful context using helper language which we call “pilotfish”:

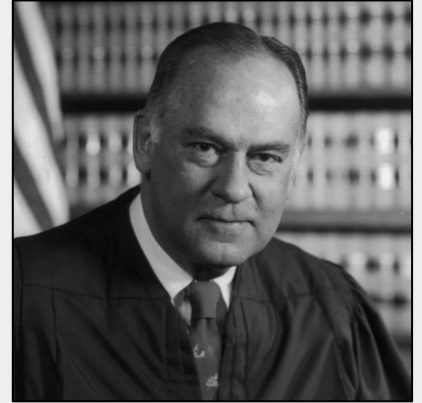
- Intensifiers
- Xenophobic references
- Targeting language
- Negative adjectives
- Emoji



# What is hate speech?

Hatebase defines hate speech as any term which broadly categorizes a specific group of people based on **malignant**, **qualitative**, and/or **subjective** attributes -- particularly if those attributes pertain to:

- **ethnicity**
- **nationality**
- **religion**
- **sexuality**
- **disability**
- **class**



"I know it when I see it."

**Justice Potter Stewart**  
Jacobellis v. Ohio, 1964





# Hate speech vs. free speech

Hatebase does not support censorship or the criminalization of speech, with a few important caveats:



Online communities have a right (and increasingly a legal responsibility) to moderate user interactions and ensure fair and respectful treatment of all users



While hate speech as an expression of opinion is (and should be) protected, hate speech which carries the threat of violence isn't (and shouldn't be)

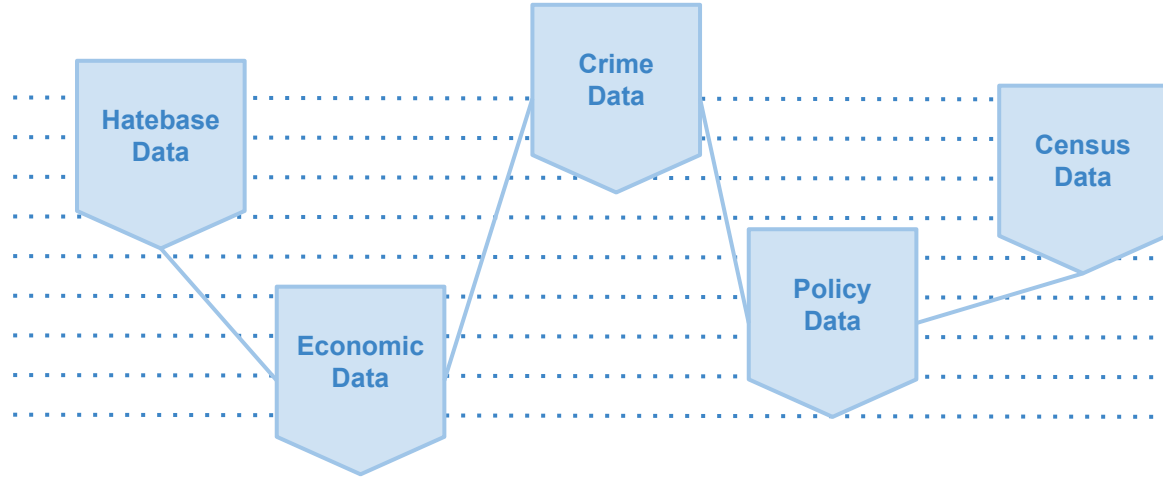


Government, law enforcement and civilian peacekeepers have a right (and increasingly a responsibility) to monitor hate speech as an early indicator of violence



# Hatebase is used by government agencies, NGOs, academic researchers and technology partners to:

- Monitor tensions across areas of concern
- Triage distribution of human, material, and financial resources
- Respond appropriately and in a timely fashion to spikes in hate speech usage
- Perform long-term analysis on underlying causes and apply predictive results to future planning efforts



