A Privacy framework for DNS big data

3 October 2014

Cristian Hesselman, Jelte Jansen, Maarten Wullink,

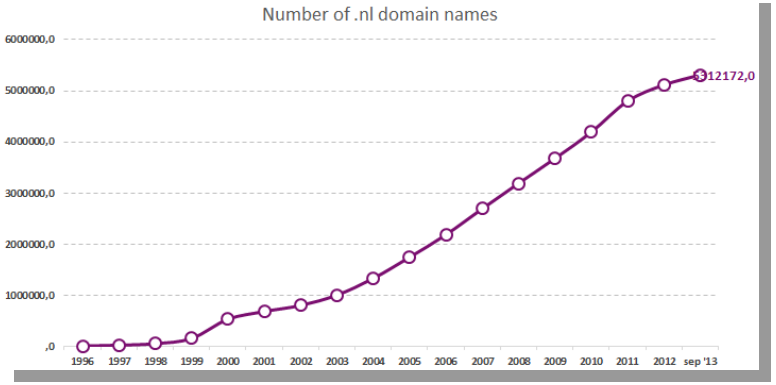
Karin Vink en Maarten Simon

SIDN labs
Internet Research & Innovation

SIDN

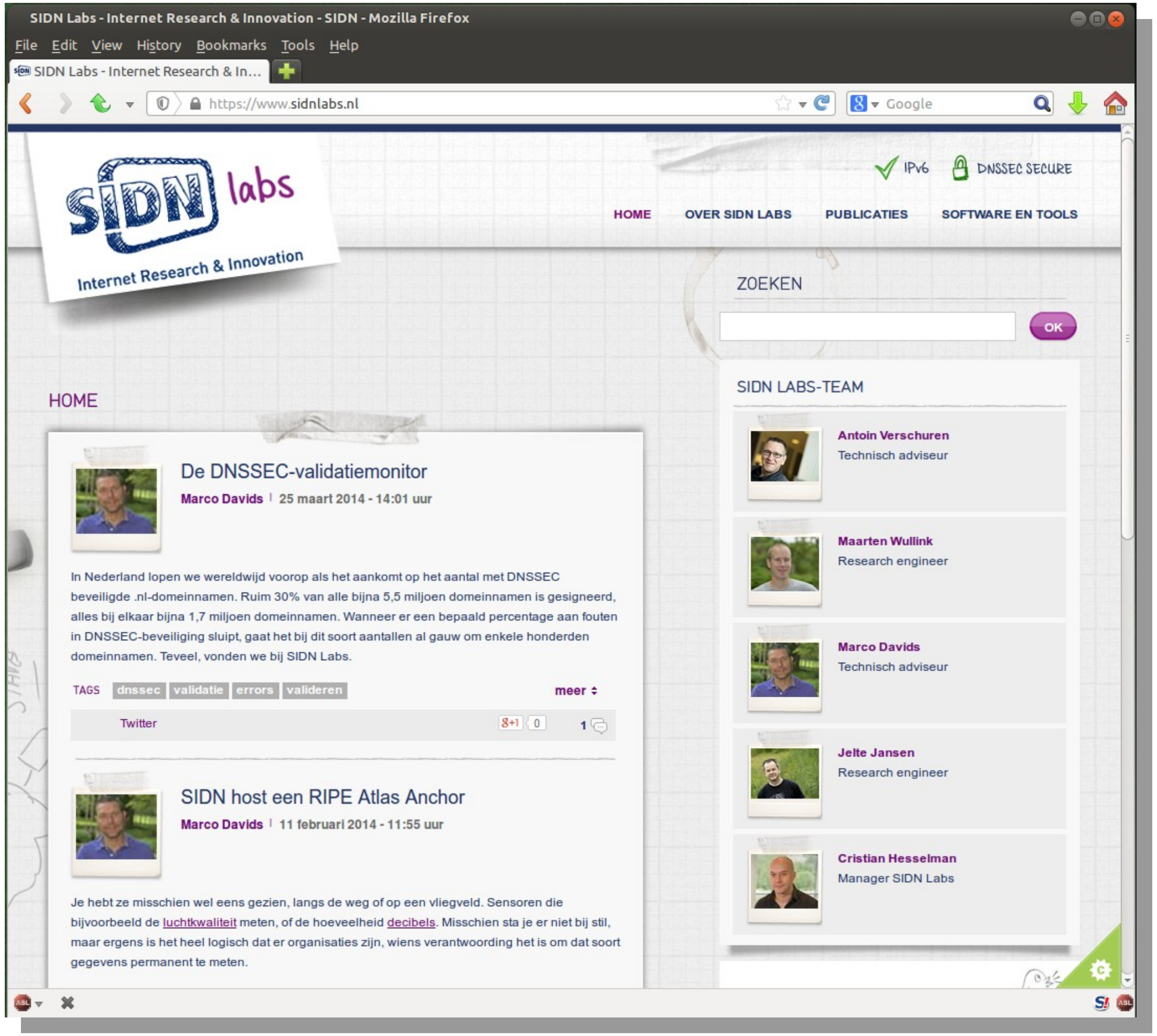
SIDN

- “.nl” (Registry voor Nederland)
- 5.5M domain names, >1.600 registrars
- > 1.000.000.000 DNS queries per day
- Private foundation with public task



SIDN Labs

- R&D team SIDN
- Improve services of SIDN
- Center of expertise
- Improve security of Internet in the Netherlands
- Facilitates external research


SIDN Labs - Internet Research & Innovation - SIDN - Mozilla Firefox
File Edit View History Bookmarks Tools Help
SIDN Labs - Internet Research & In...
https://www.sidnlabs.nl
Google
SIDN labs
Internet Research & Innovation
IPv6
DNSSEC SECURE
HOME
OVER SIDN LABS
PUBLICATIES
SOFTWARE EN TOOLS
ZOEKEN
OK

HOME

De DNSSEC-validatiemonitor

Marco Davids | 25 maart 2014 - 14:01 uur

In Nederland lopen we wereldwijd voorop als het aankomt op het aantal met DNSSEC beveiligde .nl-domeinnamen. Ruim 30% van alle bijna 5,5 miljoen domeinnamen is gesigneerd, alles bij elkaar bijna 1,7 miljoen domeinnamen. Wanneer er een bepaald percentage aan fouten in DNSSEC-beveiliging sluipt, gaat het bij dit soort aantallen al gauw om enkele honderden domeinnamen. Teveel, vonden we bij SIDN Labs.

TAGS dnssec validatie errors valideren meer

Twitter 8+1 0 1

SIDN host een RIPE Atlas Anchor

Marco Davids | 11 februari 2014 - 11:55 uur

Je hebt ze misschien wel eens gezien, langs de weg of op een vliegveld. Sensoren die bijvoorbeeld de luchtkwaliteit meten, of de hoeveelheid decibels. Misschien sta je er niet bij stil, maar ergens is het heel logisch dat er organisaties zijn, wiens verantwoording het is om dat soort gegevens permanent te meten.

SIDN LABS-TEAM

Antoin Verschuren	Technisch adviseur
Maarten Wullink	Research engineer
Marco Davids	Technisch adviseur
Jelte Jansen	Research engineer
Cristian Hesselman	Manager SIDN Labs

SIDN labs Internet Research & Innovation

SIDN

Privacy Framework

- What?

- Why?

- How?

Privacy Framework: Why?

- Public service that is vital to Dutch society and economy
- Keep trust and confidence in SIDN as the operator for .nl
- Responsibility to be proactive in the field
- SIDN wants to act transparently

Privacy Framework: Innovations

- Introduces Privacy management to the use of DNS data
- Integrates legal, technical and organisational aspects of privacy management

