*Drs. J.R.P. Jansen en dr. ir. C.E.W. Hesselman**

Ervaringen met privacybeheer voor DNS-'big data'-toepassingen

146

Trefwoorden:
privacybeheer, operationalisering, big data, Domain Name System

We bespreken de invoering van en de uitbreidingen op het privacyraamwerk waarmee we op een privacybewuste manier netwerkverkeer kunnen analyseren, om daarin vroegtijdig bedreigingen zoals phishingsites en botnets te ontdekken. De uitbreidingen zijn gebaseerd op onze ervaring met het raamwerk in de afgelopen anderhalf jaar. In deze periode hebben we het ingevoerd en actief gebruikt voor de opslag en analyse van de berichten die we verwerken als beheerder van het .nl-deel van het Domain Name System (DNS), het systeem in de internetinfrastructuur dat domeinnamen vertaalt naar IP-adressen. We bespreken ook onze belangrijkste 'lessons learned' om andere organisaties te helpen die een vergelijkbaar privacyraamwerk in willen voeren.

1 Inleiding

Systemen die grote hoeveelheden data analyseren ('big data') bieden veel mogelijkheden, bijvoorbeeld voor het verhogen van de internetveiligheid. Ze vereisen echter ook maatregelen om de persoonlijke levenssfeer van gebruikers te beschermen, bijvoorbeeld via transparantie over de persoonsgegevens die systemen verwerken en verantwoording van de gebruikte analysemethoden [1].[1]

Een voorbeeld van een 'big data'-toepassing is het analyseren van internetverkeer om daarin vroegtijdig bedreigingen voor eindgebruikers op te sporen, zoals phishingsites en botnets. Dit kan bijvoorbeeld op basis van het verkeer uit het Domain Name System (DNS) [2]. Het DNS vertaalt domeinnamen (bijvoorbeeld www.example.nl) naar IP-adressen (bijvoorbeeld 94.198.159.35) zodat browsers, e-mailprogramma's en andere internettoepassingen de server die bij een domeinnaam hoort kunnen benaderen. Deze vertaalslag is nodig, omdat het internet werkt met IP-adressen, die voor mensen vaak lastig te onthouden zijn.

Het DNS is onderdeel van de 'publieke kern van het internet' [8], samen met bijvoorbeeld het internetprotocol (IP) en het systeem dat berichten over het internet routeert. Het is een wereldwijd systeem waarvan een groot aantal partijen ieder een deel beheren. Als beheerder van de internetextensie van Nederland zijn wij verantwoordelijk voor het .nl-deel van het DNS. We verwerken in die rol dagelijks meer dan 1,3 miljard DNS-berichten die vragen om de vertaling van een .nl-domeinnaam naar een IP-adres en beheren ook de database met daarin de 5,6 miljoen geregistreerde .nl-domeinnamen.

In ons artikel van december 2014 in dit tijdschrift beargumenteerden we dat DNS-berichten persoonsgegevens kunnen bevatten en dat dit een afdwingbaar mechanisme vereist om de privacy van gebruikers te beschermen [2]. De aanpak die we voorstelden was een multidisciplinair privacyraamwerk dat juridische, technische en organisatorische aspecten van privacybeheer verenigt. Ons privacyraamwerk stelt verwerkers in staat om (i) een systematische en transparante afweging te maken tussen het vinden van bedreigingen in DNS-verkeer aan de ene kant en de privacy van internetgebruikers aan de andere kant en (ii) de vereiste privacymaatregelen af te kunnen dwingen in het technische systeem dat de analyses uitvoert. Het privacyraamwerk is cruciaal voor de vertrouwensrol die SIDN voor Nederland vervult als beheerder van de .nl-extensie, maar is ook geschikt voor andere jurisdicties en andere soorten netwerkverkeer.

We hebben het raamwerk gelijktijdig ontwikkeld met ENTRADA (ENhanced Top-level domain Resilience through Advanced Data Analysis), het technische systeem waarmee we DNS-berichten opslaan en analyseren [10, 11, 12]. ENTRADA is ontworpen om grote hoeveelheden DNS-berichten op te slaan en snel te analyseren [10]. Wij zetten het systeem in voor .nl, maar het is ook geschikt voor andere extensies en andere typen DNS-beheerders. ENTRADA is een experimenteel systeem dat ontwikkeld is door SIDN Labs, ons R&D-team. De software is open source en te downloaden van entrada.sidnlabs.nl.

In dit artikel bespreken we de uitbreidingen op ons privacyraamwerk, gebaseerd op onze ervaring met de invoering en het actief gebruik ervan binnen SIDN de afgelopen anderhalf jaar. We bespreken ook onze 'lessons learned' om andere organisaties te helpen die een vergelijkbaar privacyraamwerk willen invoeren.

Het raamwerk is algemener dan ons specifieke gebruik ervan, maar omdat we ook willen laten zien hoe we het

* **Jelte Jansen is research engineer bij SIDN Labs, jelte.jansen@sidn.nl. Cristian Hesselman is manager SIDN Labs, cristian.hesselman@sidn.nl.**

1 De cijfers tussen [ ] verwijzen naar par. 7 Referenties.

in de praktijk toepassen, geven we eerst een kort overzicht van wat DNS-data zijn en welke DNS-berichten we opslaan voor analyse (par. 2). Daarna geven we een samenvatting van ons privacyraamwerk (par. 3) en bespreken we de uitbreidingen daarvan (par. 4). We sluiten af met geleerde lessen (par. 5) en conclusies (par. 6).

