RegCheck: risicobeoordeling van nieuwe .nl-registraties

Thijs van den Hout
SIDN TechTalk

12 april 2023

SIDN Labs = onderzoek

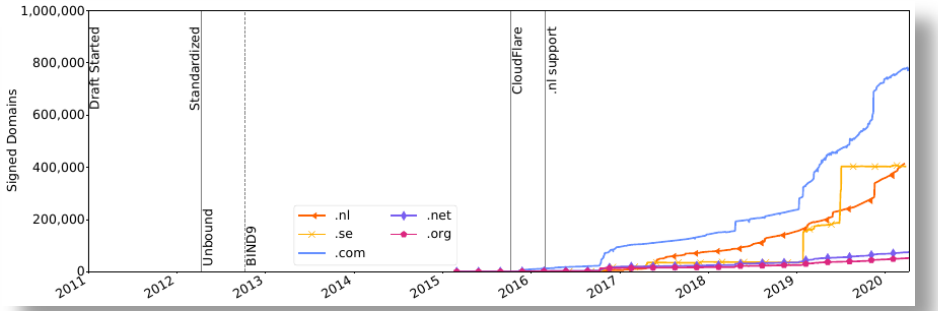
- Doel: het verhogen van de betrouwbaarheid en veiligheid van de internetinfrastructuur, voor .nl en Nederland in het bijzonder.

- Strategieën:
  - Toegepast technisch onderzoek (internetmetingen, design, prototypen, evaluatie)
  - Resultaten publiek beschikbaar en bruikbaar maken voor verschillende doelgroepen
  - Samenwerken met universiteiten, infrastructuur beheerders en andere onderzoek labs.

- Drie onderzoeksgebieden: network security (DNS, NTP, BGP), domain name & IoT security, emerging internet security technologies

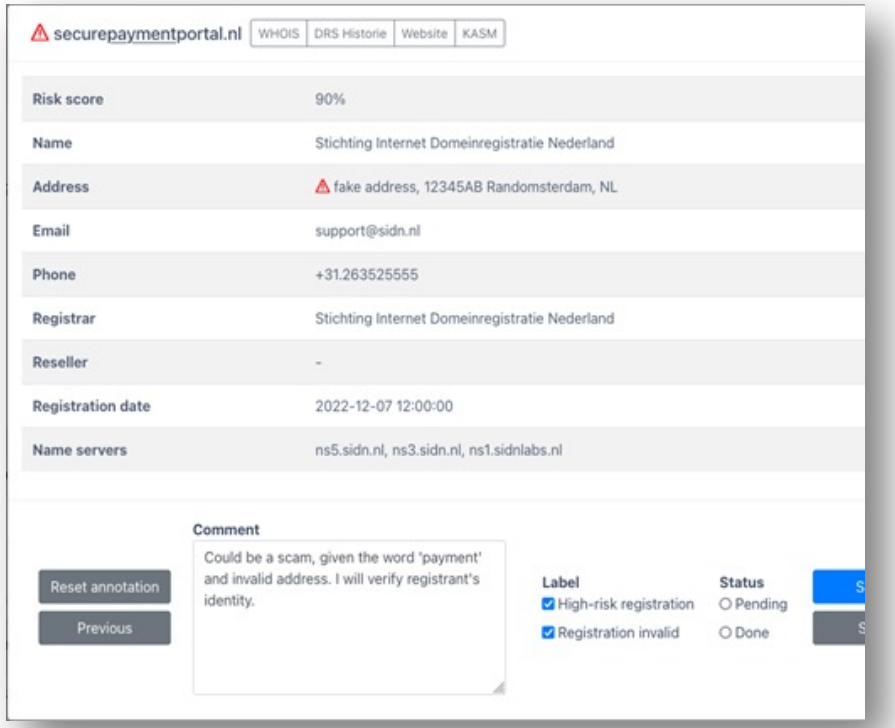
.nl = the Netherlands
~17M inwoners
6.2M domeinnamen
3.4M DNSSEC-signed
2.5B DNS queries/day
8.6B NTP queries/day

