Making Sense of Data

(collection + organization + visualization)

Final Conference – 25th November 2016



DEN-SITY GN+ About People Research Teaching Publications Opportunities Blog

Home

DENSITY DESIGN

DensityDesign is a research lab in the Design Department of the Politecnico di Milano. It focuses on the visual representation of complex social, organizational and urban phenomena read more...

Activities



Borders

Since the launch of the Internet Movie
Database in 1990 a large amount of data
about movies has been collected online,
giving us the possibility to... more...



A visual exploration of the online conversational space.

Abstract Since the last decade journalism has been going through a fast and substantial revolution. The morning paper is now replaced with almost...



Italian Social Condition

a

Elective affinities between Economic Statistic & Communication Design Economic statistic concerns understanding complex, multidimensional,... more...

See Research section

See Teaching section

News

"From Mind to Reality" Workshop

The week of May 2nd we experimented a new workshop format, engaging students

San Marino Design Workshop

Within the rich San Marino Design workshop 2015 agenda organized by the

Depiction of cultural points of view on homosexuality using Wikipedia as a proxy

Preliminary research questions Which aspects of a controversy is a

POLITECNICO DI MILANO



DensityDesign Design Research Lab

_

Via Giuseppe Candiani 72, 20158 Milano

to make complex phenomena (visually) accessible, understandable and usable

making sense of data

Collection

Organization

Visualization

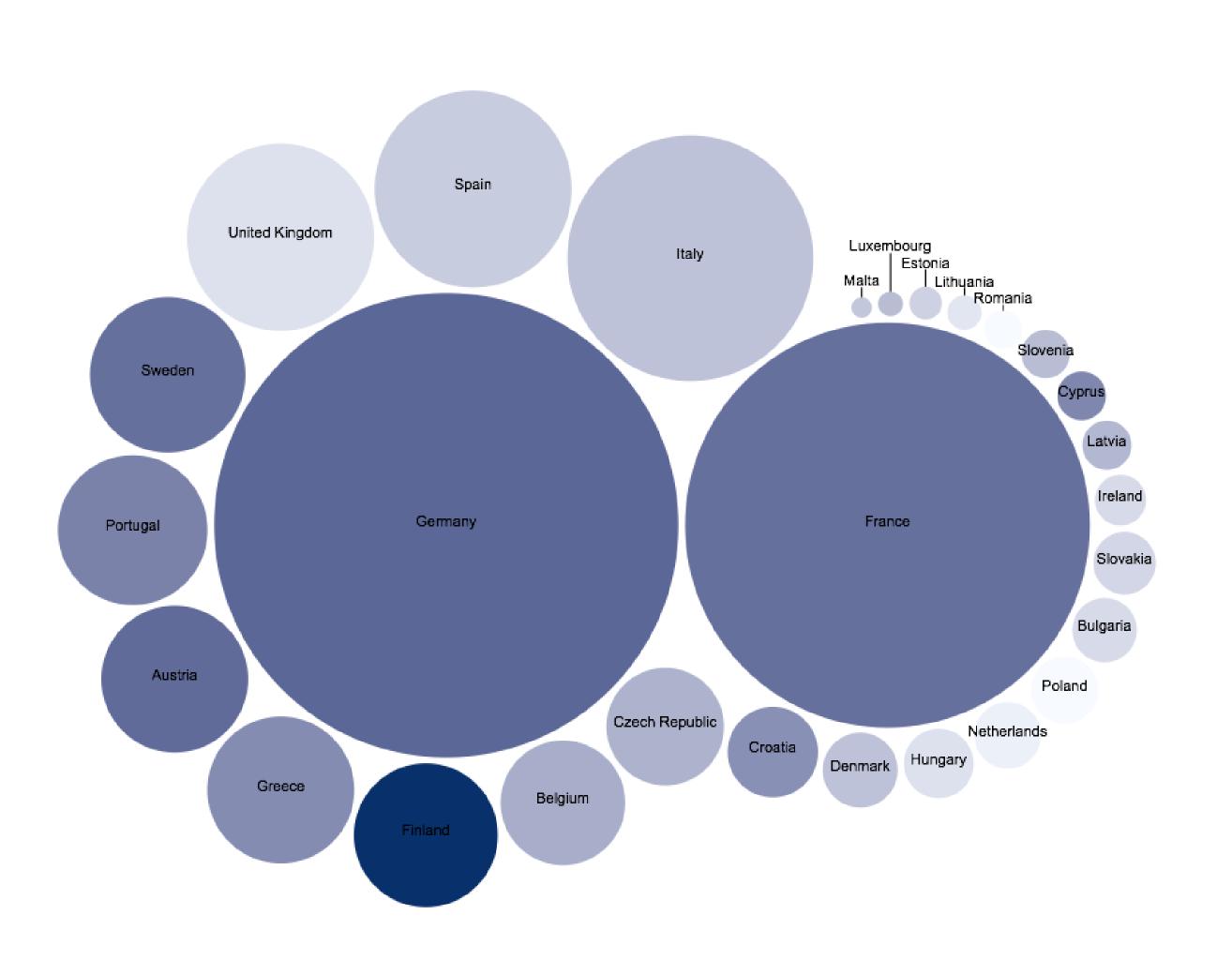
Collecting articles from EMM Newsbrief

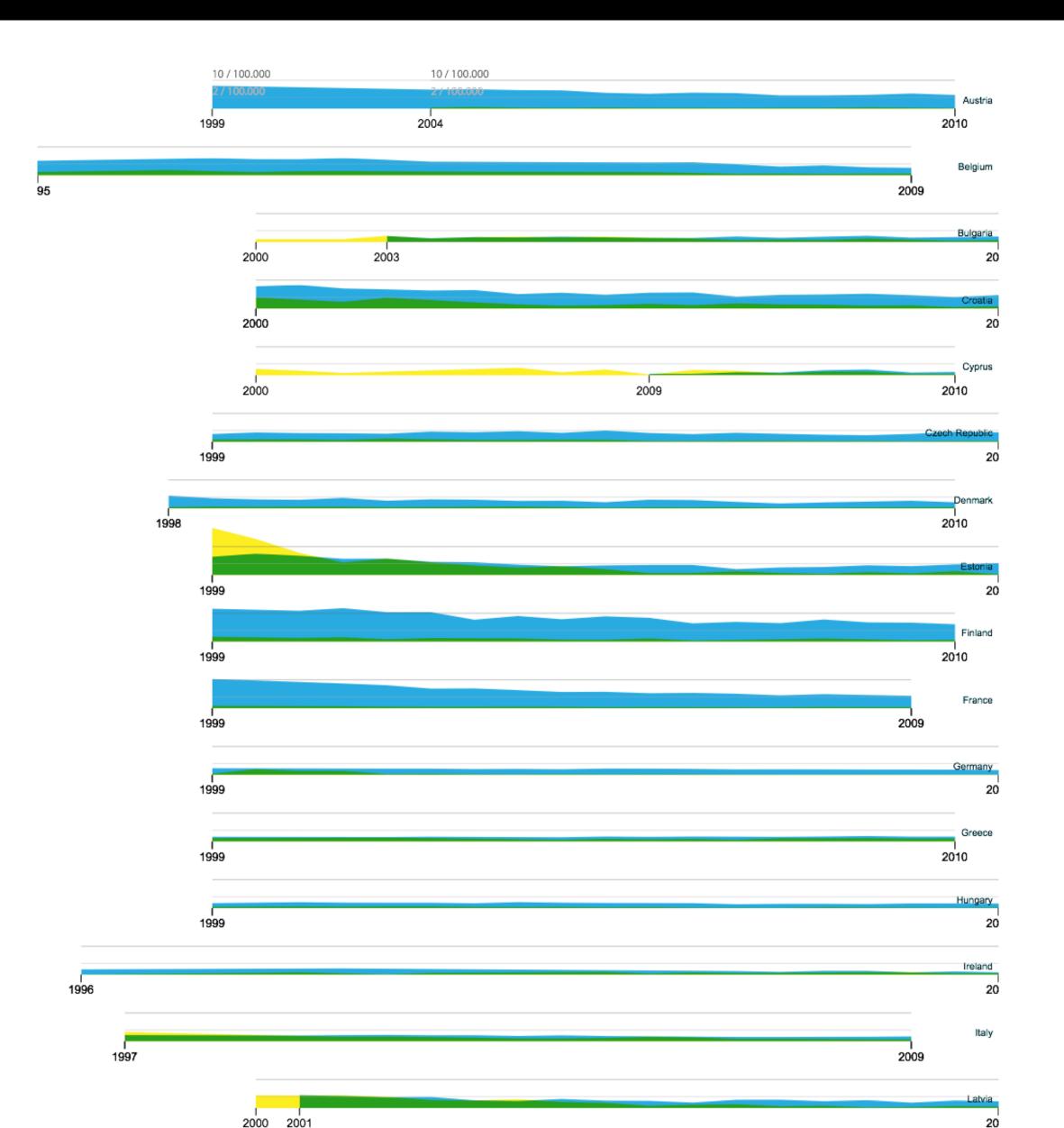
Collecting structured data from gunpolicy.org

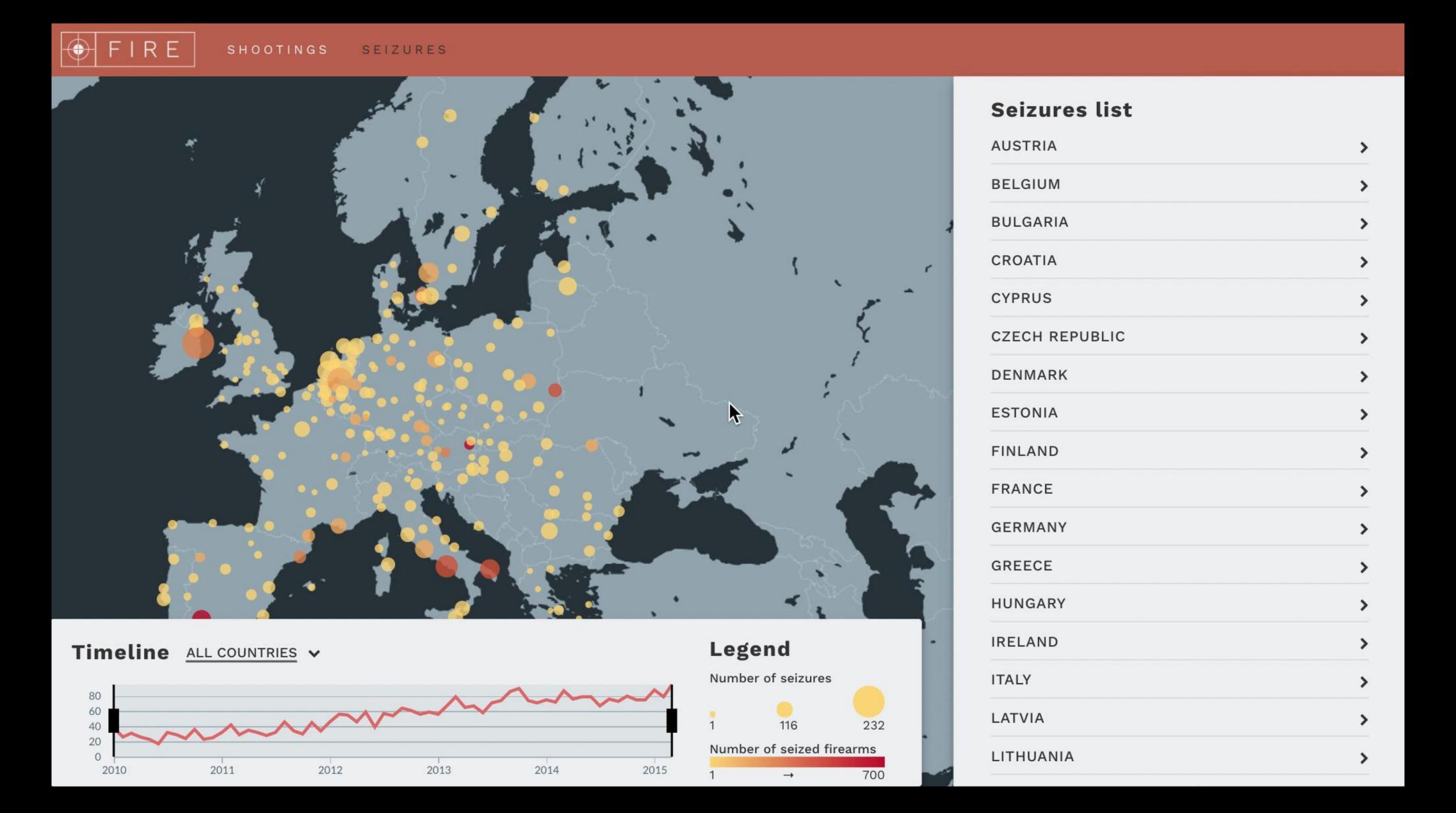
Scraping the Deep Web

Visualization as a research tool (process)

Visualization as a research result (product)







Collecting articles from the Web

Exploiting the EMM repository



collect articles from the web related to shootings and weapon seizures in the European Union (01/2010 - 03/2015).

identify events from the collection.

simplify manual analysis for finding detailed information.



