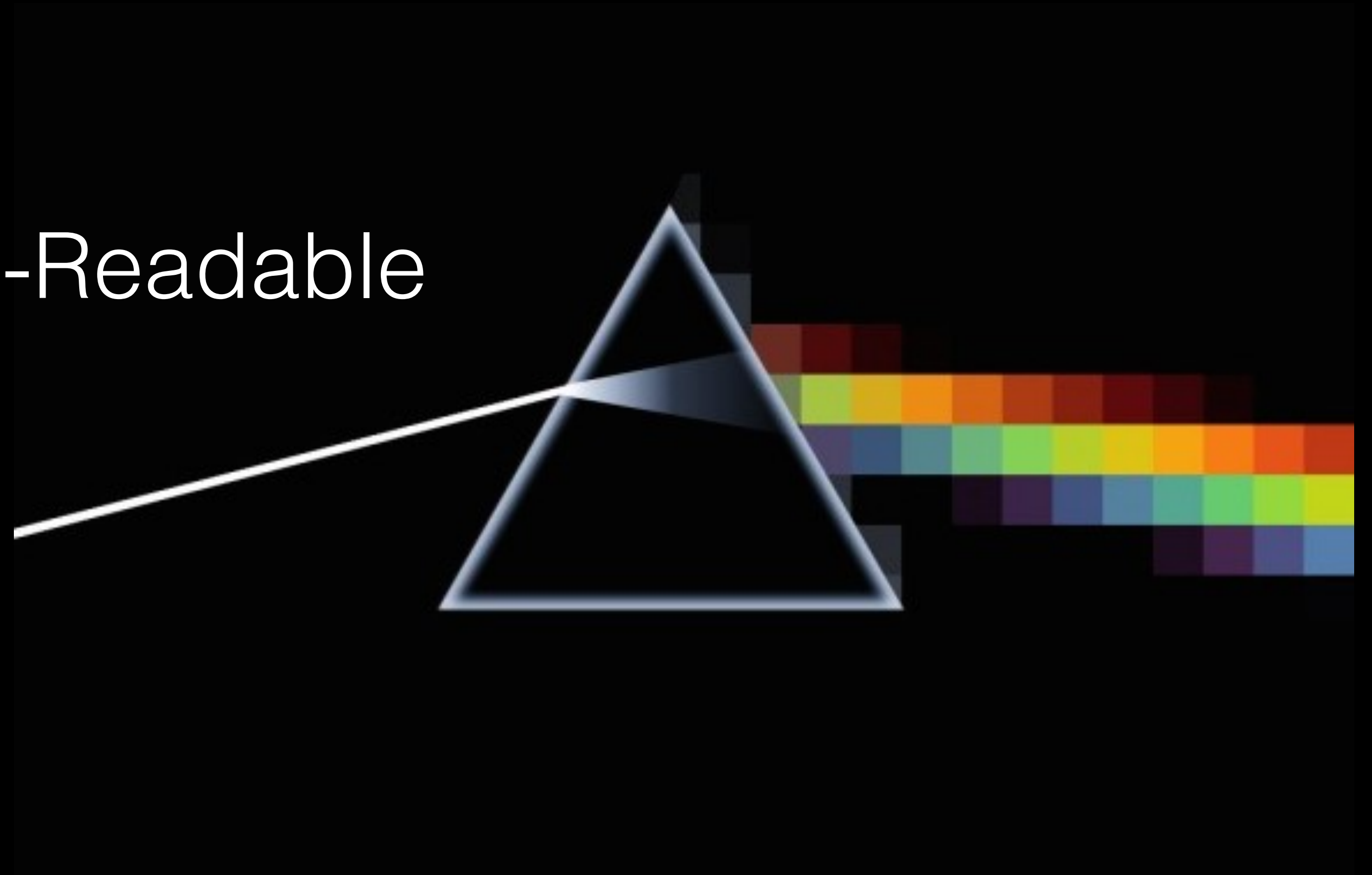


# Towards the Machine-Readable Society



What we asked for

# The Machine-Readable State



At their computer congress, the CCC also demanded the ‚machine-readable government‘ [...] „An example would be free access to all environmental and radiation data over Bildschirmtxt (BTX), so that any citizen is able to check the kind of dirt that fills the air around him“.

— Die Tageszeitung, 1987

- data collected by the government should be published under a **free license** and in an **open, computer-readable format**
- make state action transparent and controllable for citizens with computer access

What we got

...



But don't worry

# It's just so-called metadata

„[...] nobody is listening to your telephone calls. [...] what the intelligence community is doing is looking at phone numbers and durations of calls. They are not looking at people's names, and they're **not looking at content**. But by sifting through this **so-called metadata**, they may identify potential leads with respect to folks who might engage in terrorism.“

— Barack Obama

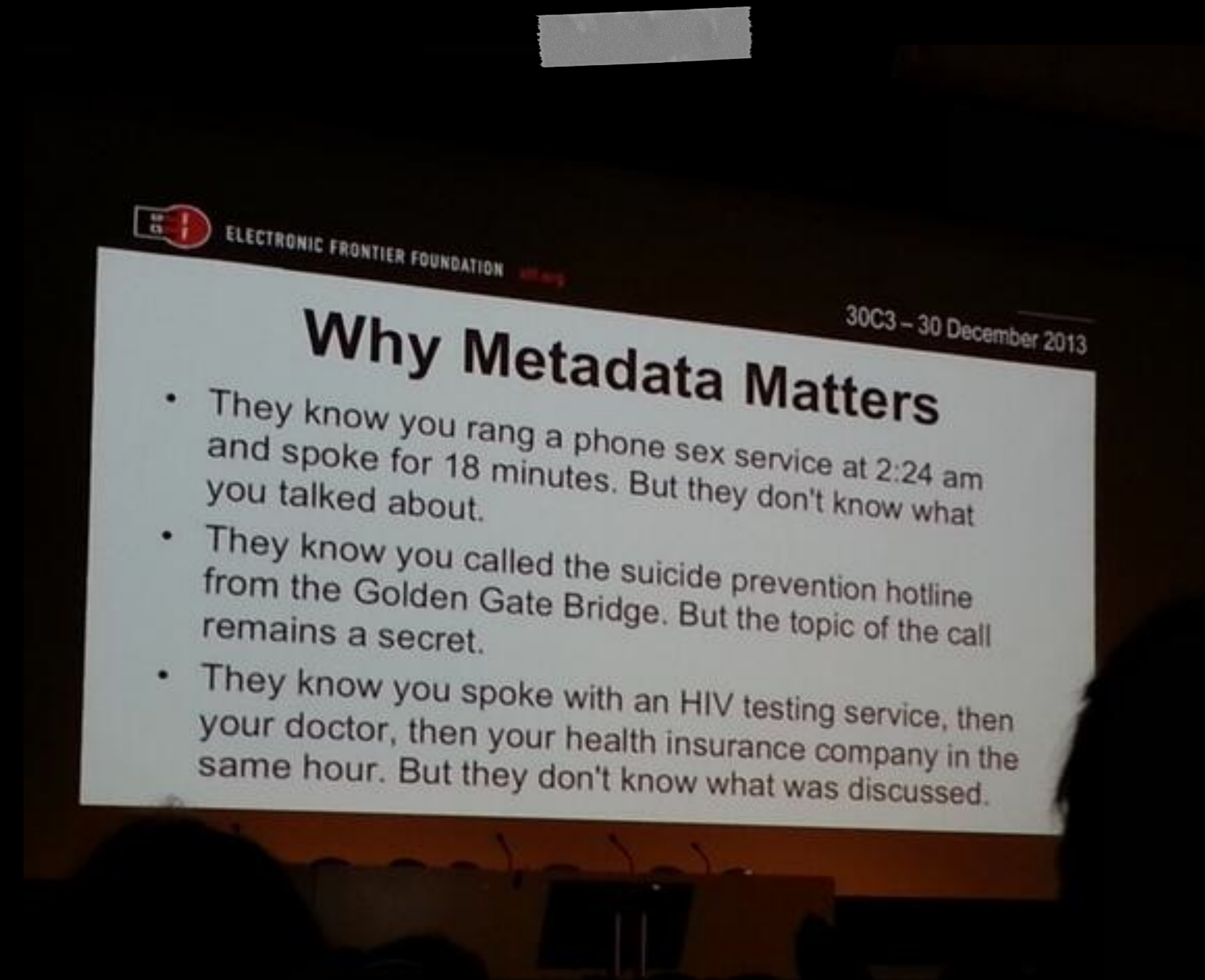


But don't worry

# It's just so-called metadata

„[...] nobody is listening to your telephone calls. [...] what the intelligence community is doing is looking at phone numbers and durations of calls. They are not looking at people's names, and they're **not looking at content**. But by sifting through this **so-called metadata**, they may identify potential leads with respect to folks who might engage in terrorism.“

— Barack Obama

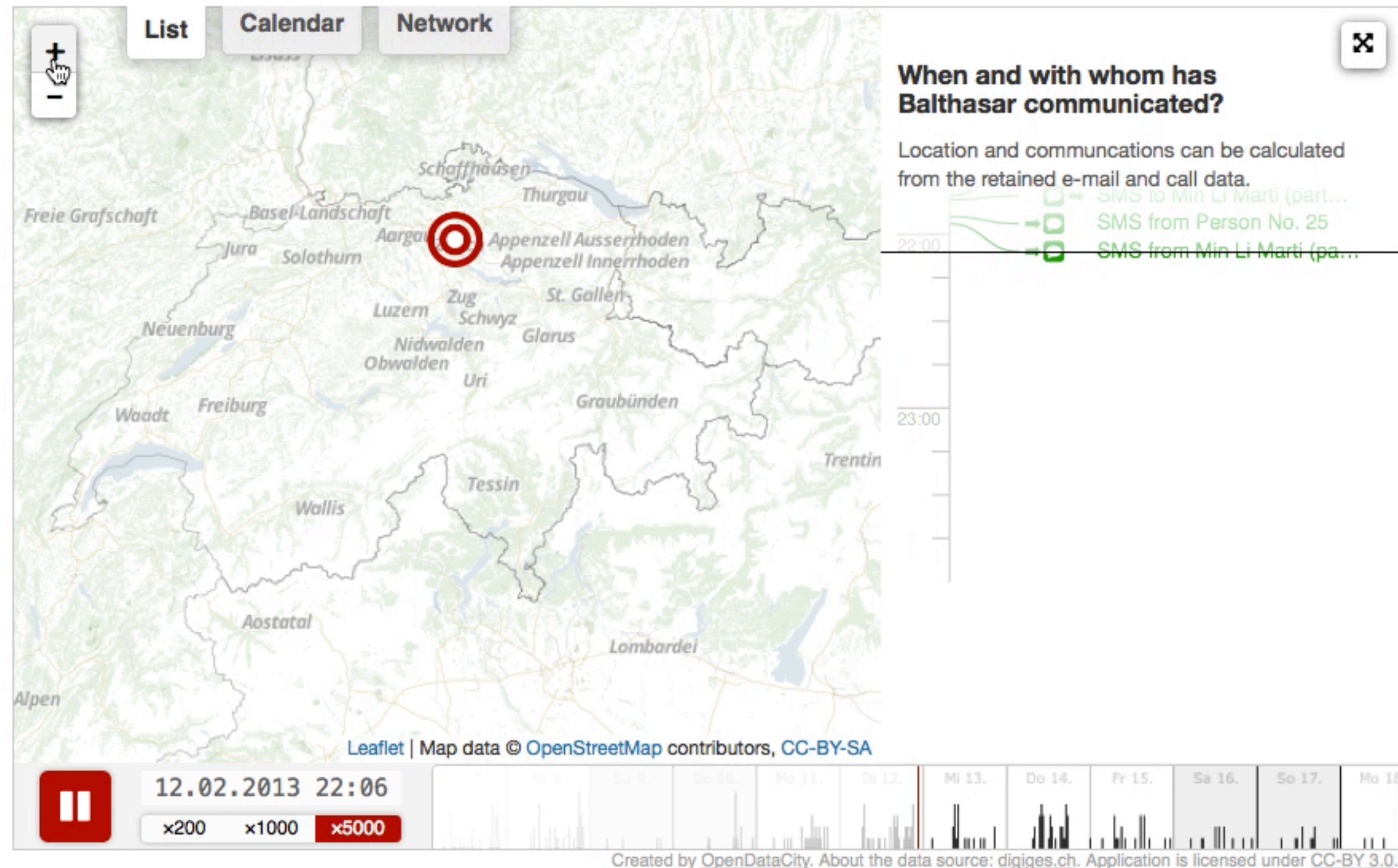


➡ 30C3 EFF Talk: [youtu.be/0Fuxb9WXIol](https://youtu.be/0Fuxb9WXIol)



