A Privacy Framework for (DNS) big data

Jelte Jansen

CENTR R&D, October 16 2017

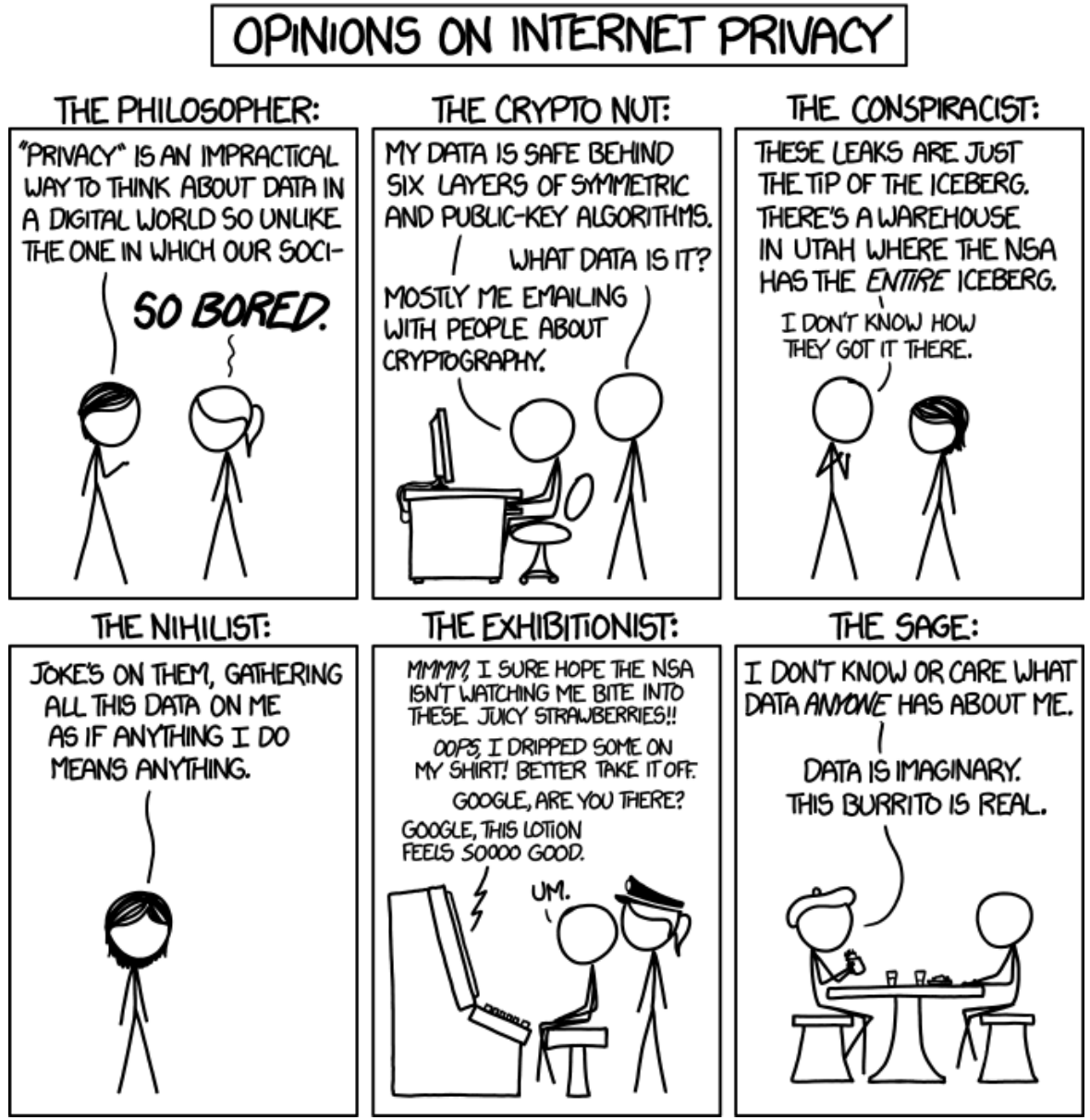
SIDN LABS

Overview

- Introduction
- Big data:

Balance privacy and security?

