.nl open DNS datasets & statistics program

RIPE71 DNS-WG

November 2015

Marco Davids, SIDN Labs

SIDN labs
Internet Research & Innovation

SIDN

SIDN

- Registry for .nl ccTLD
- ~ 5,6 million domain names
- ~ 2,45 million domain names signed
- SIDN Labs is the R&D team

SIDN labs
Internet Research & Innovation
SIDN

ENTRADA

ENhanced Top-Level Domain Resilience through Advanced Data Analysis

- > 300 GB of PCAP data daily
- > 1.3 billion query's daily
- > 3.1 million distinct resolvers
- Currently capturing some 10% of total

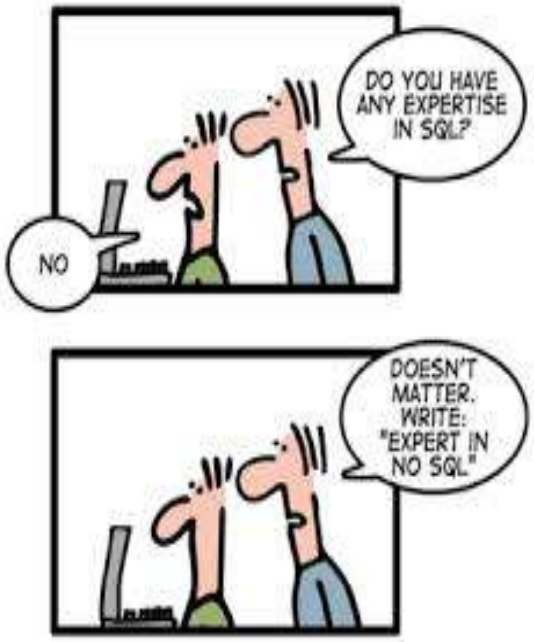
SIDN labs
Internet Research & Innovation

SIDN

Query Engine Options

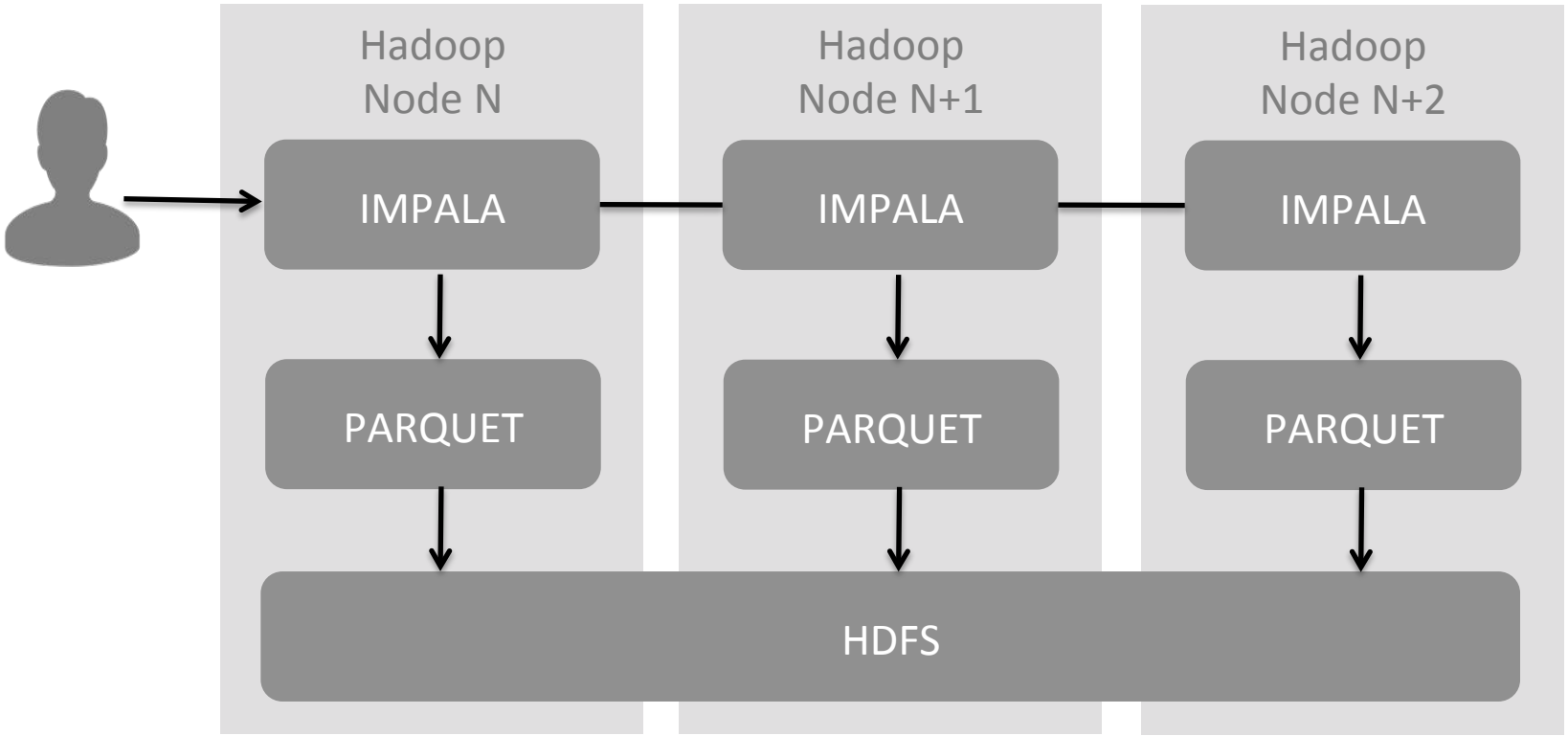
Evaluated SQL and NoSQL solutions

- Relational SQL (PostgreSQL)
- MongoDB
- Cassandra
- Elasticsearch
- Hadoop (HBASE + Apache Phoenix or Hive)
- SQL on Hadoop (HDFS + Impala + Parquet)