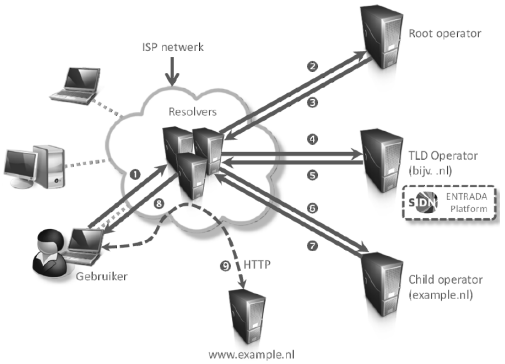
We danken Karin Vink, Lilian van Mierlo en Simon Hania voor het lezen van conceptversies van dit artikel en het geven van waardevolle feedback.

2 DNS-data

De data waarvoor we ons privacyraamwerk hebben ontwikkeld zijn berichten uit het Domain Name System (DNS). Het DNS is een wereldwijd gedistribueerd systeem dat een domeinnaam vertaalt ('resolving' in DNS-jargon) naar een IP-adres waar de server op draait (par. 2.1), bijvoorbeeld van www.example.nl naar 94.198.159.35. Het DNS bestaat uit miljoenen systemen, verspreid over de hele wereld, die worden beheerd door een groot aantal verschillende partijen. Het DNS bevat ook andere typen data, maar het opzoeken van IP-adressen is de meest gebruikte functie. SIDN beheert het .nl-deel en een gedeelte van het berichtenverkeer voor .nl slaan we op in ENTRADA (par. 2.2).

2.1 *DNS resolving*

Figuur 1 laat zien hoe resolving werkt als een gebruiker op een URL klikt of invult in de adresbalk van zijn browser. Als voorbeeld-URL gebruiken we http://www.example.nl/. Het deel tussen de slashes (www.example.nl) is de domeinnaam en verwijst naar de server waar de site op draait.



Figuur 1. DNS-resolving

Om www.example.nl te vertalen naar het IP-adres van de server stuurt de machine van de gebruiker een DNS-bericht naar een zogenaamde 'resolver' (stap 1 in figuur 1). Dit is een machine die meestal van de internetserviceprovider (ISP) is waar de gebruiker zijn of haar internetaansluiting afneemt. Deze resolver zoekt voor de browser de domeinnaam op in het mondiale DNS, te beginnen bij een vaste groep van zogenaamde 'rootservers' (stap 2). Deze rootservers verwijzen de resolver in het geval van www.example.nl door naar de nameservers van .nl (stap 3). De resolver neemt vervolgens contact op met een .nl-nameserver (stap 4), die de resolver op zijn beurt doorverwijst naar de nameservers van example.nl (stap 5). De resolver stuurt daarop een DNS-bericht naar de nameserver van example.nl (stap 6), die uiteindelijk het IP-adres van www.example.nl kent en dat terugstuurt naar de resolver (stap 7). Tot slot antwoordt de resolver met het IP-adres van www.example.nl naar de browser (stap 8), waarna de browser via HTTP de webpagina ophaalt van www.example.nl (stap 9).

De nameservers van .nl ontvangen de berichten van stap 4 en versturen die van stap 5. De binnenkomende berichten zijn een deel van het daadwerkelijke aantal DNS-berichten die clients versturen (een 'sample'), omdat resolvers gebruikmaken van caching. Dit betekent dat een resolver een antwoord een tijdlang in zijn geheugen opslaat en vragen van andere clients voor het IP-adres van dezelfde domeinnaam beantwoordt zonder opnieuw contact op te nemen met de nameservers in het DNS (de resolver slaat dan stappen 2 t/m 7 over). Caching is een belangrijk onderdeel van het DNS, omdat het een van de mechanismes is waarmee het systeem kan blijven groeien zonder dat het slechter gaat presteren (schaalbaarheid).

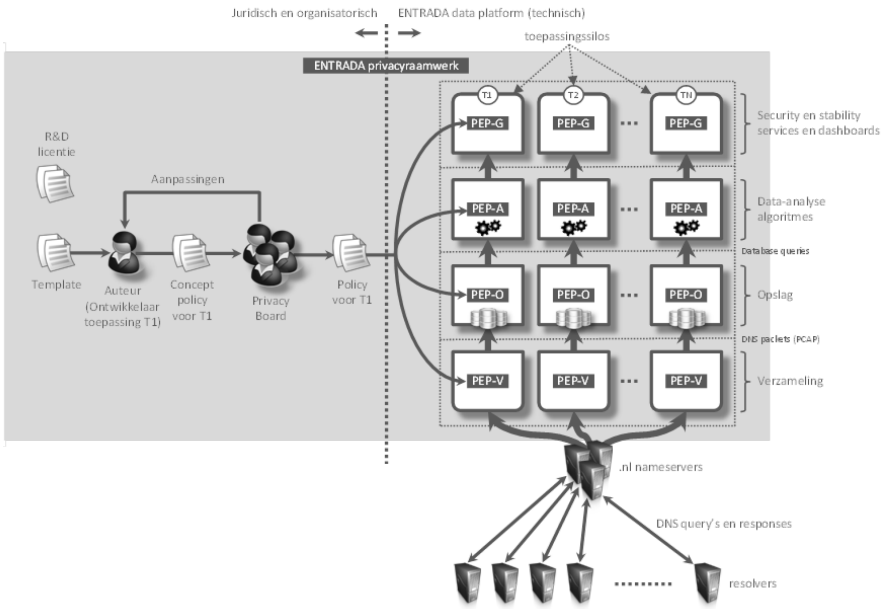
De infrastructuur die we voor .nl beheren bestaat uit 73 nameservers die geografisch verspreid zijn, zowel binnen Nederland als daarbuiten.

