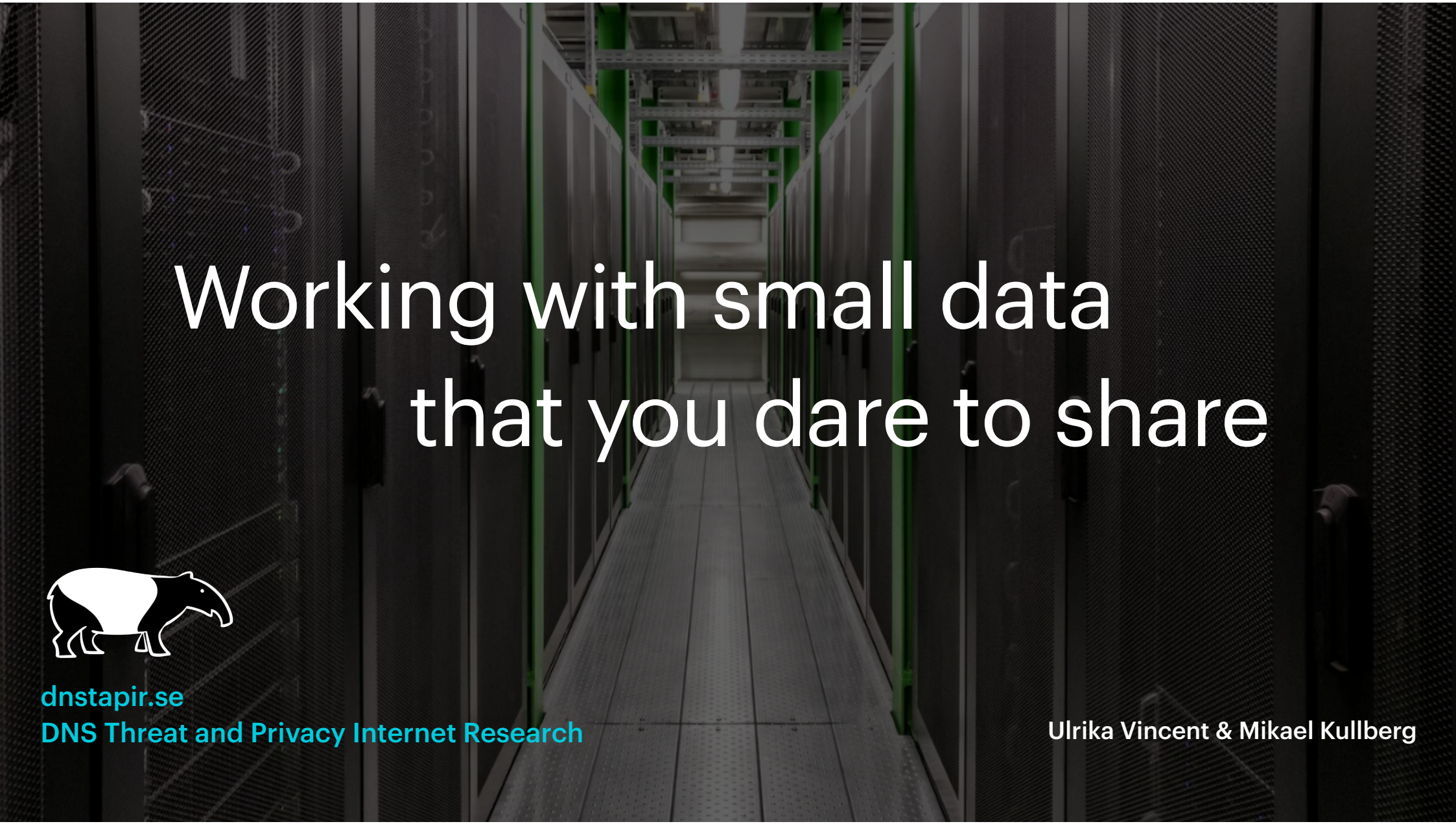


Working with small data that you dare to share

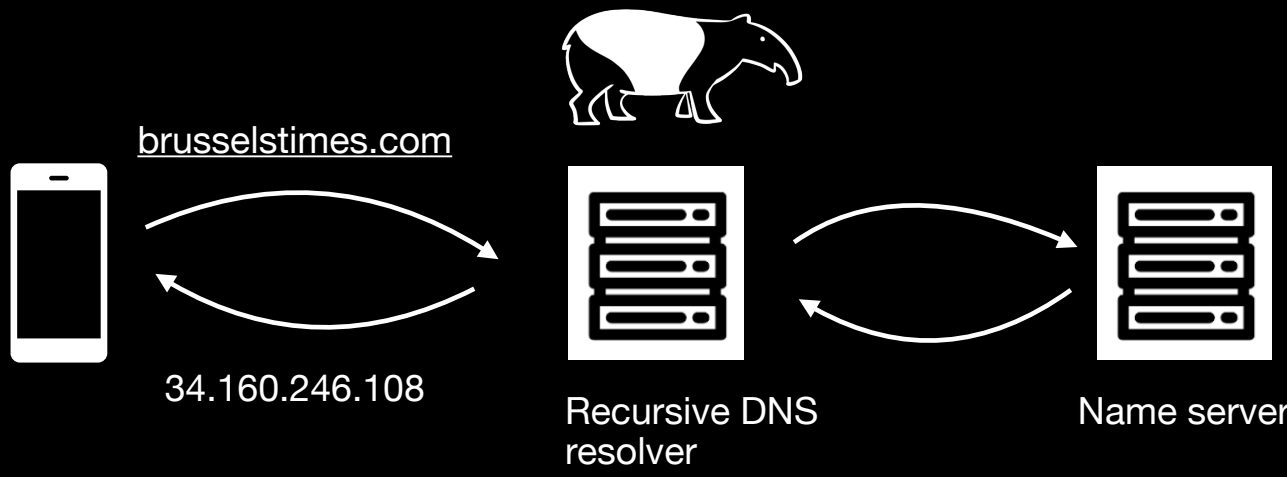
dnstapir.se
DNS Threat and Privacy Internet Research