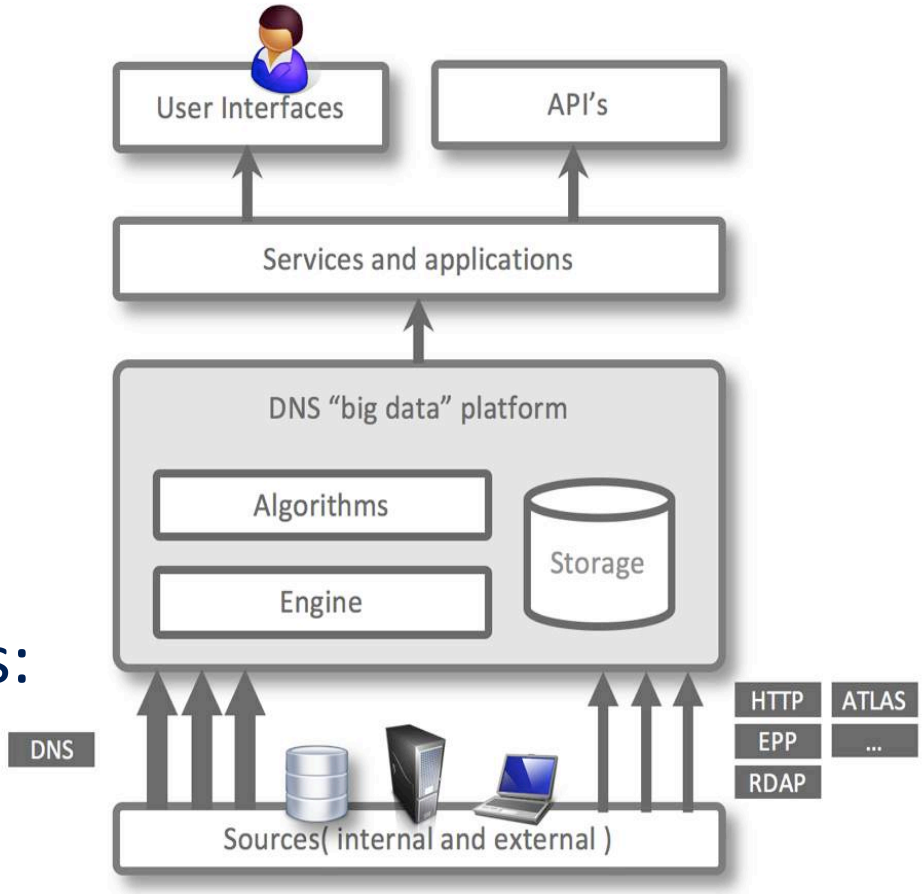
SQL on Hadoop

Best fit for our requirements



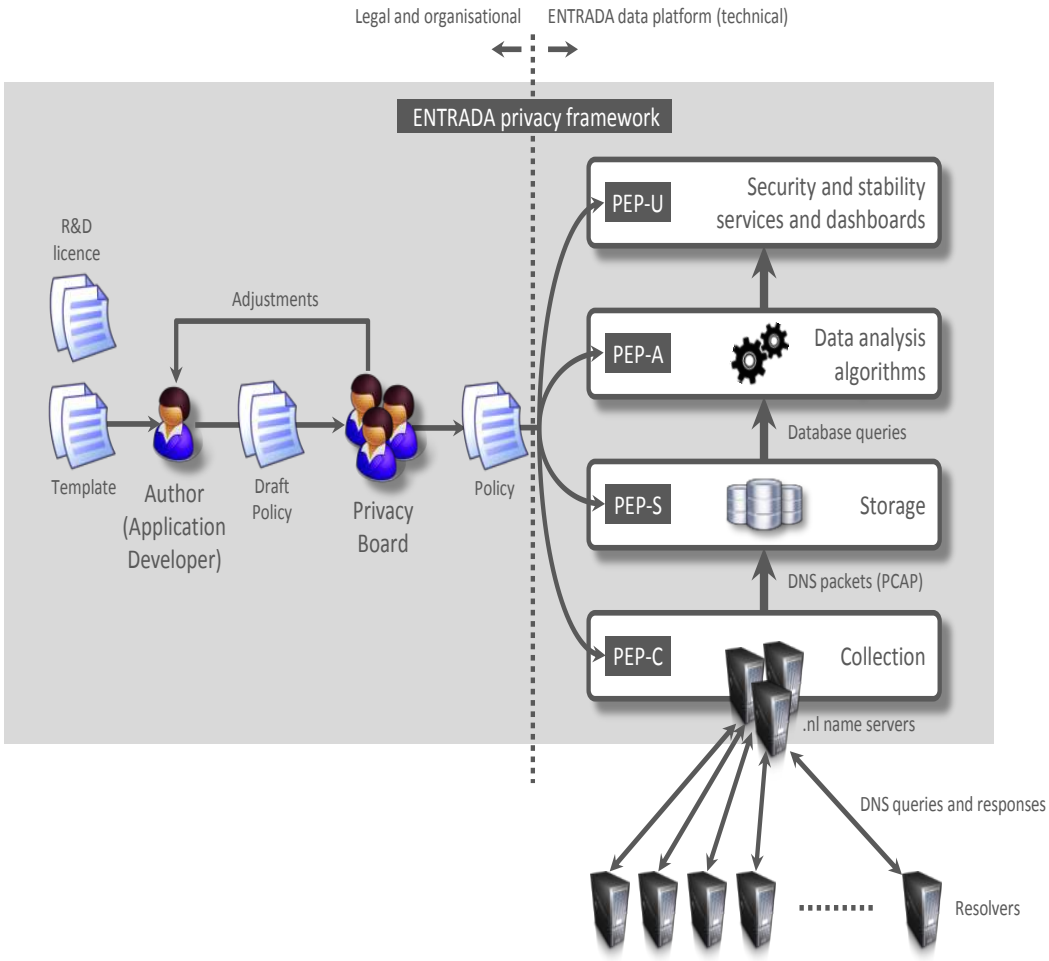
ENTRADA Architecture

- 'DNS big data' system
- **Goal**: develop applications and services that further enhance the security and stability of .nl, the DNS, and the Internet at large.
- ENTRADA main components:
  - Applications and services