2.2 *Opslag en analyse met ENTRADA*

De berichten die we opslaan in ENTRADA zijn die van de stappen 4 en 5 in figuur 1. Gemiddeld gaat dat om ongeveer 15.000 query's per seconde, wat neerkomt op 39 miljard query's (en responses) per maand. Als we deze data opslaan (inclusief IP- en ethernetheaders), dan gaat het onder normale omstandigheden om ongeveer 60 gigabyte per dag per nameserver.

ENTRADA draait op dit moment 20 maanden als experimenteel systeem in het netwerk van SIDN Labs. Het slaat data op van twee van de .nl-nameservers en bevat op het moment van schrijven 180.758.031.498 requests en responses, wat overeenkomt met 8 terabyte aan opgeslagen data. We slaan ieder bericht maximaal 18 maanden op, zodat we 6 maanden de tijd hebben om onderzoek uit te voeren op een jaar aan DNS-berichten. Na 18 maanden aggregeren we de data en verwijderen we de oorspronkelijke berichten. We verwijzen naar [1] voor een uitgebreidere toelichting op deze retentiepolicy, die ook in par. 3 nog aan de orde komt.

We gebruiken ENTRADA voor een aantal experimentele toepassingen, zoals botnetdetectie en detectie van domeinnamen die potentieel voor phishing worden gebruikt [10, 11, 12]. Ook gebruiken we het systeem voor het monitoren van de effecten van beleidswijzigingen [13], voor het delen van statistieken over .nl (via



Figuur 2. ENTRADA-privacyraamwerk

stats.sidnlabs.nl) en om toegepast en wetenschappelijk onderzoek bij derden te ondersteunen.

3 Privacyraamwerk

Het doel van ons privacyraamwerk is (i) ons in staat te stellen op een transparante manier een systematische afweging te maken tussen het vinden van bedreigingen en afwijkingen in de DNS-berichten die we verwerken voor .nl aan de ene kant en de privacy van .nl-gebruikers aan de andere kant en (ii) om vereiste privacymaatregelen technisch af te dwingen.

Figuur 2 geeft een overzicht van het privacyraamwerk dat we in dit hoofdstuk kort samenvatten. We gaan in op de belangrijkste concepten: persoonsgegevens (par. 3.1), privacypolicy's (par. 3.2), de privacyboard (par. 3.3) en policy enforcement points (par. 3.4). De componenten rechts van de verticale stippellijn vormen het technische systeem (ENTRADA). Voor een uitgebreidere uitleg van ons raamwerk en de ideeën erachter verwijzen we naar [1].

3.1 *Persoonsgegevens*

In een vroegtijdig stadium van de ontwikkeling van ENTRADA realiseerden wij ons dat een deel van de data die ENTRADA zou gaan verwerken persoonsgegevens kan bevatten. Het gaat daarbij om twee typen persoonsgegevens: de IP-adressen van resolvers (zie par. 2.1) en de domeinnamen die resolvers opzoeken namens eindgebruikers.

De IP-adressen van resolvers hoeven geen persoonsgegevens te zijn, omdat resolvers vaak namens een groep gebruikers optreden als onderdeel van de dienstverlening van een ISP. Het is echter ook mogelijk dat individuen zelf een resolver hebben, of dat zij ons rechtstreeks bevragen, bijvoorbeeld als ze besmet zijn met een botnet dat een eigen resolver aan boord heeft. Is dat het geval, dan is een IP-adres wel degelijk een persoonsgegeven. Omdat we niet van tevoren weten of een resolver optreedt voor een groep of een individu, behandelen we alle IP-adressen in principe als persoonsgegeven.

Met opgezochte domeinnamen gaan we op dezelfde manier om. Deze zijn in de meeste gevallen zeer algemeen, maar kunnen potentieel ook behoorlijk specifiek, of zelfs uniek zijn. Vergelijk bijvoorbeeld het opzoeken van 'google.nl' met het hypothetische 'klant461.mijn.bank.nl'. Het aantal van dit soort domeinnamen gaat in de toekomst overigens afnemen, omdat resolvers minder informatie gaan versturen [14]. Zo ontvangen wij straks op onze nameservers alleen nog het verzoek voor 'bank.nl' en niet meer voor 'klant461.mijn.bank.nl'. We verwachten echter dat het nog enige tijd gaat duren voordat alle resolvers op het internet deze vorm van dataminimalisatie gebruiken.

We verwijzen naar [1] voor de uitgebreide juridische onderbouwing waarom we IP-adressen van resolvers en opgezochte domeinnamen behandelen als persoonsgegevens.