Ulrika Vincent & Mikael Kullberg


BIG DATA
small data
Collect as much as possible
Store for as long as possible
Centralised data storage
Compliance by checkboxes
Protect by shields
Collect minimum
Throw away asap
Distributed storage
Compliance by design
Protect by differential privacy
and other techniques
Data you don’t have can’t be leaked
Image by Freepik

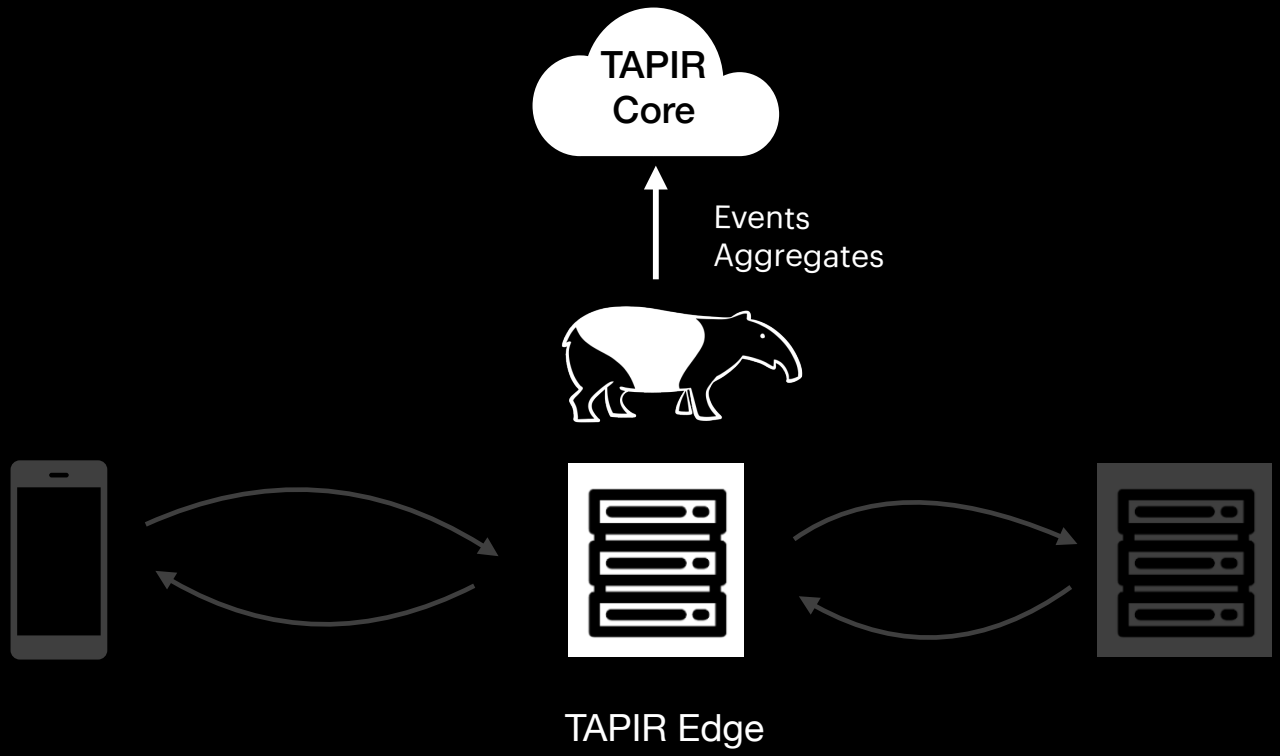
DNS TAPIR

A privacy first,
open source,
local decisions
and open data
DNS query analytics platform