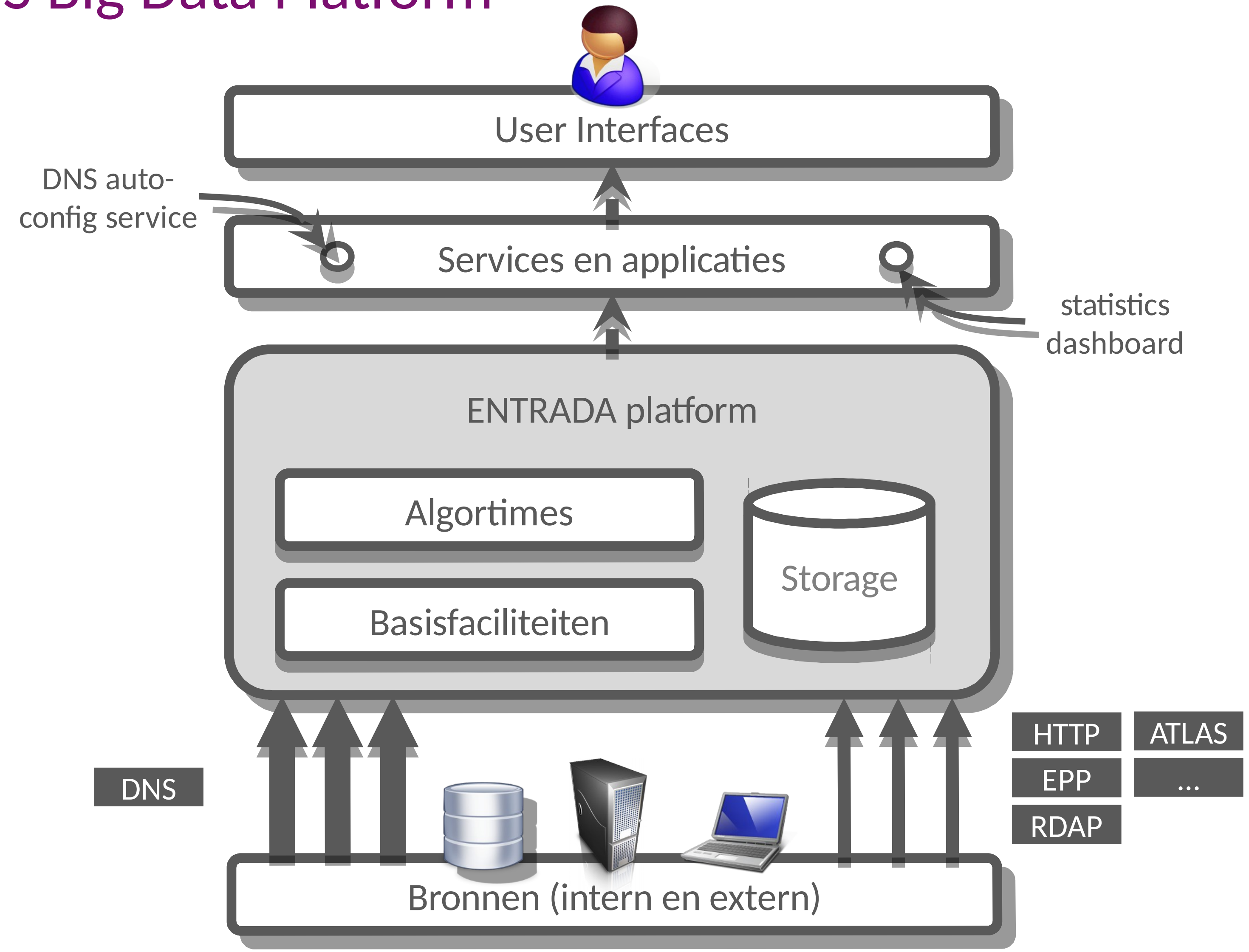
ENTRADA: DNS Big Data Platform

- ENhanced Top-level domain Resilience through Advanced Data Analysis

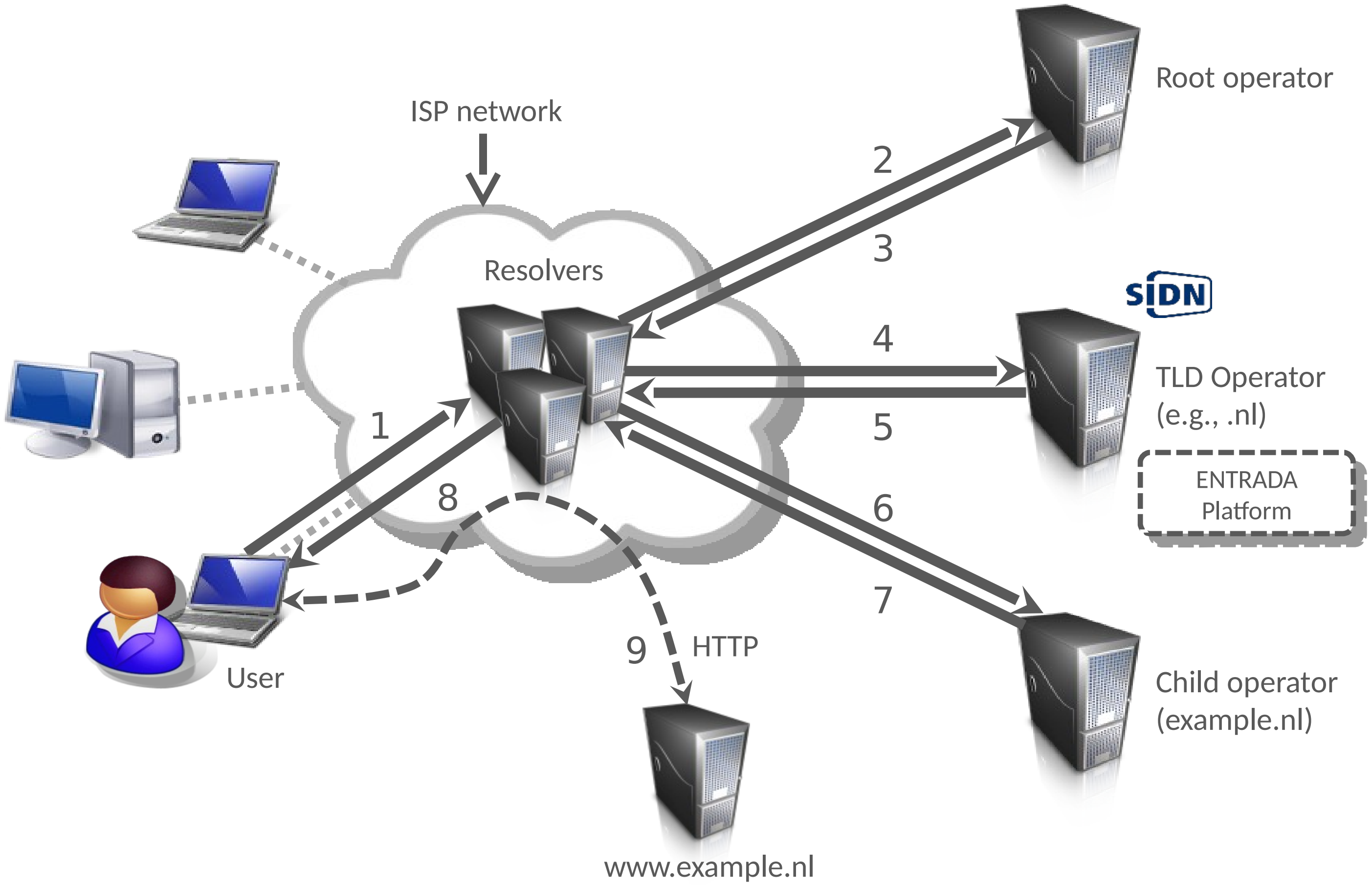
- Goal: Develop and evaluate big data applications
  - To Safeguard stability of '.nl'
  - To increase the safety of the (Dutch) Internet
  - To Detect botnets and abuse
  - **Non-goal:** commercial use

- What about privacy?

ENTRADA: DNS Big Data Platform



DNS



ISP network
Resolvers
Root operator
TLD Operator
(e.g., .nl)
ENTRADA
Platform
Child operator
(example.nl)
User
www.example.nl
HTTP
1
2
3
4
5
6
7
8
9
SIDN

(Potential) Personal Data in DNS Queries

- IP Address
- Queried name
- 'other'
  - Timestamps
  - Protocol flags
  - Etc.

WBP

- Dutch Data Protection Act (Wet Bescherming Persoonsgegevens, WBP)

- Personal Data:
  - 'any piece of information regarding an identified or identifiable natural person'

- Processing:
  - ‘any action or sequence of actions involving personal data, including but not restricted to the collection, recording, sorting, […] deletion or destruction of such data’

Requirements for Processing

- Public Function
- Contractual obligation
- Legitimate Basis
- Explicit consent
- Purpose Limitation
  - Personal data may only be used for the purpose for which it was collected
- Special Personal Data explicitely forbidden
  - Religion
  - Political views
  - Etc.