# Swiss Data Retention Visualized

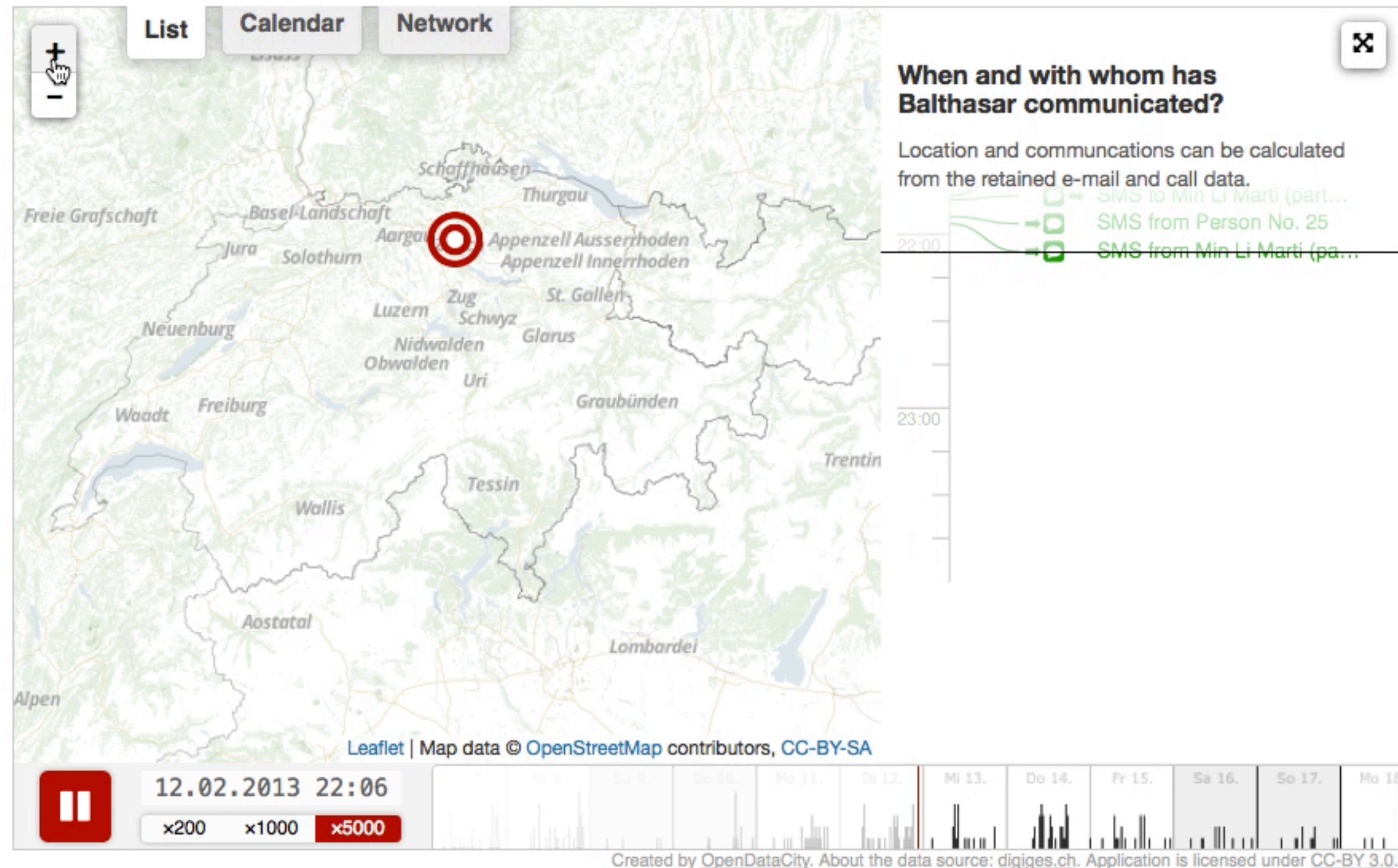
The monitored life of National Councilor Balthasar Glättli.





# Swiss Data Retention Visualized

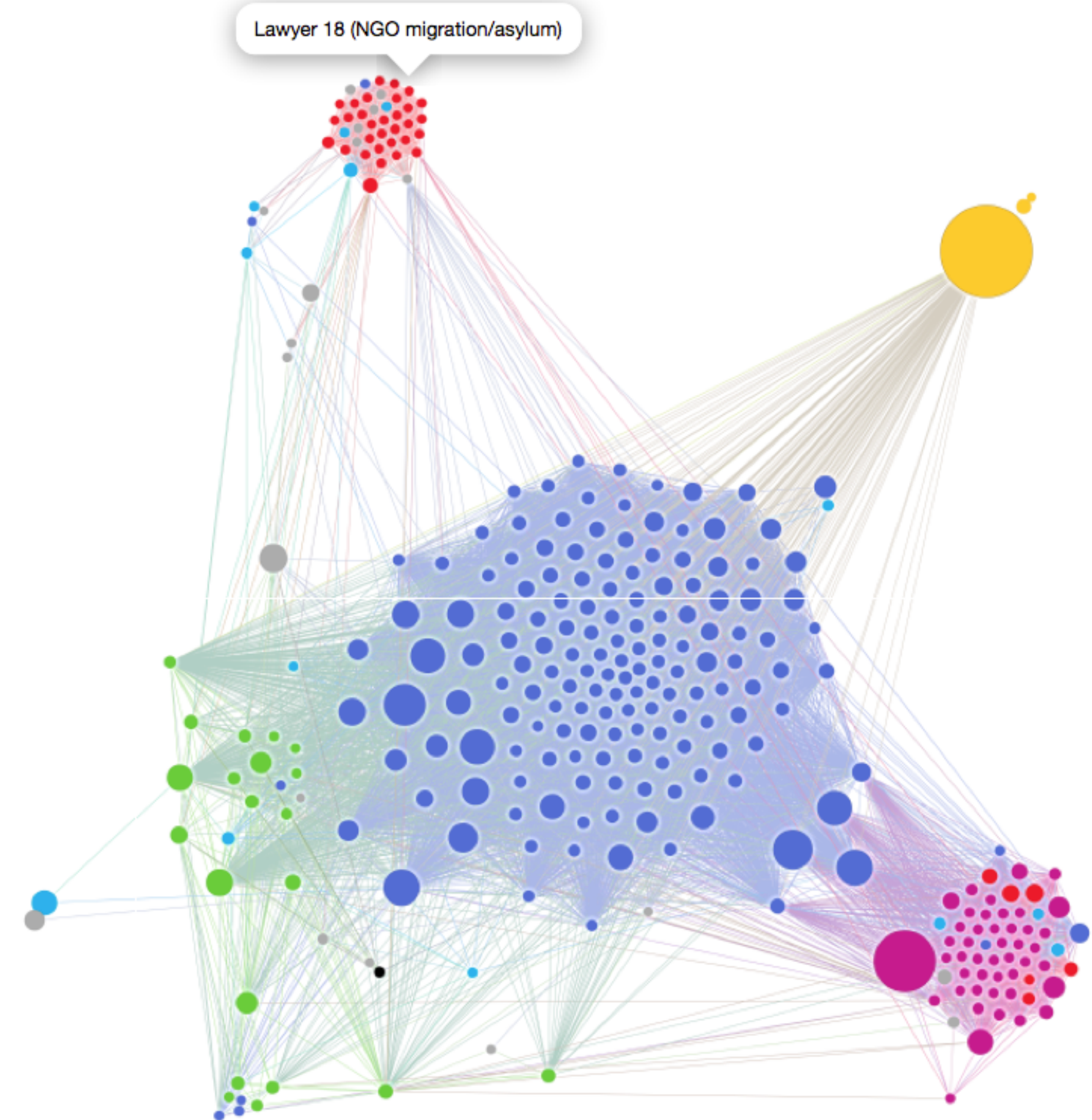
The monitored life of National Councilor Balthasar Glättli.





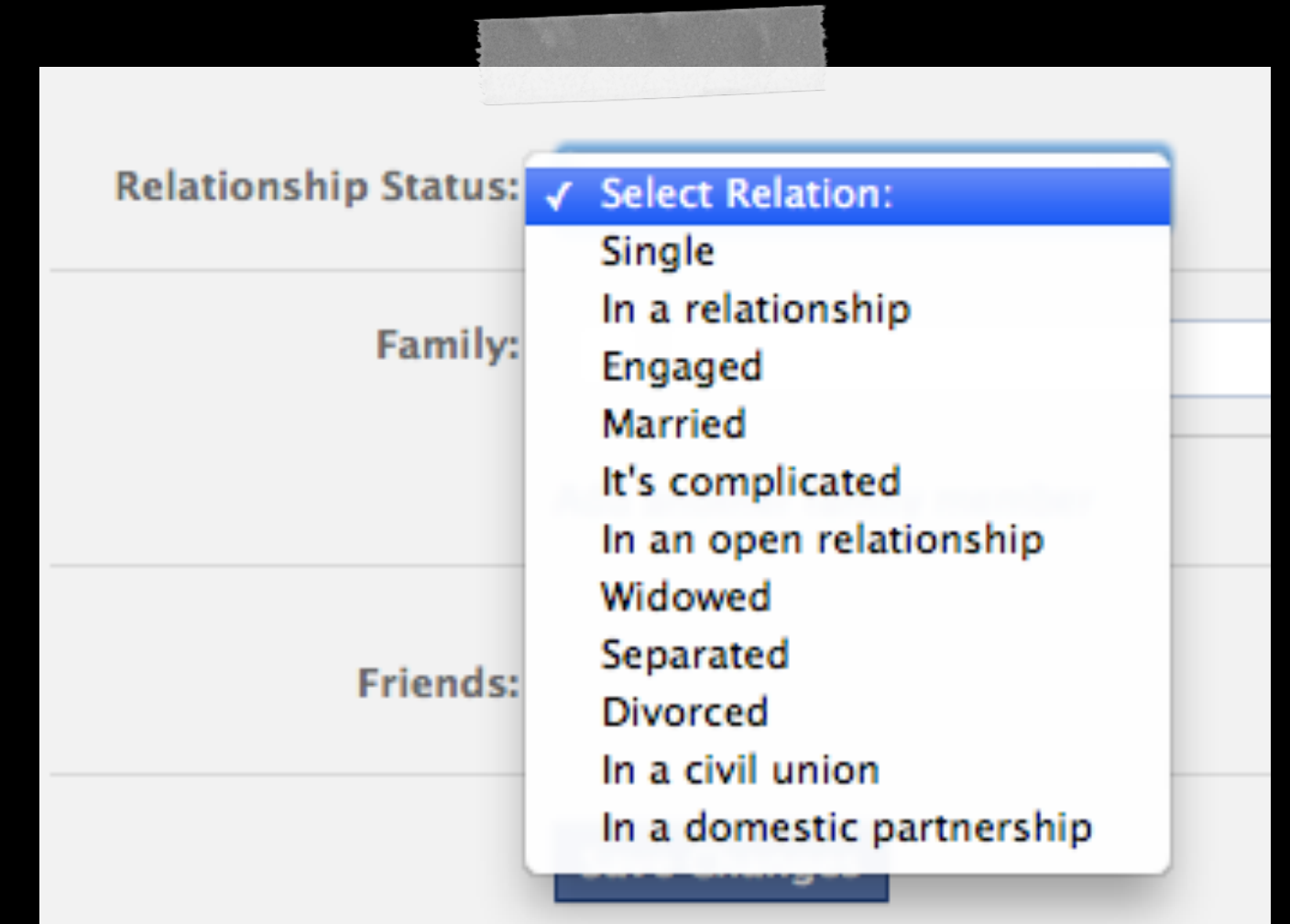
# Social Networks

- social network constructed from email metadata
  - yellow: family
  - blue: parliamentarians
  - green: party colleagues
  - red: lawyers
  - purple: NGO



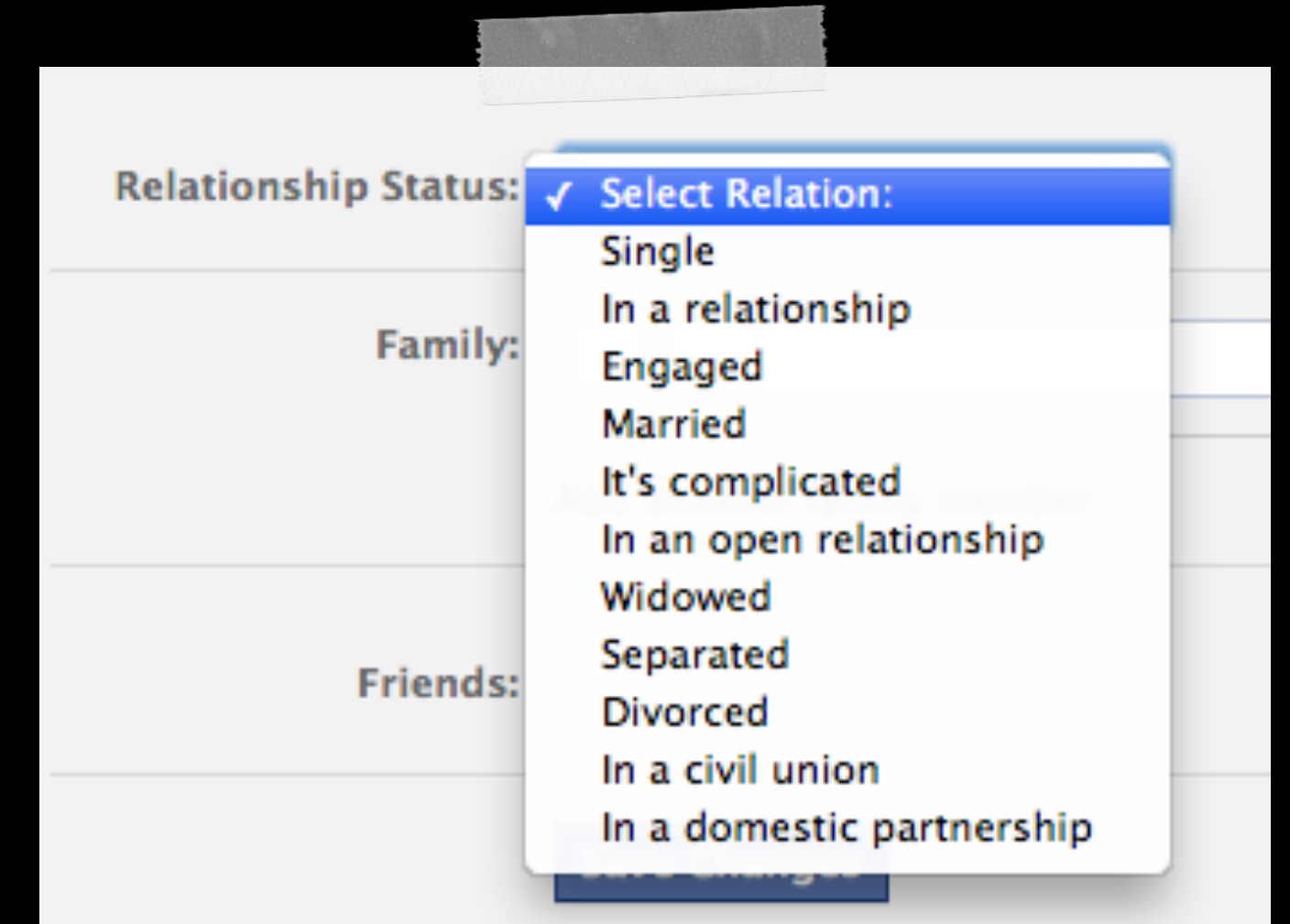
# Social Network Analysis - a Romantic Example

- researchers at Facebook asked:
  - For a given Facebook user in a relationship, can we identify his/her **romantic partner** solely from social network data?