Menu principale Notizie Principali Panoramica sulle 24 ore Identificazione di Eventi Temi più attivi Aiuto su EMM Overview Ricerca Avanzata Sources list Mappa del sito Obiettivo sull'UE

EU Policy Areas

Temi





Risultati della ricerca - At least: homicide

Man charged for murder James Russouw, 24, a month after a \$1million reward was posted dailymail mercoledi 13 aprile 2016 8.53.00 CEST | Info [it] [other]

Cold case breakthrough: Man charged with murder of James Russouw, 24, a month after a \$1 million reward was posted James Russouw, 24, was stabbed, doused in petrol and set alight in 2008; The Melbourne man's murder has remained unsolved for eight years; A mystery suspect called Mr Russouw from a pay......

Man charged for murder James Russouw, 24, a month after a \$1million reward was posted for the cold case death 🗹

dailymail mercoledi 13 aprile 2016 7.25.00 CEST | Info [it] [other]

Cold case breakthrough: Man charged with murder of James Russouw, 24, a month after a \$1 million reward was posted James Russouw, 24, was stabbed, doused in petrol and set alight in 2008; The Melbourne man's murder has remained unsolved for eight years; A mystery suspect called Mr Russouw from a pay......

Quelle: nationalpost.com, vor 11 Minuten veröffentlicht 🛃

direktbroker mercoledì 13 aprile 2016 7.00.00 CEST | Info [it] [other]

13.04.2016 - 05:57 Uhr German rail dispatcher played game on his phone shortly before horrific train collision that killed 11. The unnamed dispatcher was arrested on suspicion of negligent homicide, prosecutors in the Bavarian town of Traunstein said in a statement

Dispatcher 'played phone game' before train crash Authorities have arrested a rail dispatcher in Germany after accusing him of allowing a fatal collis 🗹

deutschewelle-en mercoledì 13 aprile 2016 5.14.00 CEST | Info [it] [other]

Accidents Dispatcher arrested: Game linked to Bad Aibling crash. Authorities have arrested a rail dispatcher in Germany after accusing him of allowing a fatal collision to happen because he was distracted by a video game. The crash killed 11 people in February....

Seattle mother 'was dismembered by her date with a 15-inch pruning saw in her own bathtub': FBI to investigate similar crimes in the region after horr 🗹

dailymail mercoledì 13 aprile 2016 2.32.00 CEST | Info [it] [other]

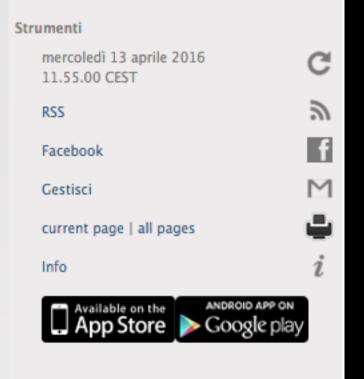
Detectives investigating the gruesome murder of a woman whose head, leg, and arm were dumped in a recycling bin in Seattle found a 15-inch pruning saw and bits of flesh in the bathroom of her Renton, Washington, home, court documents reveal. Ingrid Lyne, 40, was reported missing on Saturday morning......

LEWISTOWN, Mont (AP) — Drones might soon be dropping ... 🗹

adhoc mercoledì 13 aprile 2016 2.08.00 CEST | Info [it] [other]

voanews.com berichtet: At least eight people were killed Monday and early Tuesday in U S . At Least 8 Killed as US Airstrikes Target al-Shabab in Somalia. airstrikes targeting al-Shabab militants in southern Somalia. Pentagon spokesman Captain Jeff Davis said unmanned drones were used in the attacks....

Has man who 'butchered Seattle mother-of-three killed before'? FBI to investigate similar crimes in the region after it emerges a 15-inch pruning saw [3]





Source

EMM newsbrief: A news aggregator website collecting articles from a vast variety of online newspapers.

EMM newsbrief

http://emm.newsbrief.eu/

challenges | time & volume

~3 seconds to get full information about an article

~50.000 articles per country

~2 days for each country

challenges | noise

Many outliers not related to the starting queries.

Many duplicates of the same articles on different sites.

Many different articles talking about the same event.

challenges language

28 different countries, almost each one with its own language.

Text analysis and **results evaluation** may be difficult for less used idioms.

Methodology

Definition and preparation of the queries

Each query is a set of keywords, translated in all the languages of the EU

Α	В	С	D	E	F	G	Н	I	J	К	L	M	N	0	Р	Q
Category	English	Bulgarian	Catalan	Croatian	Czech	Danish	Dutch	Estonian	Finnish	Flemish	French	German	Greek	Hungaria	r Irish	Italian
Firearm (object)	firearm	огнестрель	arma de f	vatreno oi	střelná zb	skydevåb	vuurwape	tulirely	tuliase	vuurwape	arme à feu	Schusswaffe	πυροβόλ	lőfegyver	arm tine	arma
	revolver	револвер	revòlver	revolver	revolver	revolver	revolver	revolver	revolveri	revolver	revolver	Revolver	περίστρο	revolver	gunnán	rivoltella
	pistol	пистолет	pistol	pištolj / re	pistole	pistol	pistool	püstol	pistooli	pistool	pistolet	Pistole	πιστόλι	pisztoly	piostal	pistola
	rifle	пушка	rifle	puška	puška	riffel	geweer	vintpüss	kivääri	schietwap	fusil	Gewehr	τουφέκι	puska	raidhfil	fucile
	gun	пистолет	pistola	pištolj	pistole	gun	pistool	relv / püss / pü	ase	geweer	pistolet	Gewehr	πιστόλι	pisztoly	gunna	pistola
	weapon	оръжие	arma	oružje	zbraň	våben	wapen	relv	ase	wapen	arme	Waffe	όπλο	fegyver	arm	arma
Seizure	seize	конфискув	confiscar	zaplijeniti	chytit	beslaglæç	in beslag	haarama / kinn	takavariko	grijpen / ii	saisir	beschlagnah	προβαίνω	megraga	urghabh	sequestrare
	seized	обзет	confiscad	zaplijenje	chytil	beslaglag	in beslag	konfiskeerima	takavariko	gegrepen	saisis	beschlagnah	κατασχεμ	lefoglalt	urghabha	sequestrate
	seizure	припадък	confiscaci	zapljena	záchvat	beslaglæç	inbeslagn	konfiskeerimin	takavariko	beslagleg	saisie	Beschlagnah	κατάσχεο	elkobzás	urghabha	á sequestro
	confiscate	конфискув	confiscar	konfiscira	zabavit	konfiskere	confiscere	konfiskeerima	takavariko	aangegre	confisquer	konfiszieren	κατάσχω	lefoglal	coigistigl	confiscare

Site scraping

Each translated query has been used as input for an ad-hoc developed crawler and scraper.

Information about each article (title, publication date, source country, source publisher, full text, original link) can be saved in a **structured dataset**.

Entries filtering

A first filter is applied by looking for a set of **geographical** ("Mexico", "Syria",...), **VIP** ("Obama","Pistorius", ...), **stopwords**, in order to exclude potentially non-relevant entries.

Entries extraction

In order to further characterize each entry in our dataset, an **Entity Extraction Routine** was applied to the full text of each article.

Barclays misled shareholders and the public about one of the biggest investments in the bank's history, a BBC

Panorama investigation has found.

Words Phrases Relations Entities Meaning Dependency Parse

The bank announced in 2008 that Manchester City owner Sheikh Mansour had agreed to invest more than £3bn.

Words Phrases Relations Entities Meaning Dependency Parse

But the **BBC** found that the **money**, which helped **Barclays** avoid a **bailout** by **British taxpayers**, actually came from the **Abu Dhabi** government.

Words Phrases Relations Entities Meaning Dependency Parse

Barclays said the mistake in its accounts was "a drafting error".

Words Phrases Relations Entities Meaning Dependency Parse

CATEGORIES

- 0.77 economy, business and finance>economy (general)
- 0.74 politics>government
- 0.67 politics
- 0.67 investment
- 0.65 religion and belief>values>ethics
- 0.65 economy, business and finance>finance (general)
- 0.65 economy, business and finance>business (general)
- 0.65 crime, law and justice>laws
- 0.64 social issue>social issues (general)
- 0.63 economy, business and finance>financial and business service

TOPICS

- 1.00 Barclays
 - 00 Mansour bin Zayed Al Nahyan
 - .00 Economics
- LOO Business
- 1.00 Finance
- 1.00 Financial economics
- 1.00 Investment
- 1.00 International Petroleum

Text Razor

A Named Entity
Recognition and
extraction service that
recognizes and
categorizes semantically
relevant words in a text.

Supports 10 languages

Text Razor

https://www.textrazor.com/

Aggregation by event

Articles related to the same event where grouped by 3 main factors:

Date proximity

Title similarity

Matching entities

Manual filtering and data extraction

to obtain detailed information that can't be automatically obtained by the text, and to clean the dataset from outliers and errors.

Collecting structured data from gunpolicy.org

Exploiting the EMM repository





gunpolicy.org

A repository of detailed data about international firearms policies and facts.

Many information for each country (statistics about firearm possession, injuries and killings, seizures, etc)

http://www.gunpolicy.org

For each country, many different data formats can be found

single datum

no. of privately owned firearms: 1,800,000

time series

long gun homicides: 2010: 2

2009: 1

2008: 1

2007: 1

2005: 3

2004: 4

1999: 6

1998: 6

verbose statements

destruction and disposal policy: In Belgium, it is state policy to destroy surplus, collected and seized firearms rather than return them to the secondary arms market

single datum and statements

Rate civilian firearms possession

World ranking

N. registered firearms

Rate registered firearms

Estimated n. illicit firearms

Estimated rate illicit firearms

Smuggling Guns and Ammunition

Transparency of Small Arms Transfers

Penalty for Illicit Firearm Possession

time series

Total Number of Gun Deaths!

Rate of All Gun Deaths per 100,000 People!

Homicides (any method)!

Rate of Homicide per 100,000 People (any method)!

