



Open Source Intelligence in Data Journalism

http://bit.ly/osint_strata

Anne-Lise Bouyer

COO & project manager of Journalism++

Open Source Intelligence in Data Journalism



My goal for today:

1. Show you examples of data-driven OSINT journalism
2. Explain the tools & methods
3. Discuss the costs & benefits



// Open Source Intelligence

OSINT=use information that was already published

Opposed to
HUMINT or SIGINT

// The Migrants files project





What was the impact of the wall?



// **Body counting**

No NGO or public body had an answer.

The Greek administration counts the number of migrants who cross.

But not the number of migrants who do not make it.

// **Body counting**

We gathered all available information mentioning migrants and refugees who died on their way to or in Europe.

Lists, news archives, PDF files, databases...

List of 17306 documented refugee deaths through Fortress Europe

Documentation on 01-11-2012 by UNITED

UNITED for Intercultural Action, European network against nationalism, racism, fascism and in support of migrants and refugees

Postbus 413 NL-1000 AK Amsterdam phone +31-20-6834778, fax 31-20-6834582, info@unitedagainstracism.org, www.unitedagainstracism.org

found dead	number	name	country of origin	cause of death
25/10/12	54	N.N.	unknown	drowned, 14 found, 40 missing, boat to E sunk off Al Huceima (MA), Front
23/10/12	1	N.N. (±25, man)	Sub-Saharan Africa	drowned, body washed ashore off Punta Almina (E) reportedly fell into the
09/10/12	1	N.N. (±5, girl)	Sub-Saharan Africa	body washed ashore near Galapagos beach (E), reportedly was travelling
09/10/12	1	N.N. (±5, girl)	Sub-Saharan Africa	body washed ashore near Galapagos beach (E), reportedly was travelling
08/10/12	6	N.N.	Africa	drowned, after boat of 24 migrants from Comoros capsized off the coast of
08/10/12	10	N.N.	Africa	missing, after boat of 24 migrants from Comoros capsized off the coast of
08/10/12	1	N.N. (8 months, baby)	Sub-Saharan Africa	drowned, boat sunk off Melilla's coasts (E)
01/10/12	1	N.N. (39, woman)	Nepal	fell from 5th floor escaping police controlling for noise at birthday party, wa
18/09/12	1	N.N. (young man)	unknown	advanced state of decomposition, found off Ceuta (E) allegedly fell overbo
17/09/12	1	N.N. (young man)	unknown	drowned, small boat sunk off Mostaganem (DZ) was on the way to Spain
16/09/12	1	N.N. (30, man)	Eritrea	car accident near Alexandroupolis (GR), car driven by smuggler overturned
09/09/12	1	N.N. (20-30, man)	Angola	stowaway, body fell from a plane from Angola to Heathrow, onto a street in
08/09/12	27	N.N.	Africa	missing, after boat with 36 migrants from Anjouan (Comoros) sank off May
08/09/12	6	N.N. (5 women; 1 child)	Africa	drowned, after boat 36 migrants from Anjouan-Comoros sank off the coast
07/09/12	79	N.N.	Tunisia	reportedly drowned while trying to reach the coast of Sicily (I), boat is still r
06/09/12	30	N.N. (12 men; 18 women)	Palestine/Syria	drowned, overcrowded boat from Ahmetbeyli (TR) sank off Aegean coast,
06/09/12	31	N.N. (28 children; 3 babies)	Palestine/Syria	drowned, overcrowded boat from Ahmetbeyli (TR) sank off Aegean coast, 4
06/09/12	64	N.N.	Tunisia	drowned, 1 found, 63 missing shipwreck near the island of Lampedusa (I)
04/09/12	1	Hashemi Samir	Iran	suicide, in a refugee centre in Kirchheim (D), depression due to living conc
29/08/2012	7	N.N.	Syria	drowned, family of asylum seekers, shipwreck on the way from Latakia (Sy
26/08/12	40	N.N.	Eritrea/Nigeria/Bangladesh	drowned, 10 found, 30 missing after vessel from Tripoli (LY) capsized off