Voorbeeld projecten

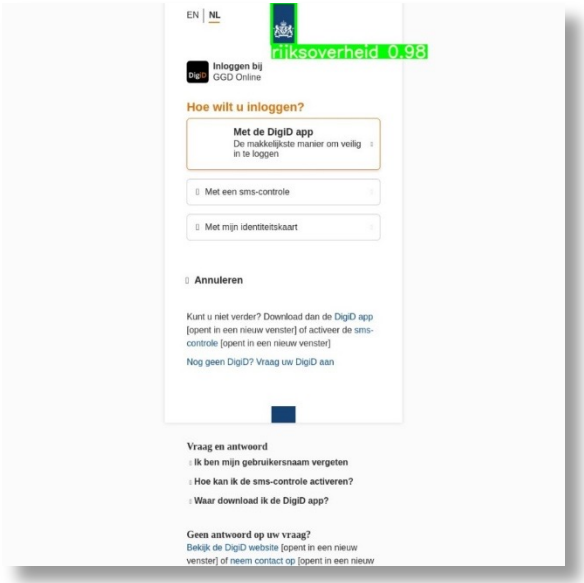


Measuring the deployment of newly standardized DNSSEC algorithms

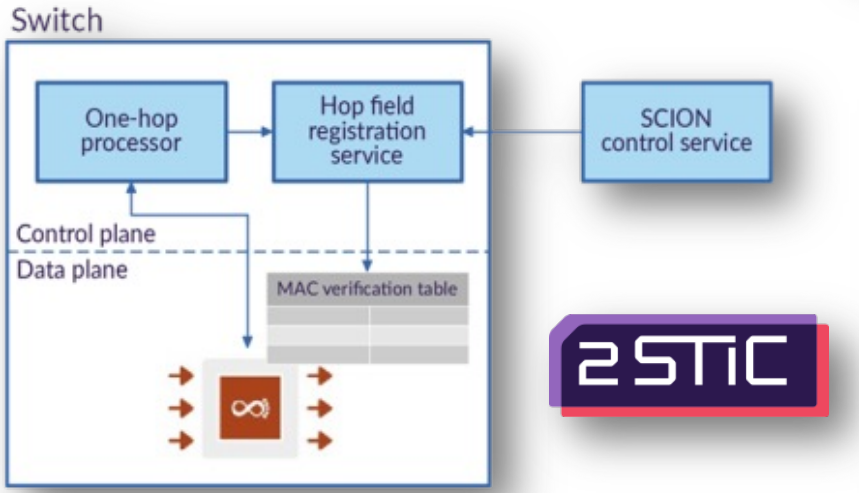
Provide well-managed and secure time services



Detecting high-risk domain name registrations

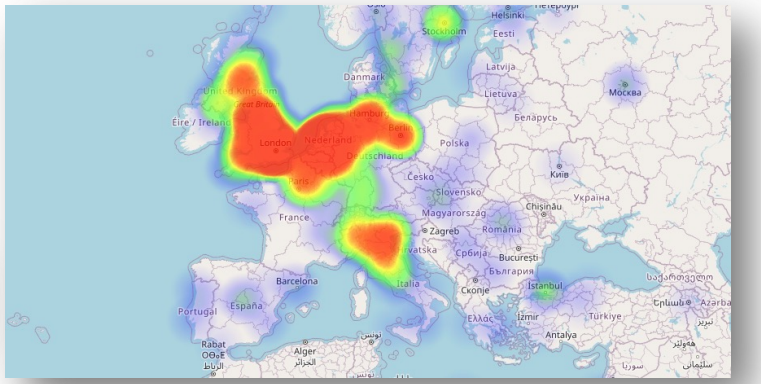


Logo detection technology to identify malicious .nl websites



2STiC

Experimenting with secure future networks and programmable networks



Optimize anycast routing

SIDN LABS

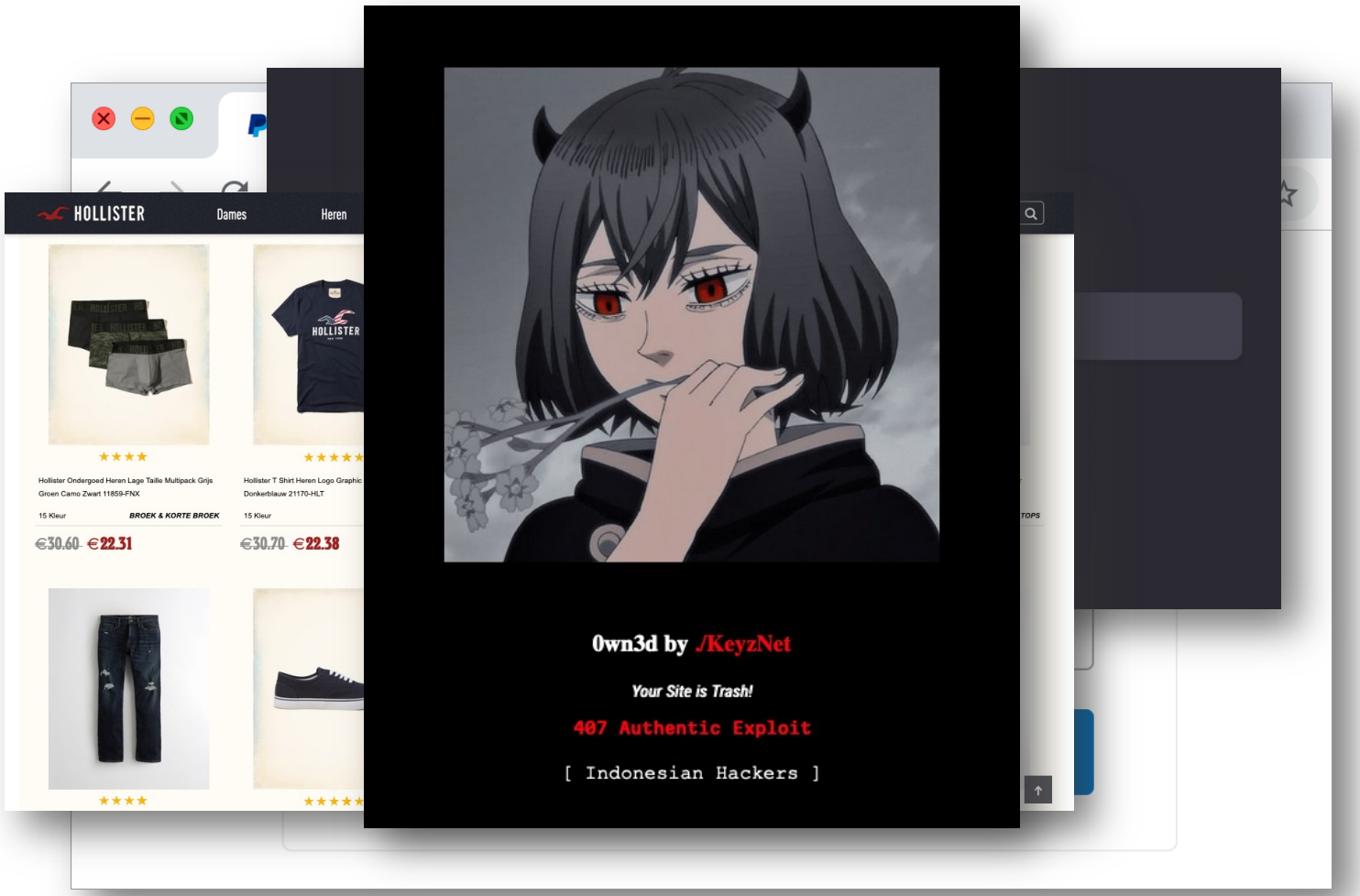
Wie heeft er ooit een domeinnaam geregistreerd?