  - Platform
  - Data sources
  - Privacy framework



SIDN labs
Internet Research & Innovation
SIDN

Privacy



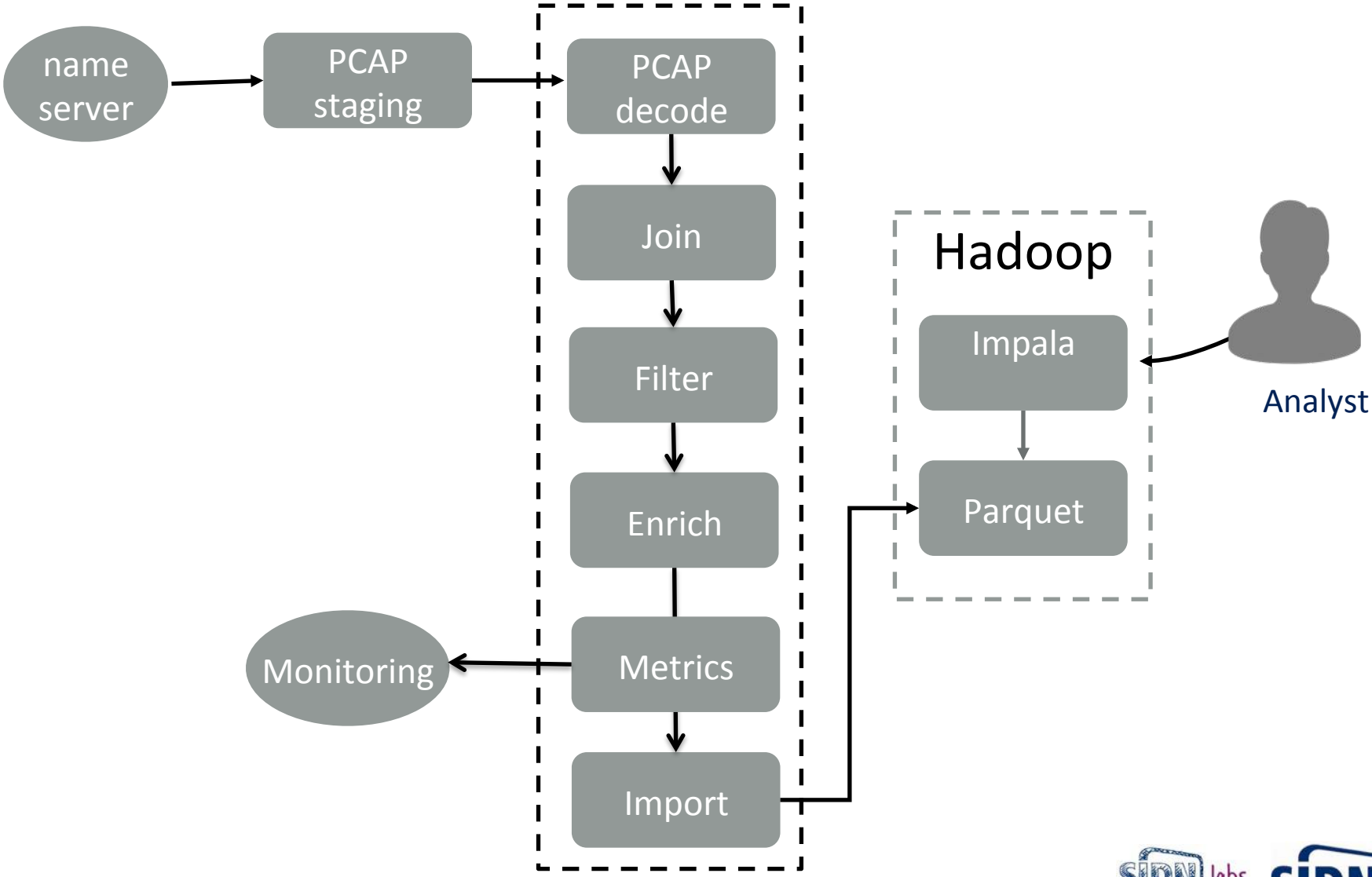
Download paper:
https://goo.gl/wec5dR

Policy elements:

- Purpose
- Data that is used
- Filters on the data
- Retention period
- Access to the data
- Type of application (Research vs. Prod.)