Gun Homicides

Rate of Gun Homicide per 100,000 People,

Handgun Homicides

Rate of Handgun Homicide per 100,000 People

Long Gun Homicides

Rate of Long Gun Homicide per 100,000 People

Gun Homicides (Other)

Rate of Gun Homicide (Other) per 100,000 People

Gun Suicides

Rate of Gun Suicide per 100,000 People

Retrieved data for each country

Handgun Suicides

Rate of Handgun Suicide per 100,000 People

Long Gun Suicides

Rate of Long Gun Suicide per 100,000 People

Gun Suicide (Other)

Rate of Gun Suicide (Other) per 100,000 People

Unintentional Gun Deaths

Rate of Unintentional Gun Death per 100,000 People

Unintentional Handgun Deaths

Rate of Unintentional Handgun Death per 100,000 People

Unintentional Long Gun Deaths

Rate of Unintentional Long Gun Death per 100,000 People

Unintentional Gun Deaths (Other)

Rate of Unintentional Gun Death (Other) per 100,000 People

Gun Deaths from Undetermined Cause

Rate of Gun Death from Undetermined Cause per 100,000 People

Handgun Death from Undetermined Cause

Rate of Handgun Death from Undetermined Cause per 100,000 People

Long Gun Death from Undetermined Cause

Rate of Long Gun Death from Undetermined Cause per 100,000 People

Gun Death (Other) from Undetermined Cause

Rate of Gun Death (Other) from Undetermined Cause per 100,000 People

Small Arms Exports (US\$) - Customs

Challenges

Data from gunpolicy.org can't be directly analyzed, since no data retrieval service is provided.

Development of an ad-hoc crawler and scraper to automatically download all the relevant data.

Value

Data regarding each country in the EU was saved in a non relational database, allowing the use of complex data structures.

Single spreadsheets and aggregations were extracted to share data and allow analysis by third party.

Open Data

Data should be **available** in a open format on the FIRE project website (*give-back*, *public engagement*).

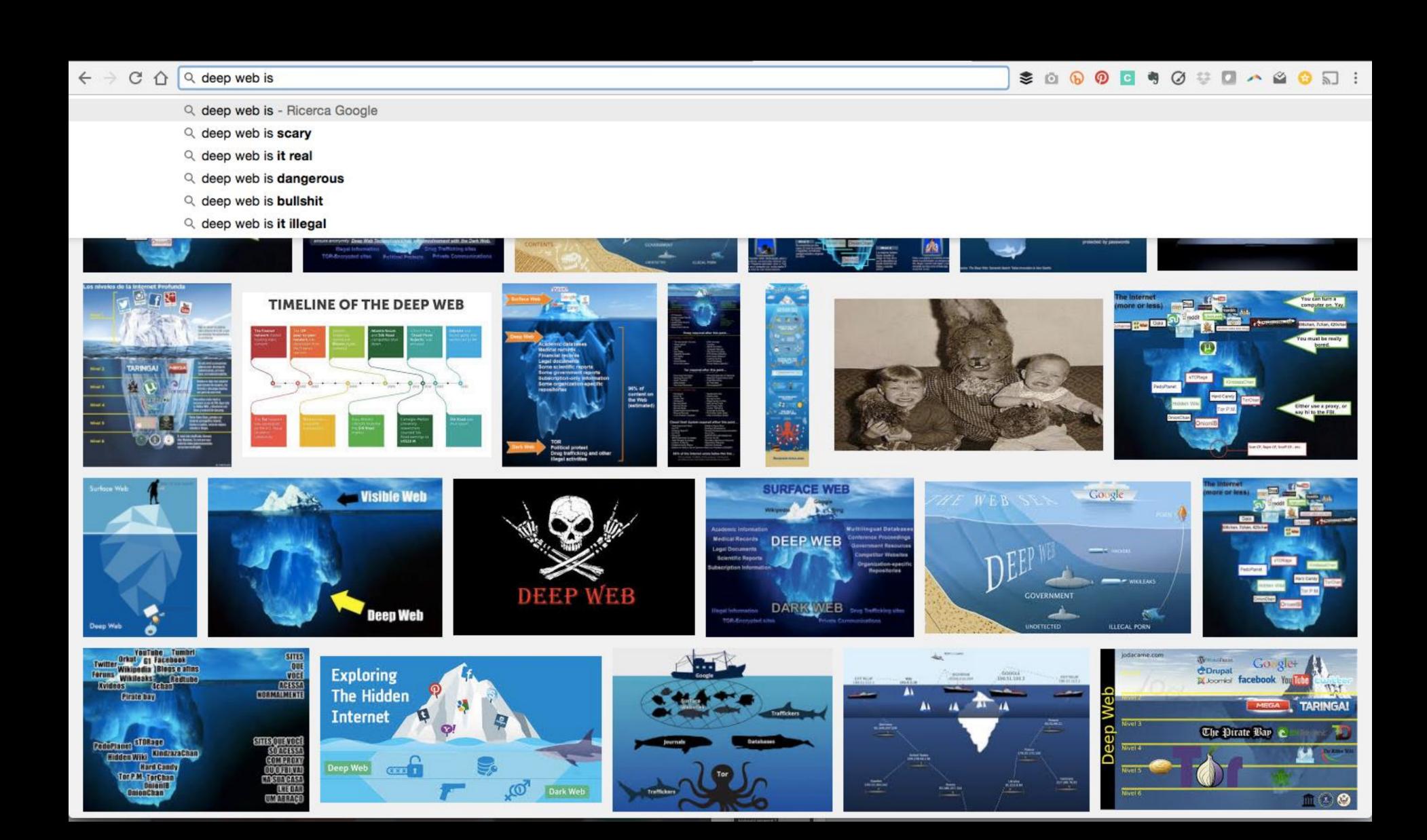
Scraping the Deep Web

An Experimental Journey on a Controversial Terrain

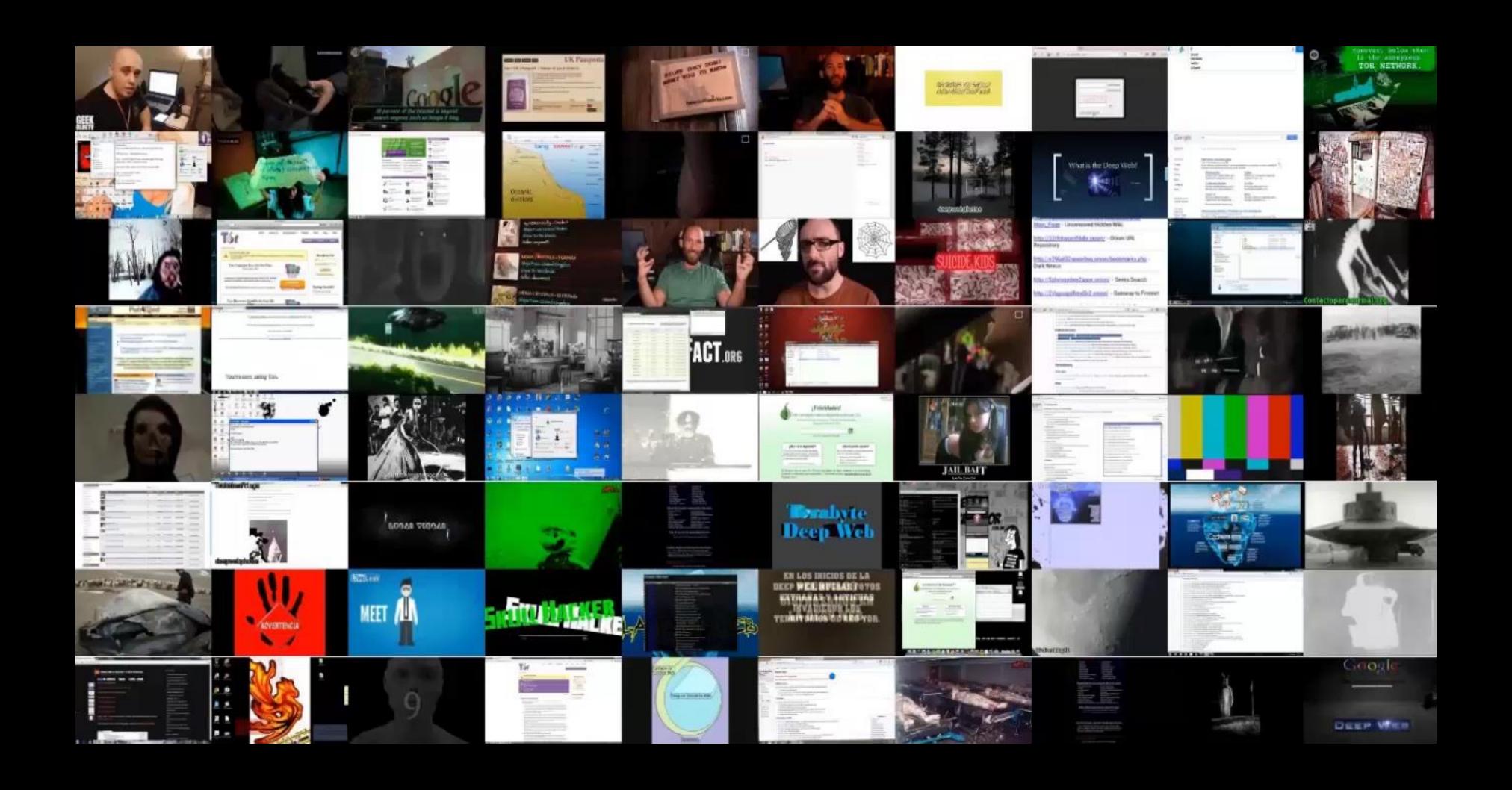


Onion, TOR and the Deep Web

behind the (dark) curtain



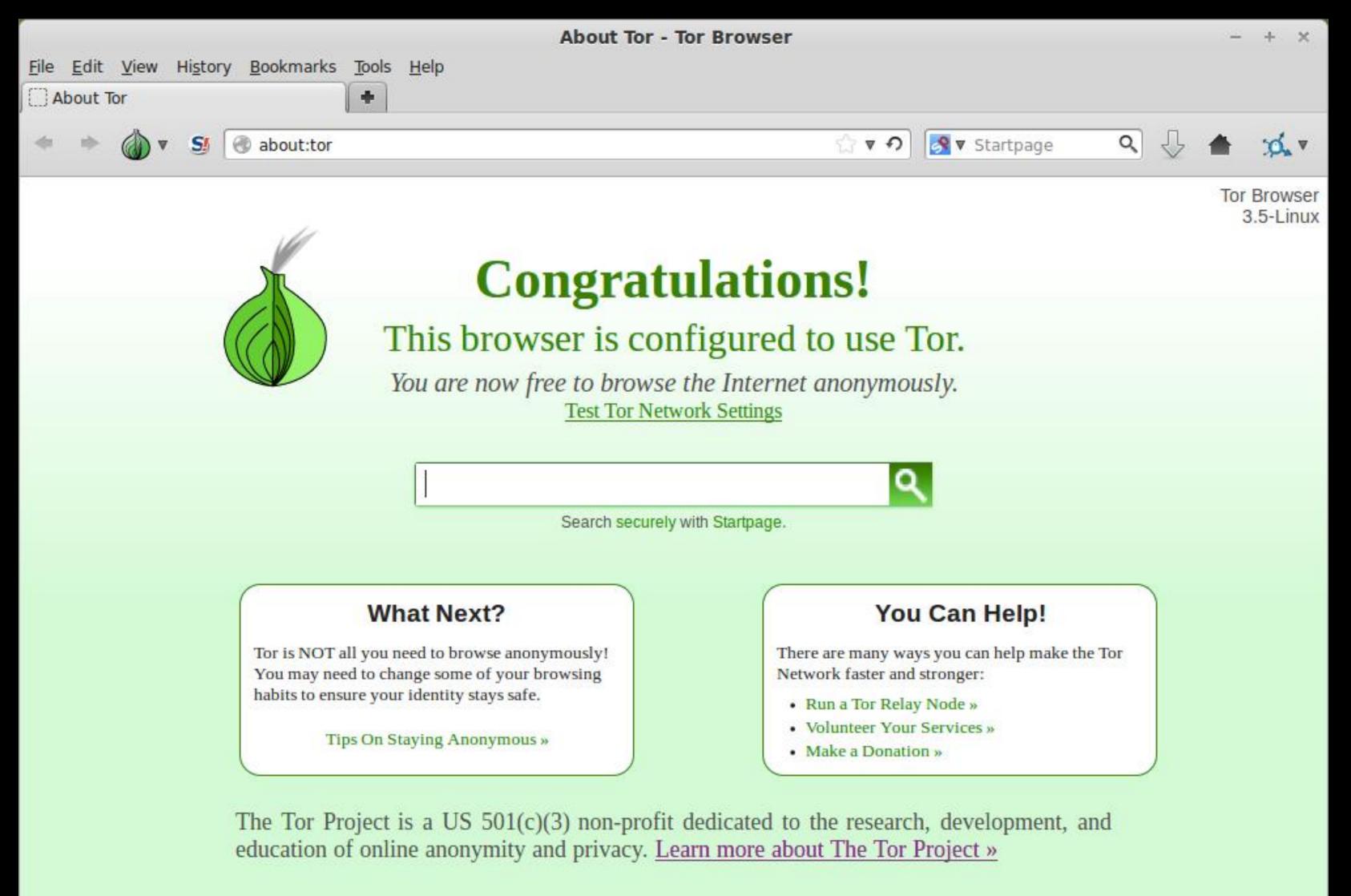
behind the (dark) curtain



The Deep Web is a World Wide Web layer that needs specific software to be accessed.