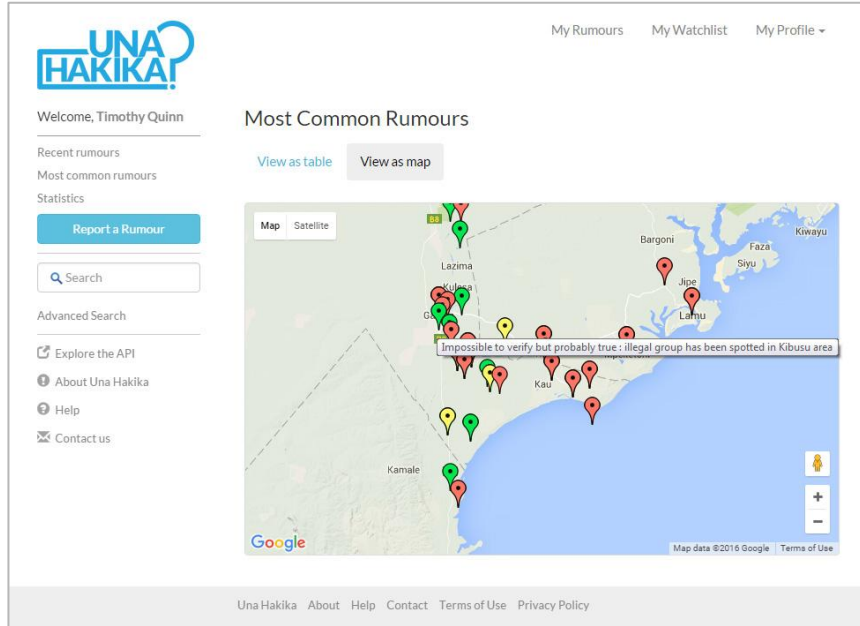
Combining data from numerous datasets can help reveal important relationships between government, citizens and external actors



# WikiRumours



# WikiRumours is a web- and mobile-based platform for moderating misinformation and disinformation



The screenshot shows the UNA HAKIKA! WikiRumours website. The header includes the logo and navigation links: "My Rumours", "My Watchlist", and "My Profile". The main content area is titled "Most Common Rumours" and features a map of the Kibisu area with various colored pins indicating locations. A text overlay on the map reads: "Impossible to verify but probably true : illegal group has been spotted in Kibisu area". The left sidebar contains a welcome message for Timothy Quinn, links to "Recent rumours", "Most common rumours", and "Statistics", a "Report a Rumour" button, a search bar, and an "Advanced Search" section with links to "Explore the API", "About Una Hakika", "Help", and "Contact us". The footer includes links to "Una Hakika", "About", "Help", "Contact", "Terms of Use", and "Privacy Policy".



# Rumours on WikiRumours go through a process of **continual annotation** and classification

“ The Mungiki (a deadly vigilante group with historically Kikuyu membership) have attacked

Occu  
Reported by **anonymous** via S

Attacked Kibera Mathare Mungiki

This rumour is **probably false** and is of **high** priority – here's why:

Despite unrest, looting, violence and clashes with police there was no evidence the Mungiki sect took place in the form of attacks on either Mathare or Kibera

The majority of sightings for this rumour were hearsay, incidents of looting or several cases, efforts by community members to mobilize security for the community to keep watch.

New / uninvestigated

**Under investigation**

Probably true

Probably false

Confirmed true

Confirmed false

Impossible to verify

Impossible to verify but probably true

Impossible to verify but probably false



# WikiRumours allows for some evidentiary collection when rumours are assessed to be verifiably true or false

Verified with

Photographic evidence  
and other file attachments

Drag or click here to upload...

☐ Delete 51904386\_788920011468543\_498363579

Tags

On the other locations and fronts especially in Morsak, parts of Mundri, Rokon, northern Juba, Lo'bonok, Kajo-keji, and around Torit and Kapoeta, the NAS forces are firm and on alert expecting attack at any given time by Mathiang Anyoor Militia who are out to attack and destroy NAS revolutionary forces and its movement.