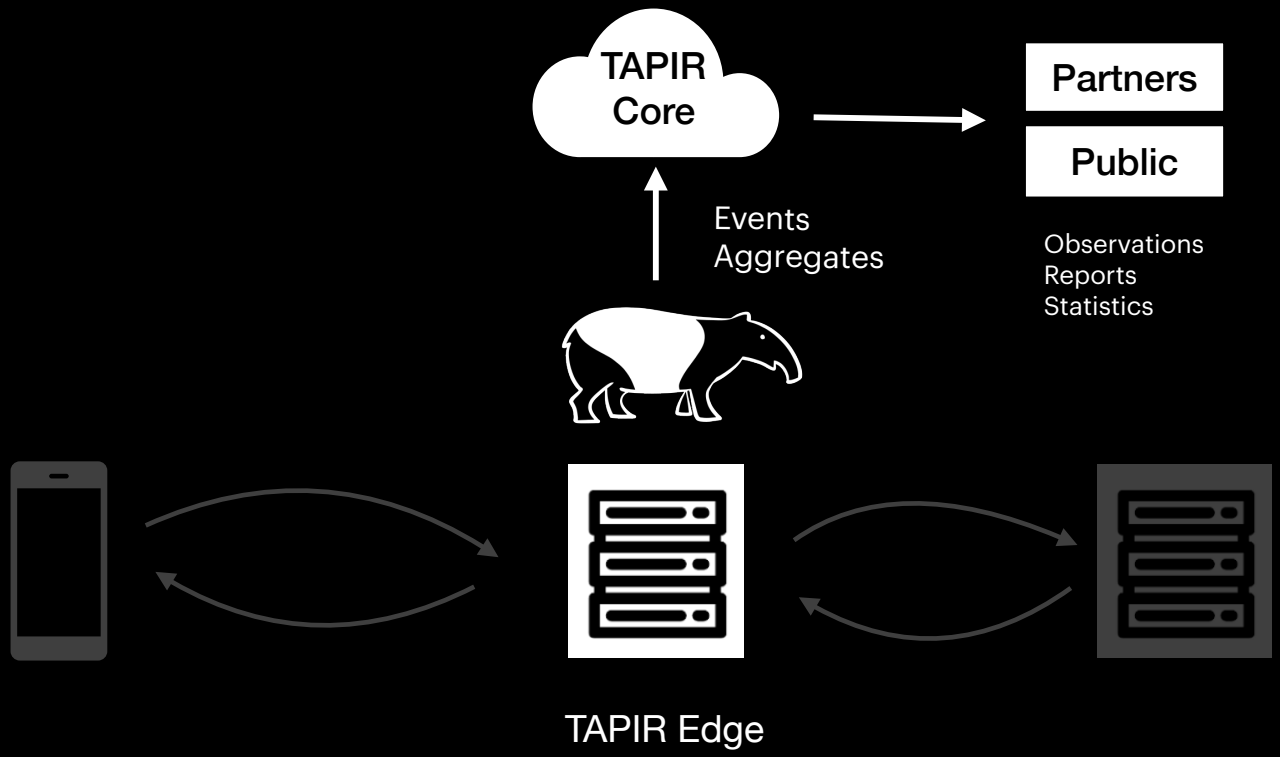
DNS TAPIR

A privacy first,
open source,
local decisions
and open data
DNS query analytics platform