Show all

Migration

Human trafficking

Smuggle

Crisis

Lan	Published ▾	Source	Type	Country	Country2	Suspect	Count
			migration				
en	2014.04.11	tradeearabia	Migration-Illegal-Entry	Iran		three Iranians	3
en	2014.04.11	aa	Migration-Illegal-Entry	Kenya	Kenya	refugees	
en	2014.04.11	aa	Migration-Illegal-Entry	Kenya	Kenya	refugees	
en	2014.04.11	egyptindependent	Migration-Illegal-Stay	Italy	Egypt	illegal Egyptian migrants	
en	2014.04.11	kenyan-post	Migration-Illegal-Entry			no refugees	
en	2014.04.11	int	Migration-Illegal-Entry	Kenya		refugees	
en	2014.04.11	int	Migration-Illegal-Entry	Kenya		48 refugees	
en	2014.04.11	int	Migration-Illegal-Entry	Kenya		48 refugees	48
en	2014.04.11	aljazeera	Migration	Europe	Africa	Many migrants	400
en	2014.04.11	dailymail	Migration	Egypt	Somalia	all the migrants	60
			Migration-				

La mort du migrant inconnu.

13 AVRIL 2001 À 00:28

L'AUTEUR

On ne sait rien de lui. Ni son âge, ni son nom. Sinon, selon les enquêteurs, qu'il était «probablement» de nationalité afghane. Immigrant clandestin, l'homme a forcé les enceintes barbelées du site de protection du terminal français de Coquelles, près de Calais. Sa destination? La Grande-Bretagne. Son objectif, emprunter cette navette poids lourd encore à quai, et traverser ainsi le tunnel sous la Manche. Son corps a été retrouvé jeudi, électrocuté et brûlé par la caténaire, avant qu'il réussisse à se hisser dans le convoi. **Le 1er février, un autre exilé avait lui aussi trouvé la mort** dans ce même terminal d'Eurotunnel, percuté par une navette. La région de Calais, aux portes de la Grande-Bretagne, est une des plaques tournantes de l'immigration clandestine en Europe. A Sangatte, commune limitrophe de Coquelles, un centre d'accueil géré par la Croix-Rouge héberge en permanence un millier d'étrangers en attente d'un passage clandestin vers Douvres, soit à bord d'un camion transitant par le port, soit à bord d'un train ou d'une navette empruntant le tunnel.