3.2 *Privacypolicy's*

Het centrale concept van het privacyraamwerk is de privacypolicy. Een privacypolicy definieert welke data een ENTRADA-toepassing verwerkt, met welk doel deze verwerking plaatsvindt, en welke filters daarbij op de

persoonsgegevens toegepast worden. Een voorbeeld van een onderzoek waar een policy voor zou zijn is een onderzoek van een universiteit naar patronen van domeinnamen die opgevraagd worden bij een landelijk computersecurity-incident. Voor een dergelijk onderzoek zijn IP-adressen niet nodig, en deze zouden dan uitgefilterd worden. Een voorbeeld van een dienst die een privacypolicy zou krijgen is het detecteren van met een virus geïnfecteerde computers bij internetserviceproviders, om deze aan de ISP te melden zodat zij hun netwerk kunnen opschonen. In dat geval is juist alleen het IP-adres van belang, en niet zozeer welke domeinnamen er vanuit dat adres werden opgevraagd. De policy zou in dat geval omschrijven dat de data alleen met de betreffende ISP gedeeld worden, en dat de filters de opgevraagde domeinnamen zouden verwijderen.

Een filter is een operatie op de persoonsgegevens (bijvoorbeeld pseudonimisering of aggregatie) om te voldoen aan de beginselen van proportionaliteit en subsidiariteit door het vermijden van een overmatige of onnodige verwerking van persoonsgegevens. De filters vormen een essentieel element in ons privacyraamwerk, omdat ze policy's op een toetsbare wijze afdwingen door middel van technische maatregelen (zie par. 3.4).

Een applicatie krijgt alleen de data die de bijbehorende policy beschrijft, en zonder een privacypolicy krijgt een applicatie geen toegang tot persoonsgegevens. Een privacypolicy zorgt er daarnaast voor dat zaken als doelbinding, grondslag, en beveiliging toetsbaar zijn, zowel intern als publiek. In het ENTRADA-privacyraamwerk heeft iedere toepassing een privacypolicy.

Een privacypolicy is een tekstdocument waarvan de structuur lijkt op die van het 'meldingsformulier verwerking persoonsgegevens' van de Autoriteit Persoonsgegevens (AP) [15]. De belangrijkste onderdelen zijn:

- Doel: beschrijving van het doel van de ENTRADA-toepassing die de persoonsgegevens verwerkt. Een voorbeeld hiervan is een applicatie die automatisch malafide botnets detecteert binnen het dataverkeer van de .nl-zone, om de veiligheid van het .nl-domein te verbeteren.
- Persoonsgegevens: de persoonsgegevens die de toepassing verwerkt. Bij ENTRADA betreft dit IP-adressen, opgevraagde domeinnamen, of beide (zie par. 3.1).
- Filters: een beschrijving van de gegevensfilters die bij deze policy horen, de omstandigheden waar ze gebruikt worden, en de persoonsgegevens waar ze op toegepast worden. Voorbeelden van filters zijn pseudonimisatie of het volledig verwijderen van persoonsgegevens.
- Retentie: de periode dat we de persoonsgegevens die nodig zijn voor een applicatie bewaren. Na de bewaartermijn zal het ENTRADA-platform de gegevens verwijderen of anonimiseren.
- Toegang: vermelding van de personen of systemen die toegang hebben tot de gegevens, en onder welke voorwaarden deze toegang mogelijk is. Als het om andere systemen gaat die toegang hebben moet de policy ook beschrijven welke veiligheidsmaatregelen er worden genomen, en vermelden hoe de toegang technisch plaatsvindt.
- Type: hier zijn twee opties mogelijk: onderzoek of productiedienst. Privacypolicy's voor onderzoek vereisen mogelijk minder strenge maatregelen, omdat het soms nog niet geheel duidelijk is welke gegevens precies nodig zullen zijn. Daarbij worden dan wel strengere eisen gesteld aan de toegang tot en het delen van de gegevens. Wanneer het om een productiedienst gaat zal precies bekend zijn welke gegevens de dienst nodig heeft, en kan de minimalisatie van gegevenstoegang strenger toegepast worden.
- Andere beveiligingsmaatregelen: een beschrijving van beveiligingsmaatregelen die niet in de eerdere secties vielen, indien van toepassing.

Policy-auteurs gebruiken de ENTRADA-policysjabloon om een nieuwe policy te schrijven. De sjabloon zorgt ervoor dat ENTRADA-policy's dezelfde structuur hebben en inhoudelijk gelijkvormig zijn. Dit vergemakkelijkt het schrijven van een policy, de evaluatie daarvan door de privacyboard (zie par. 3.3) en de publicatie van de policy. Voorbeelden van policy-auteurs zijn ontwikkelaars van ENTRADA-toepassingen en onderzoekers.

Het privacyraamwerk is ontworpen op infrequent gebruik, gewoonlijk alleen bij het ontwikkelen van een nieuwe ENTRADA-toepassing of een aanpassing daarvan. Het invullen van de policysjabloon is om die reden bijvoorbeeld een handmatige actie. Daarnaast zijn we van mening dat enig denkwerk op het gebied van privacy belangrijk is en dat een al te automatische opzet daar niet aan bijdraagt.