- GDPR
- Our approach

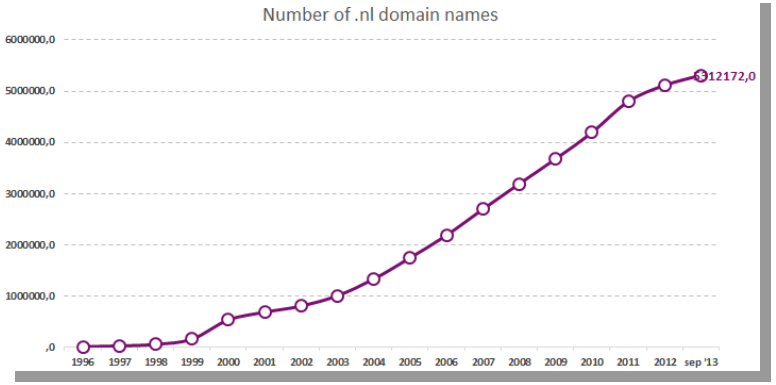


Source: http://xkcd.com

SIDN LABS

SIDN

- ".nl" (Registry of the Netherlands)
- > 5.5M domain names, > 1.600 registrars
- > 1.300.000.000 DNS queries per day
- Private organisation with public task



SIDN Labs

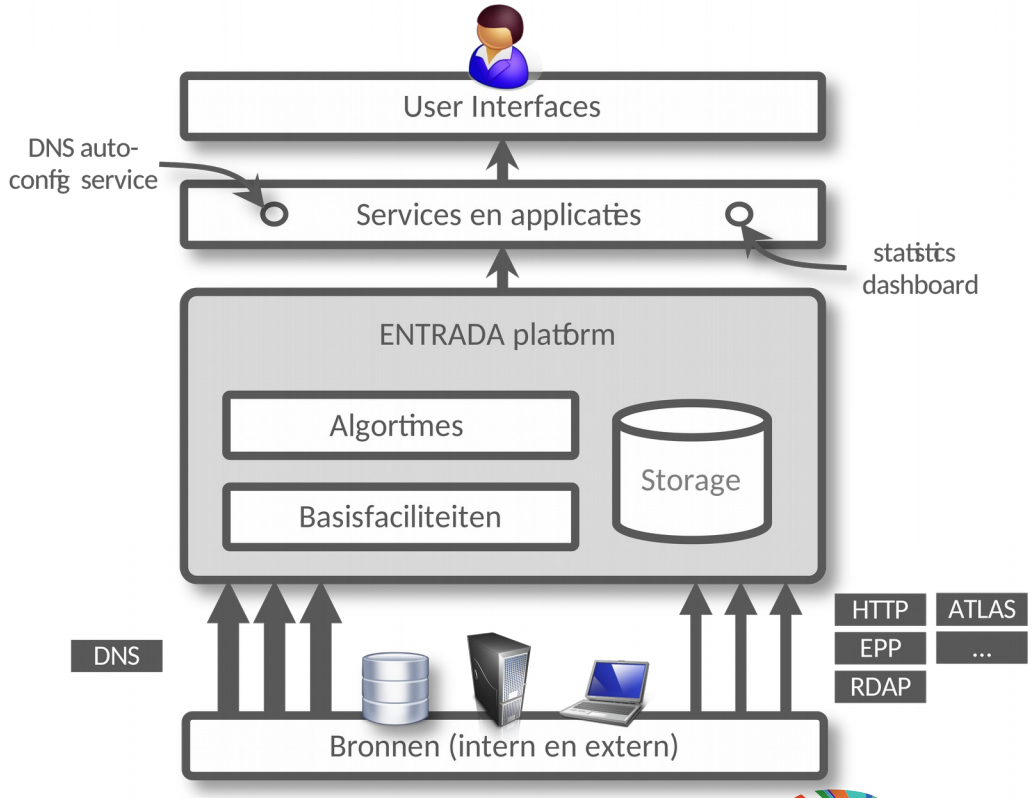
- R&D team SIDN
- Improve services of SIDN
- Center of expertise
- Improve security of the Internet in the Netherlands
- Facilitate external research

Domain names and abuse

- Malware
- Botnets
- Spam
- DDoS
- Etc.