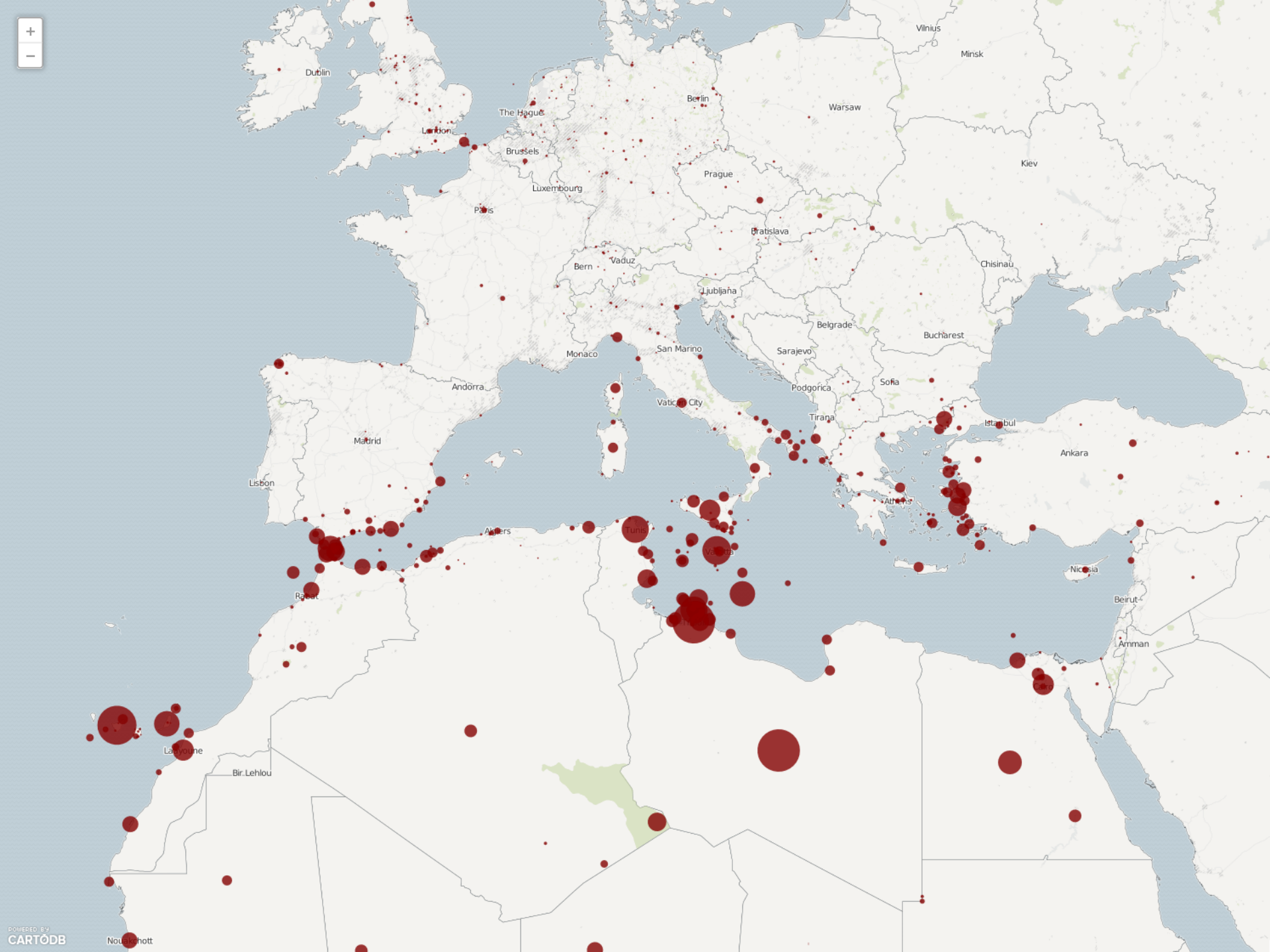
// **Body counting**

After deduplication, we had structured information about approximately 2,800 events.

We then geocoded them using a simple heuristic.

// Body counting

9	Kurds, drowned near the island of Kos (GR) reportedly after vessel capsized (Mar 24, 2000). From Del Grande's data set (translated): Boat capsizes in the waters of Kos, 8 dead, including 3 children (Mar 25, 2000)	36.799851	kos	27.102943	0	Event at Kos on Mar 24, 2000	Eastern Mediterranean route Sea	MNS/taz
1	found dead, floating near the port of Ceuta (E/MA) (Mar 24, 2000)	35.888384	ceuta	-5.324636	0	Event at Ceuta on Mar 24, 2000	Western Mediterranean route	MNS
2	drowned trying to reach Spain crossing the Strait of Gibraltar in small boat (Mar 31, 2000)	36.140751	gibraltar	-5.353585	0	Event at Gibraltar on Mar 31, 2000	Western Mediterranean route	MNS/EF



// **Body counting**

With this information, we could assess the dangerousity of each migration route.

// Body counting



The single biggest entry route for migrants into the EU is via international airports: most of those who currently reside in the EU illegally, originally entered in possession of valid travel documents and a visa whose validity period they have since overstayed.

Main migratory routes into the EU / land & sea

- ▶ Western African route
- ▶ Western Mediterranean route
- ▶ Central Mediterranean route
- ▶ Apulia and Calabria route
- ▶ Circular route from Albania to Greece
- ▶ Western Balkan route
- ▶ Eastern Mediterranean route
- ▶ Eastern Borders route

- Schengen area
- Schengen associate countries

Canary Islands



// **Body counting**

With this information, we could assess the dangerousity of each migration route.

The wall made the journey to Greece 10 times more dangerous for refugees.

The Migrants Files

A database on the more than 25,000 migrants who died on their way to Europe since 2000.

🏆 Winner of the [Datajournalism Awards](#).