3.3 *Privacyboard*

De privacyboard is het orgaan dat verantwoordelijk is voor het beoordelen van de privacypolicy's. De board beoordeelt of de privacypolicy toepasbaar is. Daarbij oordeelt de board onder andere of het doel welbepaald is, of de grondslag juist is, of de gebruikte persoonsgegevens echt nodig zijn voor het gestelde doel, en of de filters en beveiligingsmaatregelen afdoende zijn.

De privacyboard beoordeelt ook of het doel van de toepassing de middelen rechtvaardigt. De board maakt daarbij een afweging tussen de bijdrage van de ENTRADA-toepassing aan de stabiliteit en veiligheid van .nl aan de ene kant en de privacy van .nl-gebruikers aan de andere kant.

Als weerspiegeling van de combinatie van technische, juridische, en organisatorische maatregelen is de privacyboard van SIDN opgericht met deelname van een technisch expert, een juridisch expert, en iemand van

het managementteam om de organisatorische aspecten te overzien.

3.4 *Policy enforcement points*

Een Policy Enforcement Point (PEP) is een softwarecomponent in het ENTRADA-platform dat privacypolicy's technisch afdwingt, in het bijzonder de filters van een privacypolicy (zie par. 3.2). Deze filters kunnen op verschillende momenten hun werk doen: voor het verzamelen, voor het opslaan, voor het verwerken (binnen ENTRADA zelf), en voor het doorgeven aan applicaties of onderzoekers. Een voorbeeld van een filter dat we hebben geïmplementeerd is het verwijderen van IP-adressen. Applicaties die deze informatie niet nodig hebben krijgen alleen toegang tot de gefilterde data en kunnen de adressen daarmee ook niet zien.

4 Uitbreidingen

In 2015 hebben we het ENTRADA-privacyraamwerk ingevoerd en actief in gebruik genomen bij SIDN. We hebben op het moment van schrijven 2 policy's afgerond en in werking gesteld en de privacyboard heeft 5 ingediende policy's in overweging. Voor 1 policy bleek dat er geen policy nodig was, omdat er geen persoonsgegevens verwerkt werden.

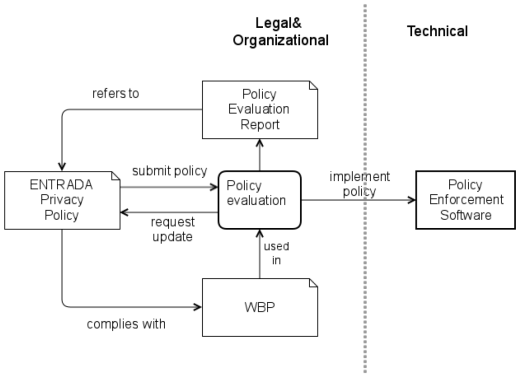
Op basis van onze ervaringen van het afgelopen jaar hebben we het raamwerk onder aanvoering van de privacyboard op enkele punten uitgebreid: we hebben een proces voor het beoordelen van privacypolicy's uitgewerkt (par. 4.1), nieuwe onderdelen toegevoegd aan onze blauwdruk voor privacypolicy's (par. 4.2), de verantwoordelijkheden van de privacyboard uitgebreid (par. 4.3), het concept van een evaluatierapport geïntroduceerd (par. 4.4) en we hebben ons werk en het privacyraamwerk aangemeld bij de AP (par. 4.5).

4.1 *Evaluatieproces*

Figuur 3 geeft een overzicht van het uitgewerkte beoordelingsproces voor privacypolicy's en de samenhang tussen de verschillende concepten uit het raamwerk.

Het proces begint met het schrijven van een privacypolicy door een policy-auteur, zoals een onderzoeker die een nieuw onderzoek wil uitvoeren, of een ontwikkelaar die een nieuwe applicatie voor het ENTRADA-platform wil maken. De policy-auteur schrijft de policy met behulp van de sjabloon uit par. 3.2 en dient die policy in bij de privacyboard.

De privacyboard beoordeelt de policy vervolgens aan de hand van het policy-evaluatierapport (par. 4.4) en informeert de policy-auteur. Bij goedkeuring van de policy gaat de policy-auteur over tot implementatie van de bijbehorende filters in het ENTRADA-platform in de vorm van een PEP (zie par. 3.4). Daarnaast publiceert de privacyboard de policy op ons intranet. Onderdeel hiervan is de publicatie van het evaluatierapport, omdat deze in-



Figuur 3. beoordelingsproces

formatie bevat over waarom de privacyboard van mening is dat de policy acceptabel is. Binnenkort gaan we goedgekeurde policy's ook op onze publieke internetsite publiceren.

Als de board een policy afkeurt, dan ontvangt de indiener het beoordelingsrapport met details over waarom de policy is afgekeurd. De indiener kan vervolgens de policy aanpassen en deze opnieuw indienen. Dit is een iteratief proces dat eindigt als de board de policy goedkeurt of definitief afkeurt. De privacyboard helpt de policy-auteur om de policy te verbeteren indien nodig.

4.2 *Privacypolicy's*

Tijdens de beoordeling van de eerste ingediende policy's kwam de privacyboard tot het inzicht dat er nog informatie ontbrak aan de policyformulieren. We hebben de ontbrekende elementen toegevoegd (zie verderop), waarbij we een afweging hebben gemaakt tussen onze eis van bondigheid en het toevoegen van de extra velden voor een hogere mate van volledigheid.