SIDN labs
Internet Research & Innovation
SIDN

Workflow



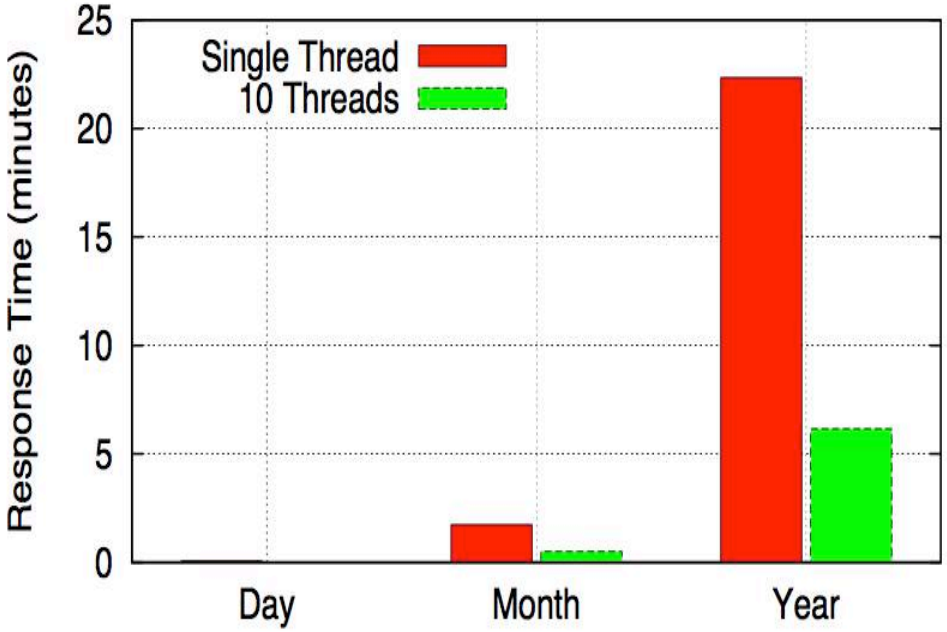
Query data available for analysis within 10 minutes

SIDN labs
Internet Research & Innovation
SIDN

Performance

Example query, count # IPv4 queries/day.

```
select
concat_ws('-',day,month,year),
count(1)
from dns.queries
where ipv=4
group by
concat_ws('-',day,month,year)
```



Query response-times

1 year of data is 2.2TB Parquet ~ 52TB of PCAP

SIDN labs
Internet Research & Innovation
SIDN

Status

Name server feeds	2
Queries per day	~150M
Daily PCAP volume(gzipped)	~33GB
Daily Parquet volume	~6GB
Months operational	18
Total # queries stored	> 71B
Total Parquet volume	> 3TB
HDFS (3x replication)	> 9TB
Cluster capacity	~150B-200B tuples

SIDN labs
Internet Research & Innovation
SIDN

Use Cases

Focussed on increasing the security and stability of .nl

- Visualize DNS patterns (like traffic patterns for phishing domain names)
- Detect botnet infections
- Real-time Phishing detection
- Statistics (https://stats.sidnlabs.nl)
- Scientific research (collaboration with universities)
- Operational support for DNS operators

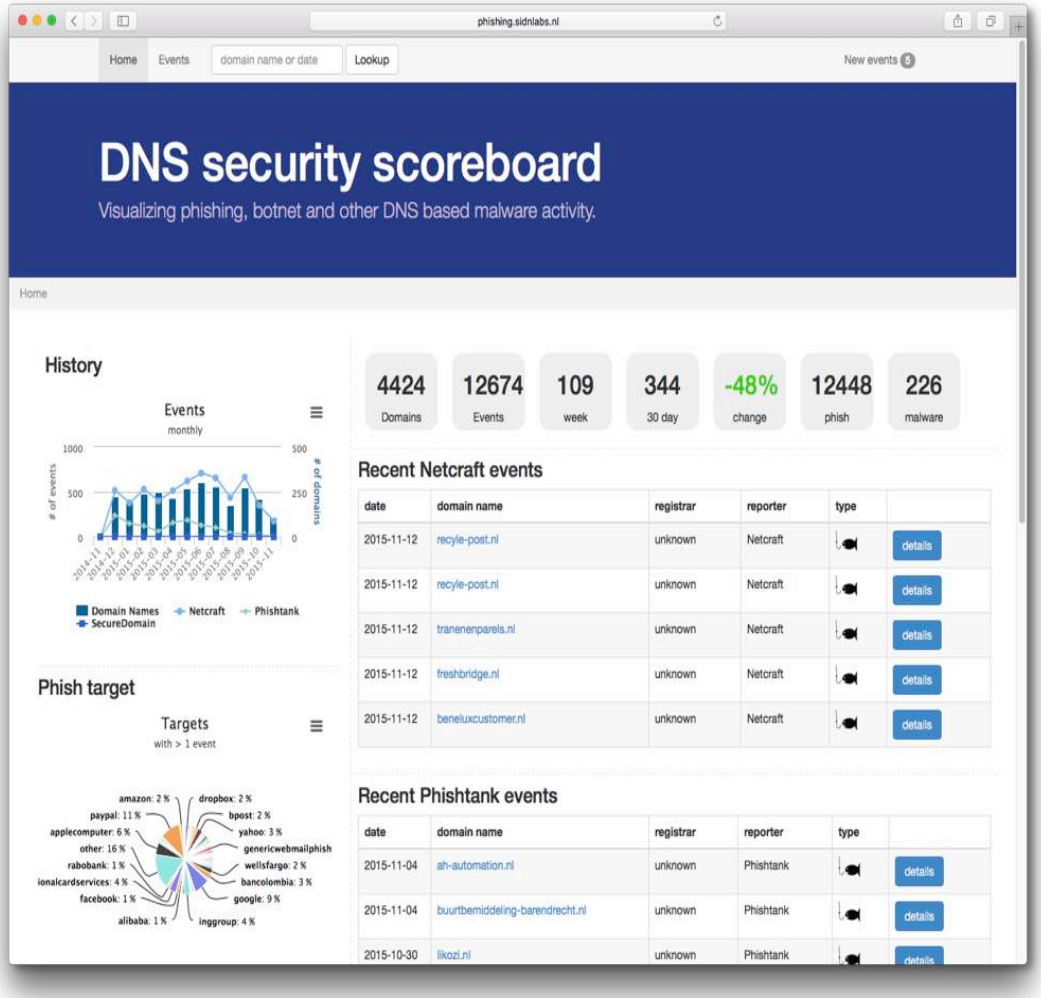
SIDN labs
Internet Research & Innovation
SIDN