WBP and ENTRADA

- We are not using 'Public Function': too weak
  - besides, we are not government

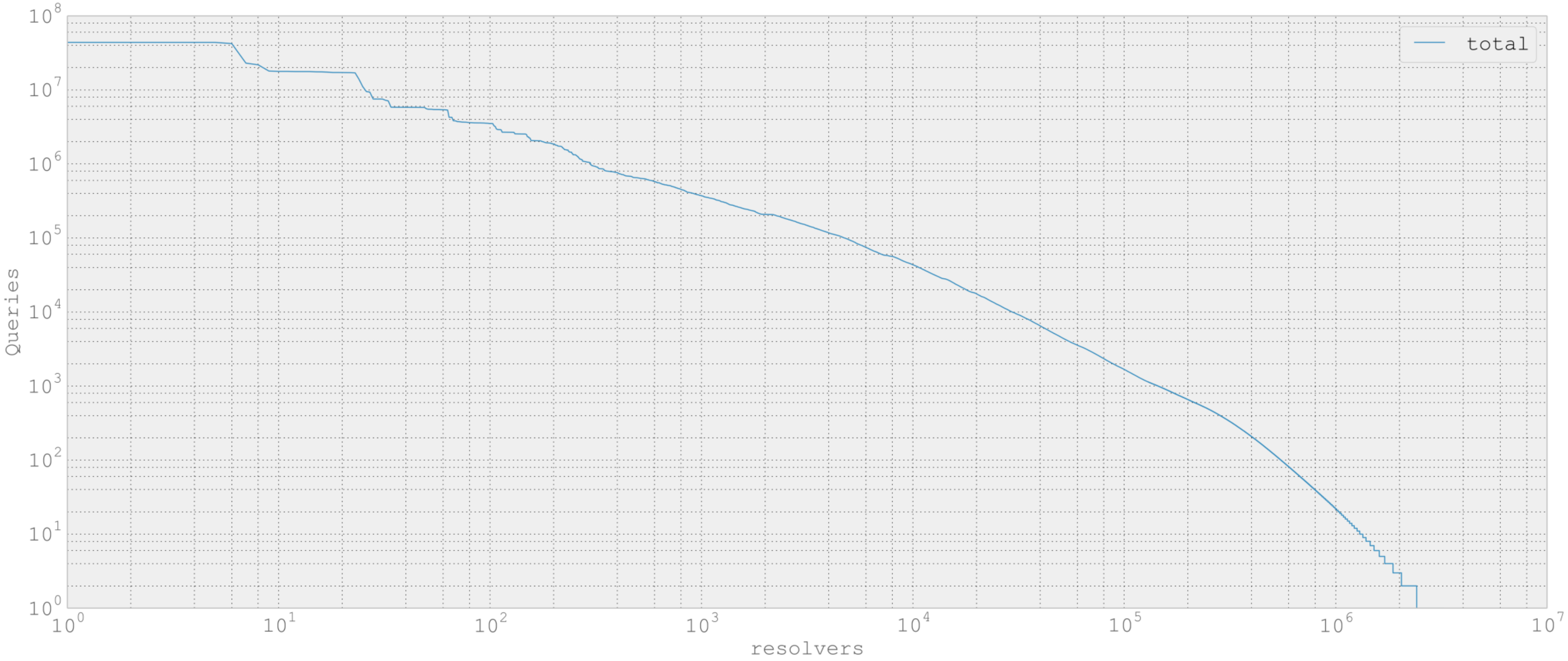
- Explicit consent not possible
  - So we need to be completely transparent

- Legitimate basis + Purpose Limitation
  - The goal is for the benefit of the users themselves