Onion routing is a technique for anonymous communication over a computer network.

In an onion network, messages are encapsulated in layers of encryption.



Tor Browser

A web browser implementing the onion routing protocol, thus enabling users to anonymously access deep web sites.

https://www.torproject.org/projects/!torbrowser.html.en

Weapons trafficking on the Deep Web





main address: gunsjmzh2btr7lpy.onion mirror: gunsdarkhzyowteb.onion

ASSAULT WEAPONS RIFLES SHOTGUNS SUBMACHINE GUNS AMMO

International Weapons Store

FULL ESCROW

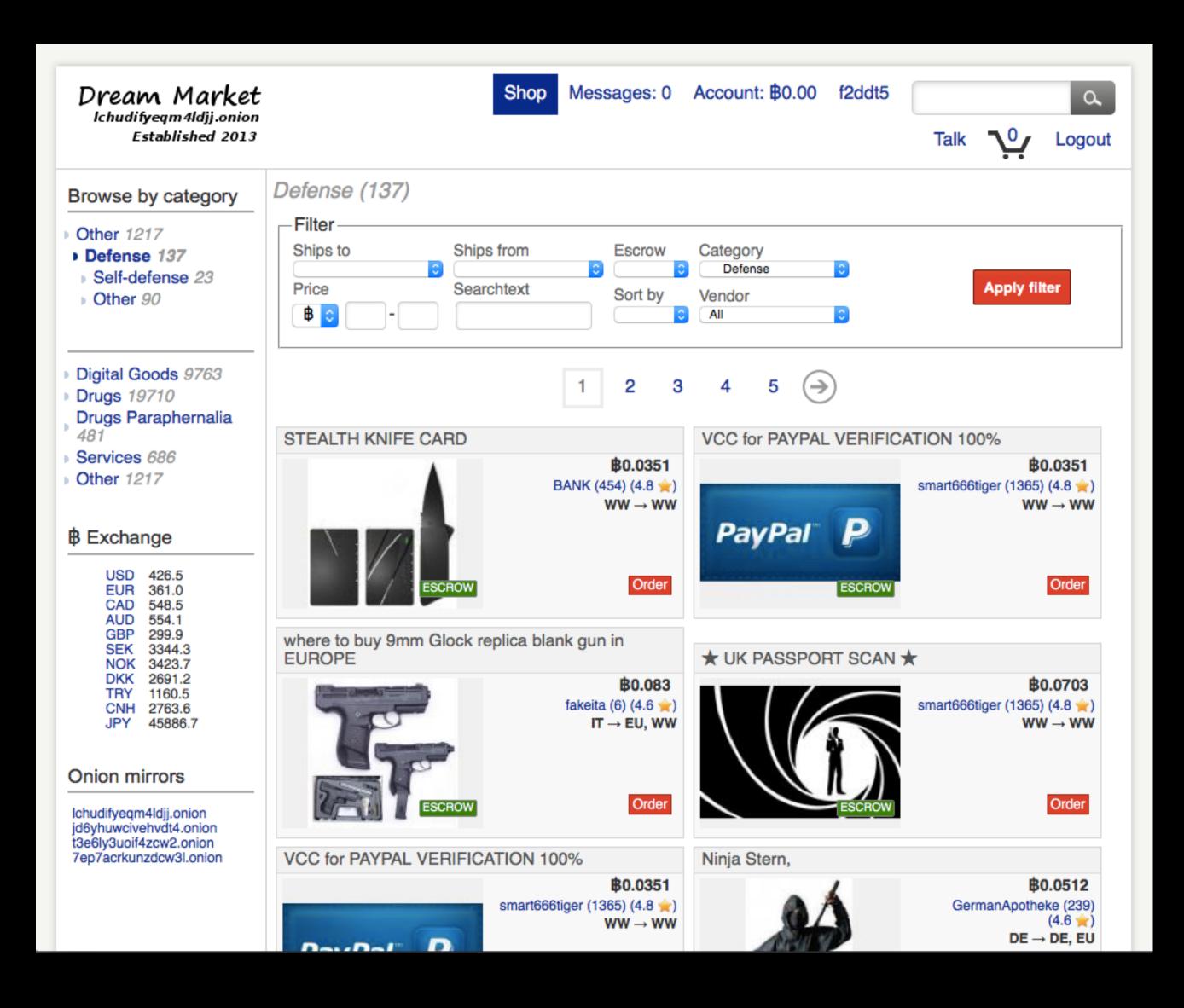
1280x768 desktop minimum resolution

Personal shops

Individual e-shops where a single seller shows available weapons for purchase

guns dark market

http://gunsjmzh2btr7lpy.onion/



"eBay" style shops

E-commerce websites where sellers and buyers can register an account to open their own shop or find wares to buy

dream market

http://lchudifyeqm4ldjj.onion/

Scraping a Deep Web site

Onion routing is a technique for anonymous communication over a computer network.

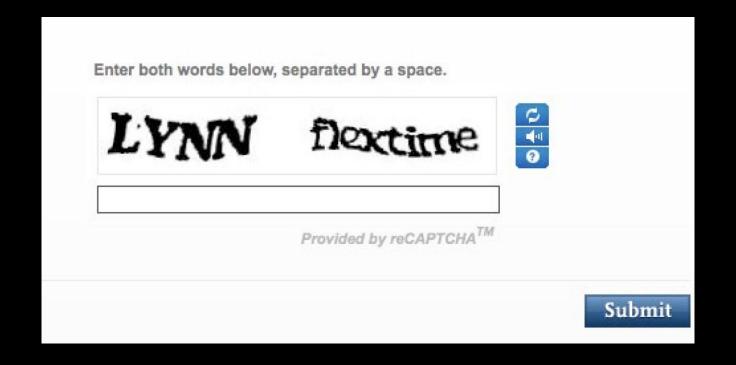
In an onion network, messages are encapsulated in layers of encryption.

deep web sites can be scraped as any other website, with some additional difficulties.

It is possible to pass through the **onion routing** by using **Tor's proxy server**.

many .onion sites require the user to fill in a captcha code to prove that they are human and prevent automated attacks.

Services that **solve captchas** for you do exist, but they don't work as expected.



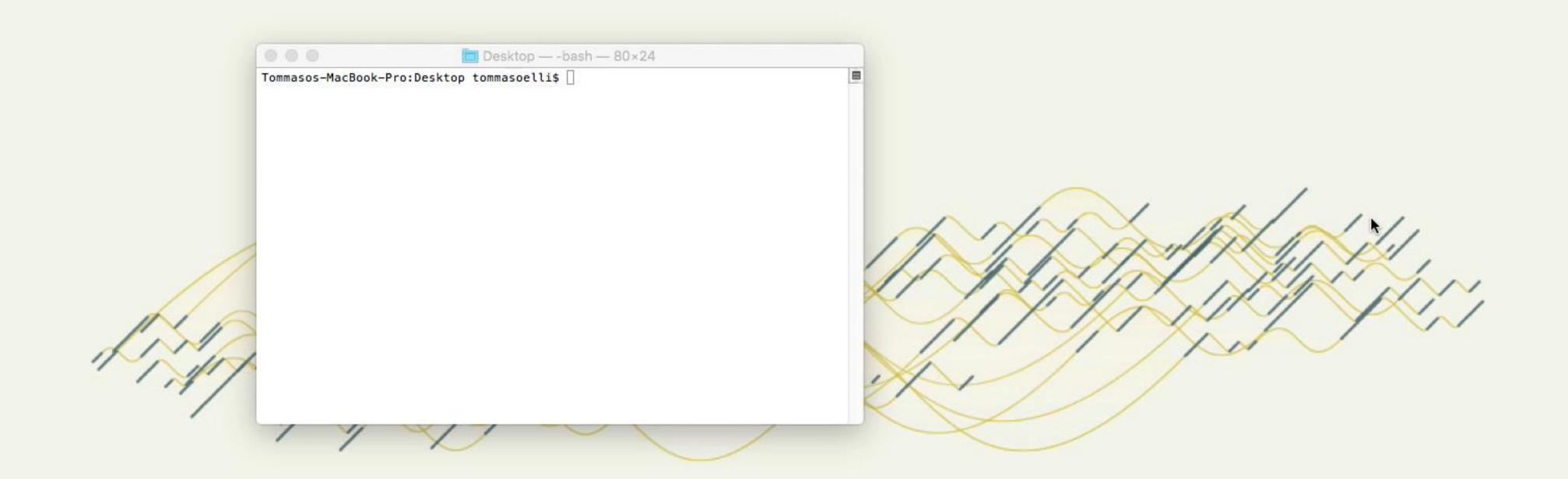
Solution

A semi-automated deep web crawler that waits for the user input on the captcha page and automatically takes every relevant information.

QuickTime Player File Edit View Window Help

Demo





Don't try this at home!

Results

3 months from 6 sites (Black market guns, Oasis, Valhalla, Outlaw, Alphabay market, Dream market)

name, price, description + caliber, provenience, destination -

Challenges

heterogeneity (analysis)

instability (consistency)

reliability (validation)

Value

experimental approach

unprecedented collection (and analysis)