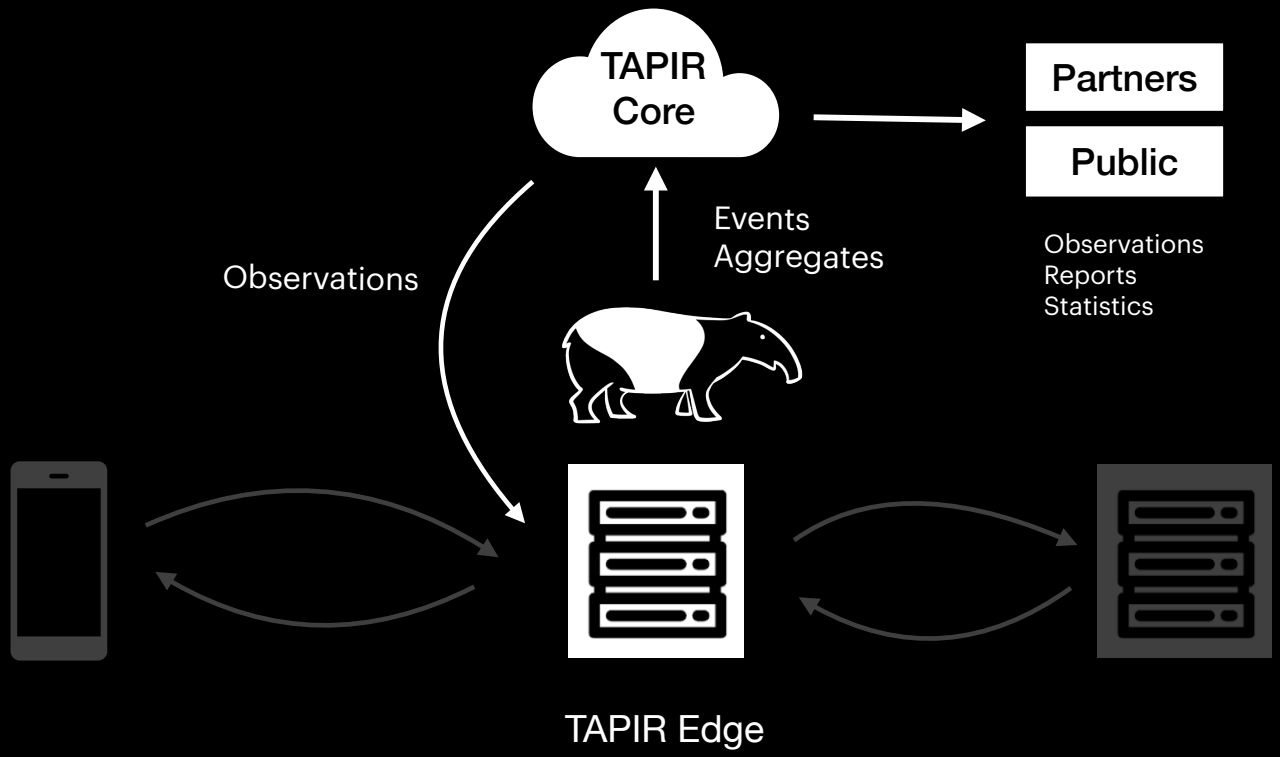
DNS TAPIR

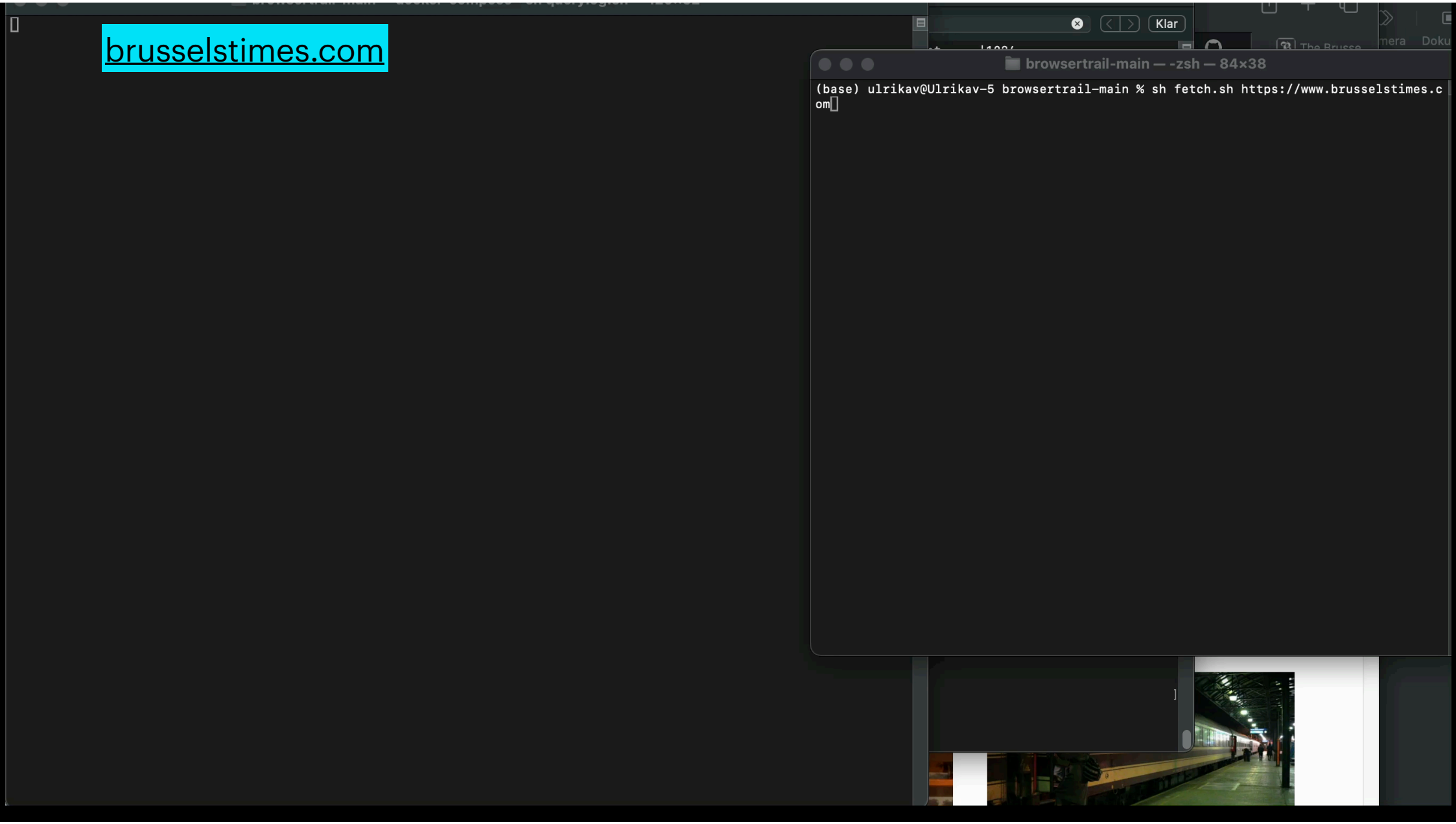
A privacy first,
open source,
local decisions
and open data
DNS query analytics platform



DNS TAPIR

A privacy first,
open source,
local decisions
and open data
DNS query analytics platform




brusselstimes.com
Klar
browsertrail-main — -zsh — 84×38
(base) ulrikav@Ulrikav-5 browsertrail-main % sh fetch.sh https://www.brusselstimes.com