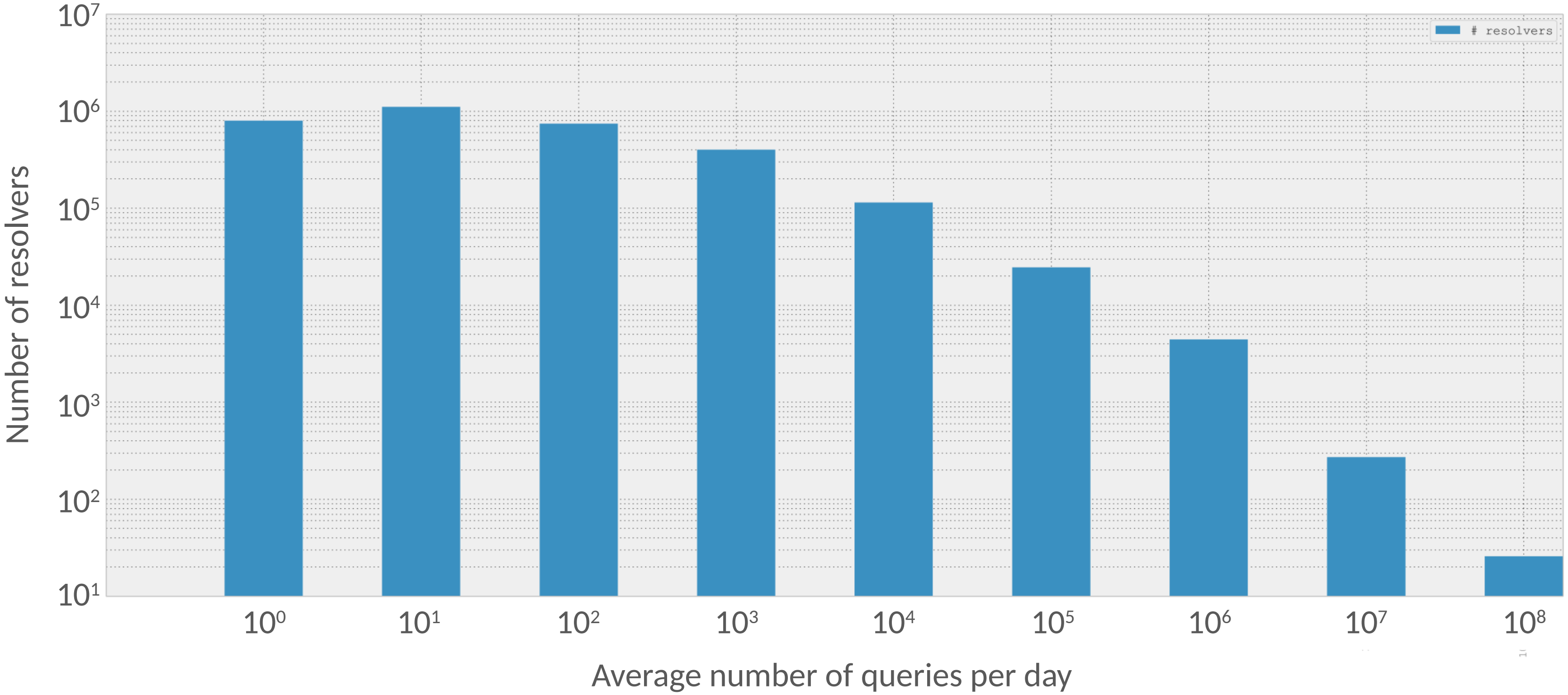
WBP and ENTRADA: IP Addresses

- Can't simply anonimize them
- Most addresses are from resolvers (shared by users)
- BUT: Resolvers may be 'home' systems

Number of Queries per Resolver per Day



Number of Queries per Resolver per Day



WBP and ENTRADA: IP Addresses

- Most individual resolvers are 'home' resolvers
  - Few users, so addresses likely to be personal data