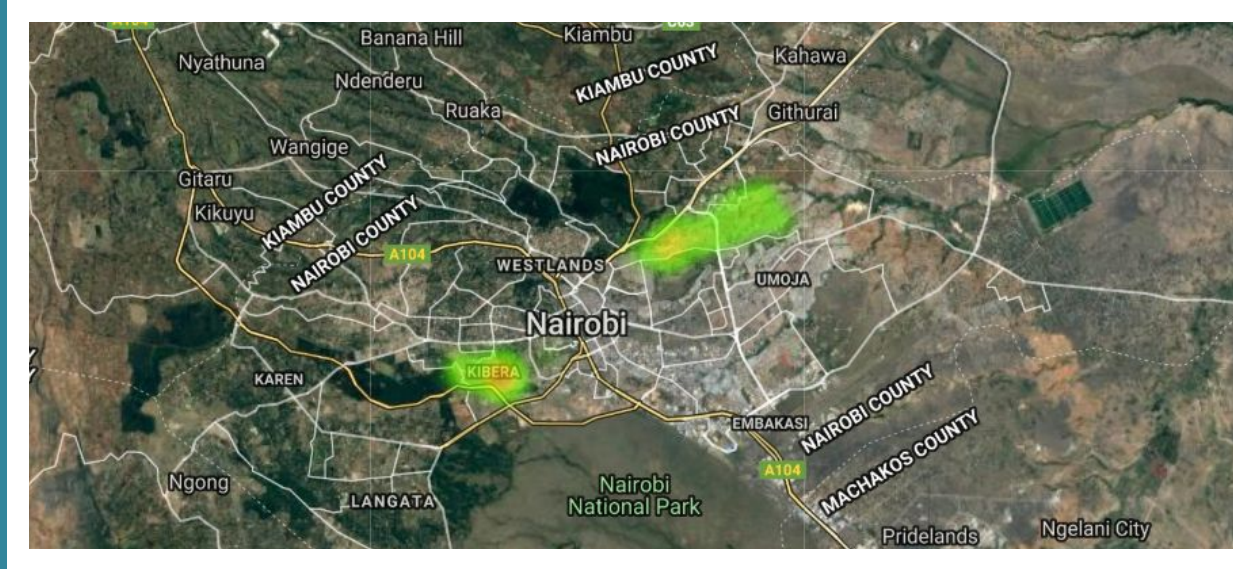
NAS leadership therefore reassure its members, supporters and sympathizers that your movement will not waver in the face of juba regime reign of terror. Your revolutionary forces are resilient and committed to fight to defend themselves and the innocent civilians of South Sudan against these barbaric militia and their affiliates.

While NAS is committed to the Cessation of Hostilities (CoHA), signed in December 2017, in Addis-Ababa Ethiopia; it however, reserves the natural right of self-defense should our forces come under attack.

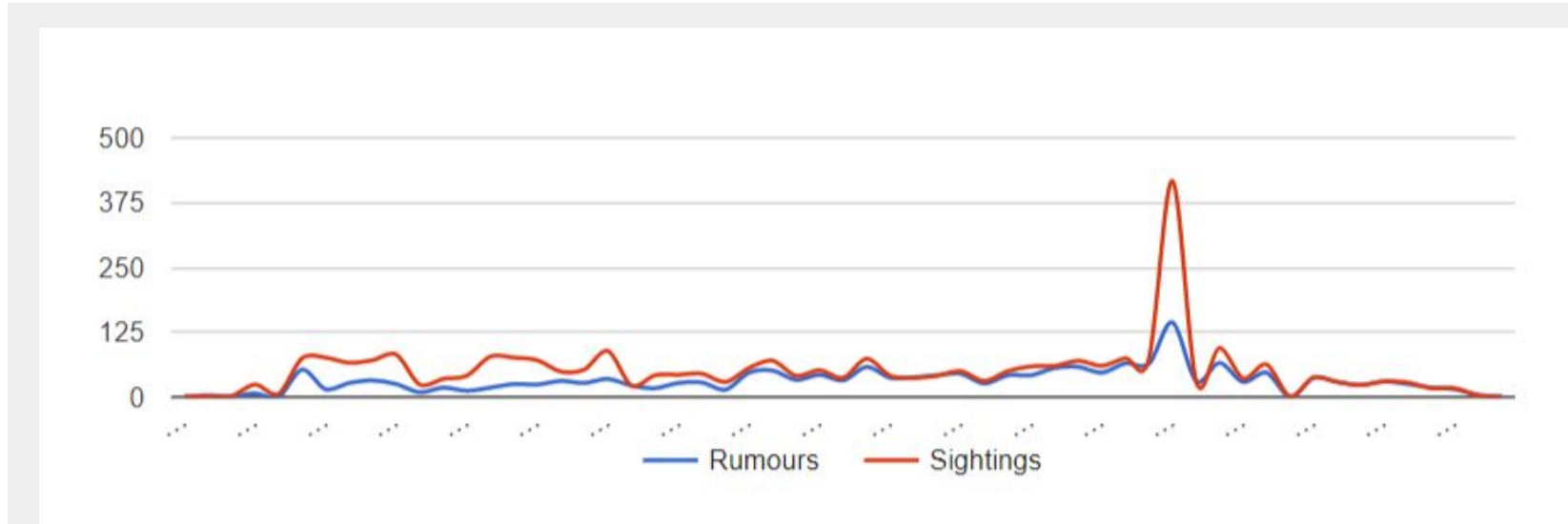
Suba Samuel Manase  
NAS Spokesman.