Add data

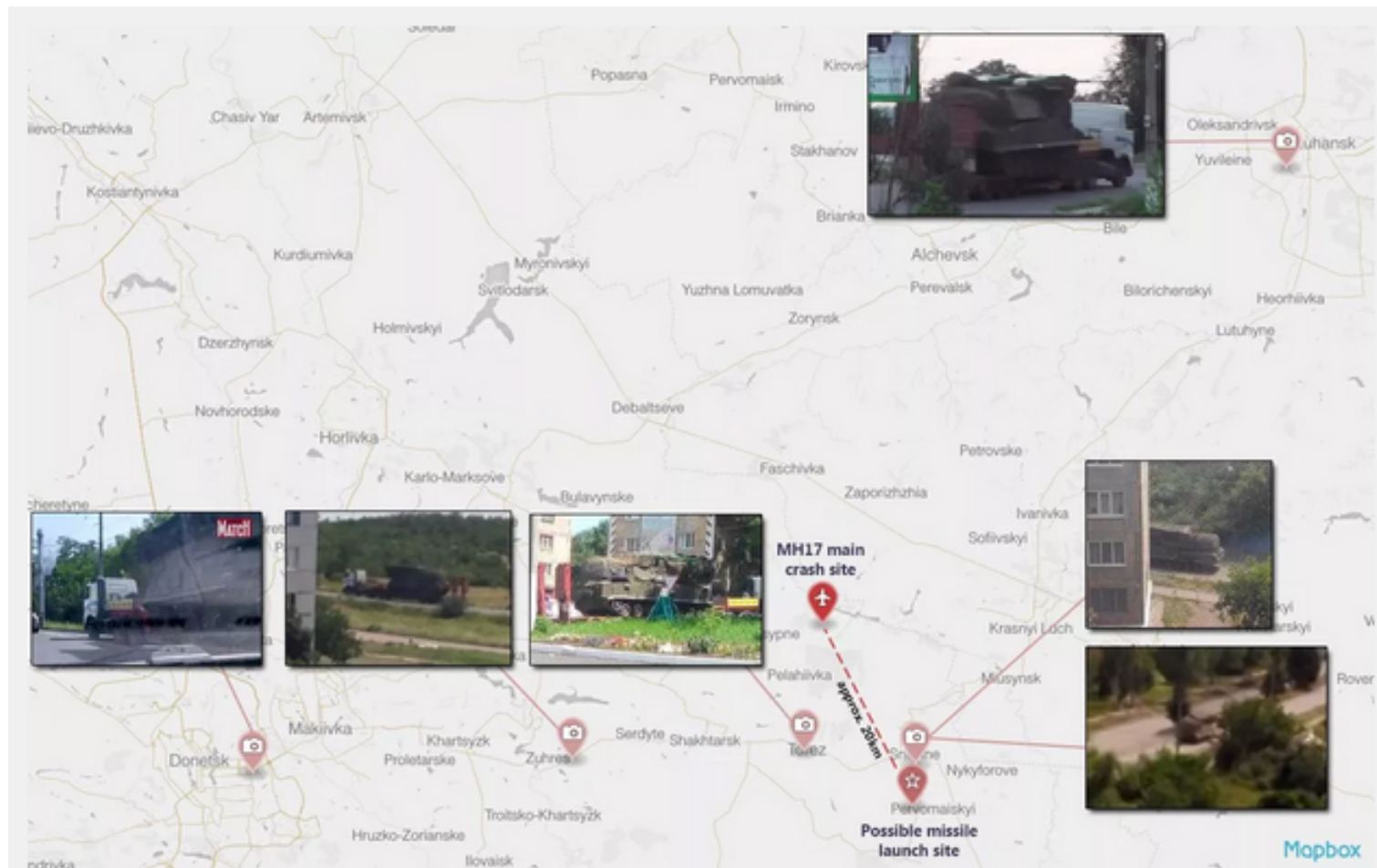
Explore

// OSINT and data

OSINT isn't just for data.

Bellingcat shows that you can investigate using any kind of open sources.

// OSINT and data



Map of the six sightings of a Buk missile launcher on July 17 and 18 in separatist-controlled territories of eastern Ukraine.