# Social Network Analysis - a Romantic Example

- researchers at Facebook asked:
  - For a given Facebook user in a relationship, can we identify his/her **romantic partner** solely from social network data?
- yes, easily
  - out of on average >100 Facebook friends, partner is correctly pointed out in 50% of all cases





# Social Network Analysis - a Romantic Example

- researchers at Facebook asked:
  - For a given Facebook user in a relationship, can we identify his/her **romantic partner** solely from social network data?

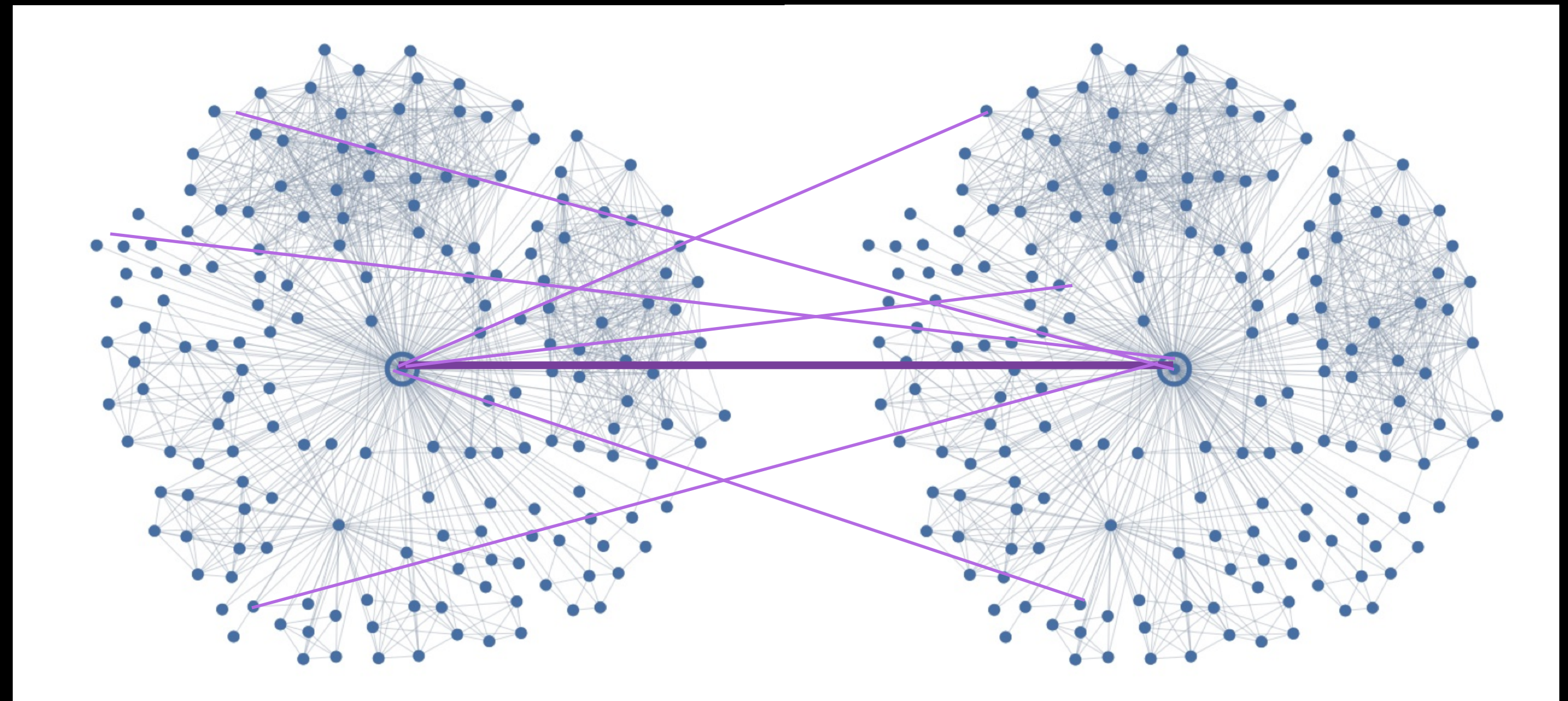
Relationship Status: ☒ Select Relation:

Family:

Friends:

- Single
- In a relationship
- Engaged
- Married
- It's complicated
- In an open relationship
- Widowed
- Separated
- Divorced
- In a civil union
- In a domestic partnership

- yes, easily
  - out of on average >100 Facebook friends, partner is correctly pointed out in 50% of all cases





Facebook?

# That's so 2012

Facebook?

# That's so 2012



**Introducing a brand new  
way to share everything.**

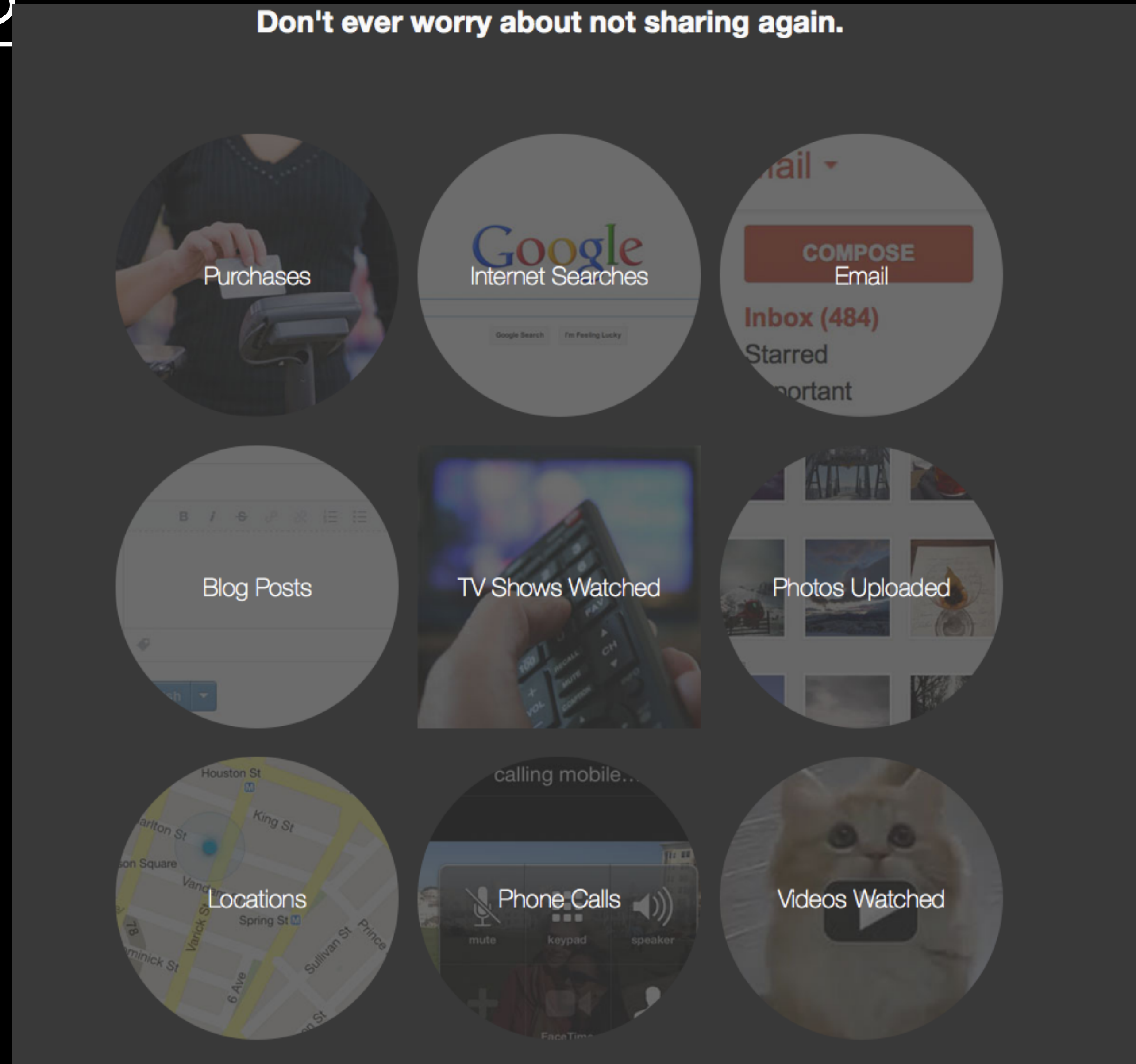
[Learn more](#)

Facebook?

# That's so 2012

Facebook?

# That's so 2012



Facebook?

# That's so 2012

Facebook?

# That's so 2012

Instantly upload trillions of megabytes of data.

## Really fast computers

Our Titan Supercomputer is capable of handling one quadrillion requests per second.



## Really big computers

Our datacenter can store up to 5 zettabytes of information.



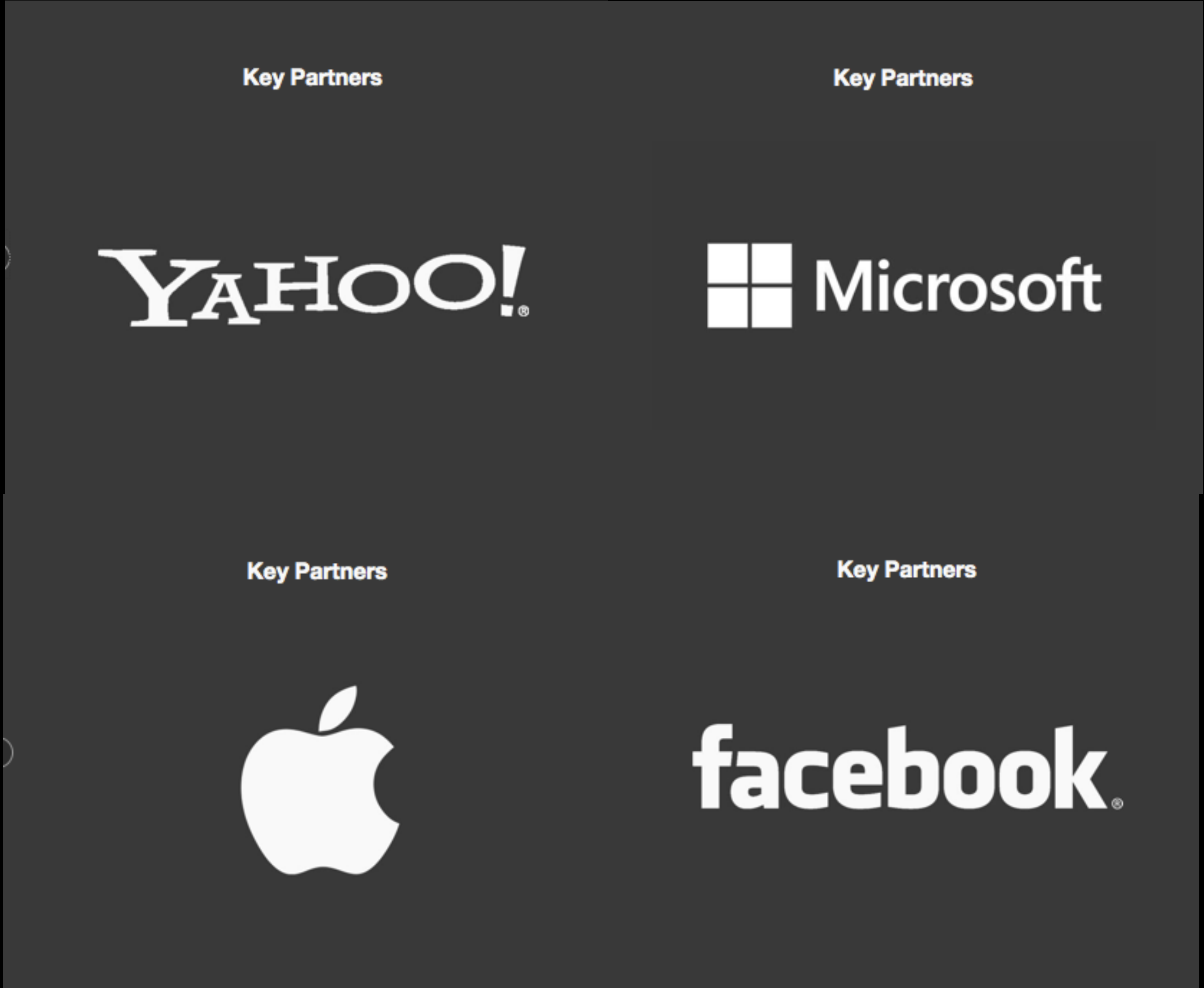
Facebook?

# That's so 2012



Facebook?

# That's so 2012



Facebook?