Example Applications

- DNS security scoreboard
- Resolver reputation

SIDN labs
Internet Research & Innovation
SIDN

DNS Security Scoreboard

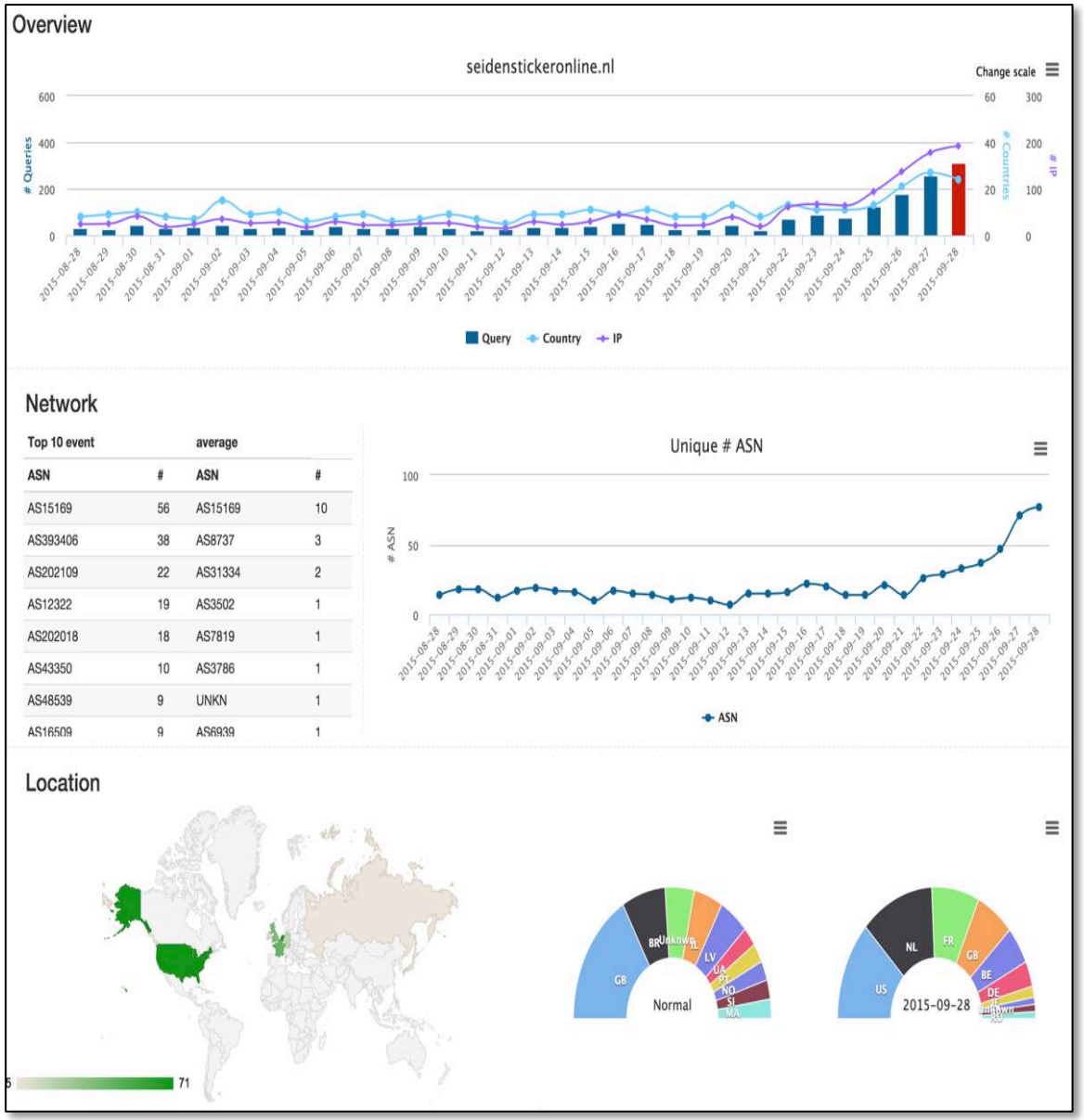


Goal: Visualize DNS patterns for malicious activity.

How: Combine external phishing feeds with DNS data.

SIDN labs
Internet Research & Innovation
SIDN

Traffic Visualization



SIDN labs
Internet Research & Innovation

SIDN

Resolver Reputation (ResRep)

Goal:

Try to detect malicious activity by assigning reputation scores to resolvers.