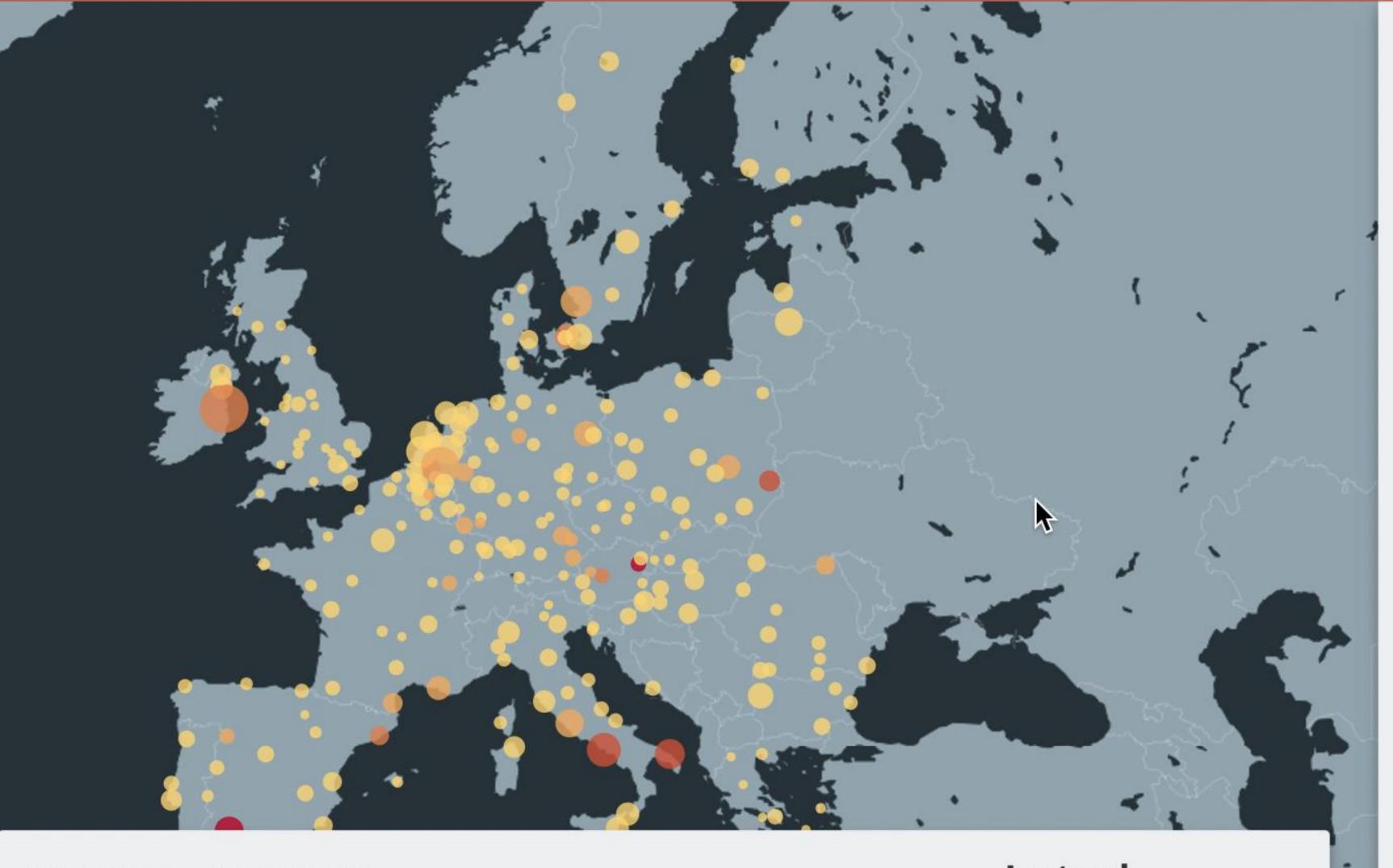
knowledge (limits & legacy)

Interfaces for visual explorations

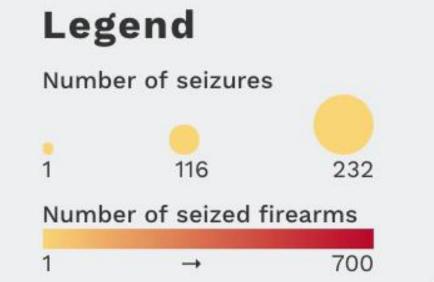
from collection to visualization (sense-making)



Available at: http://fireproject.eu/explore



Timeline ALL COUNTRIES 80 60 40 20 2010 2011 2012 2013 2014 2015

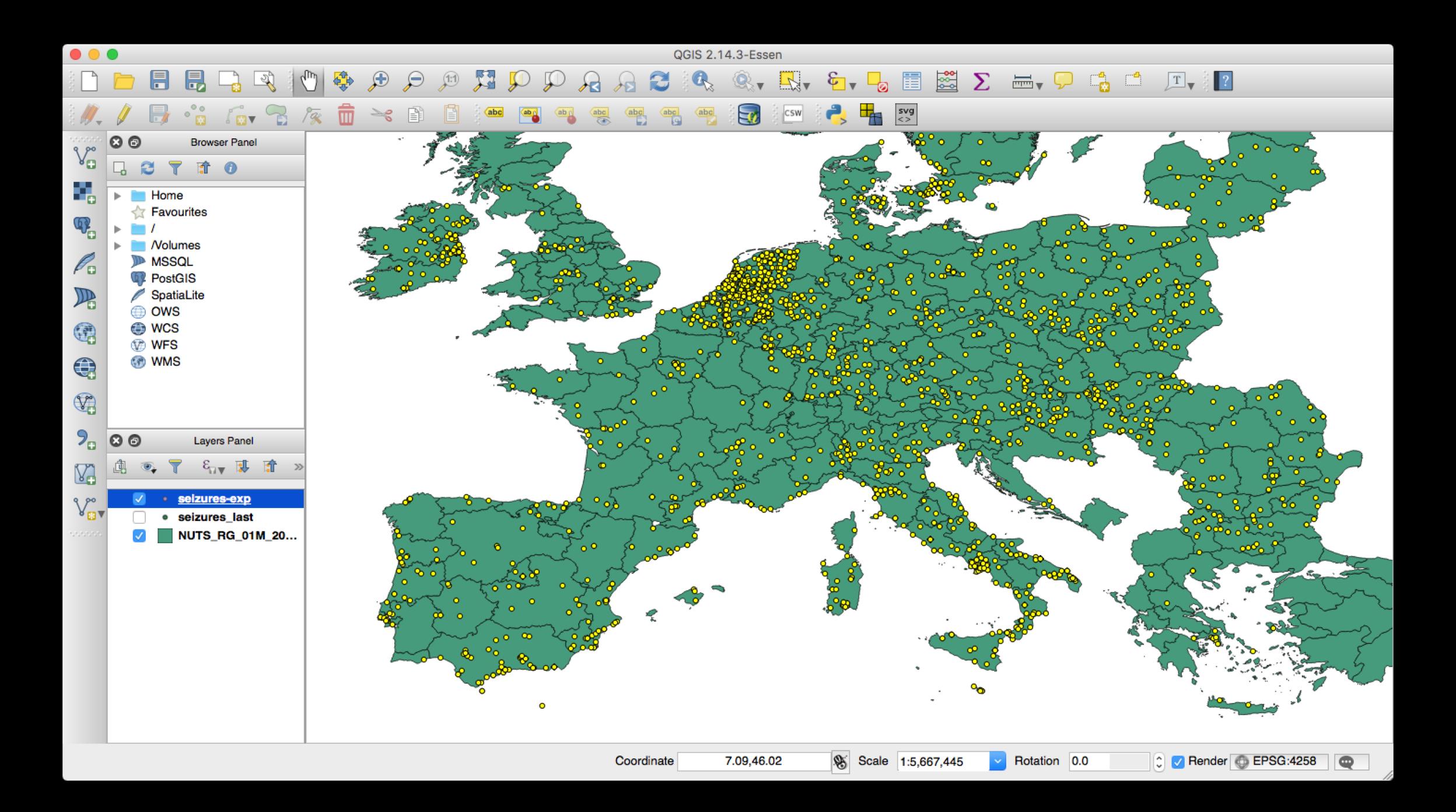


AUSTRIA	>
BELGIUM	>
BULGARIA	>
CROATIA	>
CYPRUS	>
CZECH REPUBLIC	>
DENMARK	>
ESTONIA	>
FINLAND	>
FRANCE	>
GERMANY	>
GREECE	>
HUNGARY	>
IRELAND	>
ITALY	>
LATVIA	>
LITHUANIA	>

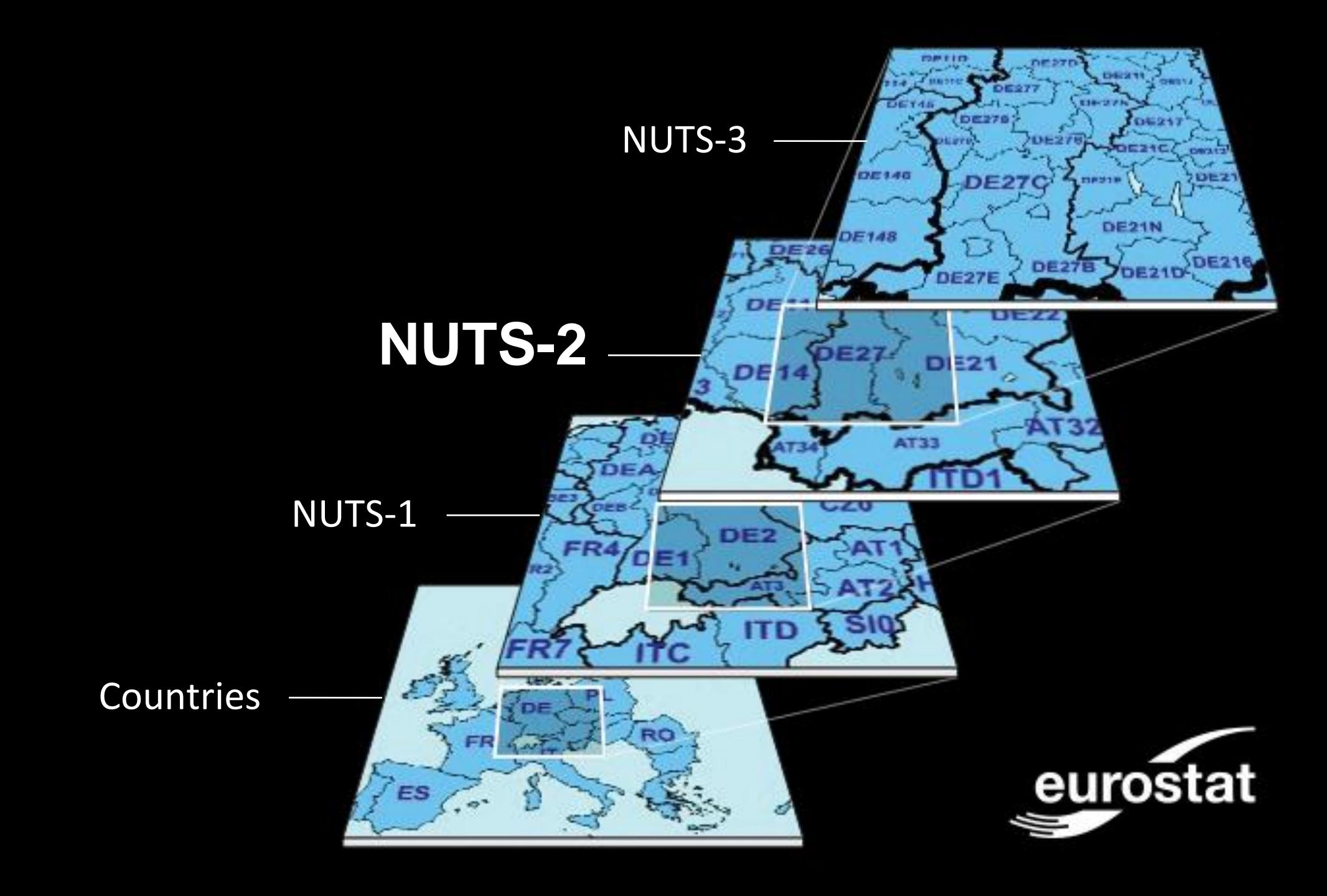
web-app compatible with modern browsers.

Based on open standard and libraries.