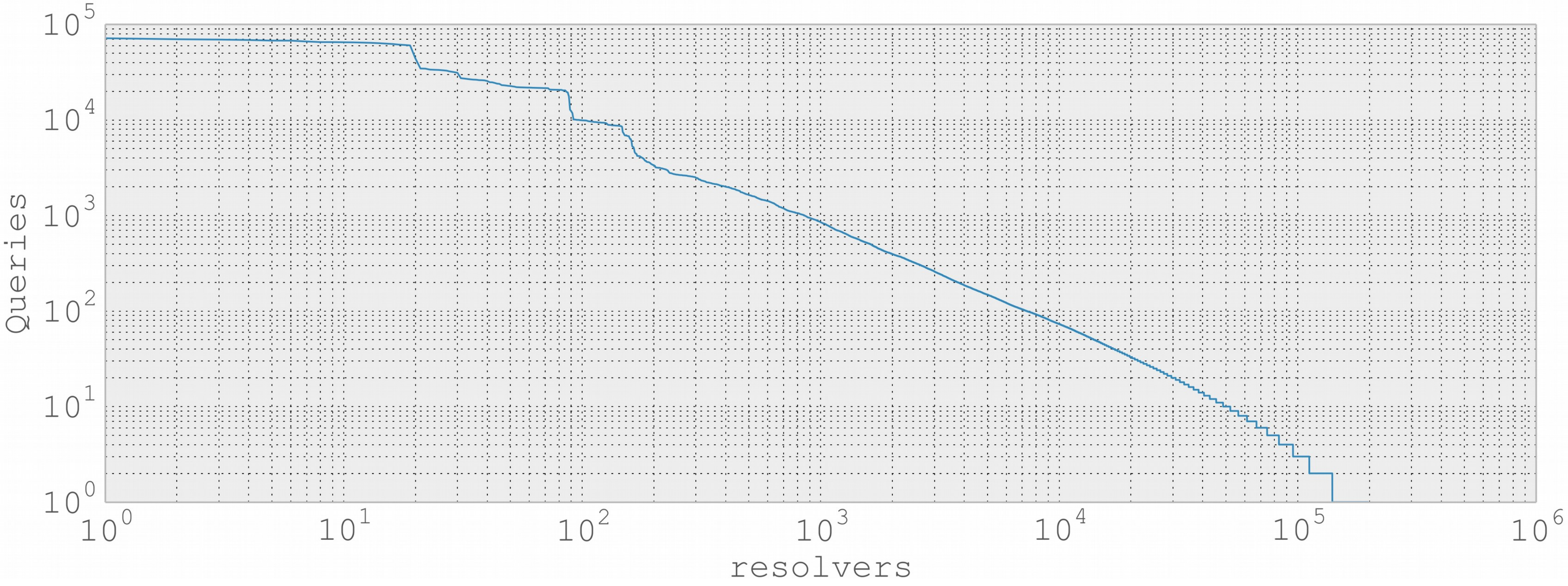
- 'Big' resolvers either ISP or domainers
  - In the second case, still personal data

- Better metrics are future work
  - Problem: to decide whether the address is personal data, you need to process it

WBP and ENTRADA: Queried names

- Single query does not say much
  - 'www.universityoftilburg.nl'
  - Not even associated with Tilburg University ;)
- Combined data can be considered personal
  - Query patterns, pre-fetching
- Query names may include other personal data
  - Personal names (firstname.lastname.mycloud.nl)
  - IP addresses (192.0.2.1.customer-adsl.example.nl)
- Can also be combined with IP address of resolver (previous slides)