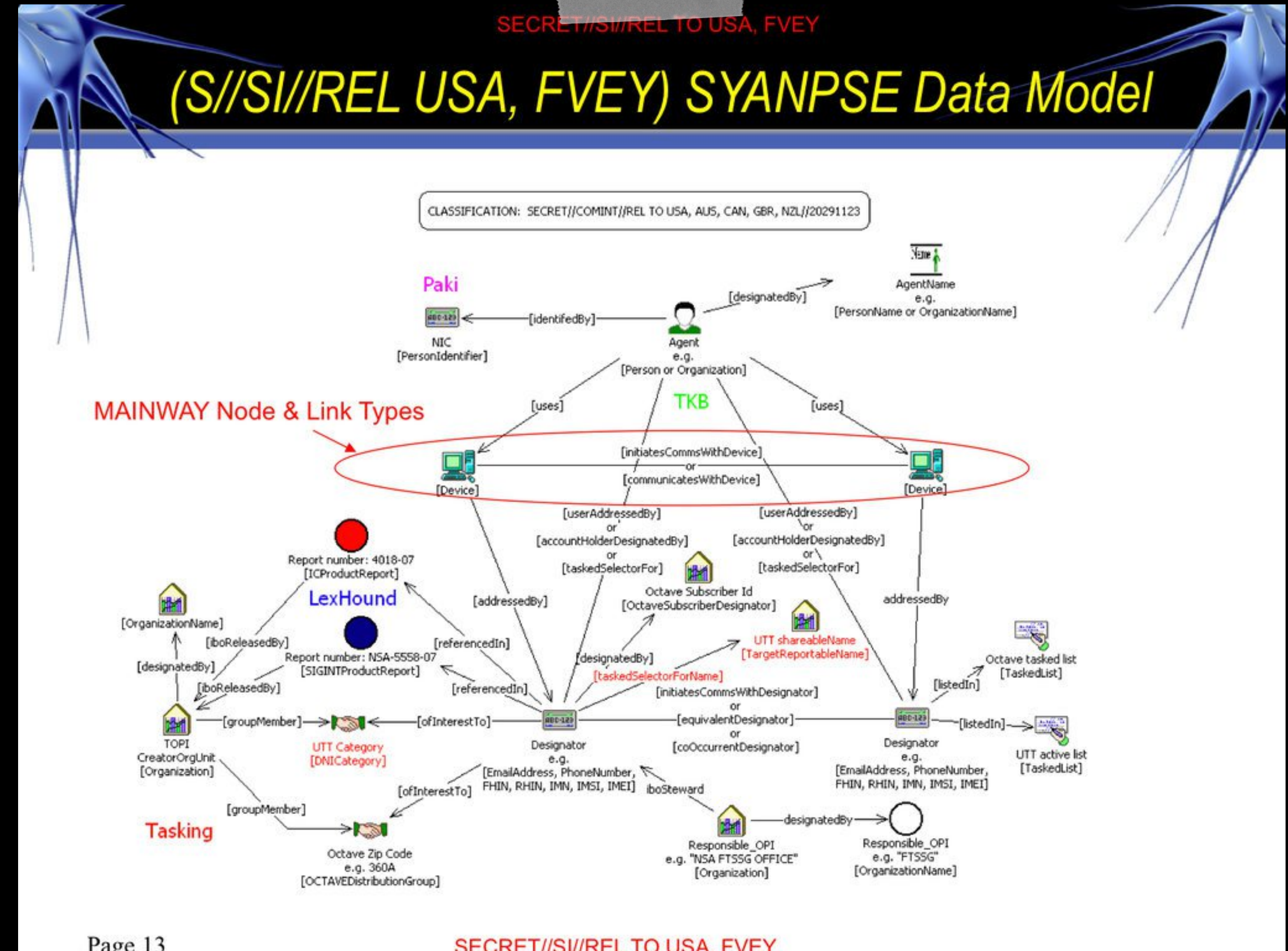
# That's so 2012

Presenting...

# The NSA Social Graph

- complex **graph data structure** connecting persons, organizations, identifiers, devices...
- under construction since 2010
- NSA conducts
  - “large-scale graph analysis on very large sets of communications metadata”
- data sources
  - phone
  - email
  - Facebook
  - bank data
  - tax and population registers
  - passenger lists
  - ...

➡ NYT: [nyti.ms/TdQJVM](https://nyti.ms/TdQJVM)



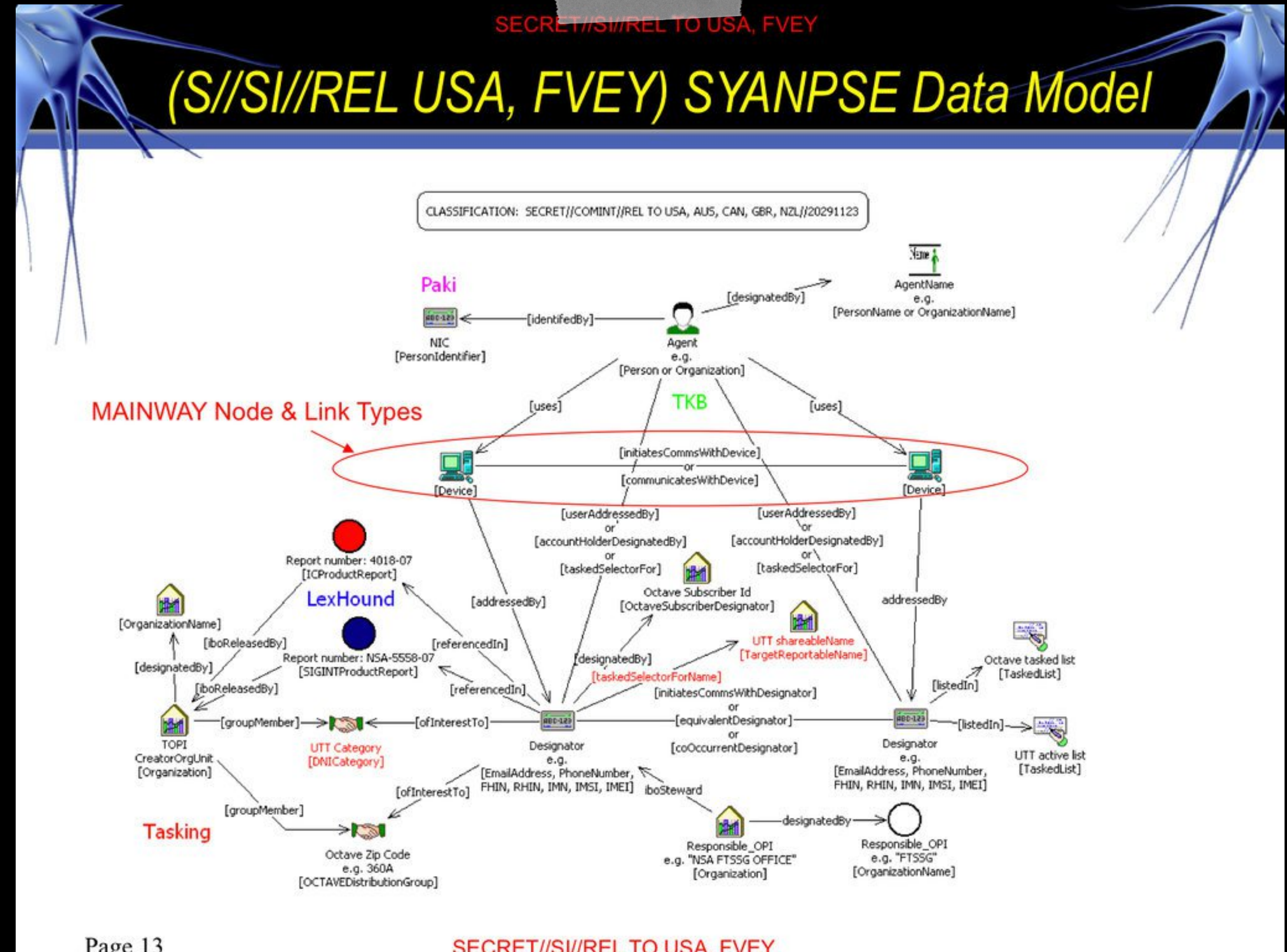


Presenting...

# The NSA Social Graph

- complex **graph data structure** connecting persons, organizations, identifiers, devices...
- under construction since 2010
- NSA conducts
  - “large-scale graph analysis on very large sets of communications metadata”
- data sources
  - phone
  - email
  - Facebook
  - bank data
  - tax and population registers
  - passenger lists
  - ...

➔ NYT: [nyti.ms/TdQJVM](https://nyti.ms/TdQJVM)



## Featured Career Opportunities

Computer Scientists

Computer Engineers

Electrical Engineers

Cyber Professionals

Where Intelligence Goes to Work







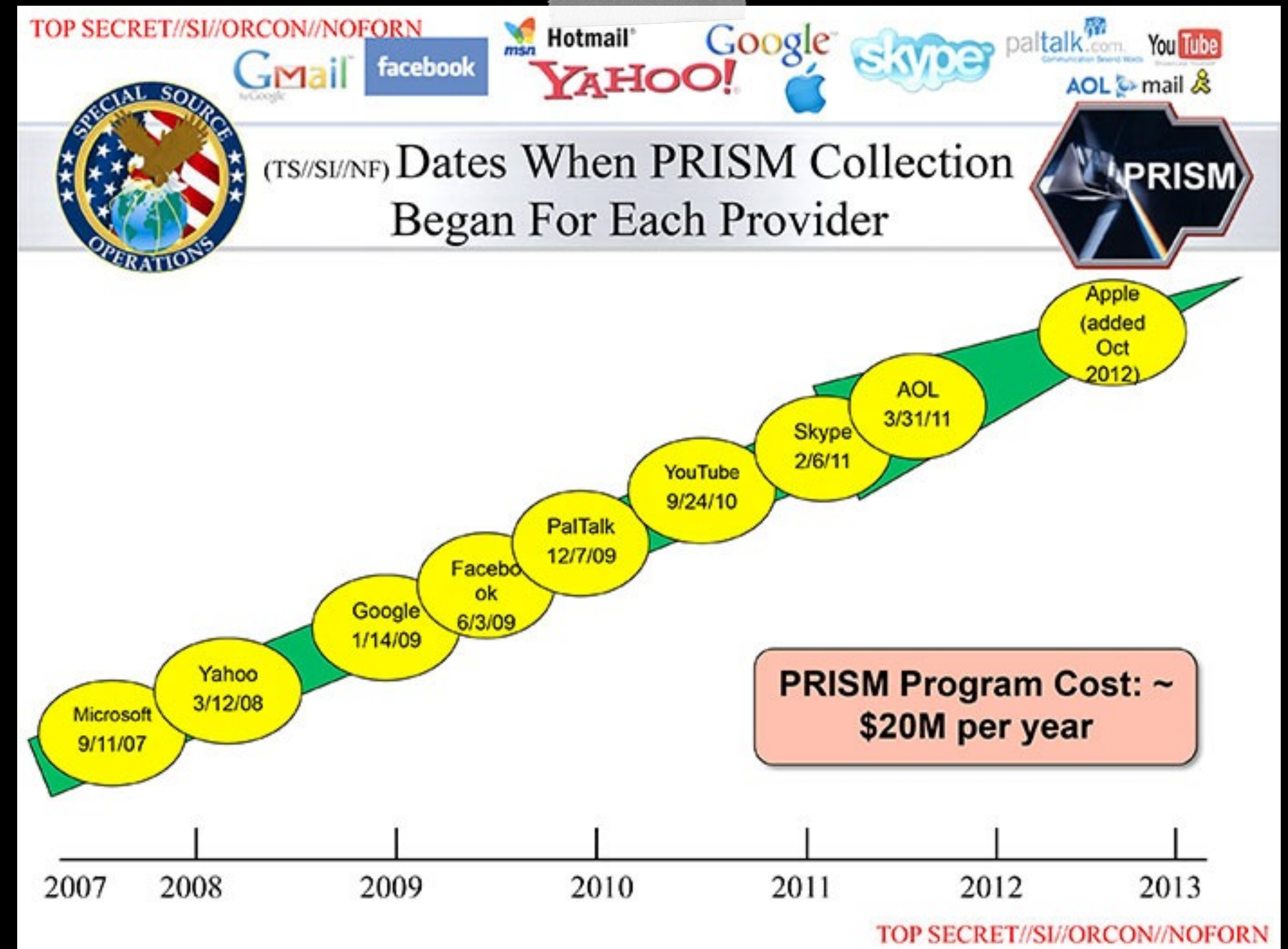
# Collection PRISM



- collects **communications stored by internet companies**
  - email
  - video/voice chat
  - videos
  - photos
  - social network data
  - browsing history
- operated by NSA with GCHQ participation
- analysts can query for search terms in content and target persons
  - and look **2-3 hops** far

„PRISM **is** about content“  
— Edward Snowden

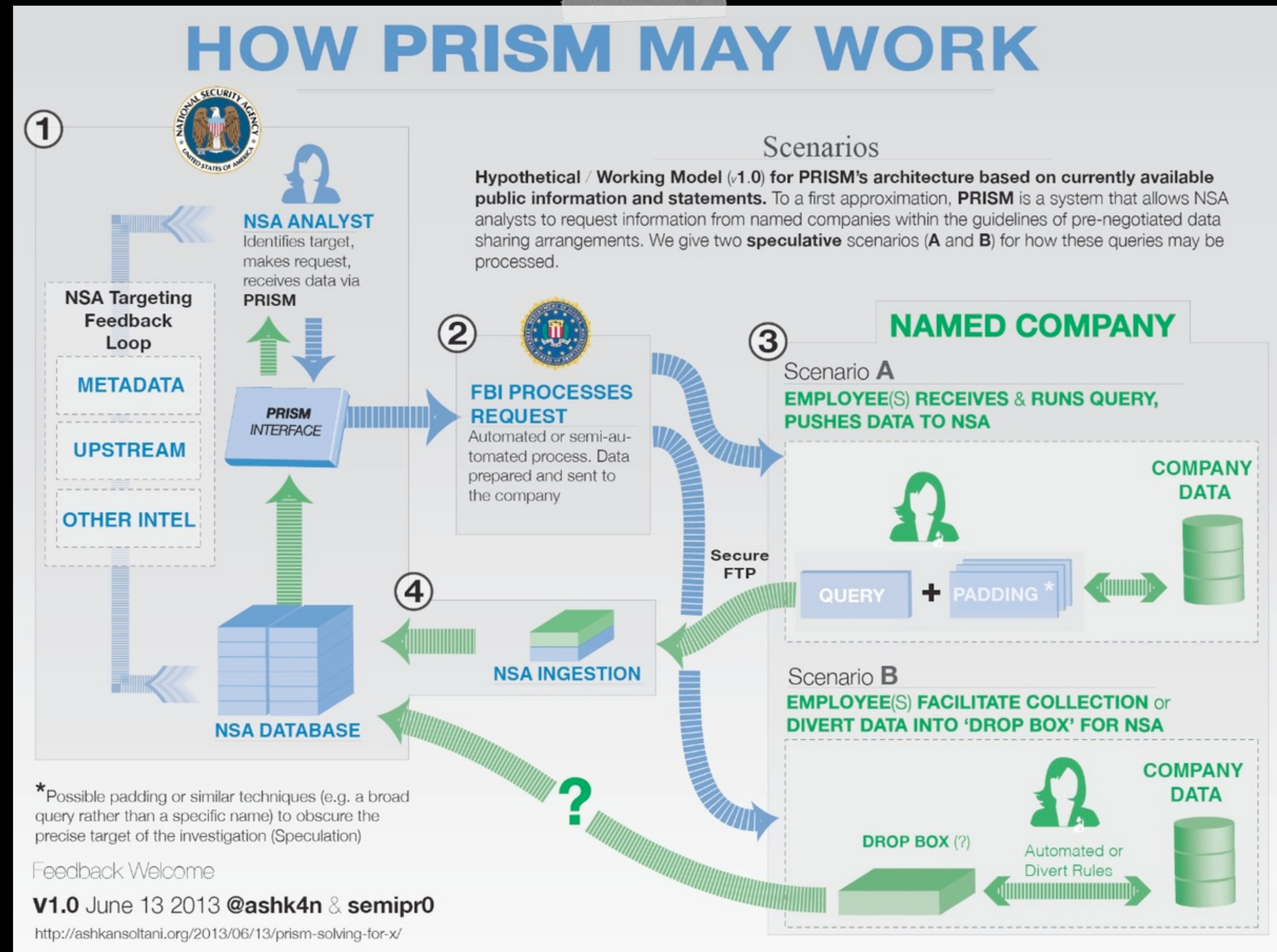
➡ Snowden @ TED: [bit.ly/1nCdWhS](http://bit.ly/1nCdWhS)