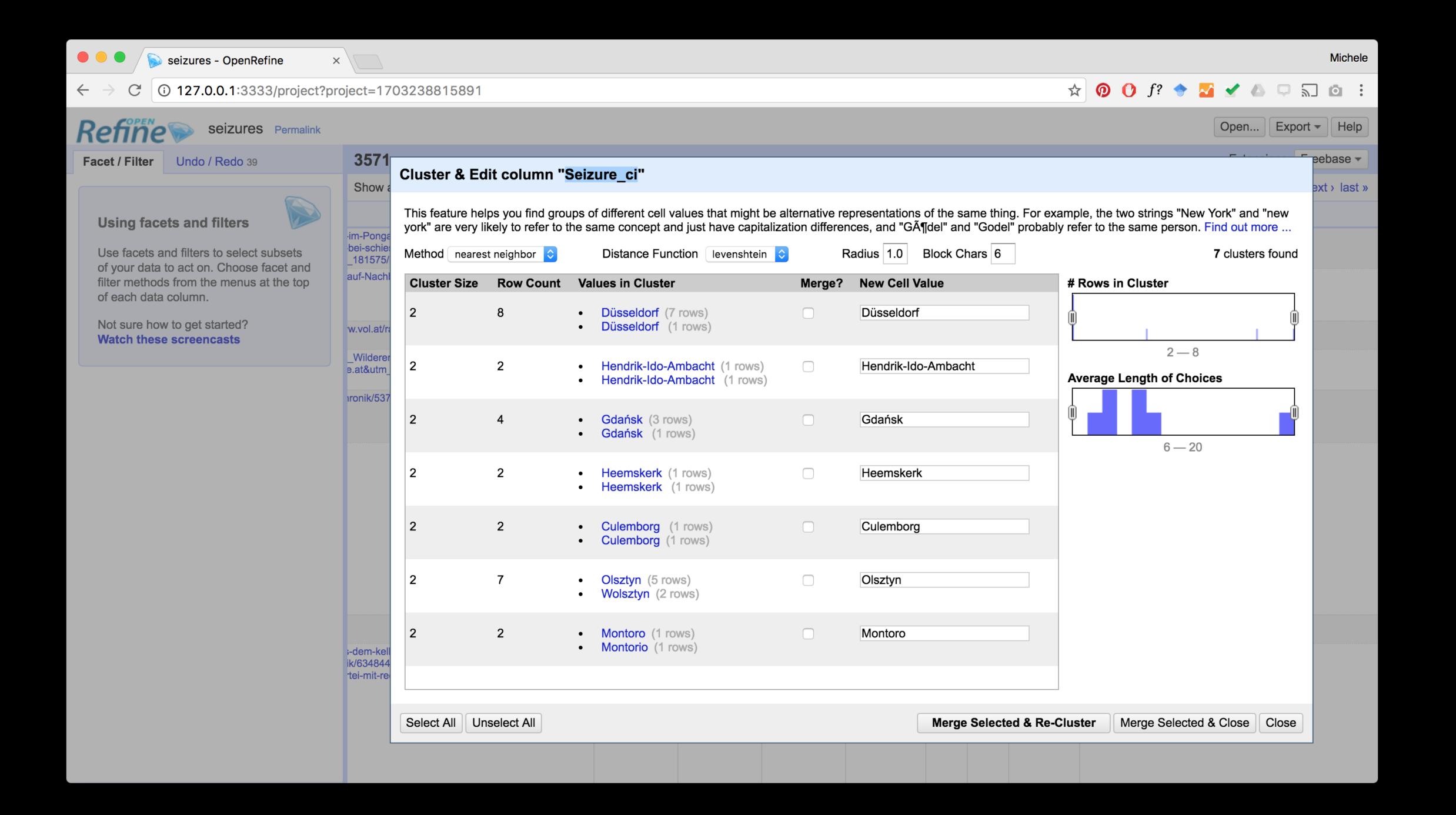
Collected data had been geo-located



and clustered by region.



Data were then checked and refined to avoid inconsistencies.



Visualization of seizures

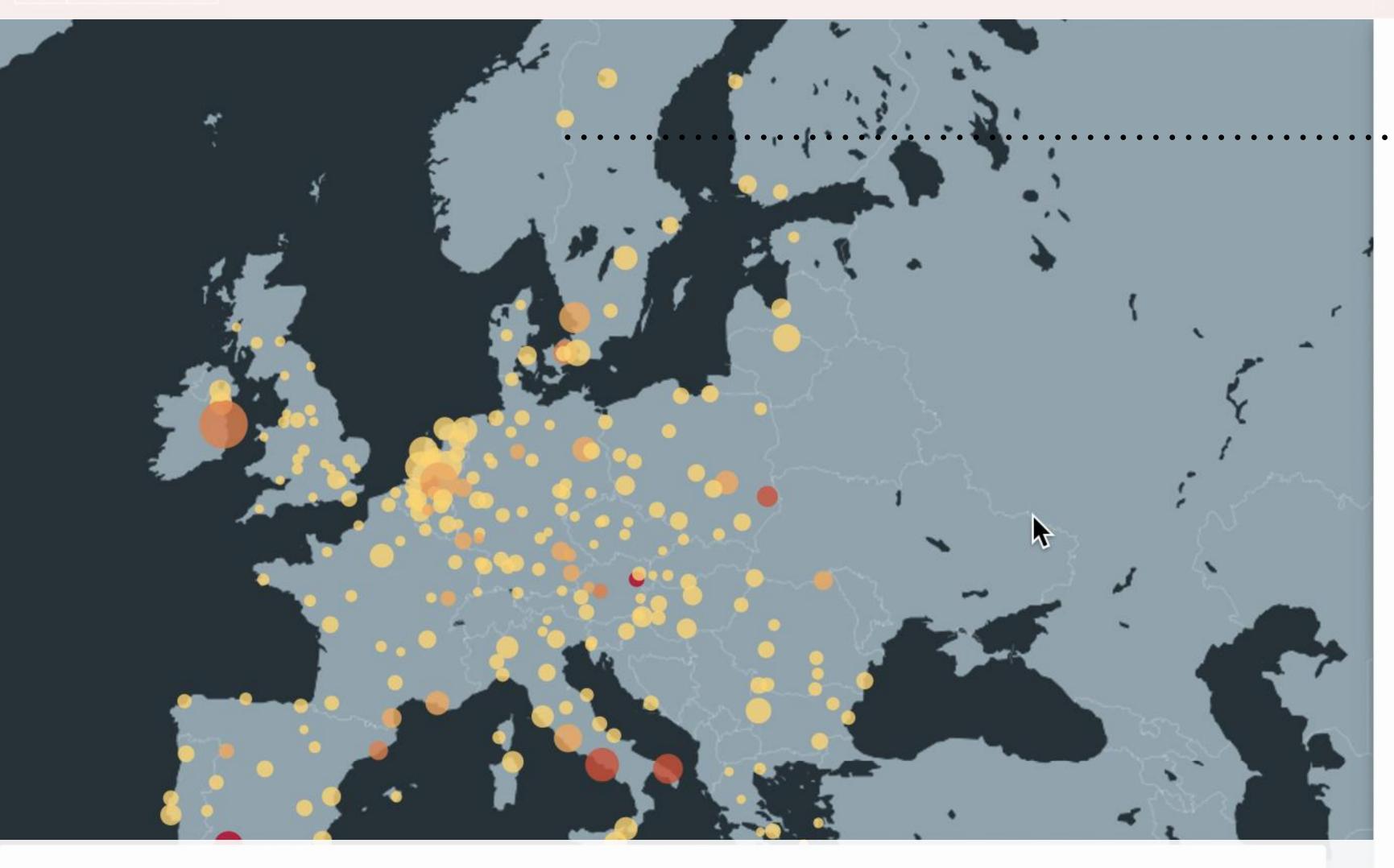
http://fireproject.eu/explore/#/seizures



focus on seizures events

data is aggregated:

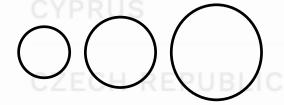
spatially
temporally



Map showing the locations of seizures

Close location are clustered together in one circle for better clarity

Size



Number of seizures

Color





FINLAND

FRANCE

GERMANY

GREECE

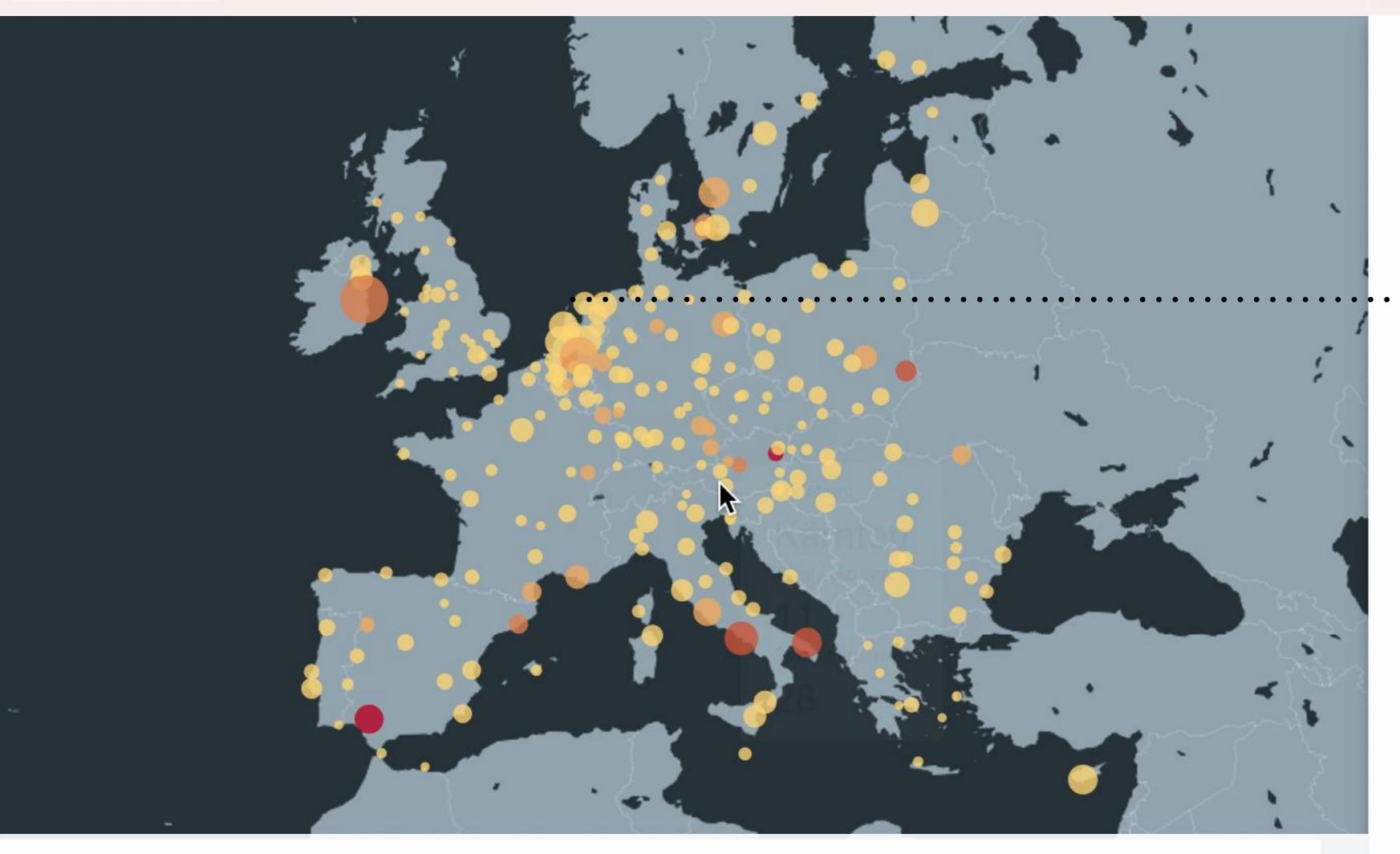
HUNGARY

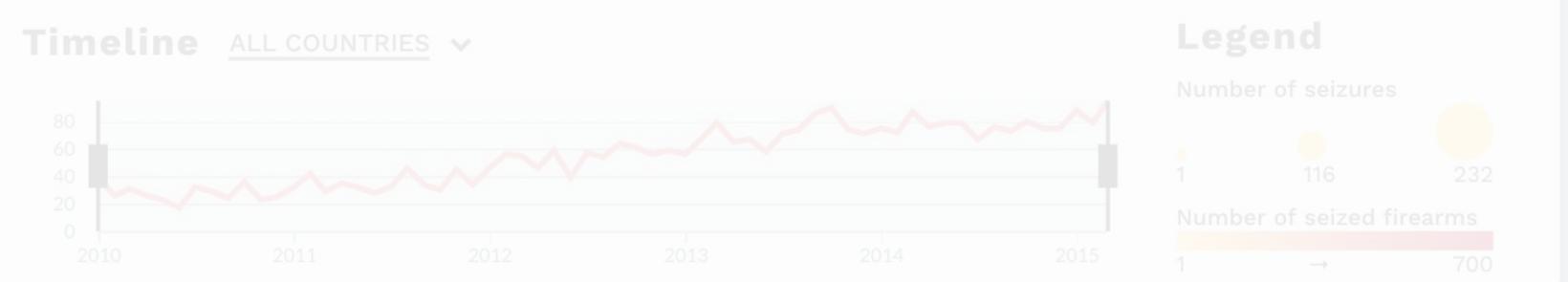
IRELAND

ITALY

LATVIA

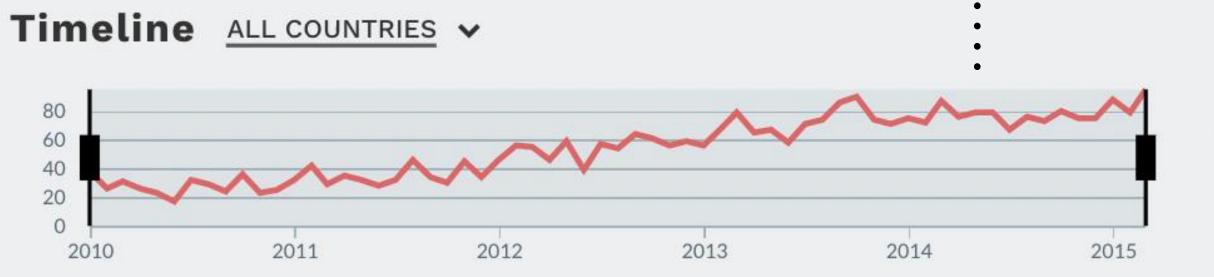
LITHUANIA





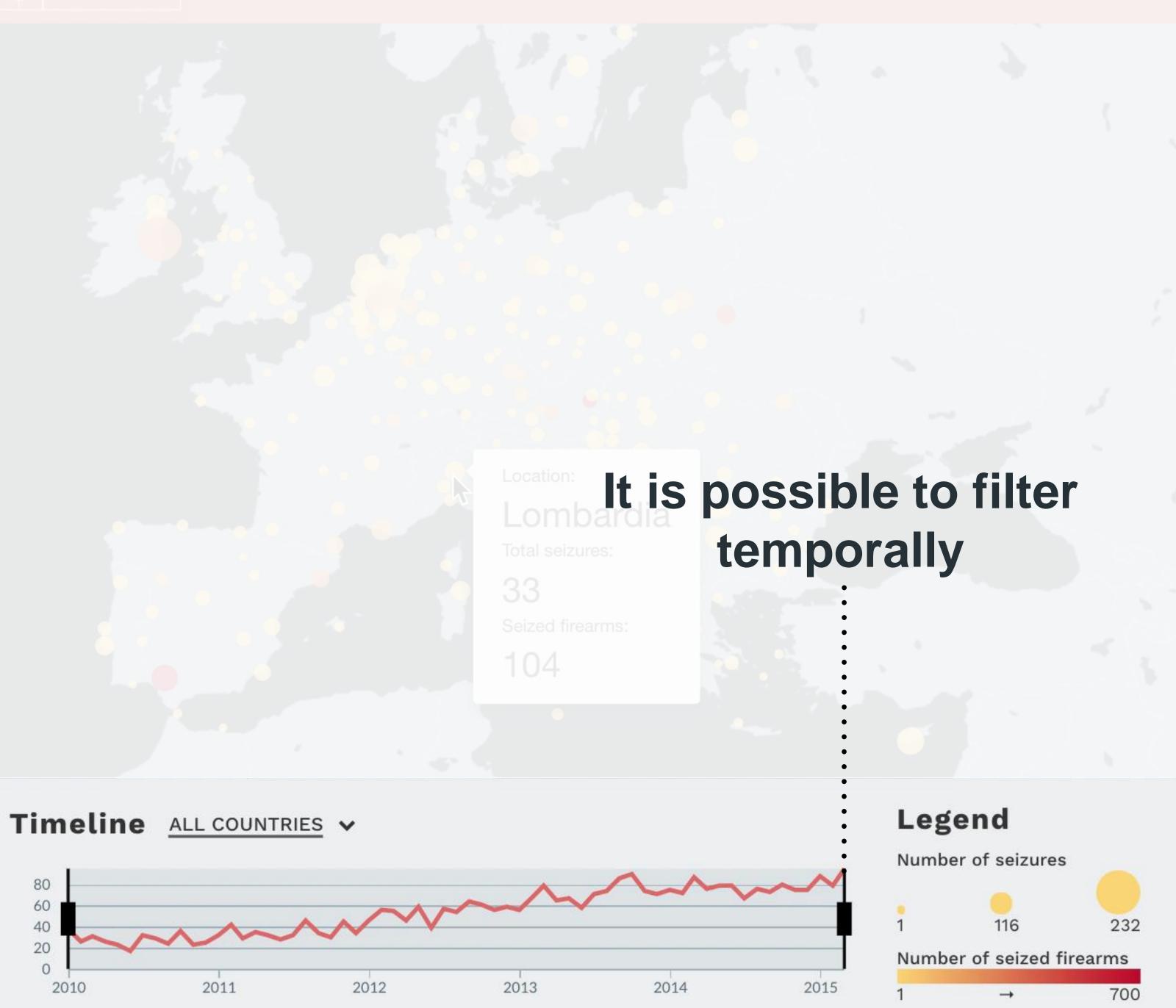
Timeline representing the number of firearms seized

33
Seized firearms:



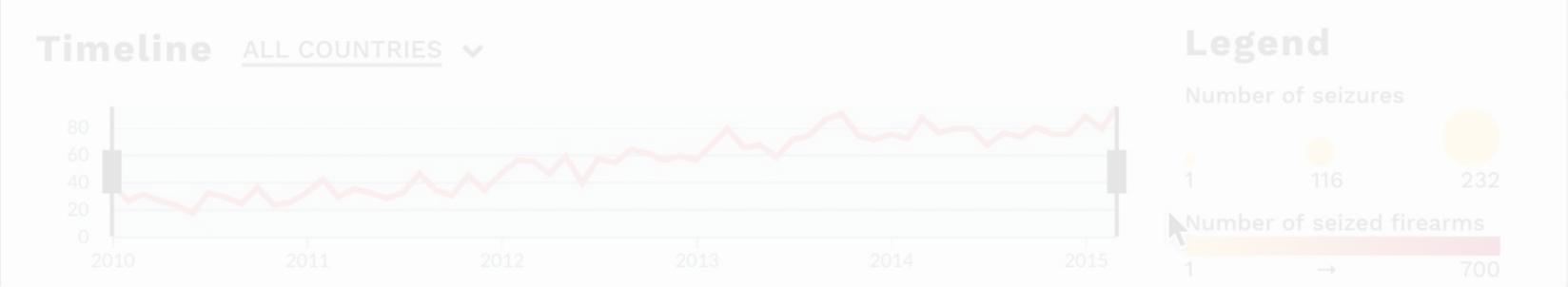
Legend

Numb	er of seizures	
1	116	232
Numb	er of seized f	irearms
1	→	700



Or on a specific country Legend Timeline ALL COUNTRIES > Number of seizures 80 60 116 232 Number of seized firearms 0 7 2010 2014 2015 2012 2013 2011 700

The panel on the right provides a list of all the seizures



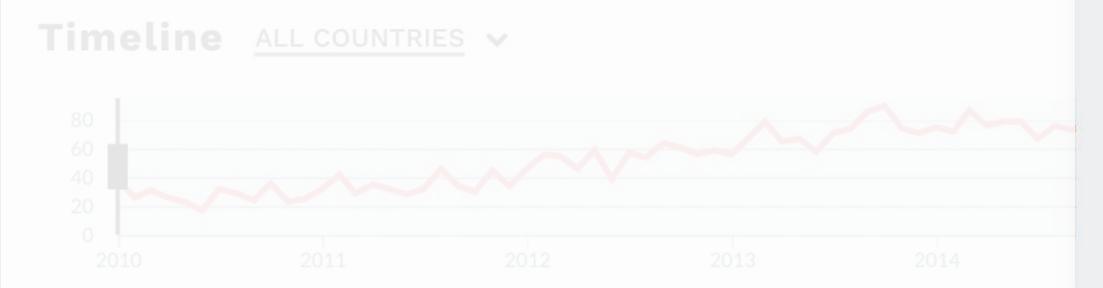
AUSTRIA	>
BELGIUM	>
BULGARIA	>
CROATIA	>
CYPRUS	>
CZECH REPUBLIC	>
DENMARK	>
ESTONIA	>
FINLAND	>
FRANCE	>
GERMANY	>
GREECE	>
HUNGARY	>
IRELAND	>
ITALY	>
LATVIA	>
LITHUANIA	>

Selecting a single case, it is possible to see all the related details

where firearms were found · · ·

involved people, their nationality, gender and age · · ·

the list of seized firearms · · ·



close 🗙

Krumpendorf am Wörthersee 10/05/2012

Storage	Amn	nunition
House	Yes	
Actors	1	
Nationality	Gender	Age
German	Ť	35
Firearms	7	
type		number
Other		3
Pistol		4

	AUSTRIA	
		•
	City	Number of firearms
9	Sankt Johann im Pongau	1
90	Spittal an der Drau	1
90	Bregenz	1
%	Bad Ischl	1
90	Götzis	N/A
90	Villach	3
90	Liezen	80
%	Hartberg	N/A
90	Vienna	N/A
90	Judenburg	1
9	Salzburg	1
90	Vienna	1
9	Baden	1
90	Krumpendorf am Wörthersee	7
9	Sankt Pölten	5
90	Hallein	N/A
90	Graz	N/A
90	Sankt Veit an der Glan	1