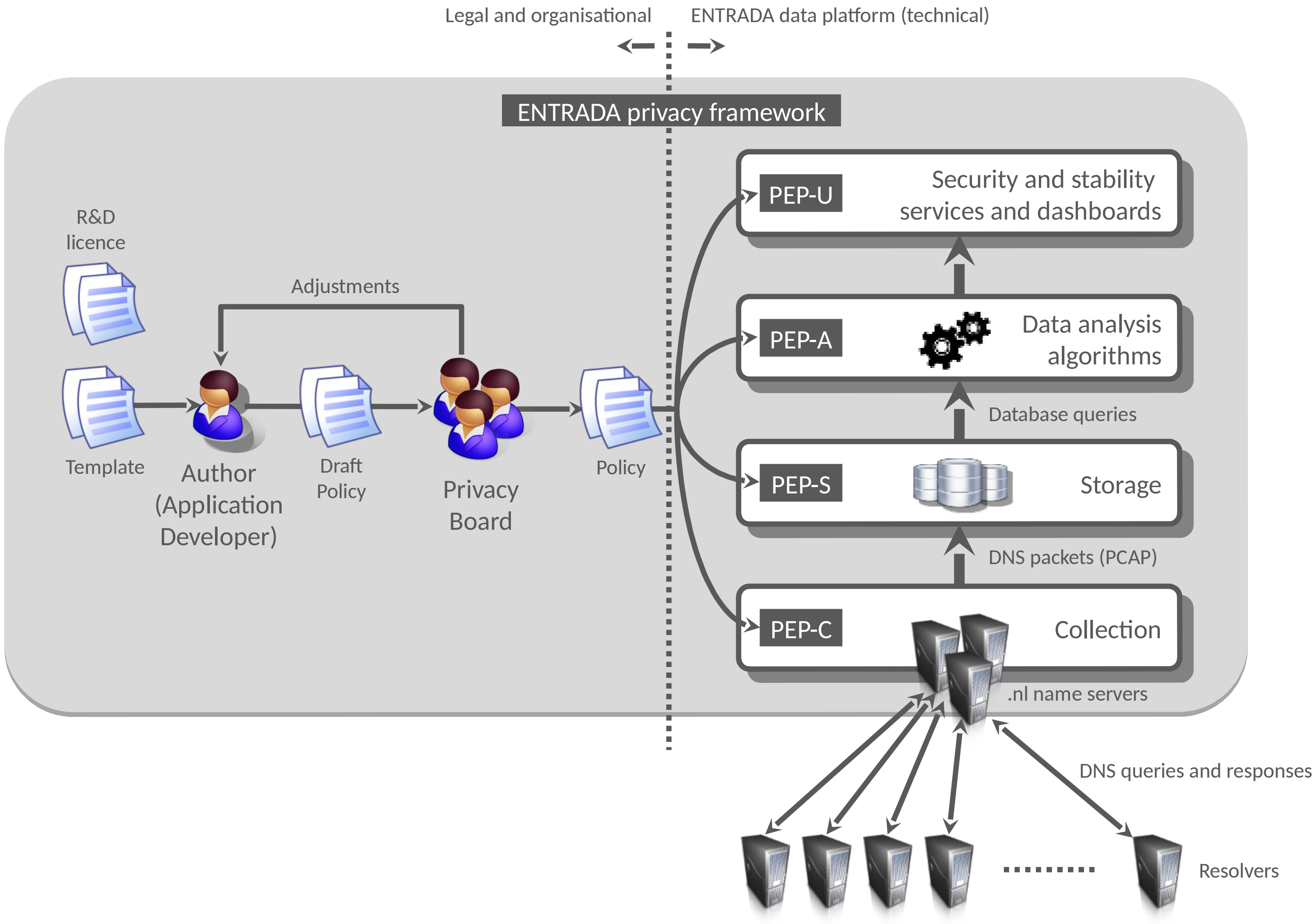
WBP and ENTRADA: IP Addresses in Queried names



Privacy Framework: Requirements

- Purpose limitation
  - Per type of use of the data (i.e. per application)
- Verifiable
  - Transparency
- Simple
- Extensible

Privacy Framework: Overview



Privacy Framework: Policies

- One policy per application
- Policy describes:
  - Purpose
  - Data that is used
  - Filters on the data
  - Access to the data
  - Type of application (Research vs. Production)
  - Other security measures

Privacy Framework: Data Filters at PEPs

- Anonimization
- Pseudonimization
- Aggregation
- Etc.

Privacy Framework: Privacy Board

- Reviews and approves policies
- Members:
  - Legal
  - Technical
  - Organisational
- Publishes approved policies

Privacy Framework: Position Paper

- Currently at https://sidnlabs.nl

Privacy Framework: Conclusions

- DNS Data can be personal data
- DNS Data processing needs privacy-protecting measures
- Not 'just' technical

Privacy Framework: Future work

- Solicit feedback and discussion (hi!)
- Better metrics for 'public' vs 'private' resolvers
- How to incorporate the policy system when sharing data
- Keep eye on new laws (EU Data protection regulation, for one)
- Apply the framework to other types of data

Got questions?

Jelte Jansen

jelte.jansen@sidn.nl

sidn.nl | sidnlabs.nl