brusselstimes.com

```
11:48:45.733591 IP 172.18.0.2.33105 > 192.1                    941+ A? pixel.quantserve.com. (38)
                30517 IP 172.18.0.2.56177 > 192.1              963+ Type65? region1.google-analytics.com. (46)
                30521 IP 172.18.0.2.48284 > 192.1              744+ A? region1.google-analytics.com. (46)
                70754 IP 172.18.0.2.46641 > 192.1              59+ Type65? cdn.cxense.com. (32)
11:48:45.870755 IP 172.18.0.2.53548 > 192.1                    277+ A? cdn.cxense.com. (32)
11:48:45.884957 IP 172.18.0.2.56119 > 192.1                    975+ Type65? platform.twitter.com. (38)
11:48:45.885104 IP 172.18.0.2.54637 > 192.1                    263+ A? platform.twitter.com. (38)
11:48:45.916410 IP 172.18.0.2.33070 > 192.1                    230+ Type65? ping.chartbeat.net. (36)
11:48:45.916691 IP 172.18.0.2.36010 > 192.1                    927+ A? ping.chartbeat.net. (36)
11:48:45.941609 IP 172.18.0.2.42788 > 192.1                    925+ A? fundingchoicesmessages.google.com. (51)
11:48:45.941609 IP 172.18.0.2.52459 > 192.1                    713+ Type65? fundingchoicesmessages.google.com. (51)
11:48:45.948397 IP 172.18.0.2.44277 > 192.1                    925+ Type65? imasdk.googleapis.com. (39)
11:48:45.948417 IP 172.18.0.2.57523 > 192.1                    599+ A? imasdk.googleapis.com. (39)
11:48:45.975681 IP 172.18.0.2.41211 > 192.1                    450+ Type65? onesignal.com. (31)
11:48:45.975681 IP 172.18.0.2.46113 > 192.1                    408+ A? onesignal.com. (31)
11:48:45.990014 IP 172.18.0.2.37579 > 192.1                    518+ Type65? p.brid.tv. (27)
11:48:45.990506 IP 172.18.0.2.60739 > 192.1                    578+ A? p.brid.tv. (27)
11:48:46.015831 IP 172.18.0.2.38714 > 192.1                    192+ Type65? api-2-0.spot.im. (33)
11:48:46.015846 IP 172.18.0.2.41685 > 192.1                    58+ A? publisher-assets.spot.im. (42)
11:48:46.015966 IP 172.18.0.2.37146 > 192.1                    927+ Type65? publisher-assets.spot.im. (42)
11:48:46.016134 IP 172.18.0.2.41449 > 192.1                    242+ A? api-2-0.spot.im. (33)
11:48:46.035864 IP 172.18.0.2.36754 > 192.1                    06+ Type65? syndication.twitter.com. (41)
11:48:46.036460 IP 172.18.0.2.43307 > 192.1                    308+ A? syndication.twitter.com. (41)
11:48:46.043259 IP 172.18.0.2.45384 > 192.1                    396+ Type65? c2-eu.piano.io. (32)
11:48:46.043259 IP 172.18.0.2.43593 > 192.1                    776+ A? c2-eu.piano.io. (32)
11:48:46.110634 IP 172.18.0.2.39573 > 192.1                    857+ Type65? vm.target-video.com. (37)
11:48:46.110650 IP 172.18.0.2.56190 > 192.1                    732+ A? vm.target-video.com. (37)
11:48:46.118274 IP 172.18.0.2.41539 > 192.1                    486+ Type65? imasdk.googleapis.com. (39)
11:48:46.118440 IP 172.18.0.2.48209 > 192.1                    113+ A? imasdk.googleapis.com. (39)
11:48:46.133185 IP 172.18.0.2.43698 > 192.1                    325+ Type65? s0.2mdn.net. (29)
11:48:46.133313 IP 172.18.0.2.45363 > 192.1                    252+ A? s0.2mdn.net. (29)
11:48:46.134703 IP 172.18.0.2.58638 > 192.1                    778+ Type65? pagead2.googlesyndication.com. (47)
11:48:46.134703 IP 172.18.0.2.42681 > 192.1                    432+ A? pagead2.googlesyndication.com. (47)
11:48:46.135375 IP 172.18.0.2.45898 > 192.1                    + A? stats-dev.brid.tv. (35)
11:48:46.135704 IP 172.18.0.2.43809 > 192.1                    944+ Type65? stats-dev.brid.tv. (35)
11:48:46.168918 IP 172.18.0.2.40329 > 192.1                    853+ Type65? cdn.cxense.com. (32)
11:48:46.169085 IP 172.18.0.2.56991 > 192.1                    791+ A? cdn.cxense.com. (32)
11:48:46.259965 IP 172.18.0.2.60559 > 192.1                    991+ A? p1cluster.cxense.com. (38)
11:48:46.260056 IP 172.18.0.2.39837 > 192.1                    85+ Type65? p1cluster.cxense.com. (38)
11:48:46.357988 IP 172.18.0.2.36062 > 192.1                    999+ Type65? comcluster.cxense.com. (39)
11:48:46.358004 IP 172.18.0.2.55845 > 192.1                    569+ A? comcluster.cxense.com. (39)
11:48:46.359180 IP 172.18.0.2.40539 > 192.1                    537+ Type65? id.cxense.com. (31)
11:48:46.359237 IP 172.18.0.2.41307 > 192.1                    509+ A? id.cxense.com. (31)
11:48:46.631190 IP 172.18.0.2.55829 > 192.1                    132+ Type65? static-cdn.spot.im. (36)
11:48:46.631358 IP 172.18.0.2.45751 > 192.1                    803+ A? static-cdn.spot.im. (36)
11:48:46.728194 IP 172.18.0.2.57376 > 192.1                    202+ Type65? pagead2.googlesyndication.com. (47)
11:48:46.728242 IP 172.18.0.2.47082 > 192.1                    710+ A? pagead2.googlesyndication.com. (47)
11:48:46.738142 IP 172.18.0.2.50366 > 192.1                    739+ Type65? direct-events-collector.spot.im. (49)
11:48:46.738251 IP 172.18.0.2.33890 > 192.1                    432+ A? direct-events-collector.spot.im. (49)
11:48:46.786029 IP 172.18.0.2.35335 > 192.1                    702+ Type65? csi.gstatic.com. (33)
11:48:46.786183 IP 172.18.0.2.35204 > 192.1                    778+ A? csi.gstatic.com. (33)
```

”Something that seems anonymous, more often than not, is not anonymous, even if it’s designed with the best intention”

(Matt Blaze)

Design principles in TAPIR

Aggregation

Separation of datasets - Crowd vs Unique. Treat them differently!