ENTRADA: DNS Big Data Platform

- ENhanced Top-level domain Resilience through Advanced Data Analysis
- Purpose: Create applications and perform research to:
  - Safeguard stability of '.nl'
  - Improve security of (Dutch) Internet
  - Detect botnets and abuse
- What about privacy?



SIDN LABS

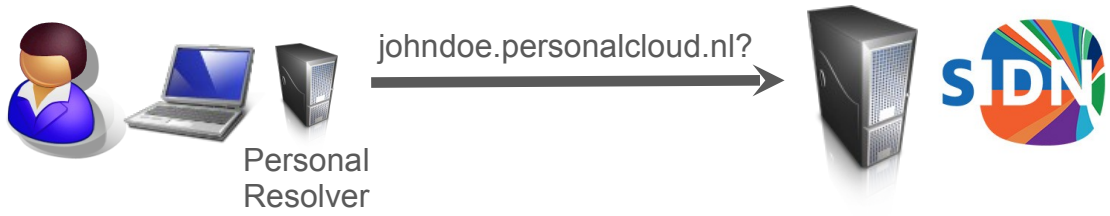
Personal data?

- Sometimes (direct queries, domain name that is queried)
- Sometimes not (shared resolvers, general queries)

Personal data?

- Sometimes (direct queries, domain name that is queried)

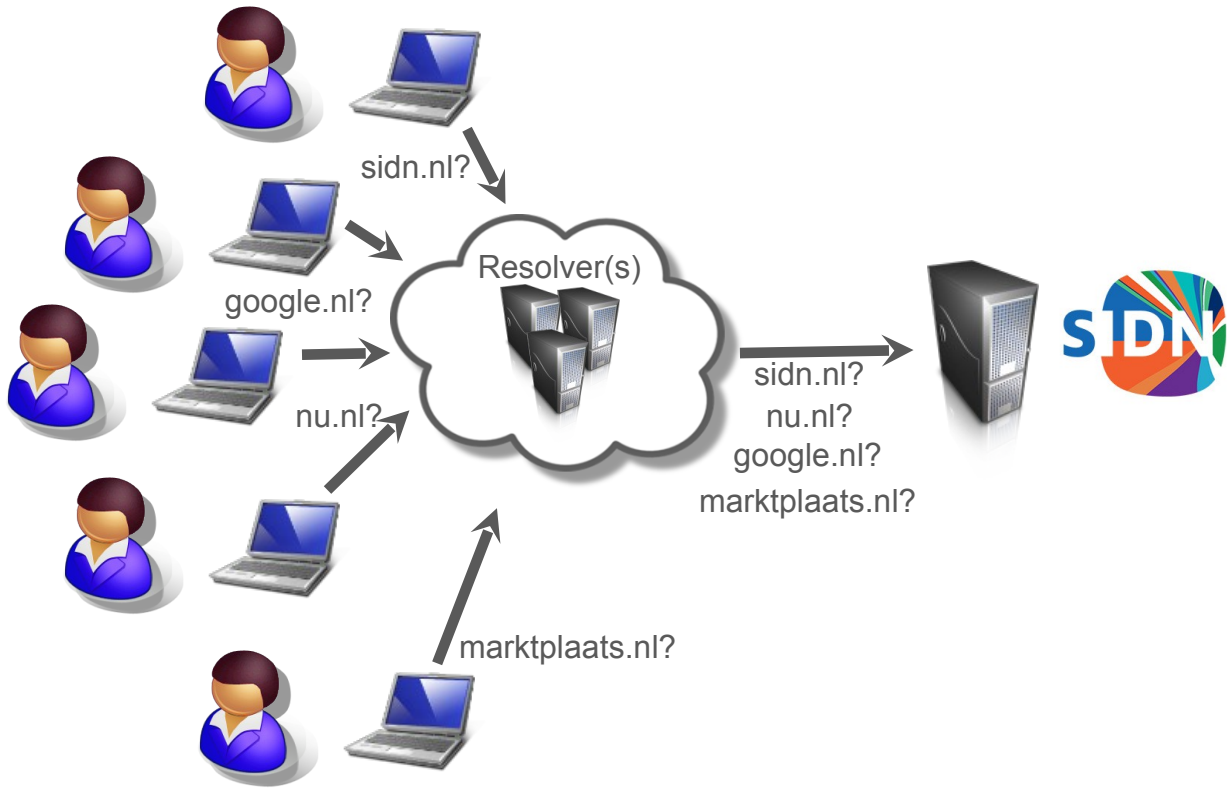
- Sometimes not (shared resolvers, general queries)



Personal data?