# PRISM: Implementation

- probably not...
  - direct access to company servers
  - complete ingestion of all data flowing to/from the provider
- more likely...
  - an „API that allows for streamlined processing of **FISA requests**“
  - a system for requesting, transferring, and ingesting of responses from service providers
- **FISA**
  - legislature that governs PRISM
  - **foreign** intelligence, but
    - surveillance legitimate if 51% chance that target is non-US citizen
  - assumed foreign unless proven otherwise



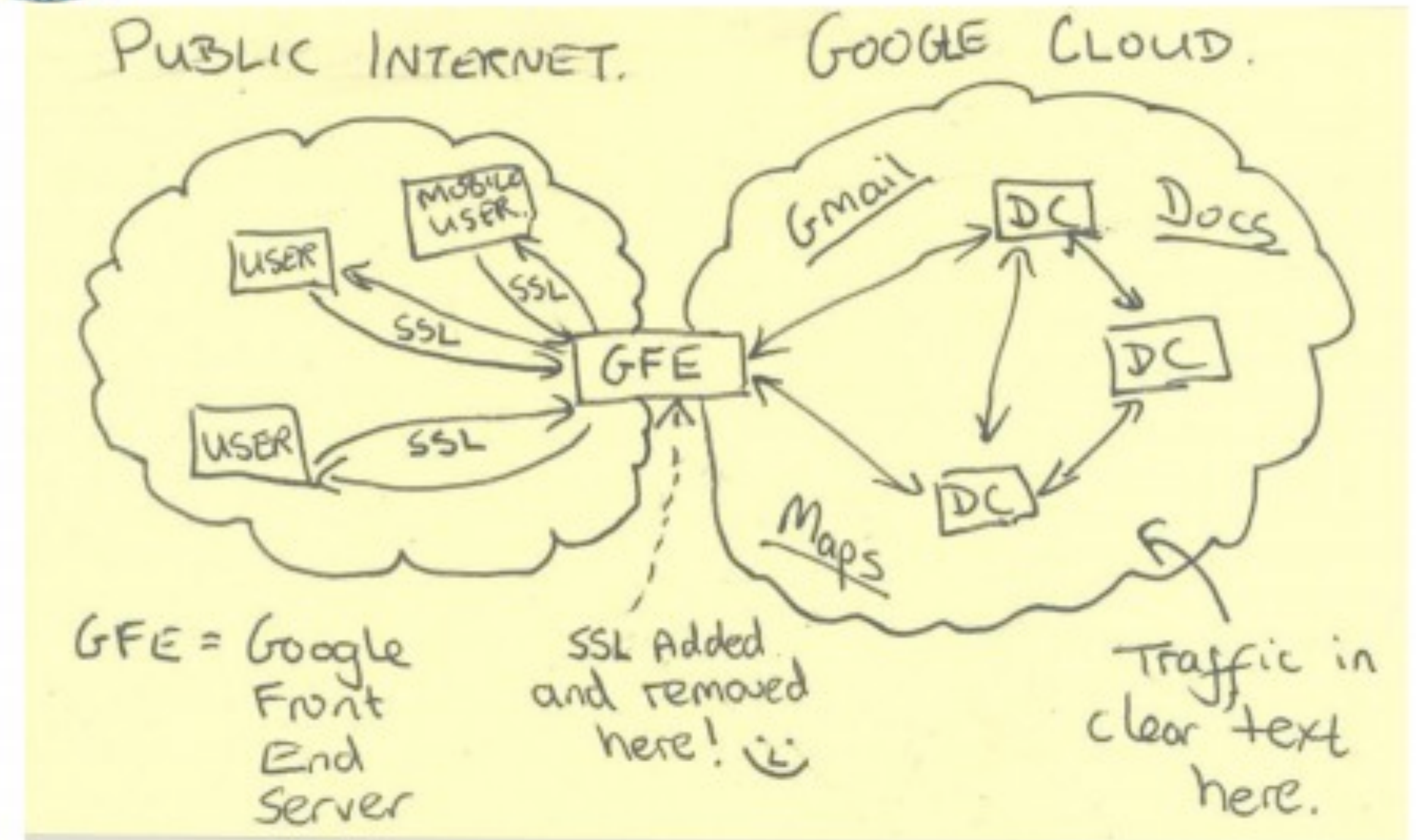


# MUSCULAR

- taps the internal networks of Google and Yahoo!
- to obtain **communication content and metadata** (e.g. email)
- joint operation of GCHQ & NSA
- requires no warrants at all
- collects 2x as many data points as PRISM



## Current Efforts - Google



TOP SECRET//SI//NOFORN

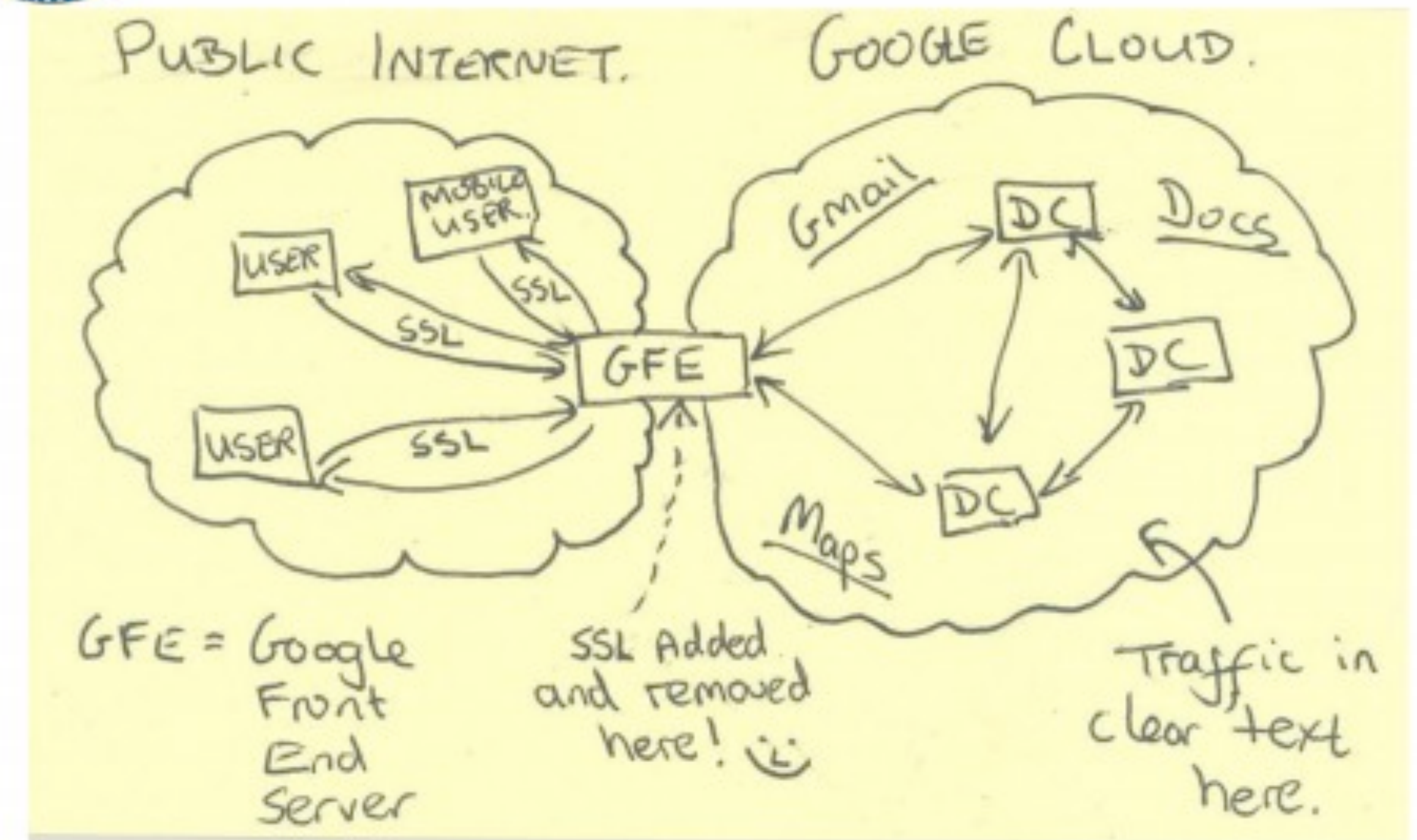


# MUSCULAR

- taps the internal networks of Google and Yahoo!
- to obtain **communication content and metadata** (e.g. email)
- joint operation of GCHQ & NSA
- requires no warrants at all
- collects 2x as many data points as PRISM



## Current Efforts - Google



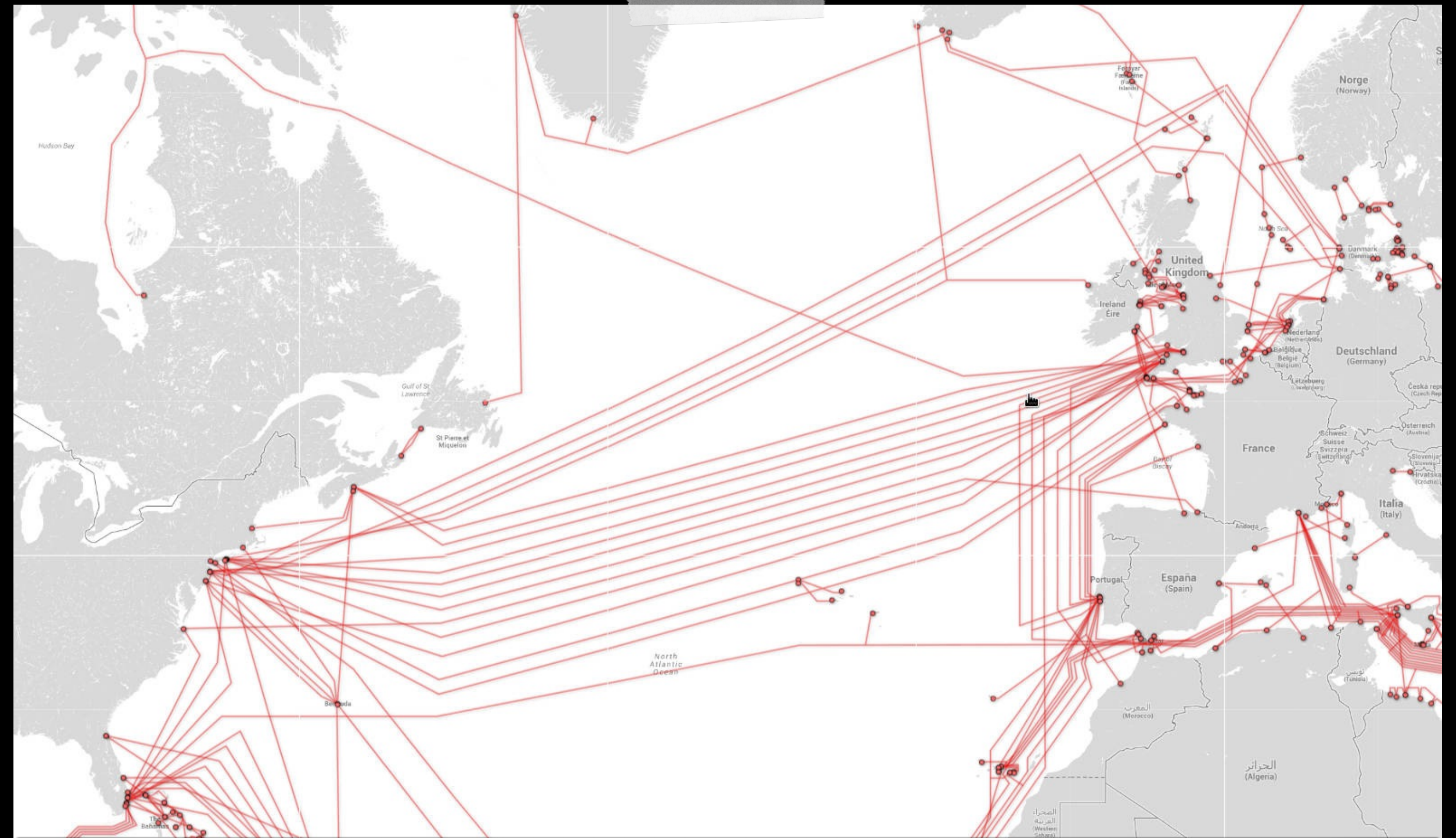
TOP SECRET//SI//NOFORN

„Two engineers with close ties to Google exploded in profanity when they saw the drawing.“  
- Washington Post



# Tempora

- **taps undersea fibre-optic cables** between UK and US
- **full take of internet backbone traffic**
  - > 21 petabytes / day
  - content kept for 3 days, metadata for 30 days
- operated by GCHQ, shared among the „Five Eyes“ (UK, US, Canada, Australia, New Zealand)



# NSA Databases

## MAINWAY



## NUCLEON



## MARINA



## PINWALE



- **telephone metadata**
  - > 2 **trillion** call-detail records
  - receives > 2 billion phone records daily
- used for
  - traffic analysis
  - social network analysis

- **telephone content**
  - recorded calls
- covering ca. 1 million target persons

- **internet metadata**
  - browser settings & history, account logins etc.
- used to create profiles of internet users
- stored for 1 year

- **email archive**
  - and other internet usage data



Storage

# NSA Data Center

- brand new data center in Bluffdale, Utah
  - 10 000 m<sup>2</sup> server space
  - 12 exabytes storage capacity
  - 70 MW energy consumption





# Big Data...

## Stasi versus NSA

How much space would the filing cabinets of the Stasi and the NSA use up, if the NSA would print out their 5 Zettabytes?