Approximate counts (HyperLogLog) - gives signals. How many, NOT who!

Make individual tracking **architecturally impossible.**

Aggregation is **irreversible**

Design principles in TAPIR

Aggregation

Separation of datasets - Crowd vs Unique. Treat them differently!

Approximate counts (HyperLogLog) - gives signals. How many, NOT who!

Make individual tracking **architecturally impossible.**

Aggregation is **irreversible**

Minimisation

Transform at the local source before extract.

No "data lake" of queryable DNS records. The **aggregated** sketch IS the data product

Time accuracy and sequences only at the **local** Edge resover

very short or ZERO retention regardless of what compliance allows

Design principles in TAPIR

Aggregation

Separation of datasets - Crowd vs Unique. Treat them differently!

Approximate counts (HyperLogLog) - gives signals. How many, NOT who!

Make individual tracking **architecturally impossible.**

Aggregation is **irreversible**

Minimisation

Transform at the local source before extract.

No "data lake" of queryable DNS records. The **aggregated** sketch IS the data product

Time accuracy and sequences only at the **local** Edge resover

very short or ZERO retention regardless of what compliance allows

Data you don’t have you can’t loose

Design principles

≈ Differential Privacy

Nearly identical results **with or without your** browsing/query data included

Deniability.
Even with full access to TAPIR Core data, you **can't definitively prove** any specific user queried a specific domain.

Design principles

≈ Differential Privacy

Nearly identical results **with or without your** browsing/query data included

Deniability.
Even with full access to TAPIR Core data, you **can't definitively prove** any specific user queried a specific domain.

Design for open data

Design for sharing.
An open data system gain operators (ISPs) **trust** to give us access to their data.

From "trust us to delete data" to "the data **will be open** and shared.”

It’s structure **fundamentally cannot represent** individuals."


Stop the pathological hoarding!
Toledo
CASIO
AUERBACH
nicht stürze
do not drop
ne laisser pa
(121)
compact dis
Image by: https://en.wikipedia.org/wiki/GNU_Free_Documentation_License

(the current)

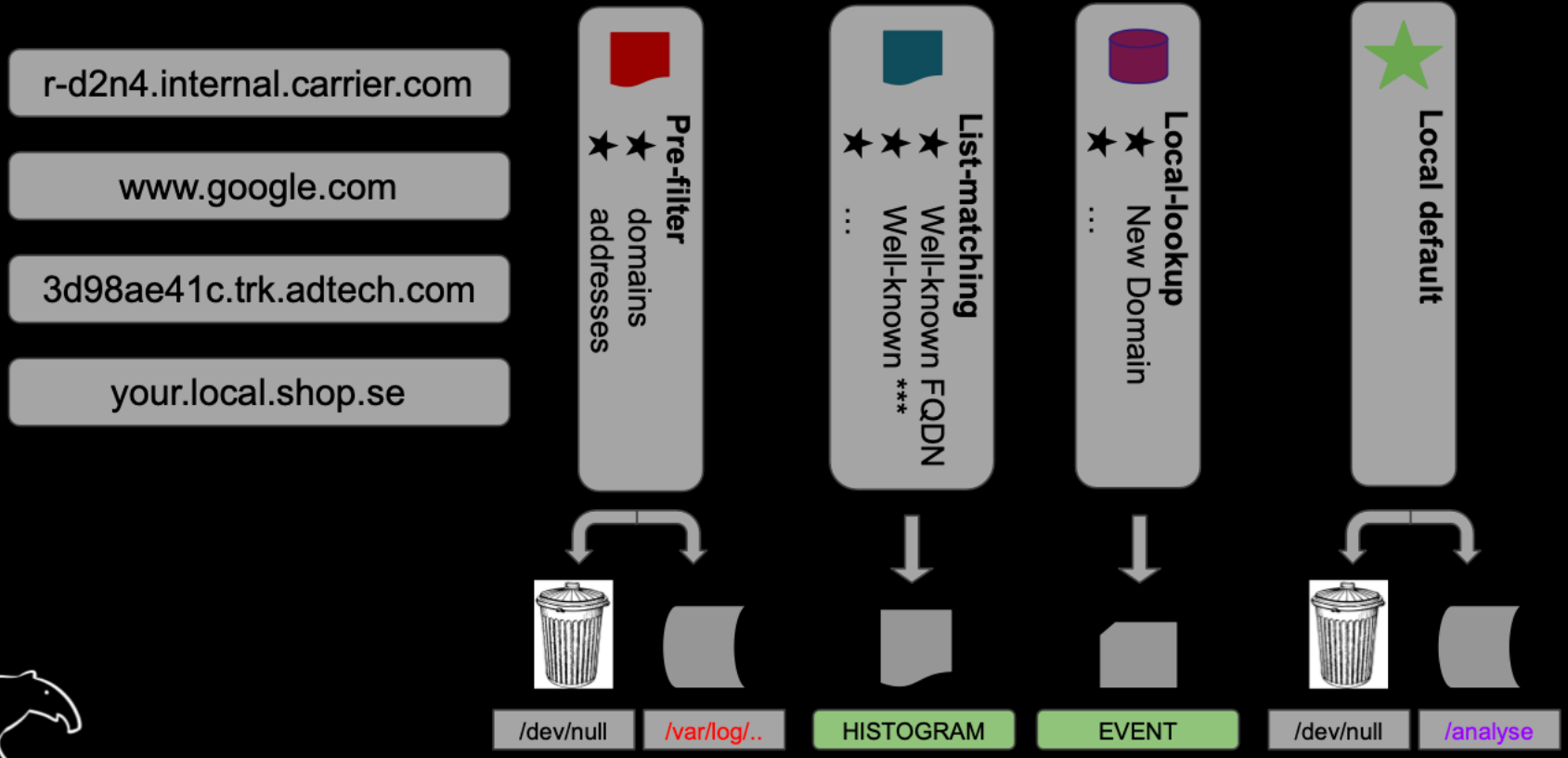
DNS TAPIR Analysis platform

...or, how the (current) sausages are made.