Wie heeft er ooit een .nl domeinnaam geregistreerd?

Wie is ooit een phishing website tegengekomen?

Domeinnaam abuse

- Typo squatting
- Phishing website
- Malware host
- Fake webshop
- Bekladde website



rabo-betaalverzoek.nl

bakkerijhans.nl

arendslogistiek.nl

arendslogistiek.nl

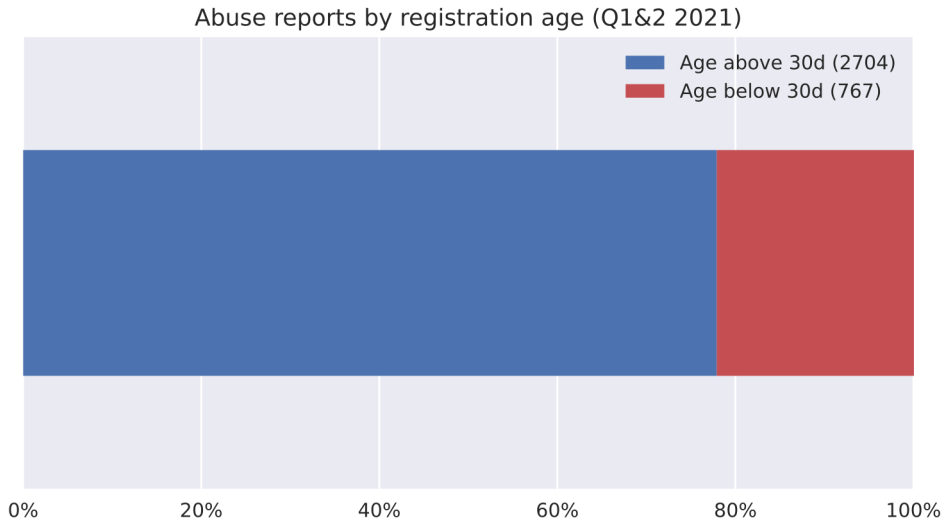
Registratiegegevens:
Naam: sdfgi dfgsda
Email: adsservices20d02fa137@protonmail.com
Datum: 2023-04-12 02:33
Adres: Yuying rd 528231, Guangdong, China

marktpaats.nl

Aanleiding RegCheck

- SIDN staat voor een veilig .nl-domein

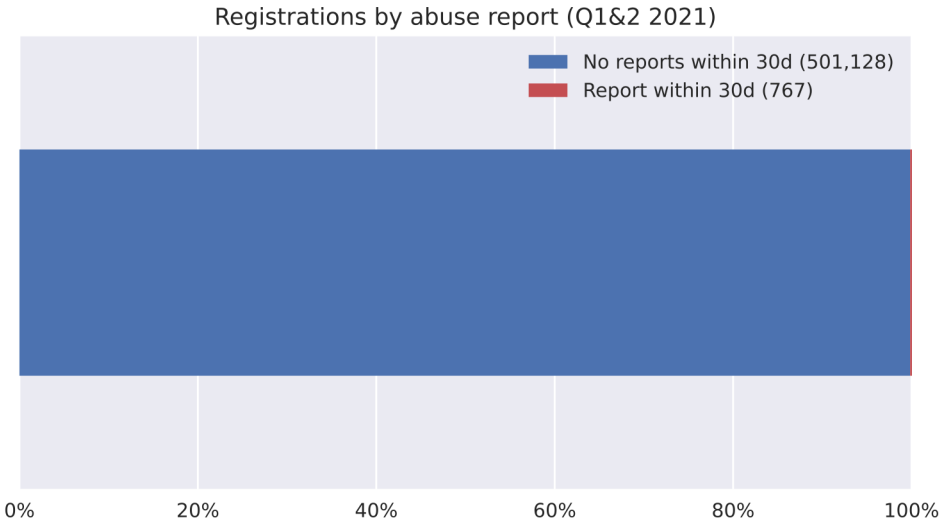
- Malafide intenties soms vrij duidelijk
  - Risicovolle domeinnaam
  - Ongeldige houdergegevens