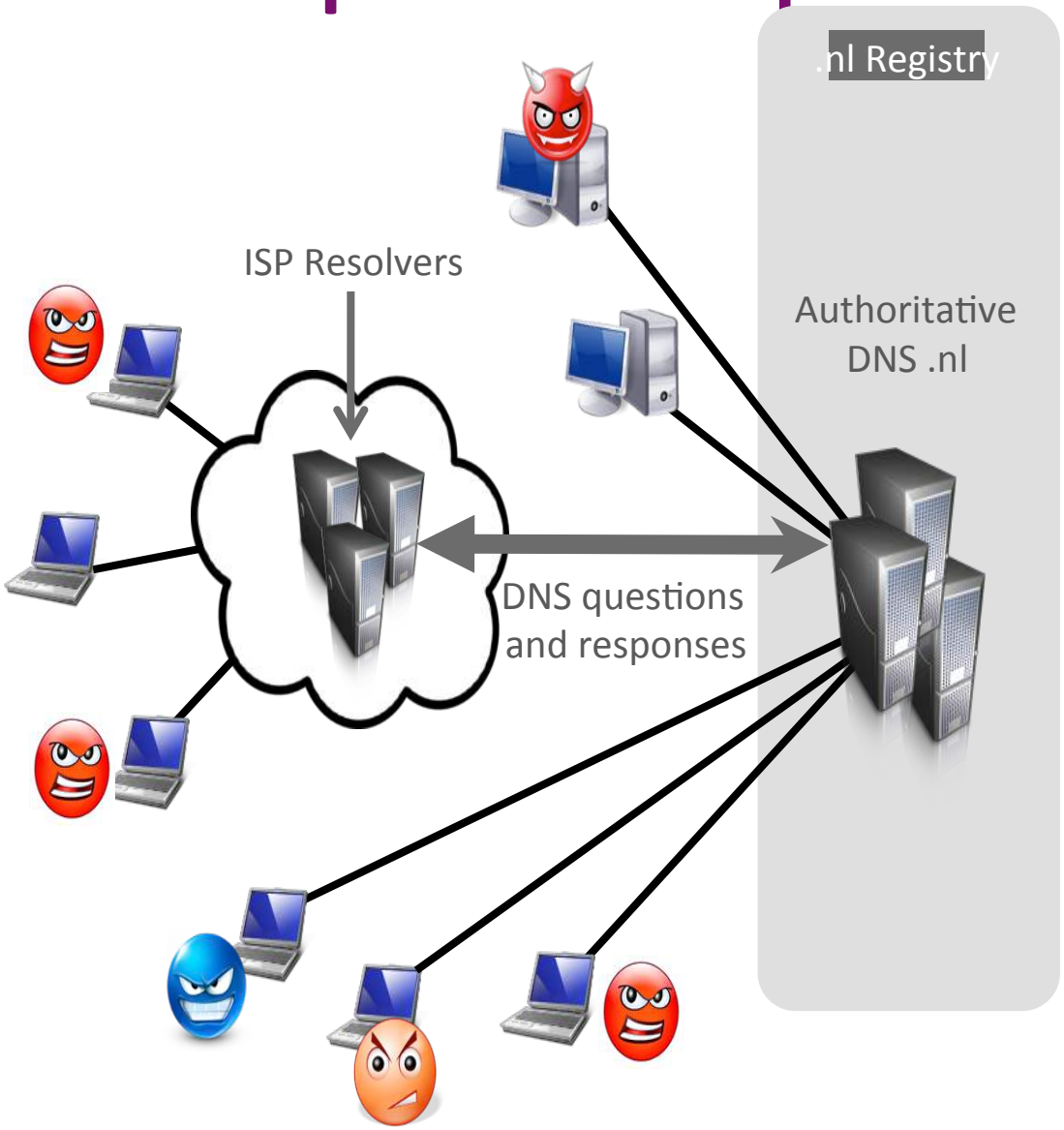
How:

“Fingerprinting” resolver behaviour.