As rumours spread and are sighted in different communities,  
WikiRumours heatmaps each datapoint

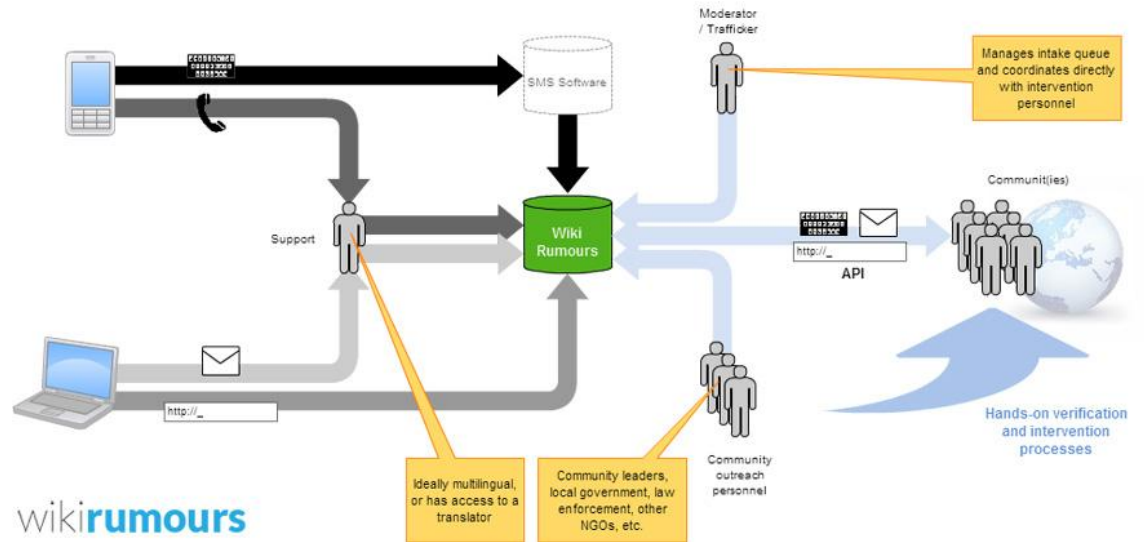


This data can then be correlated to reveal patterns in misinformation creation and dissemination