- Sometimes (direct queries, domain name that is queried)

- Sometimes not (shared resolvers, general queries)



Needed data differs per application

- Example: Detecting botnets
  - Queried domain name not important
  - But IP Address is

- Example: measuring DDoS patterns
  - IP Address not relevant
  - Query name and type is

Needed data differs per application

- “Keep it all, we just might need it”
  - Does not adhere to purpose limitation
  - Or data minimization
  - Or data retention

- “Remove all IP addresses and qnames”
  - Er, yeah, no.

GDPR Important points

- Lawful basis for processing (art. 6)
  - Informed consent hard to obtain with DNS data
  - Public or legitimate interest can be appropriate
- Keep a record of all processing (art. 30)
  - What, who, why, retention, security
- Keep privacy by design in mind when setting up systems
  - Right to be forgotten (art. 17)
  - Communication of personal data breach (art. 34&35)
- Keep a law expert nearby (art. all of them)
  - Consider a DPO

SIDN LABS

SIDN's approach:
Privacy framework and implementation

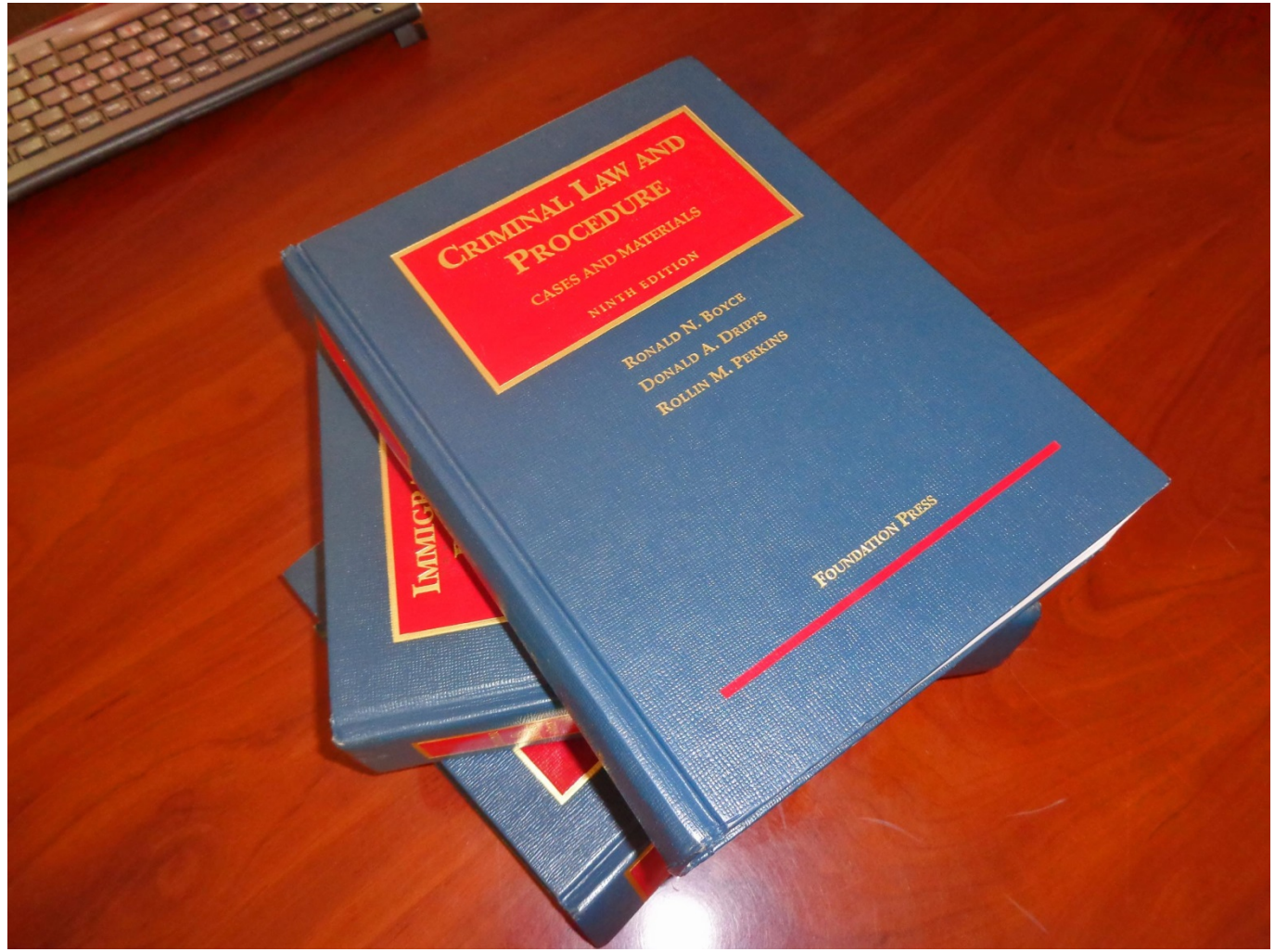
Multidisciplinary approach

- Technical
  - Filtering / Aggregation
  - Hard retention limit
  - Data silos
- Judicial
  - Dutch Data Protection law
  - EU Data Protection Regulation
- Organizational
  - Privacy board
  - Privacy Policies

SIDN LABS

Multidisciplinary approach

- Technical
  - Filtering / Aggregation
  - Hard retention limit
  - Data silos
- Judicial
  - Dutch Data Protection law
  - EU Data Protection Regulation
- Organizational
  - Privacy board
  - Privacy Policies