<https://www.bellingcat.com/news/uk-and-europe/2014/11/08/origin-of-the-separatists-buk-a-bellingcat-investigation/>

These are the people killed in traffic crashes in New York City this year. Click the red arrow to read more about each crash. The "Story" icon offers personal stories. If you know someone who has died, we'd like to hear from you. Email us at transponation@gmail.com or call (347) 352-5686.


Victim: All Types

Borough: All

197


deaths

100% of all deaths

- 


Unidentified male

October 3 Mott Haven, Bronx

Killed after being struck by a bus.
- 


Unidentified male, age 35

October 1 Flushing, Queens

Killed while walking on the Long Island Expressway.
- 


Unidentified male, age 34

September 30 Bellerose, Queens

Killed after exiting a car on a highway.
- 

David Malave, age 38

September 28 Glendale, Queens

Killed while riding a motorcycle.
- 

Doohee Cho, age 33

September 28 Union Square, Manhattan

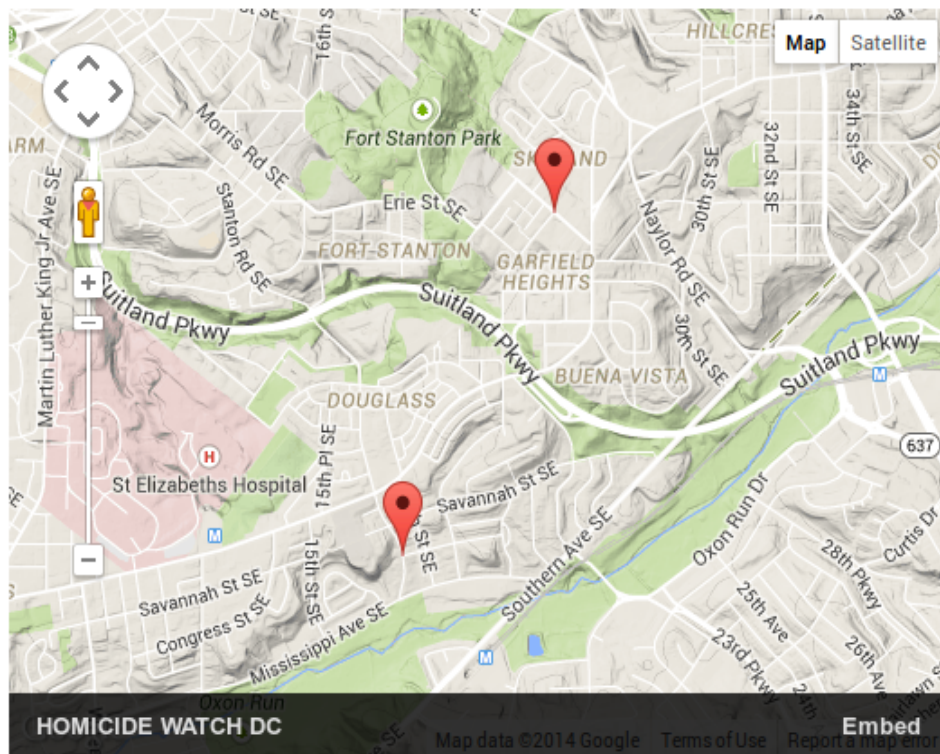
Killed while crossing the street.



Week in Review

Posted Nov. 9, 2014, 9:44 a.m. by Chris Amico | [View Comments](#)

In brief:



Special Coverage: 2013 Year in Review

What is Homicide Watch D.C.?

Homicide Watch is a community-driven reporting project covering every murder in the District of Columbia. Using original reporting, court documents, social media, and the help of victims' and suspects' friends, family, neighbors and others, we cover every homicide from crime to conviction.

[Read more...](#)

Subscribe

Support Homicide Watch D.C.