[Embed](#)

# Big Data...

## Stasi versus NSA

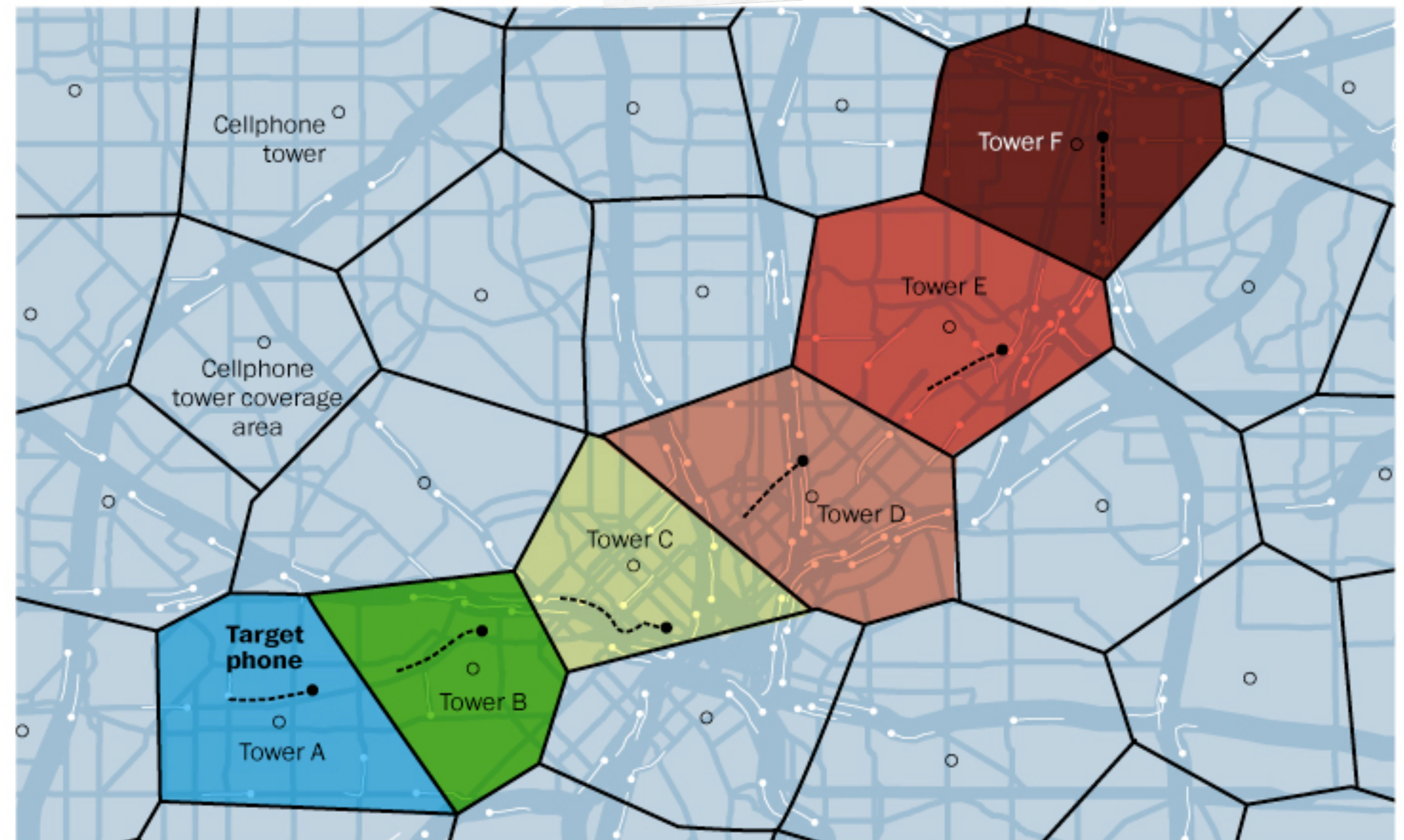
How much space would the filing cabinets of the Stasi and the NSA use up, if the NSA would print out their 5 Zettabytes?

[Embed](#)

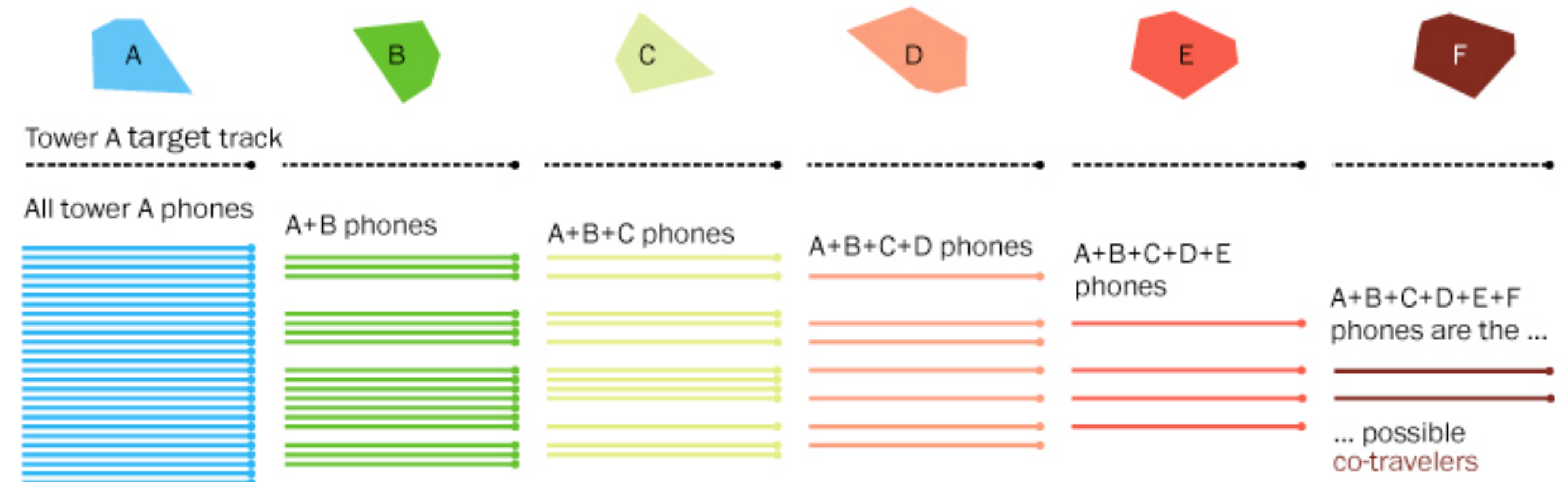


# CO-TRAVELLER

- tracks location of cell phones
- identifies groups that travel together
- NSA operation
- billions of records / day



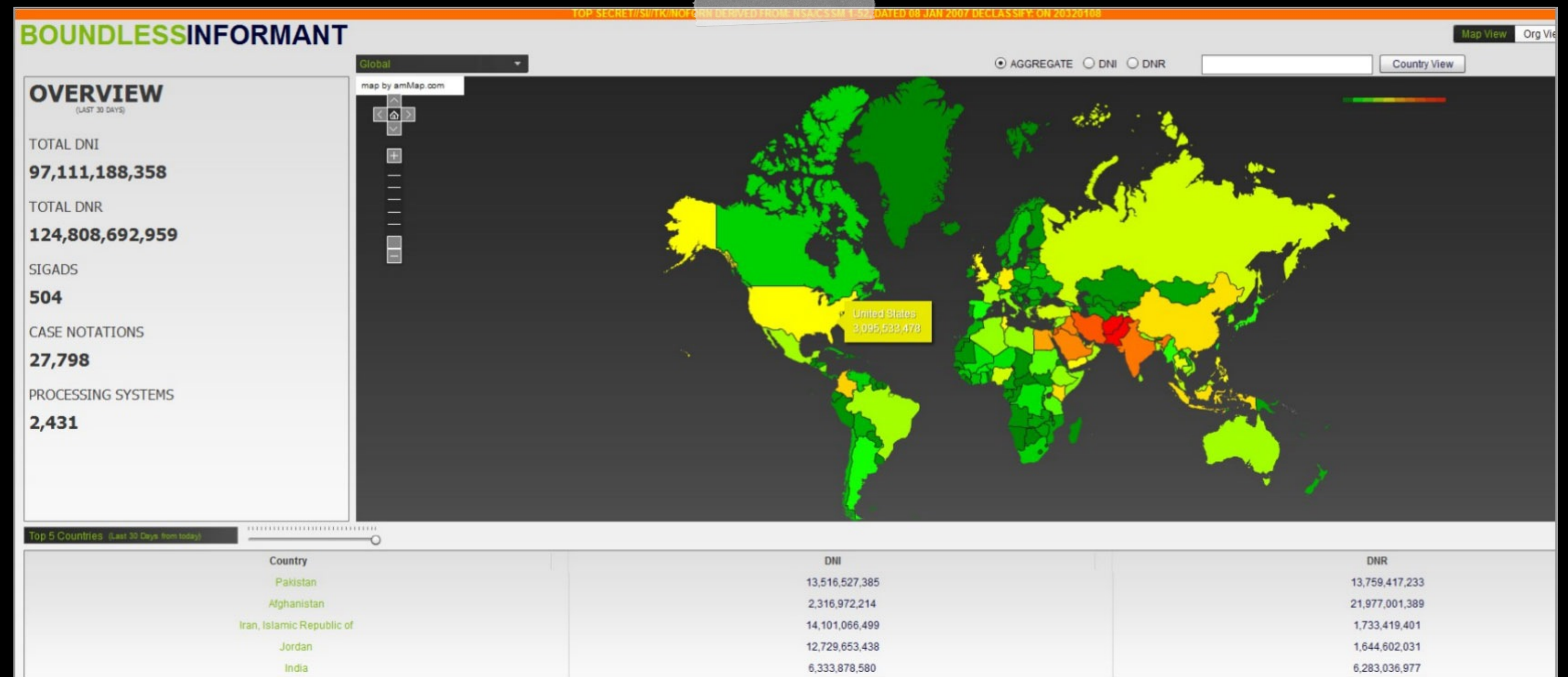
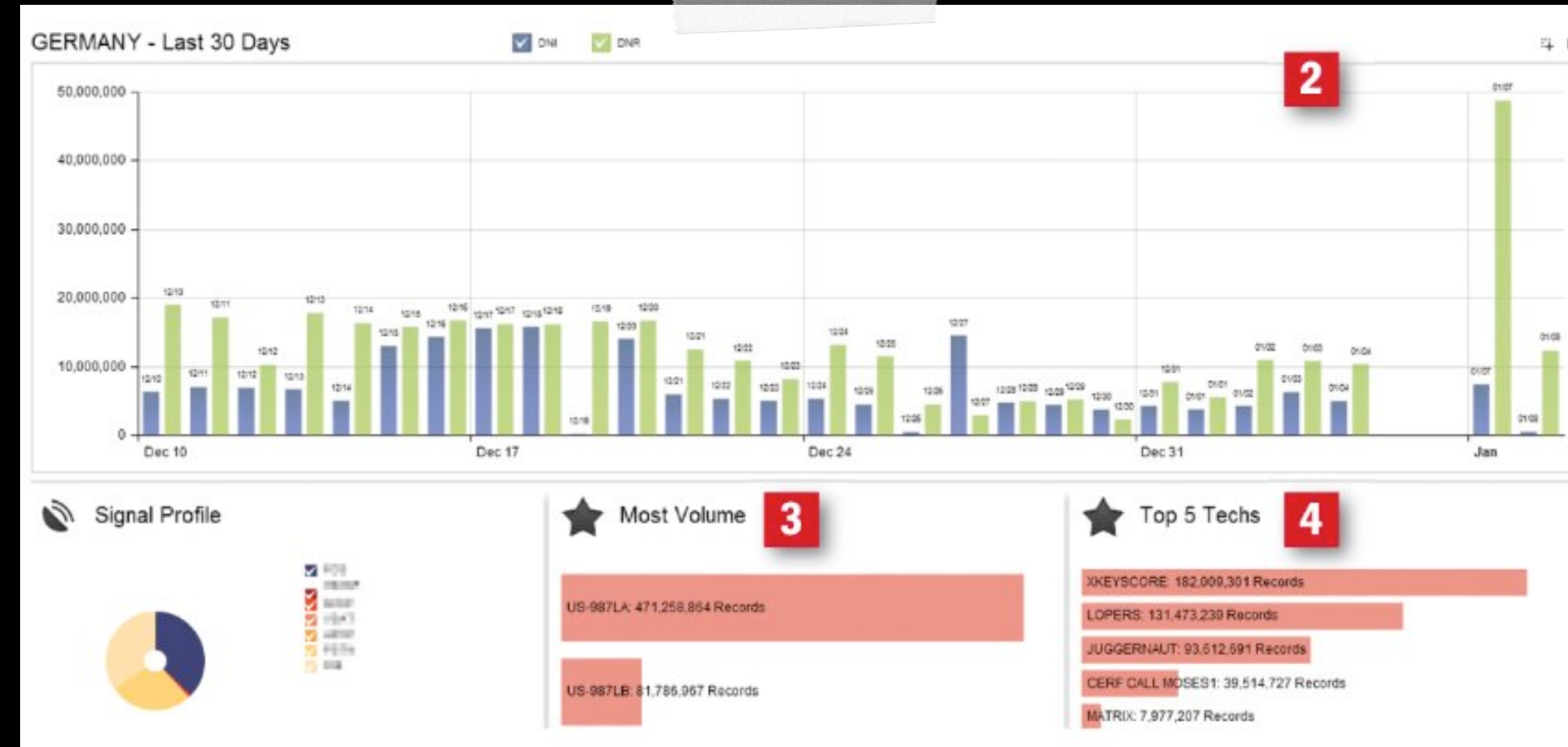
As the target phone moves from tower to tower, fewer and fewer potential co-travelers remain.