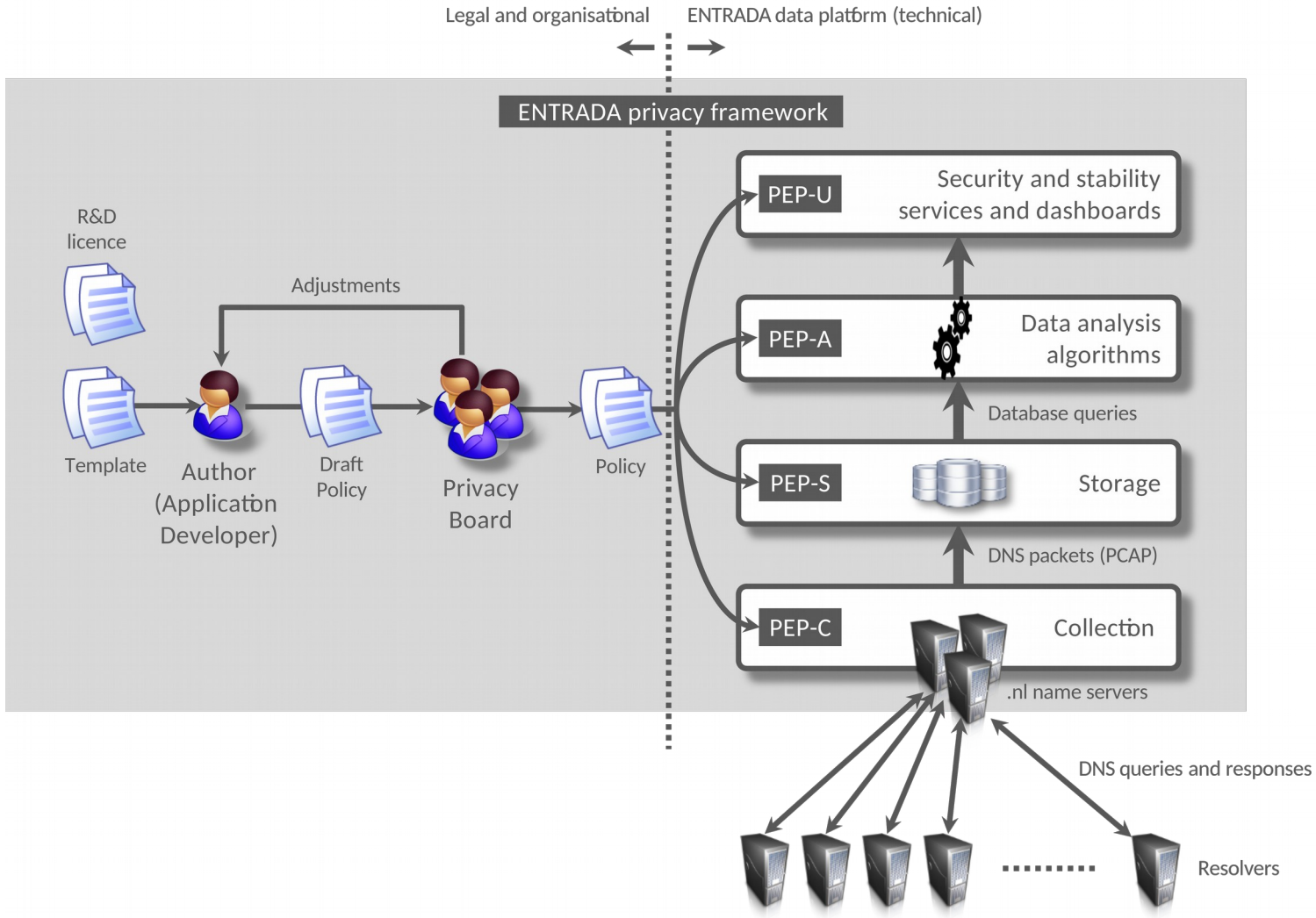
SIDN LABS

Multidisciplinary approach

- Technical
  - Filtering / Aggregation
  - Hard retention limit
  - Data silos
- Judicial
  - Dutch Data Protection law
  - EU Data Protection Regulation
- Organizational
  - Privacy board
  - Privacy Policies

SIDN LABS

Privacy framework: overview



Privacy Board and framework

- Board with 3 members
  - Chair
  - Legal expert
  - Technical expert

- Processes for use
  - Instructions for submitting policies
  - Instructions for evaluating policies

Procedures Privacy Board

- Developer or researcher submits policy by filling in template
  - Goal
  - What data
  - Which filters
  - Which security measures
  - Etc.
  - Mostly process-focused, not much law

- Board evaluates policy
  - Checklist based on Position Paper and Data Protection Law
  - Is used data necessary?
  - Are the filters applicable and effective?
  - Do we have a good basis for the use of the data?
  - Etc.
  - More law-focused

Example policy summary: JTIE (original proposal)

- Project: Joint threat intelligence Enrichment
  - Cooperation between SIDN and Fraudehelpdesk (FHD)
  - Goal: to combat fraud by marking abuse domains

- **Process:**
  - People report fraud e-mail to FHD
  - FHD takes all domain names in e-mail and sends them to SIDN
  - SIDN responds with following data:
    - Number of queries for those domains in last 7 days
    - Date of registration
    - Registrant's country of residence