SIDN labs
Internet Research & Innovation

SIDN

ResRep Concept

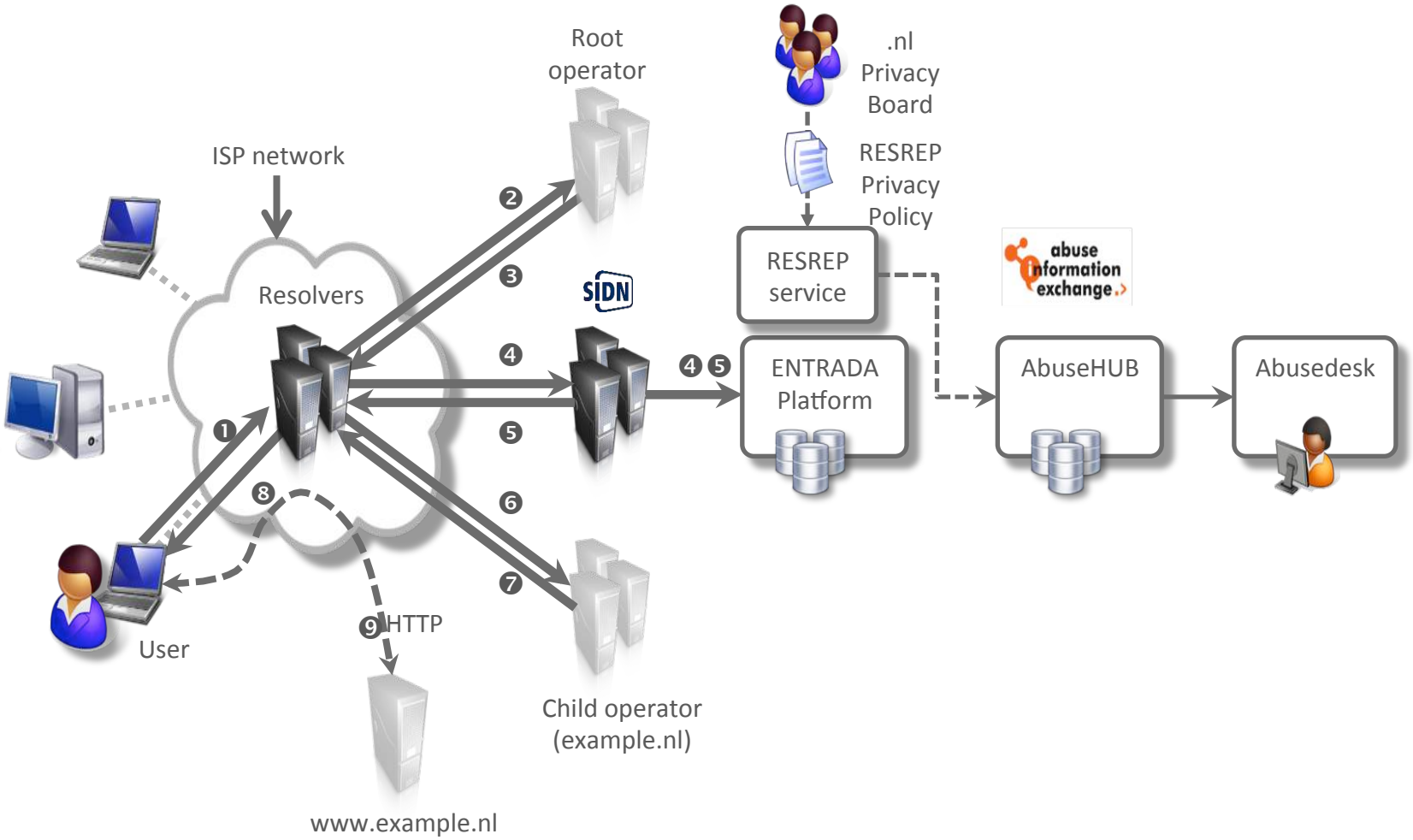


Malicious activity:

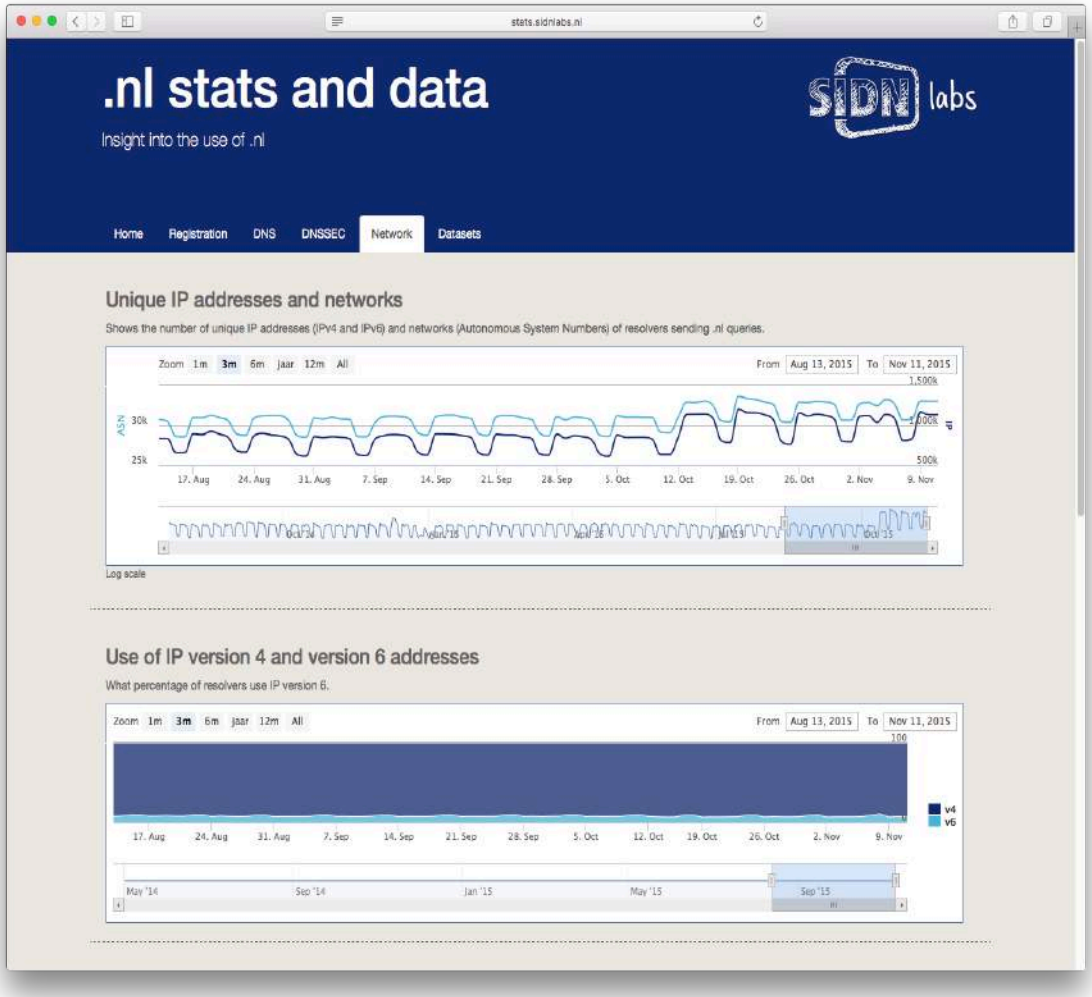
- Spam-runs
- Botnets like Cutwail
- DNS-amplification attacks