Some useless details

- aggrec, eventrec - data landing on “S3”
- Apache Spark, NATS, running on k8s
- µServices connect to µAnalysts
- JupyterHub interfaces to human analysts

Query Segmentation



```
-RECORD 0---------------------------------
 date                | 2025-11-21
 creator             | agitated-albattan...
 label0              | com
 label1              | google
 label2              | NULL
 label3              | NULL
 label4              | NULL
 label5              | NULL
 label6              | NULL
 label7              | NULL
 label8              | NULL
 label9              | NULL
 hour                | 20
 minute              | 25
 tagstring           | A
 fqdn                | google.com.
 r_fqdn              | com.google.
 idn_fqdn            | NULL
 a_count             | 5
 aaaa_count          | 0
 mx_count            | 0
 ns_count            | 0
 other_type_count    | 0
 non_in_count        | 0
 ok_count            | 5
 nx_count            | 0
 fail_count          | 0
 other_rcode_count   | 0
 deltas              | [180]
 ok                  | [5]
 nx                  | [0]
 fail                | [0]
 other_rcode         | [0]
 other_type          | [0]
 non_in              | [0]
 v4_clients          | [1]
 v6_clients          | [0]
 v4clients_hll       | [11 6A 7F]
 v6clients_hll       | [11 6A 7F]
 v4clients_avg       | 1.0
 v6clients_avg       | 0.0
 v4client_count_hll  | 0
 v6client_count_hll  | 0
only showing top 1 row
```

```
Msg(_client=<nats client v2.12.0>,
    subject='events.up.new_qname',
    reply='',
    data=b'{"flags":33152,"qclass":1,"qname":"AnYCAst.NiNJA.iHR.live.","qtype":'
         b'28,"timestamp":"2026-02-01T11:09:00Z","type":"new_qname","version":0'
         b'}',
    headers={'DNSTAPIR-Key-Identifier': 'competent-albattani.test.dnstapir.se',
             'DNSTAPIR-Key-Thumbprint': 'jvF_OG1sIbyJc45KFL6QbTWp23A8RHpRNqz0FtQuZlg',
             'DNSTAPIR-Message-Schema': 'https://schema.dnstapir.se/v1/new_qname',
             'DNSTAPIR-Mqtt-Topic': 'events/up/competent-albattani.test.dnstapir.se/new_qname'},
    _metadata=None,
    _ackd=False,
    _sid=1)
Msg(_client=<nats client v2.12.0>,
    subject='events.up.new_qname',
    reply='',
    data=b'{"flags":33152,"qclass":1,"qname":"SWCDn.APPLe.coM.","qtype":1,"time'
         b'stamp":"2026-02-01T11:09:00Z","type":"new_qname","version":0}',
    headers={'DNSTAPIR-Key-Identifier': 'competent-albattani.test.dnstapir.se',
             'DNSTAPIR-Key-Thumbprint': 'jvF_OG1sIbyJc45KFL6QbTWp23A8RHpRNqz0FtQuZlg',
             'DNSTAPIR-Message-Schema': 'https://schema.dnstapir.se/v1/new_qname',
             'DNSTAPIR-Mqtt-Topic': 'events/up/competent-albattani.test.dnstapir.se/new_qname'},
    _metadata=None,
    _ackd=False,
    _sid=1)
Msg(_client=<nats client v2.12.0>,
    subject='events.up.new_qname',
    reply='',
    data=b'{"flags":33187,"qclass":1,"qname":"1769944201.test.from-edge.looptes'
         b't.dnstapir.se.","qtype":1,"timestamp":"2026-02-01T11:09:00Z","type":'
         b'"new_qname","version":0}',
    headers={'DNSTAPIR-Key-Identifier': 'agitated-albattani.test.dnstapir.se',
             'DNSTAPIR-Key-Thumbprint': 'L5nEyauyBYOEutD6gghpDl_OoOeCHJzt6ZLAqbkZicQ',
             'DNSTAPIR-Message-Schema': 'https://schema.dnstapir.se/v1/new_qname',
             'DNSTAPIR-Mqtt-Topic': 'events/up/agitated-albattani.test.dnstapir.se/new_qname'},
    _metadata=None,
    _ackd=False,
    _sid=1)
Msg(_client=<nats client v2.12.0>,
    subject='events.up.new_qname',
    reply='',
    data=b'{"flags":33152,"qclass":1,"qname":"AAF5C491-6CDF-4761-AE4B-DD870F44E'
         b'9E7-netseer-ipaddr-assoc.xz.fbcdn.net.","qtype":28,"timestamp":"2026'
         b'-02-01T11:09:00Z","type":"new_qname","version":0}',