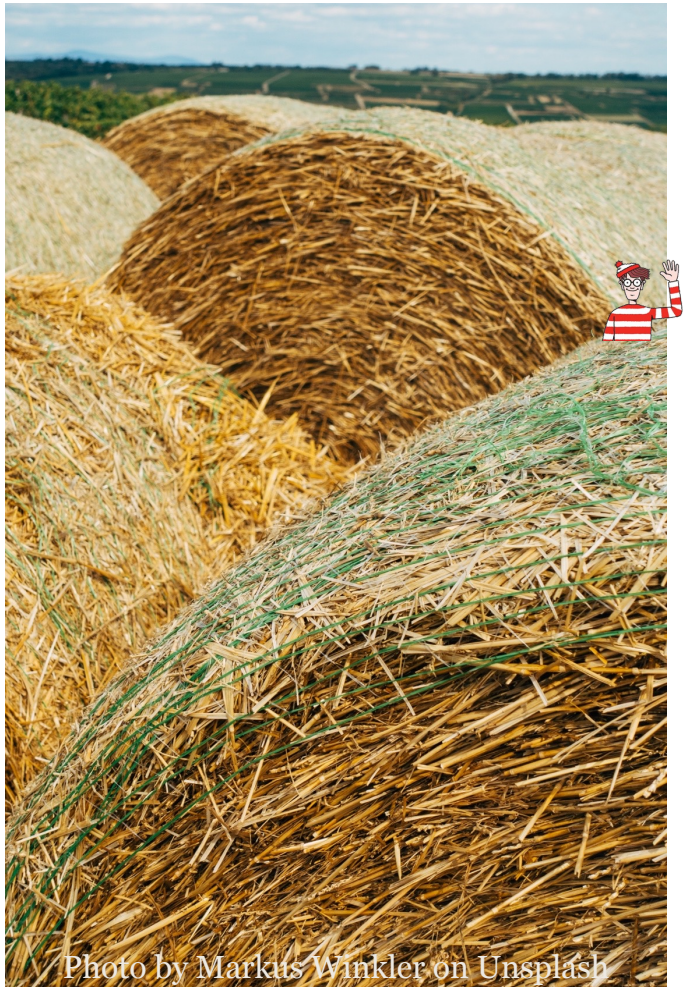


Dus… Waarom wachten tot de abusemelding?

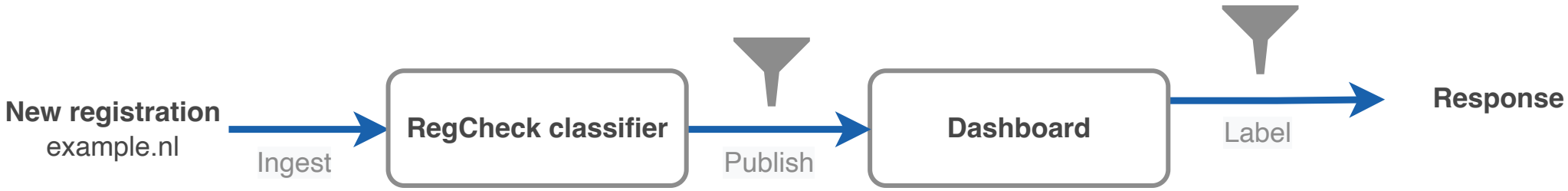
- Proactief valideren van domeinnaam registraties maakt .nl veiliger

- Handmatige alles controleren is geen optie:
  - Ruim 2.000 registraties per dag
  - Slechts 3 registraties hiervan binnen 30 dagen op Netcraft (0,15%)