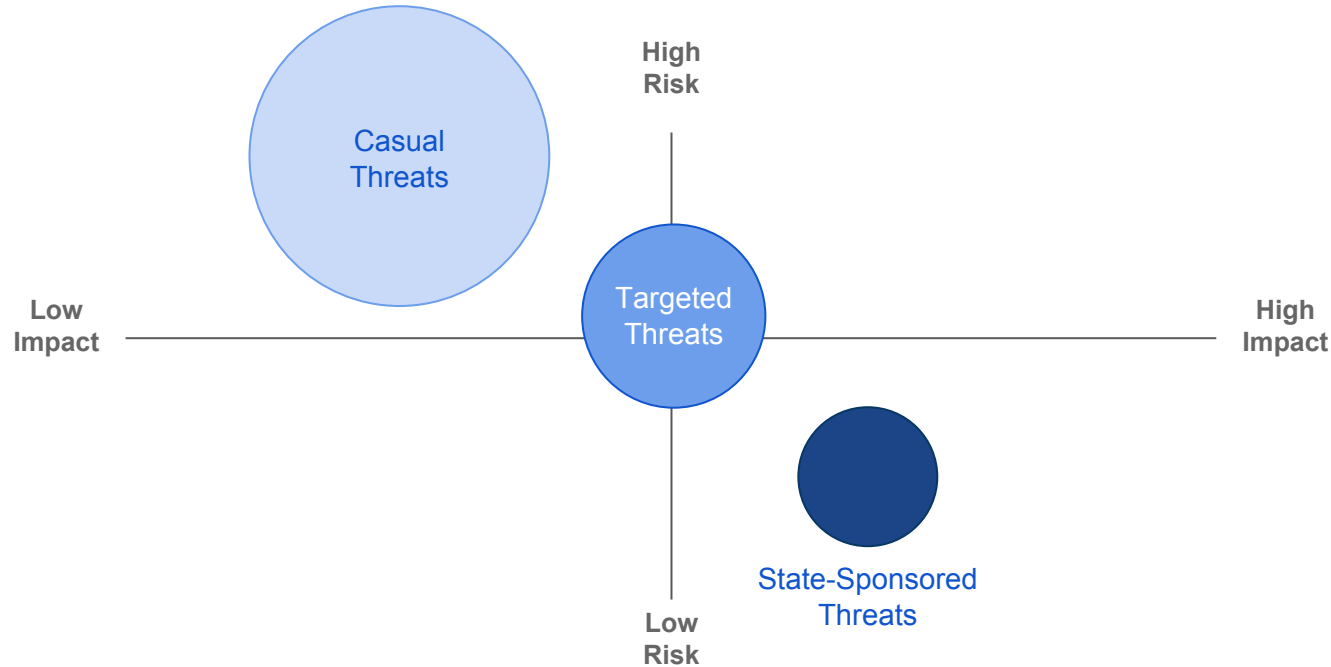




# WikiRumours is both a methodology and an open source software platform

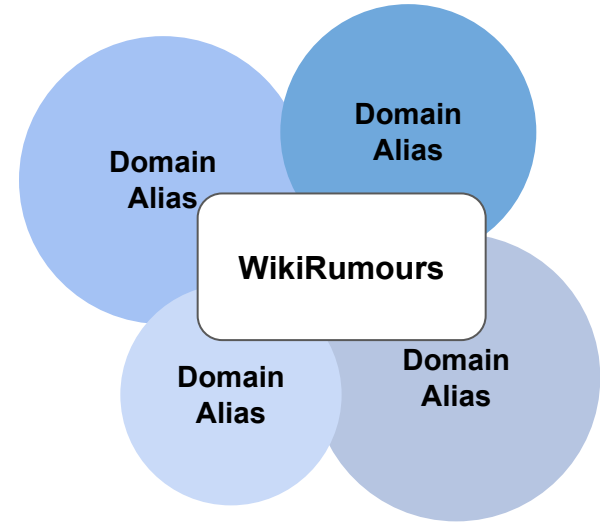


# Anonymity is a core component of both the WikiRumours workflow and the software platform



# WikiRumours incorporates **domain aliasing** functionality

This allows any organization to independently set up a regionally managed instance of the software **without need of hosting or technical resources**



# WikiRumours is currently active in...



Kenya



DRC



Myanmar



Uganda



Sarus



# UAVs address gaps in traditional conflict monitoring

- Many attacks are perpetrated at night by unknown actors
  - Attempting to monitor and report on attacks places personnel at risk of physical harm
  - Difficult terrain often impedes access and field of view
  - Handheld cameras generally lack thermal / infrared imaging capabilities
- Expensive to procure and maintain a large number of aircraft
  - Unclear and restrictive regulatory frameworks in many countries
  - Battery life and signal range can limit effectiveness
  - Vulnerable to weather and attack
  - Requires trained operators



# The Sarus fleet



3DR Aero Aircraft

~40 min flight time



Walkera TALI H500  
Hexacopter

~25 min flight time



# The attack that never was



<https://thesentinelproject.org/2015/04/01/the-attack-that-never-was-a-first-hand-case-for-expanding-una-hakika-and-using-uavs>







THE SENTINEL PROJECT

⋮ <https://thesentinelproject.org/geneva>



[timothy@thesentinelproject.org](mailto:timothy@thesentinelproject.org)



[linkedin.com/company/sentinelproject](https://linkedin.com/company/sentinelproject)



[thesentinelproject.org](https://thesentinelproject.org)



[SentinelProject](https://twitter.com/SentinelProject)