Privacypolicy's moeten kort en bondig zijn om de belasting voor policy-auteurs (bijvoorbeeld onderzoekers en ontwikkelaars) zo laag mogelijk te houden en hen zo te stimuleren het proces te volgen. Policy's moeten daarnaast kort en bondig zijn om lezers een goed overzicht te geven van welke gegevens voor welke doelen verwerkt worden, inclusief mensen van buiten de organisatie.

Daar staat tegenover dat de policy's wel alle elementen moeten bevatten die de privacyboard nodig heeft om zich ervan te kunnen verzekeren dat de policy aan de eisen van de Wet bescherming persoonsgegevens (Wbp) voldoet, zoals dataminimalisatie en limieten op retentie. Daarnaast dienen policy's duidelijk de technische maatregelen (bijvoorbeeld filters) te beschrijven zodat de privacyboard en het publiek kunnen beoordelen of deze maatregelen afdoende zijn.

De velden die we aan een privacypolicy hebben toegevoegd naast die van par. 3.2 zijn:

- **Grondslag**: de legitieme grondslag van de verwerking van de gegevens, conform artikel 8 Wbp. Dit kan bijvoorbeeld zijn om een contract met de betrokkene te vervullen. In het geval van ENTRADA zal dit meestal gerechtvaardigd belang van ofwel de betrokkene ofwel het publiek zijn. In het geval dit niet zo is, kan er een keuze gemaakt worden tussen de andere grondslagen die de Wbp benoemt.
- **Publicatie/delen**: als de uitvoer van de applicatie gedeeld wordt met derden, dan beschrijft dit onderdeel welke gegevens gedeeld worden, met wie ze gedeeld worden, en onder welke voorwaarden dit gebeurt. Merk op dat we dit onderscheiden van directe toegang tot de gegevens zoals beschreven in de oorspronkelijke policy (par. 3.2).
- **Naam toepassing**: identificeert het onderzoek of de applicatie waar de policy voor geldt.
- **Datum**: datum van het indienen van de policy.

In Appendix A geven we een voorbeeld van een ingevulde privacypolicy, die zowel de oorspronkelijke velden bevat als bovenstaande aanvullingen.

4.3 *Privacyboard*

Bij de introductie van ons privacyraamwerk kwamen we erachter dat de privacyboard extra verantwoordelijkheden nodig heeft om het privacyraamwerk in de organisatie in te regelen. Dit waren:
- Organisatorisch: het uitwerken, documenteren en inrichten van de processen en interne communicatiekanalen (bijvoorbeeld een e-mailadres) van het raamwerk.
- Technisch: het overzien van de haalbaarheid en impact van technische maatregelen van persoonsgegevensbescherming binnen het ENTRADA-platform.
- Juridisch: het inlichten van het College bescherming persoonsgegevens (nu Autoriteit Persoonsgegevens) over de verwerking (zie par. 4.5).

Daarnaast heeft de privacyboard terugkerende functies:
- Het evalueren van ingediende privacypolicy's, en het herevalueren van bestaande policy's. Uiteraard is het evalueren van nieuwe policy's de belangrijkste doorlopende taak van de privacyboard. Omdat projecten en diensten echter een nieuwe richting in kunnen slaan, of geheel kunnen stoppen, moeten de bestaande policy's ook regelmatig geüpdatet worden. Daarbij hoort dan ook een herevaluatie door de privacyboard.
- Het publiceren van actieve privacypolicy's. Omdat transparantie een van de doelen was zullen we niet alleen een globale beschrijving van de gegevensverwerking publiceren, maar elke specifieke policy wordt ook gepubliceerd.
- Het regelmatig herevalueren van het raamwerk zelf. Omdat het raamwerk een nieuw concept is, en omdat wetgeving verandert, zullen we het raamwerk zelf ook regelmatig moeten herevalueren, en waar nodig aanpassen.

4.4 *Beoordelingsrapport*

Het doel van het beoordelingsrapport is vast te leggen waarom de privacyboard een policy heeft goedgekeurd of afgekeurd. Dit is belangrijk voor auteurs om hun policy's te verbeteren, voor de organisatie om een archief aan te leggen, alsook voor het bredere publiek om het rationale achter goedgekeurde policy's te kunnen volgen.

Het beoordelingsrapport is aanvullend op de bijbehorende policy. Sommige van die onderdelen hebben als doel de precieze overweging en uitleg van de privacyboard te beschrijven, en kunnen verwijzingen naar de Wbp bevatten. Andere onderdelen zijn simpelweg ja/nee-antwoorden, die de board heeft toegevoegd om zeker te zijn dat alle aspecten van de policy worden overwogen. De privacyboard beschrijft bij elk onderdeel waarom het van mening is dat aan de eisen voldaan wordt, dan wel wat er ontbreekt of afgekeurd wordt. Als een lezer alleen geïnteresseerd is in welke data voor welk doel verwerkt worden is het lezen van de privacypolicy voldoende. Levert dat vragen op over eventuele afwegingen die zijn gemaakt dan kan deze informatie gevonden worden in het beoordelingsrapport.