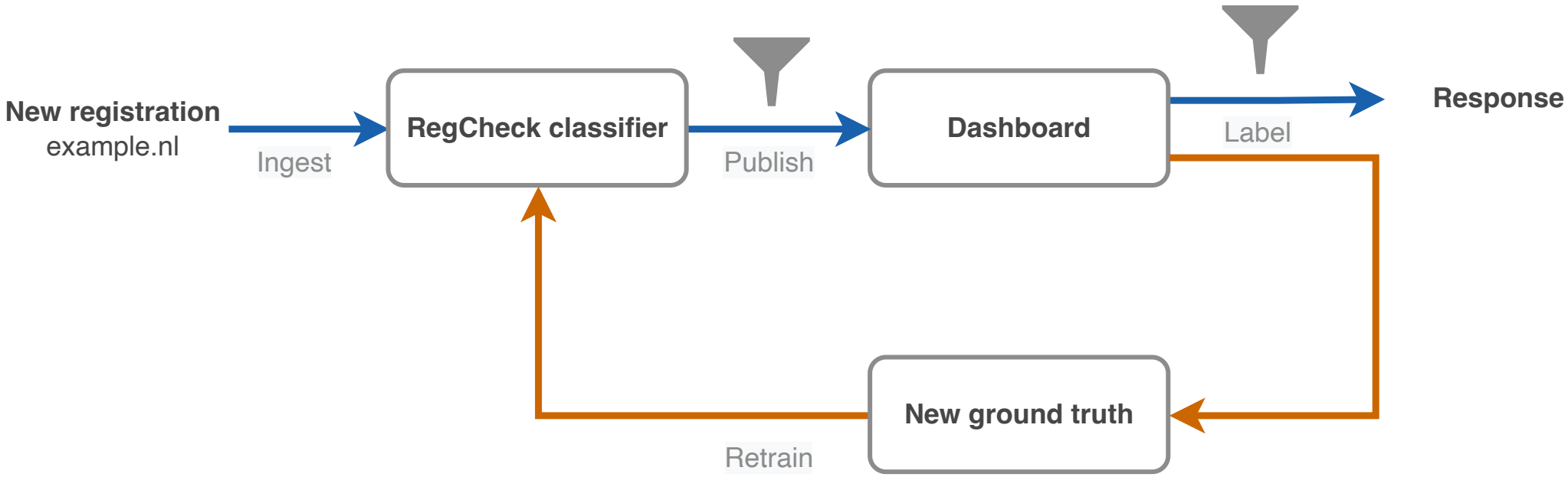




RegCheck: potentieel malafide registraties filteren



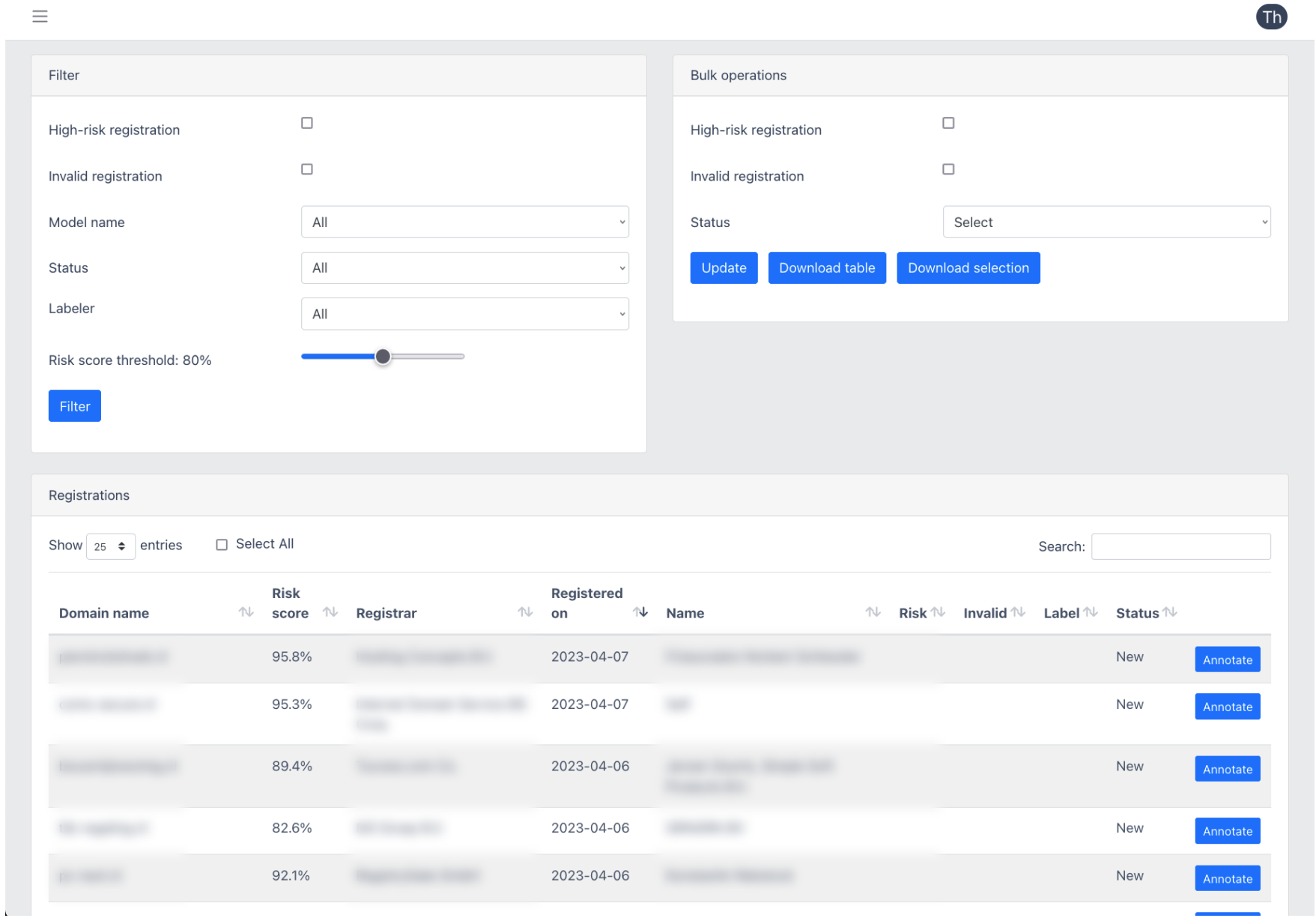
RegCheck: potentieel malafide registraties filteren