SIDN labs
Internet Research & Innovation
SIDN

ResRep Architecture



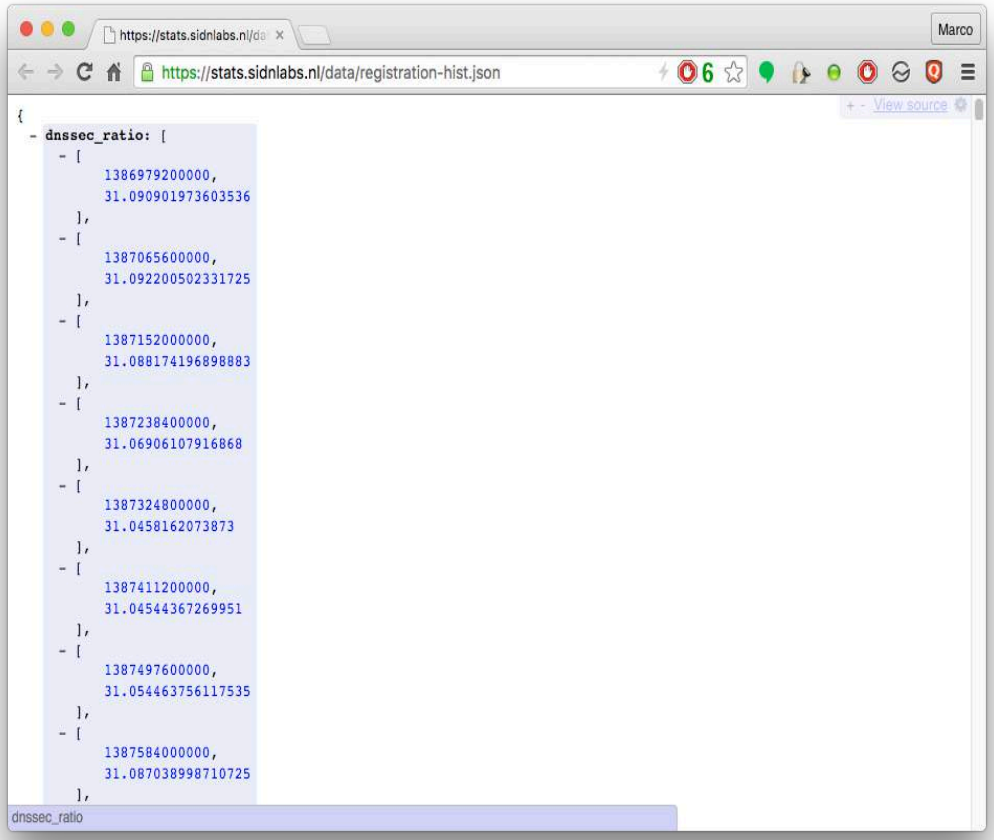
Open Data program



https://stats.sidnlabs.nl

SIDN labs
Internet Research & Innovation
SIDN

Open Data program (JSON files)



https://stats.sidnlabs.nl

SIDN labs
Internet Research & Innovation
SIDN

Open Data program (JSON files)

https://stats.sidnlabs.nl

Please let us know:

- Is it, or can it be useful?
- How, what for?
- For whom?
- What can we do to improve things?

SIDN labs
Internet Research & Innovation
SIDN

Questions and Feedback

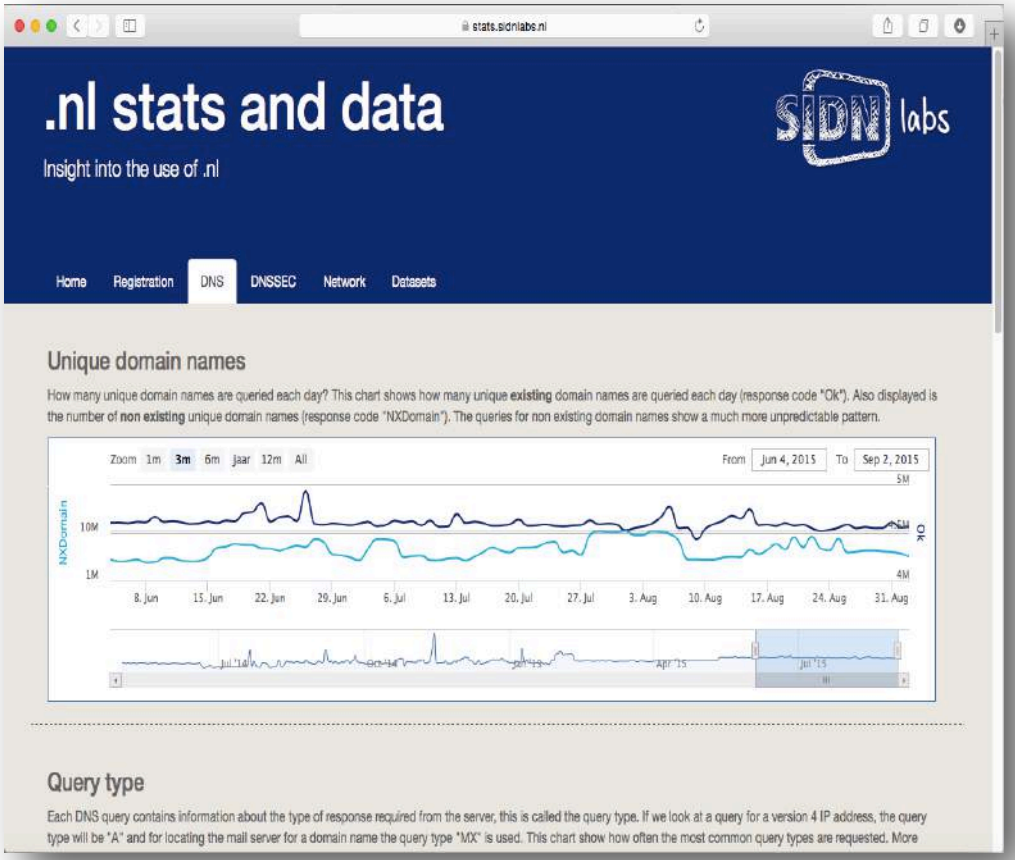
Marco Davids

Senior Research Engineer

marco.davids@sidn.nl

@marcodavids

www.sidnlabs.nl



https://stats.sidnlabs.nl

SIDN labs
Internet Research & Innovation
SIDN