    - Registrar

Example policy summary: JTIE (problems)

- Board evaluated proposal, found some problems
  - Proportionality

- **Specifics:**
  - Country of residence was not nessecary
  - Registrar name can be PII!
  - Included many domain names SIDN couldn't use anyway

Example policy summary: JTIE (currently)

- Project: Joint threat intelligence Enrichment
  - Cooperation between SIDN and Fraudehelpdesk (FHD)
  - Goal: to combat fraud by marking abuse domains

- **Process:**
  - People report fraud e-mail to FHD
  - FHD takes all 2nd-level domain names for .nl in e-mail and sends them to SIDN
  - SIDN responds with following data:
    - Number of queries for those domains in last 7 days
    - Date of registration
    - Pseudonym of registrar

Example policy summary: ENTRADA general

- Project: ENTRADA
  - 'Main' policy for use of ENTRADA
  - Goal: to enable research & development by collecting DNS traffic data

- All DNS traffic data is stored!

- But with some measures:
  - Extra-strict security measures and instructions for people with access
  - Data can't be used or shared without separate policy that specifies goal, ground and filters
  - Data is hard-anonimized after 18 months

Position paper

https://www.sidnlabs.nl/downloads/whitepapers/SIDN_Labs_Privacyraamwerk_Position_Paper_V1.4_ENG.pdf

SIDN LABS

Questions

Jelte Jansen

Research Engineer, SIDN Labs

@twitjeb

jelte.jansen@sidn.nl

@sidnlabs

sidn.nl | sidnlabs.nl