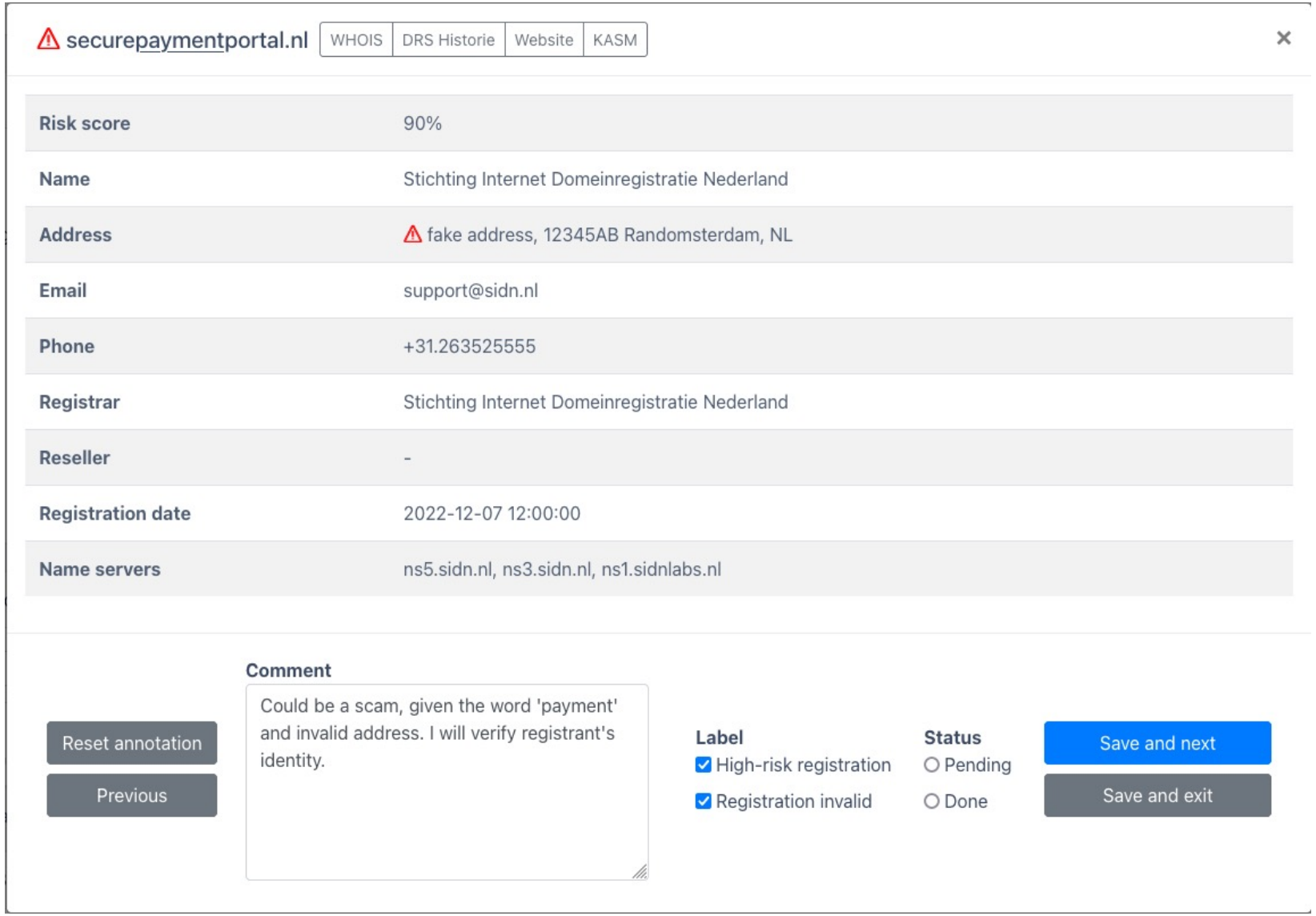
Verantwoord toepassen van ML

- Human-in-the-loop
- Simpele en interpreteerbare modellen
- Samenwerken en publiceren

Dashboard


Th
Filter
High-risk registration
Invalid registration
Model name
All
Status
All
Labeler
All
Risk score threshold: 80%
Filter
Bulk operations
High-risk registration
Invalid registration
Status
Select
Update
Download table
Download selection
Registrations
Show 25 entries
Select All
Search:
Domain name
Risk score
Registrar
Registered on
Name
Risk
Invalid
Label
Status
95.8% 2023-04-07 New Annotate
95.3% 2023-04-07 New Annotate
89.4% 2023-04-06 New Annotate
82.6% 2023-04-06 New Annotate
92.1% 2023-04-06 New Annotate