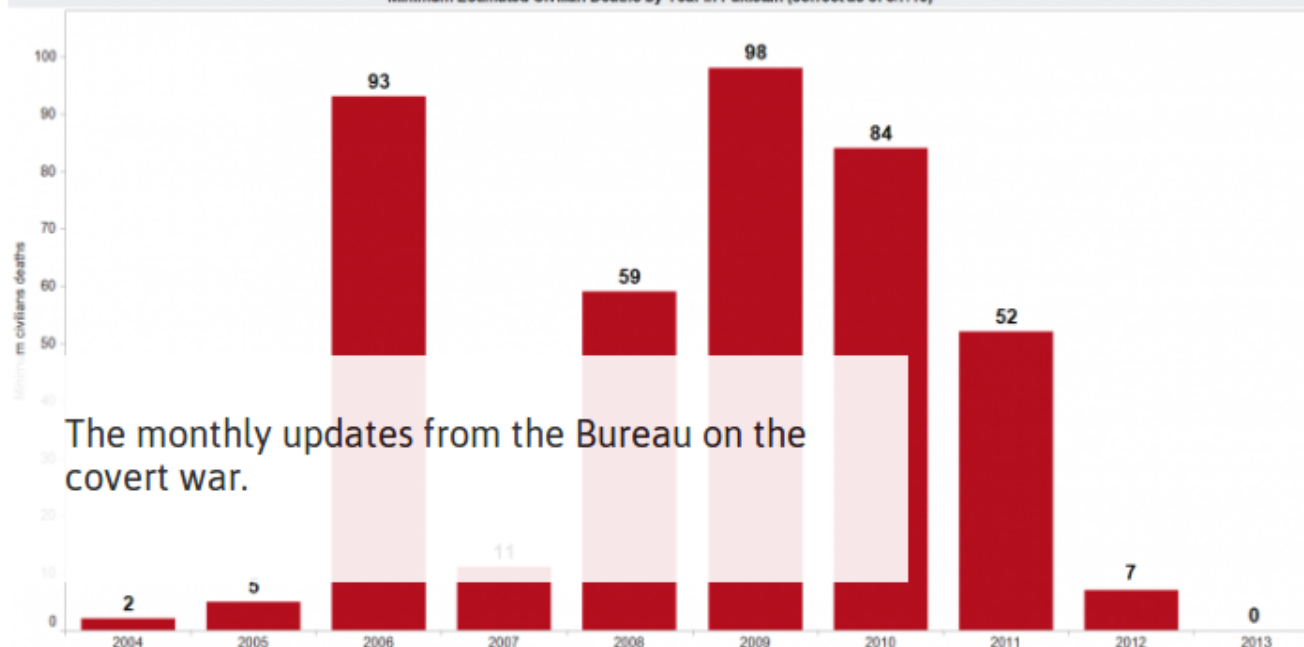
Your donation helps ensure Homicide Watch D.C. can continue to provide crime coverage the District needs.

[Subscribe in a reader](#)

Monthly Updates on the Covert War

THE BUREAU OF INVESTIGATIVE JOURNALISM

Minimum Estimated Civilian Deaths by Year in Pakistan (correct as of 3/7/13)



Latest from this investigation

Latest stories

"It's very odd" – a former UK drone operator speaks

October 3 @ 8:51am

September 2014 Update: US covert actions in Pakistan, Yemen and Somalia

October 2 @ 5:34pm

Get the court documents: The Bureau's ECHR application on government surveillance

September 30 @ 4:01pm

Podcast: Brown Moses on what's flying over Iraq and Syria

September 19 @ 3:34pm



// Tech pitfalls

All these projects are long-term and collaborative.

Most of them started on a spreadsheet, then converted into a SQL database or JSON files for the web, then again to spreadsheets for editing.