Het beoordelingsrapport bevat de volgende onderdelen:
- **Naam**: naam van de geëvalueerde policy.
- **Datum van beoordeling**: datum waarop de policy geëvalueerd is.
- **Wbp van toepassing**: onderbouwing van waarom de privacyboard van mening is dat de Wbp van toepassing is. Dit omvat welke gegevens er gebruikt worden, en waarom het persoonsgegevens zijn, en of het een geautomatiseerde verwerking dan wel een verwerking van persoonsgegevens die opgenomen zijn in een bestand betreft.
- **Doel**: onderbouwing waarom de privacyboard vindt dat het doel van de policy welbepaald, uitdrukkelijk omschreven, en gerechtvaardigd is.
- **Grondslag**: er moet een legitieme grondslag zijn voor de verwerking van persoonsgegevens. De privacyboard beschrijft hier of de grondslag in de policy van toepassing is, en om welk artikel uit de Wbp het hier gaat.
- **Borging doelbinding**: de privacyboard beschrijft hier of het van mening is dat er voldoende maatregelen zijn genomen om er zeker van te zijn dat de gegevens niet gebruikt worden voor een ander doel dan het doel dat in de policy beschreven staat.
- **Bewaartermijn**: de bewaartermijn mag niet langer zijn dan nodig is om het doel van de applicatie of het onderzoek te bereiken. De privacyboard beschrijft hier of de gestelde termijn voldoende onderbouwd en niet te groot is.
- **Minimale set gegevens**: de privacyboard beschrijft hier of het van mening is dat er geen gegevens

worden verwerkt die niet nodig zijn om het doel van de applicatie of het onderzoek te bereiken.
- **Correctheid gegevens**: dit is een controlevraag over welke waarborgen er zijn dat de verwerkte data correct zijn. Aangezien het om data gaat die voortkomen uit de softwarematige rapportages van een dienst zijn deze in principe correct; het zijn de data die gebruikt zijn om de dienst te verlenen.
- **Gegevensverwerkers**: dit is een controlevraag om er zeker van te zijn dat er geen verwerkers zijn die niet in de policy genoemd worden.
- **Beveiliging gegevens**: dit is een controlevraag om er zeker van te zijn dat de veiligheid van de gegevens voldoende is gewaarborgd.
- **Bijzondere persoonsgegevens**: dit is een controlevraag om er zeker van te zijn dat er geen bijzondere persoonsgegevens verwerkt worden.
- **Melding bij AP**: in principe vallen alle applicaties en onderzoeken die gebruikmaken van ENTRADA onder de melding aan de AP (zie par. 4.5). Dit is een controlevraag om er zeker van te zijn dat ook deze verwerking onder deze melding valt.
- **Rechten betrokkene**: dit is een controlevraag om er zeker van te zijn dat de rechten van de betrokkenen, zoals vermeld in de Wbp, niet overtreden worden, inclusief verwijzingen naar de relevante artikelen.
- **Data buiten EU**: dit is een controlevraag om er zeker van te zijn dat er geen persoonsgegevens gedeeld worden met mensen of organisaties buiten de EU zonder dat daar de juiste maatregelen en afspraken genomen en gemaakt zijn.
- **Beoordeling**: dit onderdeel bevat het uiteindelijke oordeel van de privacyboard, en eventuele additionele eisen die de privacyboard stelt om de policy te kunnen goedkeuren. Een voorbeeld van een extra eis is dat er voor het delen van de gegevens een overeenkomst gesloten moet worden.

In Appendix B geven we een voorbeeld van een ingevulde beoordeling.

Het beoordelingsrapport hebben we ontwikkeld door een lijst te maken van alle artikelen uit de Wbp die relevant zijn voor dit onderwerp. We hebben deze artikelen omgezet in een checklist van criteria voor de beoordeling van een policy. Daarnaast hebben we technische overwegingspunten aan deze lijst toegevoegd, om er ook zeker van te zijn dat de technische beveiligingsmaatregelen degelijk en compleet zijn.

Oorspronkelijk waren deze punten simpelweg een lijst van overwegingen voor de privacyboard bij de beoordeling van een policy, maar we kwamen er al snel achter dat het resultaat van het doorlopen van deze lijst zelf ook een nuttige exercitie was: het beschrijft niet simpelweg dat een policy is goedgekeurd, maar ook waarom de privacyboard al dan niet tot deze conclusie is gekomen. Daarom hebben we deze checklist veranderd in een tweede documentsjabloon, die dient als beoordelingsrapport en die door de privacyboard ingevuld wordt.

De privacyboard publiceert het beoordelingsrapport naast de bijbehorende policy, waardoor de policy's zelf kort en bondig kunnen blijven (zie par. 4.2), terwijl verdere redeneringen en verwijzingen naar de Wbp in het beoordelingsrapport gezet kunnen worden. Als een ingediende privacypolicy daarentegen niet wordt goedgekeurd, en derhalve niet gepubliceerd, zit de meerwaarde van deze aanpak erin dat de indiener van de policy precies weet waarom deze is afgekeurd, en er intern een archief van beoordelingen bijgehouden kan worden.

4.5 *Melding AP*

De laatste toevoeging aan ons privacyraamwerk is een melding bij de AP. Het is verplicht om de verwerking van persoonsgegevens te melden bij de AP (artikel 27 Wbp [4]), tenzij deze verwerking vermeld staat in een specifieke lijst uitzonderingen. Deze lijst staat in het Vrijstellingsbesluit Wbp [5], en bevat bijvoorbeeld het verwerken van adresgegevens voor het bijhouden van een ledenbestand van een vereniging, of het uitvoeren van een direct contract met de betrokkene.