Dashboard


securepaymentportal.nl
WHOIS
DRS Historie
Website
KASM
Risk score
90%
Name
Stichting Internet Domeinregistratie Nederland
Address
fake address, 12345AB Randomsterdam, NL
Email
support@sidn.nl
Phone
+31.263525555
Registrar
Stichting Internet Domeinregistratie Nederland
Reseller
-
Registration date
2022-12-07 12:00:00
Name servers
ns5.sidn.nl, ns3.sidn.nl, ns1.sidnlabs.nl
Comment
Could be a scam, given the word 'payment' and invalid address. I will verify registrant's identity.
Reset annotation
Previous
Label
High-risk registration
Registration invalid
Status
Pending
Done
Save and next
Save and exit

Berekening risico

- Risicofactor: kenmerk dat risico verhoogt (21 momenteel)

- Risicofactoren individueel bekijken
  - Voordeel: risicoscores interpreteren
  - Nadeel: non-lineaire verbanden modelleren onmogelijk (geen probleem in de praktijk)

- Regel-gebaseerde classifier en machine learning classifier
  - Registry onafhankelijke code
  - Op basis van scikit-learn interfaces

SIDN LABS

Regel-gebaseerd

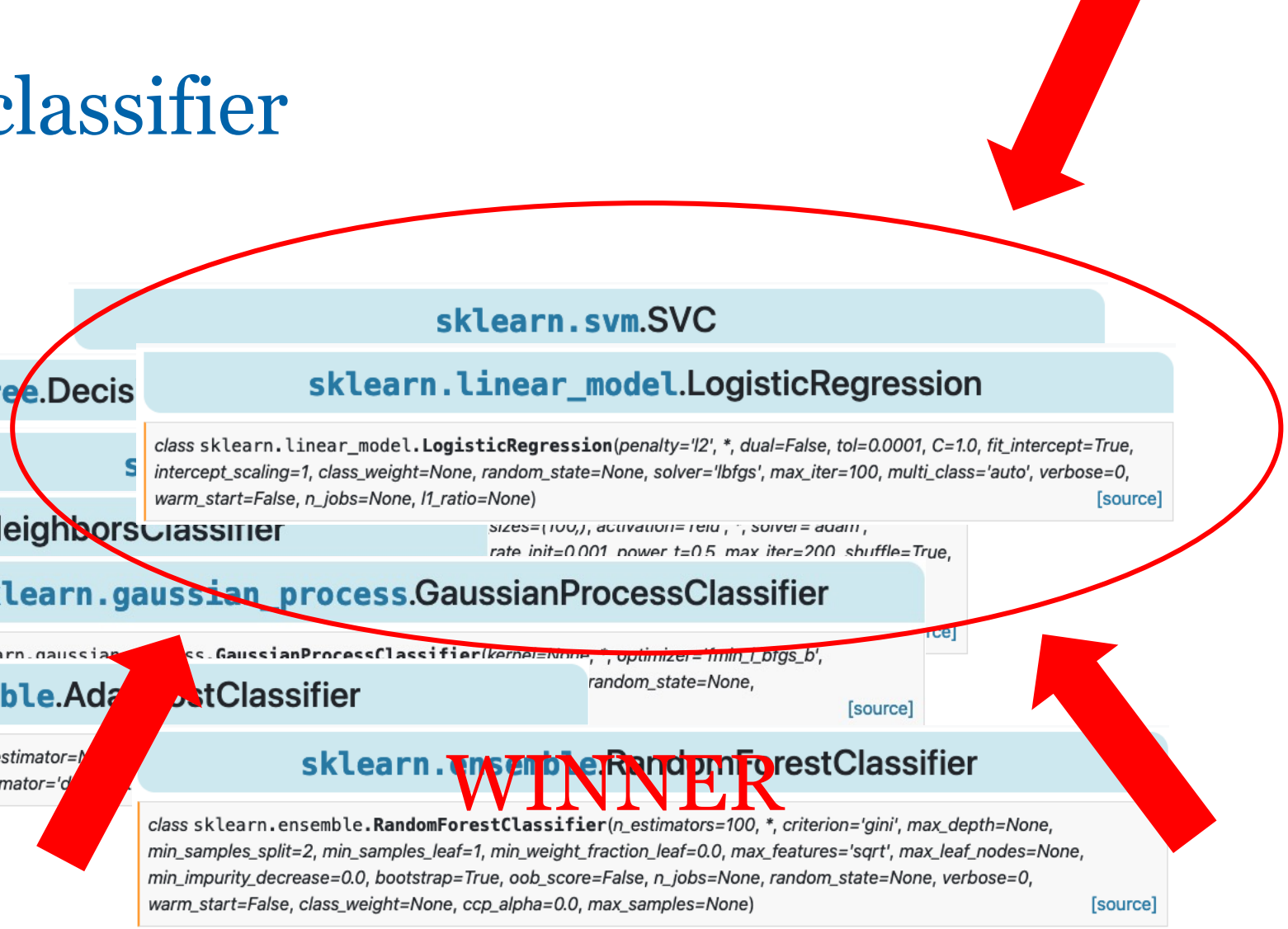
- Kennis-gedreven features:
  - “domeinnaam bevat meer dan één streepje”
  - “domeinnaam is ‘s nachts geregistreerd”
  - “domeinnaam bevat ‘bank’”

- Risicoscore is een optelsom

- Interpreteerbaar

Machine learning classifier

- Lineaire verbanden
- Interpreteerbaar
- Ve
- Non-lineaire verbanden
- Moeilijk te interpreteren
- Vergelijkbare performance



Offline en online resultaten

	Machine learning	Rule based
Recall	48%	9%
PPV (precision)	22%	0.55%

Table 1: RegCheck's results on historical data (August to November 2022).