# Boundless Informant

- big data analysis and visualization tool used by the NSA
- **a surveillance Google Analytics**
- 97 billion internet data records (DNI) and 124 billion telephony data records (DNR) collected per month
- various data sources, including PRISM



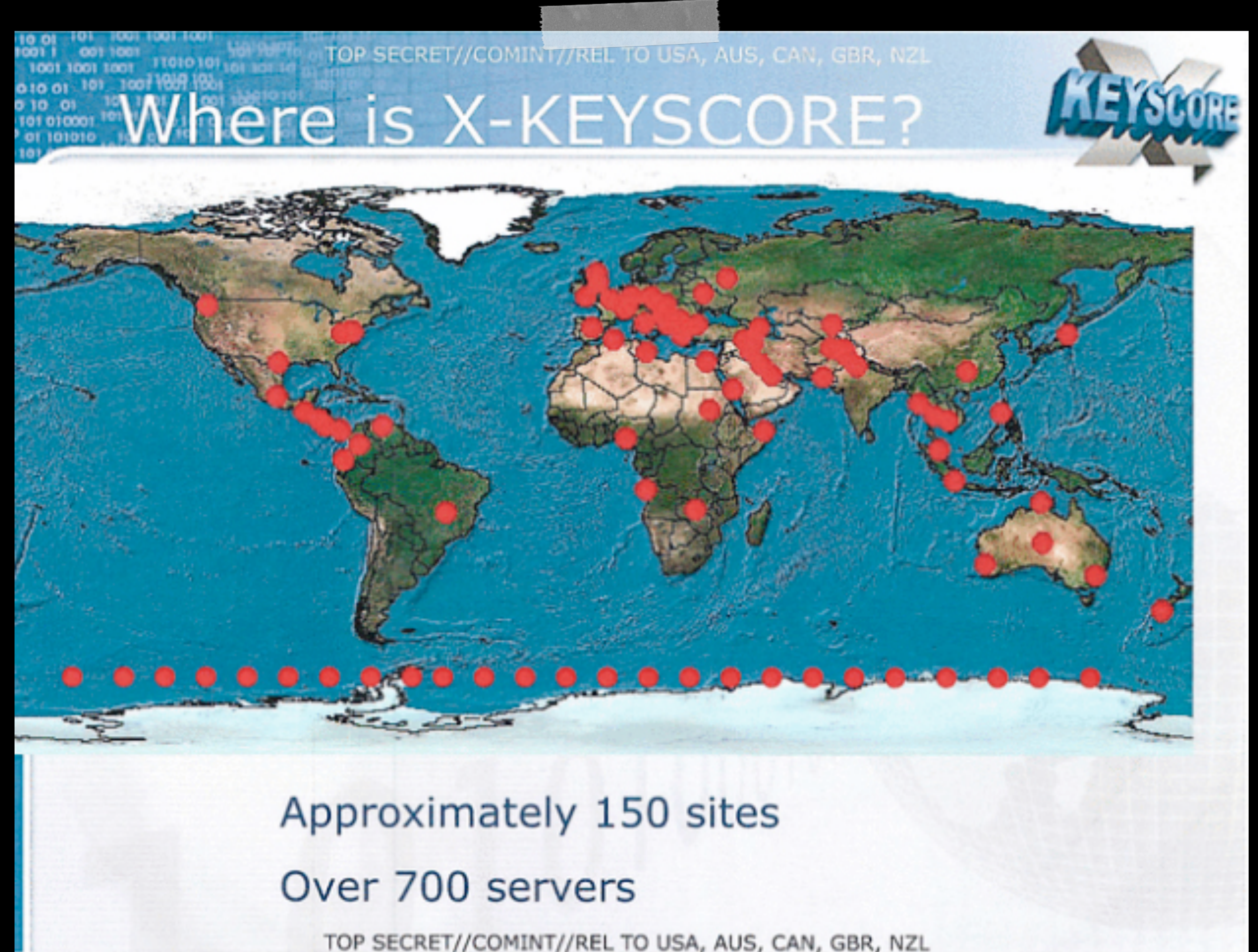


# XKeyscore

- **a surveillance search engine**
- operated by NSA
  - shared with other IAs, including **German BND**
- cluster of many servers distributed around the world

"You could read anyone's email in the world [...] Any website: You can watch traffic to and from it. Any computer that an individual sits at: You can watch it. Any laptop that you're tracking: you can follow it [...]. It's a one-stop-shop for access to the NSA's information."

— Edward Snowden

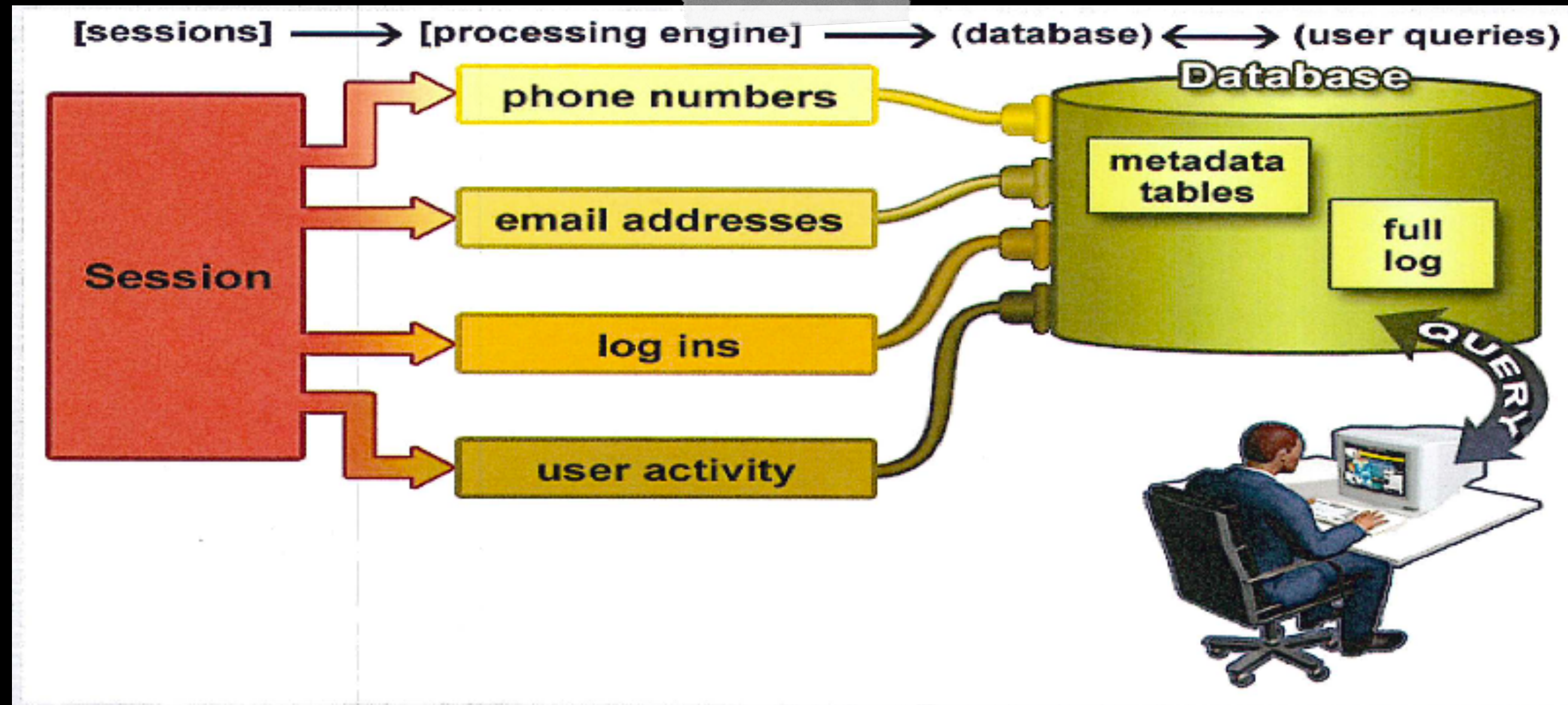




# XKeyscore

## store

- data sources
  - telco providers
  - CIA espionage
  - satellite communication interception
  - spy planes, drones, satellites
  - "Tailored Access" = hacking
  - all surveillance rubber-stamped by the FISA court (including **PRISM**)
  - intelligence agencies of Belgium, Denmark, France, Germany, Italy, Japan, the Netherlands, Norway, Sweden....
  - ....



## process

- plugins
  - index
  - extract metadata
  - enable selectors

## query

- strong selectors = identifiers
  - email adress
  - phone number
  - login
  - ...
- weak selectors, e.g.
  - keywords
  - „language anomalies“
  - usage of encryption



... and finally

# Metadata-guided Missiles

- program **Geo Cell**: geolocation of cell phones to guide drone missile strikes
  - program **Gilgamesh**
    - drone poses as cell phone tower
  - program **Shenanigans**
    - receiver on drone sniffs any wireless network in range



... and finally

# Metadata-guided Missiles

- program **Geo Cell**: geolocation of cell phones to guide drone missile strikes
  - program **Gilgamesh**
    - drone poses as cell phone tower
  - program **Shenanigans**
    - receiver on drone sniffs any wireless network in range

„We kill people based on metadata“  
— Michael Hayden, former NSA/CIA director





... and finally

# Metadata-guided Missiles

- program **Geo Cell**: geolocation of cell phones to guide drone missile strikes
  - program **Gilgamesh**
    - drone poses as cell phone tower
  - program **Shenanigans**
    - receiver on drone sniffs any wireless network in range



„We kill people based on metadata“  
— Michael Hayden, former NSA/CIA director

„Absolutely“  
— anonymous former drone operator, asked  
whether innocents have been killed

... and finally

# Metadata-guided Missiles

- program **Geo Cell**: geolocation of cell phones to guide drone missile strikes
  - program **Gilgamesh**
    - drone poses as cell phone tower
  - program **Shenanigans**
    - receiver on drone sniffs any wireless network in range



„We kill people based on metadata“  
— Michael Hayden, former NSA/CIA director

„Absolutely“  
— anonymous former drone operator, asked  
whether innocents have been killed

„We Track 'Em, You Whack 'Em.“  
— unofficial *Geo Cell* motto



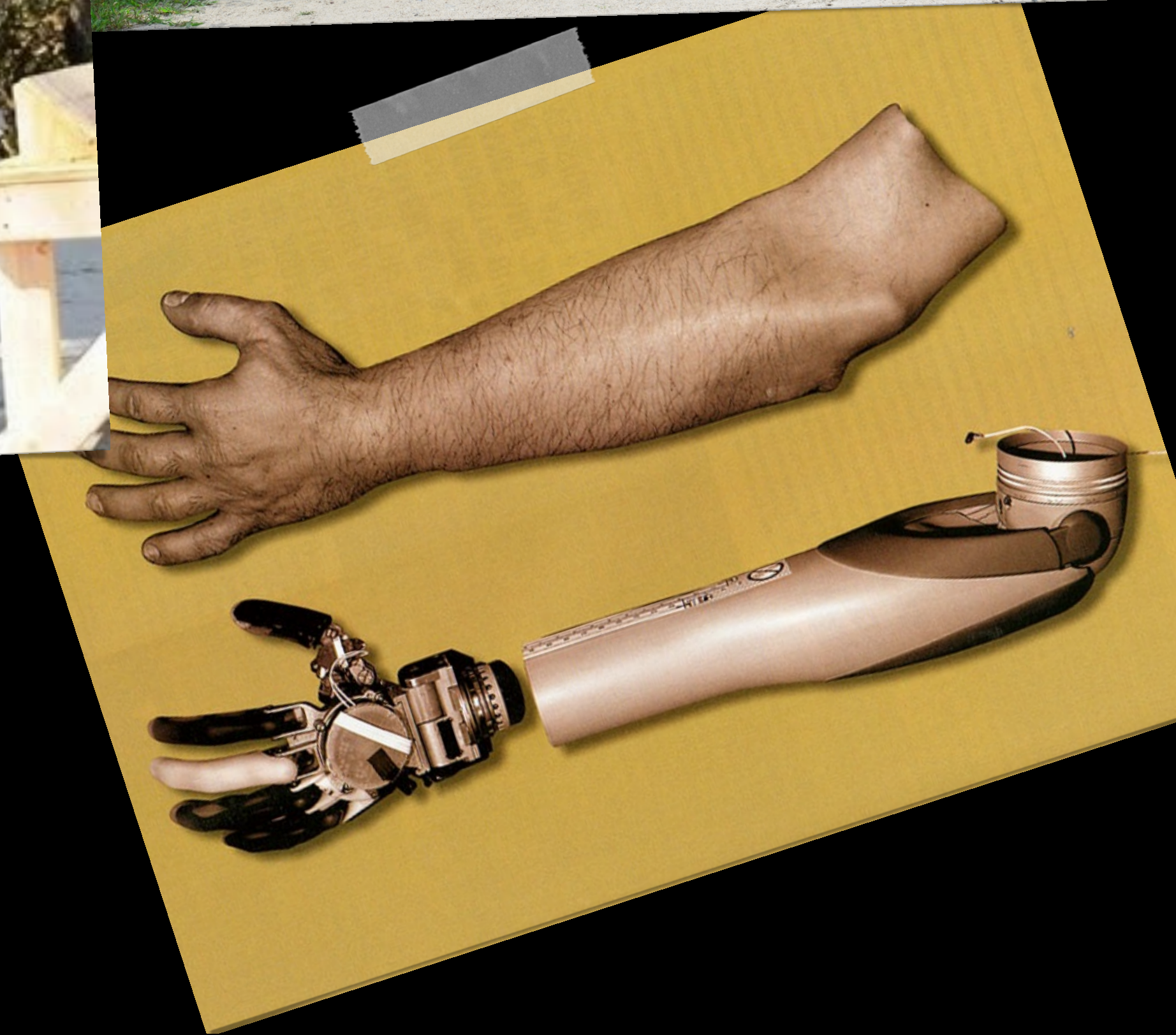
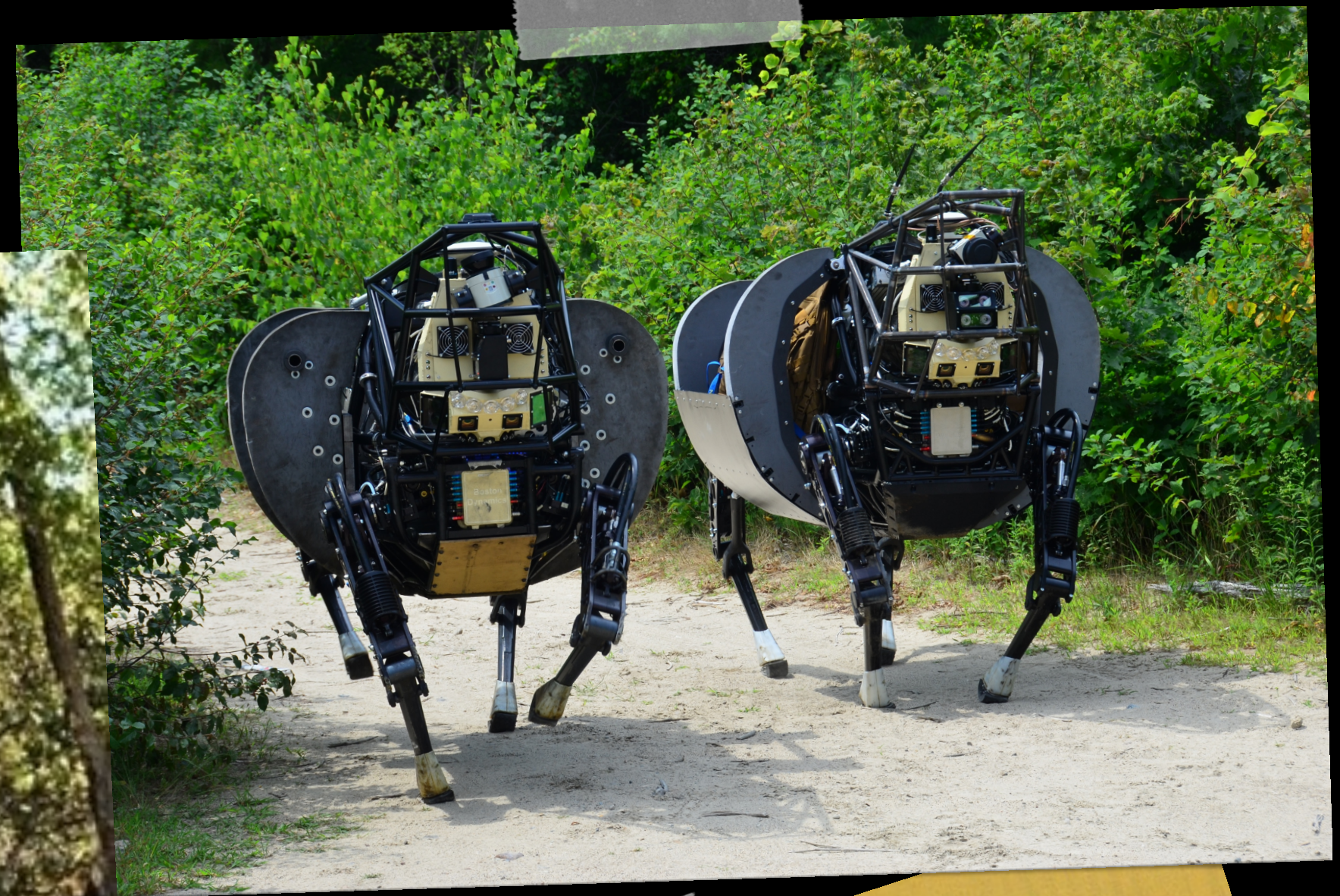
Surveillance Capabilities