// **Tech solutions**

Ad-hoc solutions

// Tech solutions

Ad-hoc solutions
Google Docs

// Tech solutions

Ad-hoc solutions

Google Docs

Visual Investigative Scenarios



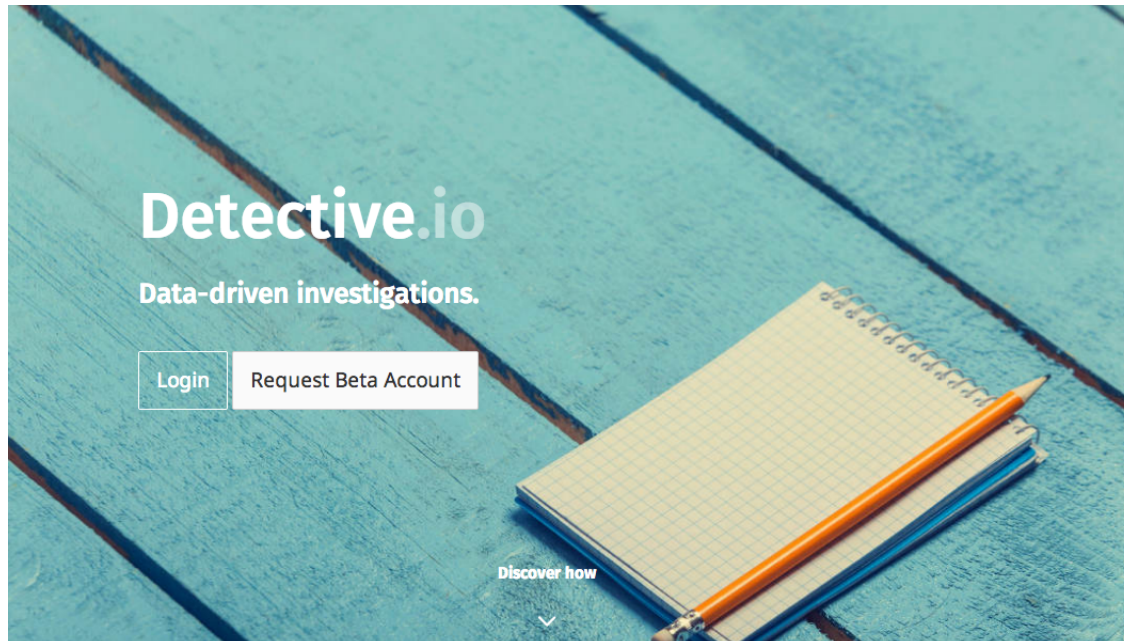
// Tech solutions

Ad-hoc solutions

Google Docs

Visual Investigative Scenarios

Detective.io



// Costs / Benefits analysis

All of these projects take time.

Investment in time is huge and is best split across a team.

For each story, you still need to analyze the collected data.

// Costs / Benefits analysis

Investment in time is largest at the beginning.

With time, information flows to you and your listening capabilities increase (social media searches, Google alerts).



Can you recoup the investment?



// **Costs / Benefits analysis**

The cost of OSINT data gathering is extremely low compared to other methodologies.

The results are exclusive.

// Selling data

Some companies might need the data.

ProPublica makes around \$6,000 a month from its datastore ([source](#)).



PUBLICA Data Store

The ProPublica Data Store

ProPublica is making available the datasets that power our data journalism. The raw data we received as the result of a FOIA request is available for free, and datasets that reflect substantial cleaning and processing by our staff are available for a one-time fee. Journalists and academic researchers can purchase premium datasets, and interested commercial users can contact us for pricing, by clicking the "Purchase" button on any dataset. We also provide a pass-through link when a data download is available on another site. [Related Story »](#)

Premium Datasets (Purchase)

Cleaned up, categorized and often created from multiple sources, these Premium

FOIA Data (Free)

Because ProPublica received these datasets from a FOIA request, we're

External Data

ProPublica frequently uses datasets that are free and available online. So instead

// Summing up OSINT and data journalism

- Know what you want to measure

// Summing up OSINT and data journalism

- Know what you want to measure
- Tech solutions are cheap

// Summing up OSINT and data journalism

- Know what you want to measure
- Tech solutions are cheap
- A steep investment in time is needed

// Summing up OSINT and data journalism

- Know what you want to measure
- Tech solutions are cheap
- A steep investment in time is needed
- Pay-off increases with time

Any
questions ?

@AnneLiseBouyer on Twitter

Links

Projects

<https://www.bellingcat.com/news/uk-and-europe/2014/11/08/origin-of-the-separatists-buk-a-bellingcat-investigation/>

https://twitter.com/bulletpoints_

<http://homicidewatch.org/>

<http://project.wnyc.org/traffic-deaths/>

<http://drones.pitchinteractive.com/>

Tools

<https://vis.occrp.org/>

<https://www.detective.io/>