	Machine learning
Registrations	43k
High-risk classifications	181 (0.4%)
True positives	38 (21%)

Table 2: RegCheck's results on new registrations (17 November to 8 December 2022).

Les 1: Spreek dezelfde taal

Photo by Tim Schmidbauer on Unsplash

- Probleemdefinitie [1]
  - Domeinnamen verdacht van abuse detecteren?
  - Domeinnamen met ongeldige houdergegevens detecteren?

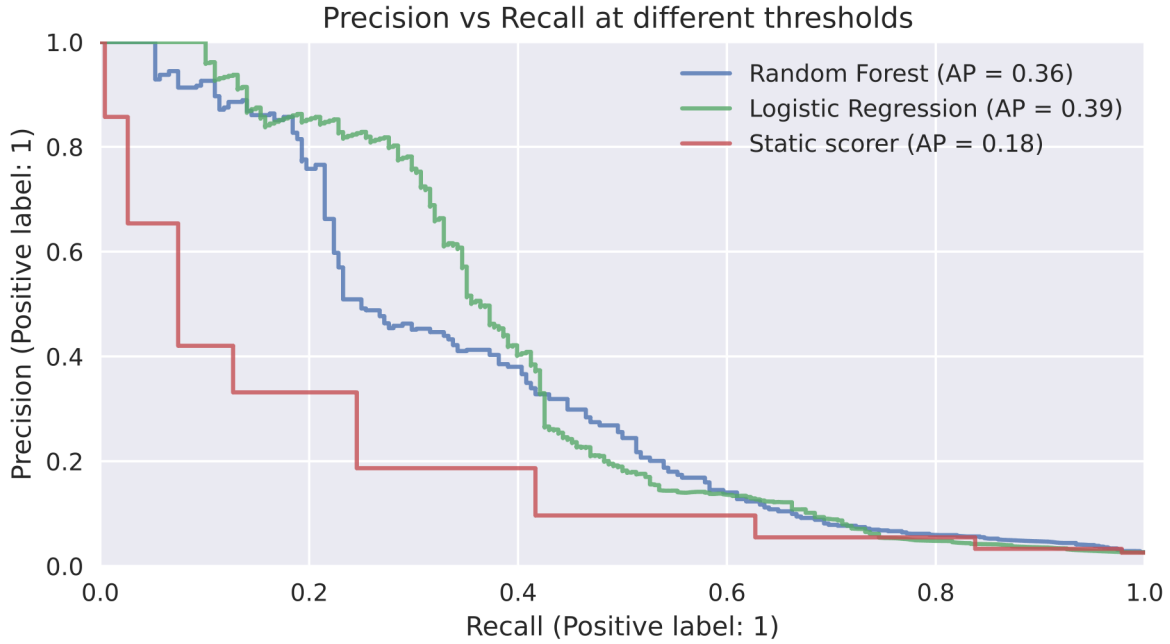
- Goed gedefiniëerde uitkomsten
  - Wat betekent “verdacht”?
  - Maakt iedereen dezelfde beslissingen?

[1] Problem Formulation and Fairness - https://arxiv.org/abs/1901.02547

SIDN LABS

Les 2: Benchmark ML tegen niet-ML baseline

- Complexe ML algoritmes zijn niet altijd beter dan simpele alternatieven (bijv. uitlegbaarheid, hogere kosten) [2]
- Een niet-ML baseline brengen voor- en nadelen van ML beter in kaart



[2] Measuring the predictability of life outcomes with a scientific mass collaboration - https://www.pnas.org/doi/10.1073/pnas.1915006117

Les 3: Discussieer technische keuzes en hun impact


Photo by Clay Banks on Unsplash

- Technische keuzes hebben invloed op processen en beleid (bijv. drempelwaardes, features)
- Verantwoord gebruik van machine learning alleen mogelijk als beslissingen expliciet worden gediscussieerd en bekeken van meerdere kanten

3 lessen

Spreek dezelfde taal

Benchmark met niet-ML baseline

Discussieer technische keuzes & impact

Plannen voor 2023+

- Prototype blijven gebruiken en verbeteren
- Mogelijk automatisch houderonderzoek starten
- Gezamenlijke ontwikkeling en evaluatie met DNS Belgium (.be)
- Andere registries helpen door code te delen

Photo by Jess Bailey on Unsplash

Q&A

https://www.sidnlabs.nl/nieuws-en-blogs/risicobeoordeling-van-nieuwe-nl-registraties-met-behulp-van-regcheck

thijs.vandenhout@sidn.nl
thymen.wabeke@sidn.nl