# What's next?



After DARPA...

... the agency that brought you...





... comes

# IARPA (Intelligence Advanced Research Projects Activity)

"IARPA's mission [is] to invest in high-risk/high-payoff research programs that have the potential to provide the United States with an overwhelming intelligence advantage over future adversaries."

— FBI National Press Release, 2009



- research towards **predictive surveillance**
  - a „data eye in the sky“

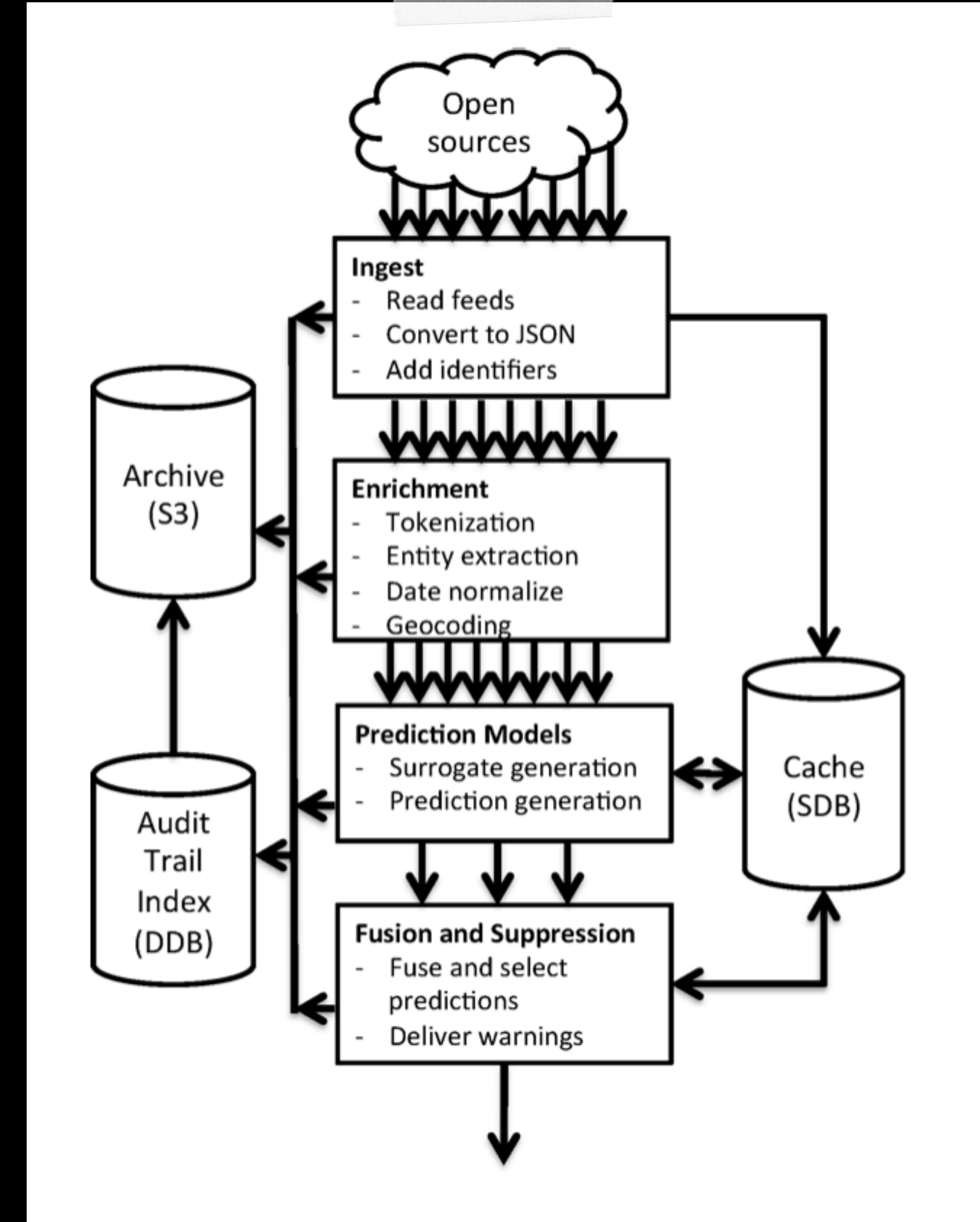
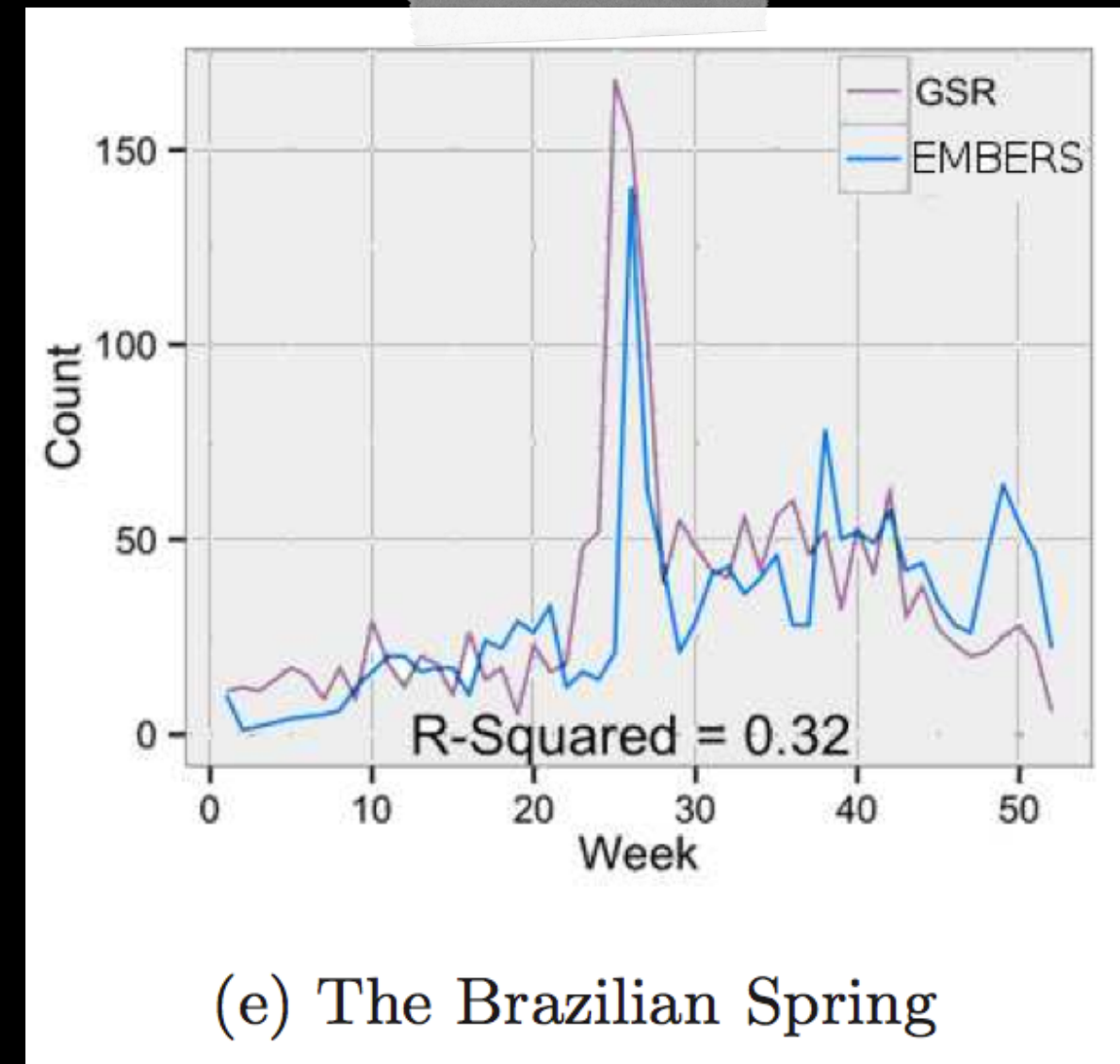
„The most optimistic researchers believe that these storehouses of ‚big data‘ will for the first time reveal **sociological laws of human behavior** — enabling them to **predict political crises, revolutions and other forms of social and economic instability**, just as physicists and chemists can predict natural phenomena.“

— New York Times

# EMBERS

- predicts „**civil unrest**“
  - strike, protest, riot, ...
- ingests **open** data sources
  - Twitter, RSS news/blog feeds, financial news, ...
- raises **alerts**
  - when, where, who, why
- technical specs
  - runs in the AWS cloud
  - 15 GB daily ingest
  - archives 3 billion messages (12 TB)
  - outputs 40 warnings / day

„Since the Arab Spring revolution began, and especially after Egypt's upheaval, many analysts have pondered: Could we have anticipated these events?“  
– EMBERS research paper





and more

# IARPA Projects

- ... develop a **search engine** that scans **video** for specific things or people
- ... incorporate social and cultural background into **natural language interpretation**
- ... geolocate photos by **recognizing places in images**
- ... research **quantum computing**
- ... detect „technical emergence“, i.e. **predict the appearance of new technologies**
- ... construct a **silently flying drone**
- ... develop machine-learning algorithms that **mimic the human brain**
- ... investigate the „potential role of **virtual reality platforms** in cyber terrorism“

Towards the Machine-Readable Society

# Some Lessons Learned...



# Some Lessons Learned...

## Technical

- **metadata** is the key to mass surveillance
  - but they are also reading your email and listening to your telephone calls
- a **full take** is going on
- the „**data eye in the sky**“ is the limit

# Some Lessons Learned...

## Technical

- **metadata** is the key to mass surveillance
  - but they are also reading your email and listening to your telephone calls
- a **full take** is going on
- the „**data eye in the sky**“ is the limit

## Political

- this is **not an „NSA scandal“**
  - it's evidence of a systematic effort to make society machine-readable, transparent, and more controllable
  - the target is the general population
- the legal framework restricting „foreign intelligence“ is completely broken
  - and it's not about „America spying on Germany“
- the system will not be reformed
  - political „solutions“ that do not include defunding and dismantling IAs are ultra-conservative



# Some Lessons Learned...

## Technical

- **metadata** is the key to mass surveillance
  - but they are also reading your email and listening to your telephone calls
- a **full take** is going on
- the „**data eye in the sky**“ is the limit

- How to raise the **cost** of mass surveillance?
  - decentralization of services
  - widespread adoption of encryption
  - ...

## Political

- this is **not an „NSA scandal“**
  - it's evidence of a systematic effort to make society machine-readable, transparent, and more controllable
  - the target is the general population
- the legal framework restricting „foreign intelligence“ is completely broken
  - and it's not about „America spying on Germany“
- the system will not be reformed
  - political „solutions“ that do not include defunding and dismantling IAs are ultra-conservative

# Some Lessons Learned...

## Technical

- **metadata** is the key to mass surveillance
  - but they are also reading your email and listening to your telephone calls
- a **full take** is going on
- the „**data eye in the sky**“ is the limit

- How to raise the **cost** of mass surveillance?
  - decentralization of services
  - widespread adoption of encryption
  - ...

## Political

- this is **not an „NSA scandal“**
  - it's evidence of a systematic effort to make society machine-readable, transparent, and more controllable
  - the target is the general population
- the legal framework restricting „foreign intelligence“ is completely broken
  - and it's not about „America spying on Germany“
- the system will not be reformed
  - political „solutions“ that do not include defunding and dismantling IAs are ultra-conservative
- How to make sure that **we** don't have to pay it?





Thank you for watching