2013

2014

80

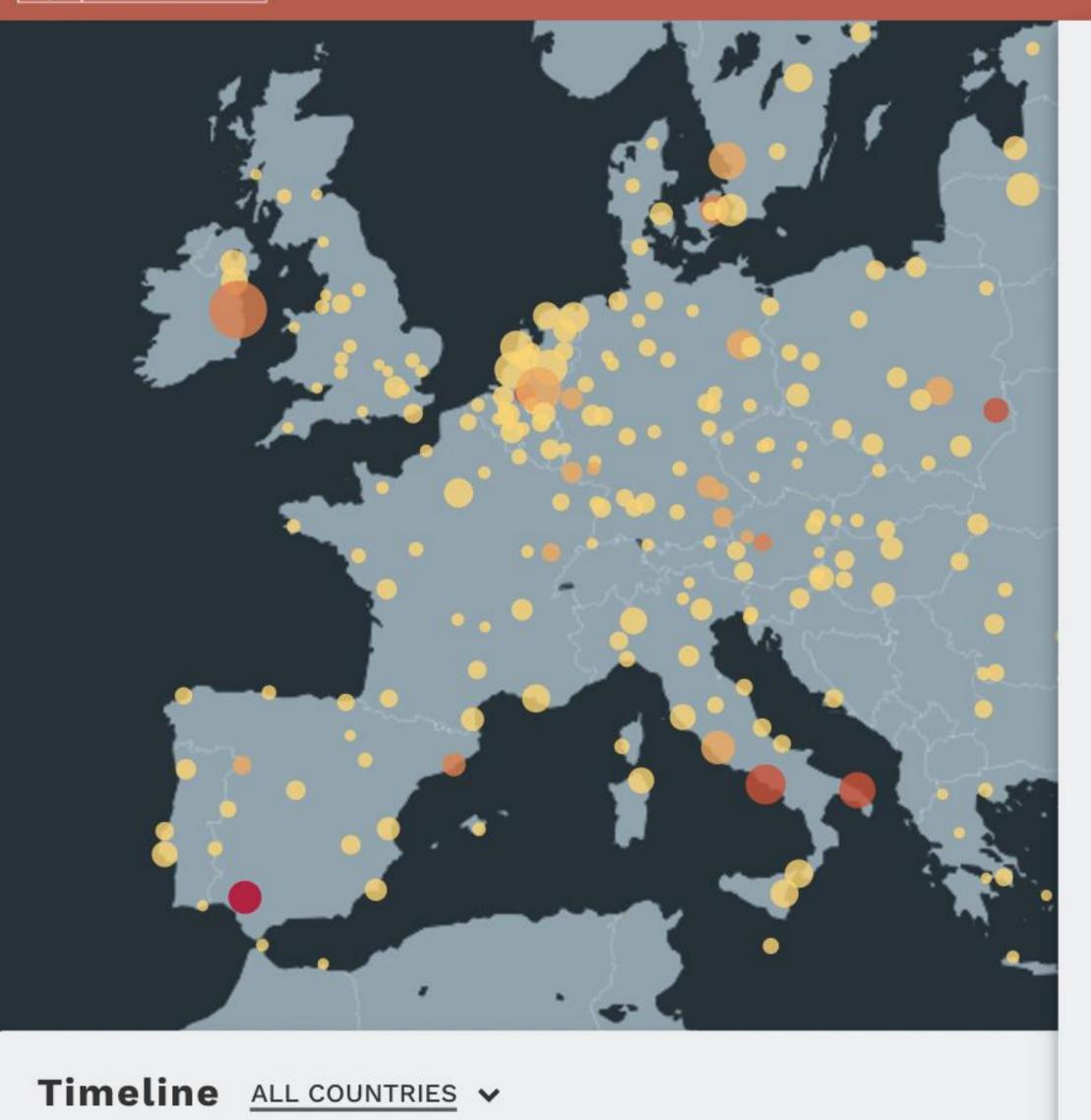
60

40

20

2010

2011



2012

Krumpend am Wörth		10/05/2012
Storage	Amn	nunition
House	Yes	
Actors	1	
Nationality	Gender	Age
German	Ť	35
Firearms	7	
type		number
Other		3
Pistol		4

Seizures list **AUSTRIA** City Number of firearms % Sankt Johann im Pongau % Spittal an der Drau % Bregenz % Bad Ischl % Götzis N/A % Villach % Liezen 80 % Hartberg N/A % Vienna N/A % Judenburg % Salzburg % Vienna % Baden % Krumpendorf am Wörthersee % Sankt Pölten % Hallein N/A

N/A

% Graz

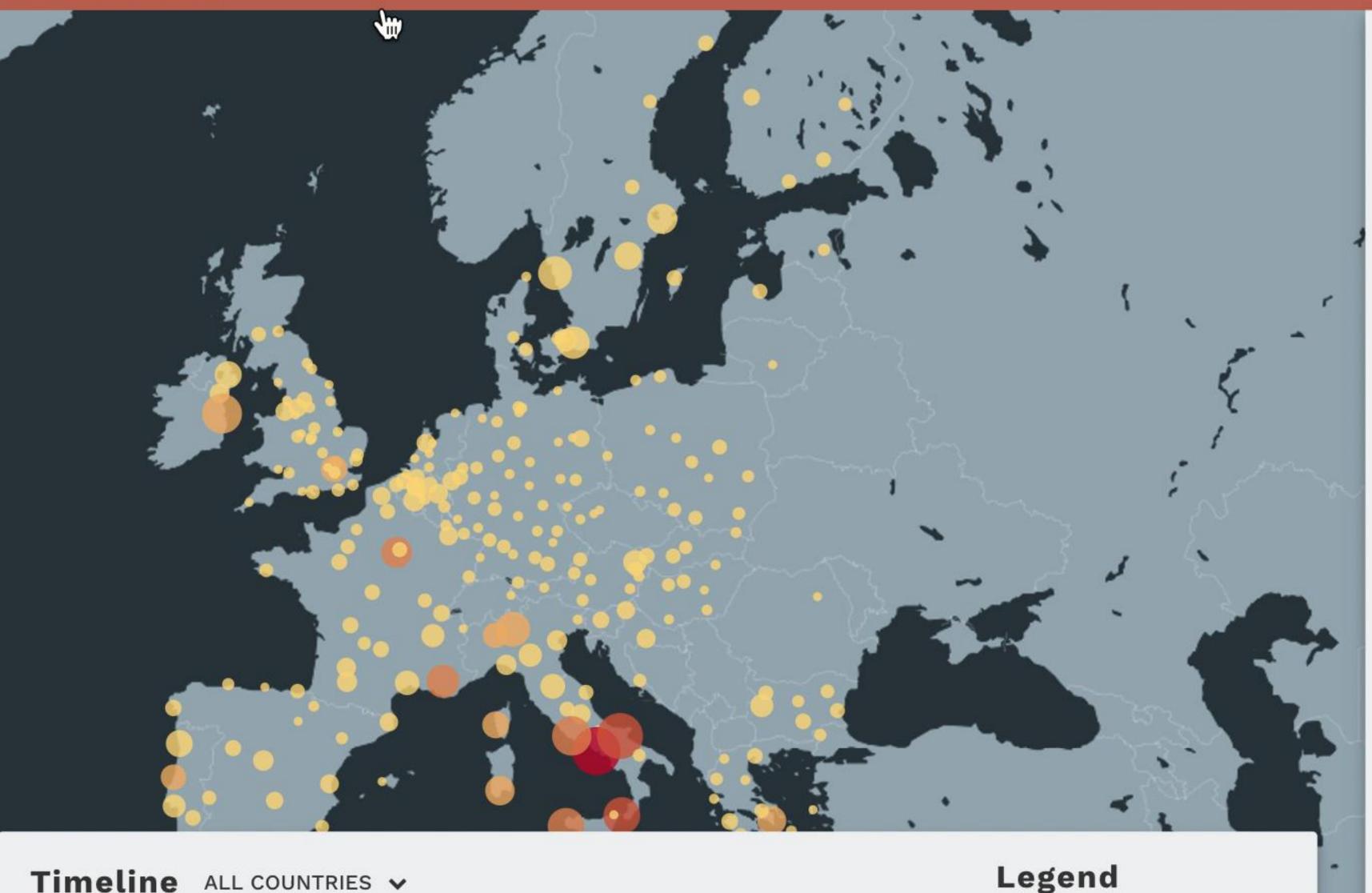
% Sankt Veit an der Glan

The interface allows to combine multiple filters

Visualization of shootings

http://fireproject.eu/explore/#/shootings





Timeline ALL COUNTRIES 100 80 40 20 0 fireproject.eu/explore/#/shootings



161

Shootings list

AUSTRIA	>
BELGIUM	>
BULGARIA	>
CROATIA	>
CYPRUS	>
CZECH REPUBLIC	>
DENMARK	>
ESTONIA	>
FINLAND	>
FRANCE	>
GERMANY	>
GREECE	>
HUNGARY	>
IRELAND	>
ITALY	>
LATVIA	>
LITHUANIA	>

focus on shooting events

where when why who shoot who were the victims data is aggregated:

spatially
temporally

Challenges

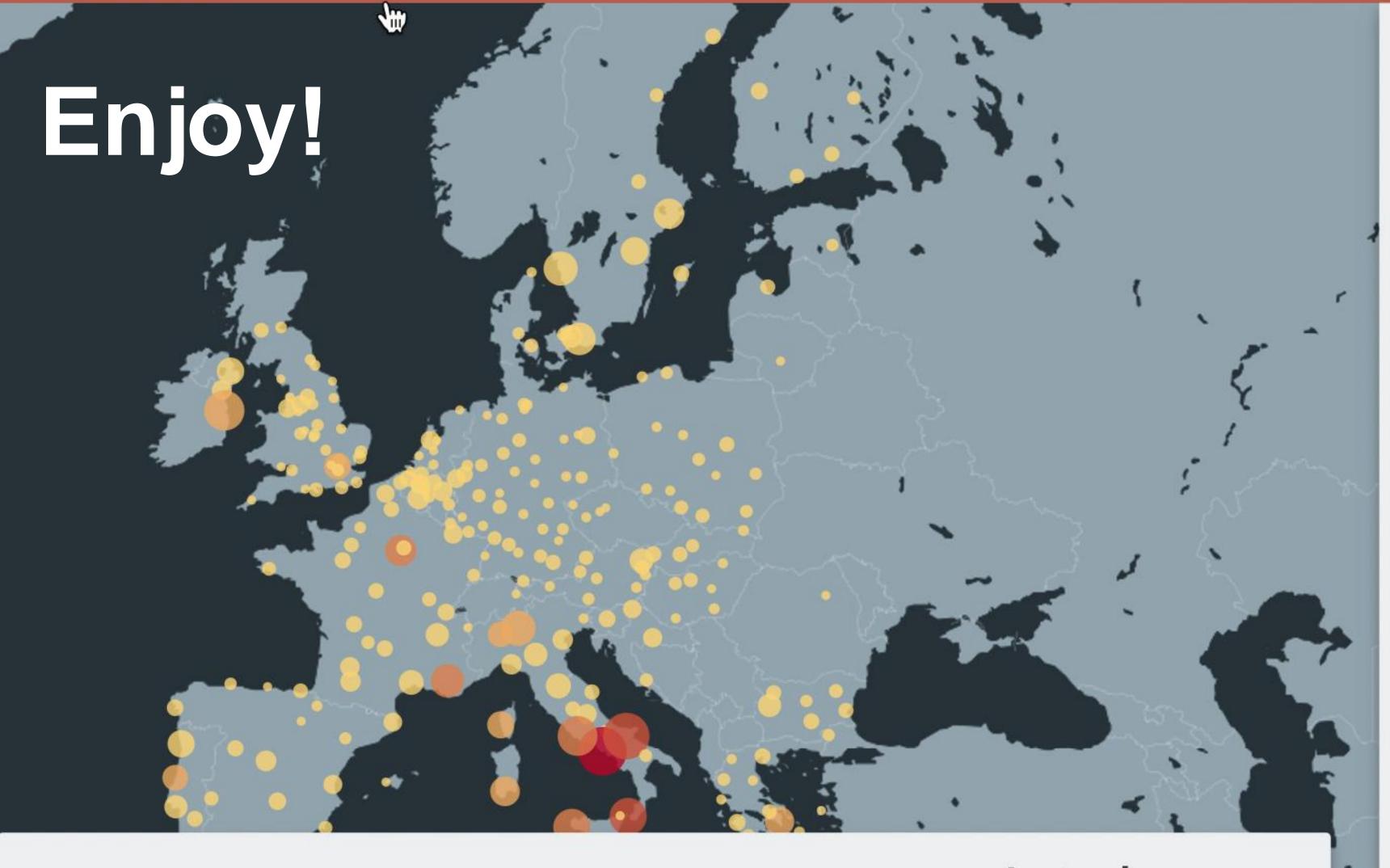
portability (devices + environments)

Value

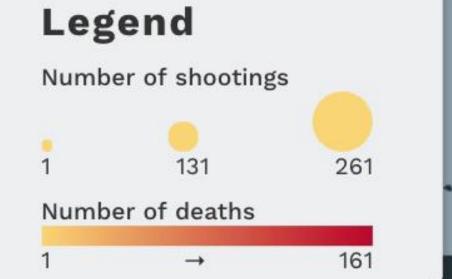
Self-service visual analytics (productivity)

deep exploration (knowledge)

dissemination (public engagement)



Timeline ALL COUNTRIES 100 80 60 40 20 0 fireproject.eu/explore/#/shootings



Shootings list

AUSTRIA	>
BELGIUM	>
BULGARIA	>
CROATIA	>
CYPRUS	>
CZECH REPUBLIC	>
DENMARK	>
ESTONIA	>
FINLAND	>
FRANCE	>
GERMANY	>
GREECE	>
HUNGARY	>
IRELAND	>
ITALY	>
LATVIA	>
LITHUANIA	>

Thank you.

Paolo Ciuccarelli - DensityDesign Research Lab (Politecnico di Milano) paolo.ciuccarelli@polimi.it - www.densitydesign.org