    headers={'DNSTAPIR-Key-Identifier': 'competent-albattani.test.dnstapir.se',
             'DNSTAPIR-Key-Thumbprint': 'jvF_OG1sIbyJc45KFL6QbTWp23A8RHpRNqz0FtQuZlg',
             'DNSTAPIR-Message-Schema': 'https://schema.dnstapir.se/v1/new_qname',
             'DNSTAPIR-Mqtt-Topic': 'events/up/competent-albattani.test.dnstapir.se/new_qname'},
    _metadata=None,
    _ackd=False,
    _sid=1)
Msg(_client=<nats client v2.12.0>,
    subject='events.up.new_qname',
    reply='',
    data=b'{"flags":33152,"qclass":1,"qname":"687035BC-5D18-49BE-88C9-C8244CE46'
         b'9D4-netseer-ipaddr-assoc.xz.fbcdn.net.","qtype":1,"timestamp":"2026-'
         b'02-01T11:09:00Z","type":"new_qname","version":0}',
    headers={'DNSTAPIR-Key-Identifier': 'competent-albattani.test.dnstapir.se',
             'DNSTAPIR-Key-Thumbprint': 'jvF_OG1sIbyJc45KFL6QbTWp23A8RHpRNqz0FtQuZlg',
             'DNSTAPIR-Message-Schema': 'https://schema.dnstapir.se/v1/new_qname',
             'DNSTAPIR-Mqtt-Topic': 'events/up/competent-albattani.test.dnstapir.se/new_qname'},
    _metadata=None,
    _ackd=False,
    _sid=1)
```

```
            F.col("edm_status_bits") == minInt64, 0
        ).when(
            F.col("edm_status_bits") < -1,
            F.try_add(
                F.col("edm_status_bits"),
                F.lit(maxint64)
            ) + 1
        ).otherwise(
            F.col("edm_status_bits")
        ).cast(T.LongType())
    ).withColumn("tagstring",
        F.concat_ws('',
            F.when(F.col("tags").bitwiseAND(pow(2, 0)) != 0, 'A'),
            F.when(F.col("tags").bitwiseAND(pow(2, 1)) != 0, 'B'),
            F.when(F.col("tags").bitwiseAND(pow(2, 2)) != 0, 'C'),
            F.when(F.col("tags").bitwiseAND(pow(2, 3)) != 0, 'D'),
            F.when(F.col("tags").bitwiseAND(pow(2, 4)) != 0, 'E'),
            F.when(F.col("tags").bitwiseAND(pow(2, 5)) != 0, 'F'),
            F.when(F.col("tags").bitwiseAND(pow(2, 6)) != 0, 'G'),
            F.when(F.col("tags").bitwiseAND(pow(2, 7)) != 0, 'H'),
            F.when(F.col("tags").bitwiseAND(pow(2, 8)) != 0, 'I'),
            F.when(F.col("tags").bitwiseAND(pow(2, 9)) != 0, 'J'),
            F.when(F.col("tags").bitwiseAND(pow(2,10)) != 0, 'K'),
            F.when(F.col("tags").bitwiseAND(pow(2,11)) != 0, 'L'),
            F.when(F.col("tags").bitwiseAND(pow(2,12)) != 0, 'M'),
            F.when(F.col("tags").bitwiseAND(pow(2,13)) != 0, 'N'),
            F.when(F.col("tags").bitwiseAND(pow(2,14)) != 0, 'O'),
            F.when(F.col("tags").bitwiseAND(pow(2,15)) != 0, 'P'),
            F.when(F.col("tags").bitwiseAND(pow(2,16)) != 0, 'Q'),
            F.when(F.col("tags").bitwiseAND(pow(2,17)) != 0, 'R'),
            F.when(F.col("tags").bitwiseAND(pow(2,18)) != 0, 'S'),
            F.when(F.col("tags").bitwiseAND(pow(2,19)) != 0, 'T'),
            F.when(F.col("tags").bitwiseAND(pow(2,20)) != 0, 'U'),
            F.when(F.col("tags").bitwiseAND(pow(2,21)) != 0, 'V'),
            F.when(F.col("tags").bitwiseAND(pow(2,22)) != 0, 'W'),
            F.when(F.col("tags").bitwiseAND(pow(2,23)) != 0, 'X'),
            F.when(F.col("tags").bitwiseAND(pow(2,24)) != 0, 'Y'),
            F.when(F.col("tags").bitwiseAND(pow(2,25)) != 0, 'Z'),
```

Data commons

Sharing data is tricky...

Fitness tracking app Strava gives away location of secret US army bases

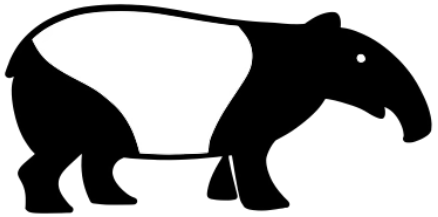
Reidentification

Netflix Cancels Contest over Privacy Concerns

Netflix canceled its second $1 million Netflix Prize after privacy concerns from the FTC

```
10  DESIGN with the intent of public data
20  THINK hard on how you would abuse the data
30  REDESIGN
40  GOTO 20
```

Be careful
what you ask for



DNS TAPIR
WWW.DNSTAPIR.SE

hula@catherd.se
ulrika@agical.se
info@dnstapir.se

@dnstapir@mastodon.social
LinkedIn