Omdat transparantie een belangrijk onderdeel is van ons raamwerk, besloten we om, of onze verwerking nu onder het vrijstellingsbesluit valt of niet, een melding te doen bij de AP. Onze melding bij de AP is te vinden op [6].

De meldingsplicht zal bij de invoering van de Algemene Verordening Gegevensbescherming komen te vervallen, maar we zijn vooralsnog uitgegaan van bestaande wetgeving.

5 Lessons learned

In het afgelopen jaar hebben we een aantal belangrijke lessen geleerd die (i) hebben geleid tot de uitbreiding van ons raamwerk (zie par. 4) en (ii) die mogelijk relevant zijn voor andere organisaties die een vergelijkbaar privacyraamwerk willen invoeren, al dan niet op basis van die van ons.

Allereerst hebben we ervaren dat het opstellen en in gebruik nemen van een privacyraamwerk samenging met een verhoging van het algemeen bewustzijn omtrent privacy binnen SIDN. Vanwege veranderende wetgeving is dit sowieso al van belang voor elke organisatie, maar ook daarbuiten heeft een privacyraamwerk een sterk bewustzijn nodig omtrent de verwerking van persoonsgegevens.

Ook hebben we gemerkt dat de privacyboard snel een centraal aanspreekpunt werd binnen SIDN en dat de board meer taken heeft dan alleen het evalueren van policy's (zie par. 4.3). Bijvoorbeeld: een initiële neventaak die al snel van groter belang bleek voor de privacyboard was het informeren van de organisatie over privacy, en het beantwoorden van vragen die er binnen de organisatie leven. De meeste werknemers bij SIDN hadden hier wel een gevoel over wat kan, mag, en verantwoord is, maar zeker op wettelijk gebied bleek dit onvoldoende.

Deze wettelijke complexiteit blijkt ook bij het invullen van een privacypolicy. Auteurs vonden de policysjabloon eenvoudig in te vullen, met uitzondering van het veld 'wettelijke grondslag voor verwerking', het enige veld dat direct de Wbp raakt. We hebben daarom besloten in een volgende versie van de policysjabloon duidelijk te maken dat er minstens een grondslag uit de lijst van artikel 8 Wbp van toepassing moet zijn. In plaats van een open invulveld veranderen we dit naar een keuzelijst.

6 Conclusies en toekomstig werk

Een afdwingbaar privacyraamwerk voor het opslaan en analyseren van 'big data' uit het Domain Name System (DNS) is cruciaal voor de vertrouwensrol die SIDN voor Nederland vervult als beheerder van de .nl-extensie. Het ENTRADA-privacyraamwerk biedt ons daarvoor de middelen, waardoor we op een transparante manier een systematische afweging maken tussen het vinden van bedreigingen en afwijkingen in DNS-verkeer aan de ene kant en de privacy van .nl-gebruikers aan de andere kant.

Door de invoering en het actief gebruik van ons privacyraamwerk het afgelopen jaar bij SIDN hebben we ons raamwerk verder kunnen verbeteren, bijvoorbeeld door het toevoegen van een beoordelingsrapport. Daarnaast hebben we lessen geleerd die mogelijk relevant zijn voor andere organisaties.

Ons raamwerk is een grondige en uitvoerbare methode om het gebruik van mogelijke persoonsgegevens, waarvoor we geen directe toestemming van de betrokkenen kunnen krijgen, te combineren met het verder verhogen van de veiligheid en stabiliteit van .nl op basis van de analyse van DNS-berichten. Ons raamwerk faciliteert transparantie in de verwerking van persoonsgegevens en is preciezer en grondiger dan het simpelweg publiceren van één algemeen privacystatement. We gaan daarmee een stap verder dan wat de Wbp vereist. Met onze aanpak denken we een voorbeeld te zijn voor andere organisaties die persoonsgegevens verwerken of overwegen die gaan te verwerken.

Op de korte termijn gaan we onze goedgekeurde privacypolicy's publiceren. Daarnaast gaan we onze bestaande policy's regelmatig herevalueren, net als het raamwerk zelf. We houden de aankomende EU Data Protection Regulation (DPR) in het oog, en hoewel we bij de ontwikkeling van ons raamwerk al wel rekening hebben gehouden met de DPR, zullen we ons raamwerk ook herevalueren wanneer deze van kracht wordt.

Het privacyraamwerk en de processen eromheen zijn inmiddels dagelijkse routine bij SIDN, maar we verwachten dat we het voortdurend aan zullen moeten blijven scherpen. We staan open voor verdere feedback en suggesties, en gaan graag in gesprek over onze aanpak.

7 Referenties

1. Wetenschappelijke Raad voor het Regeringsbeleid, 'Big Data in een vrije en veilige samenleving', april 2016, www.wrr.nl/fileadmin/nl/publicaties/PDF-Rapporten/rapport_95_Big_Data_in_een_vrije_en_veilige_samenleving.pdf.
2. C.E.W. Hesselman e.a., 'Een privacyraamwerk voor "DNS big data"-toepassingen', *P&I* 2014, afl. 6, www.sidnlabs.nl/downloads/whitepapers/PEI_2014_6_Hesselman.pdf.
3. P. Mockapetris, 'Domain Names – Concepts and Facilities', IETF, november 1987, www.ietf.org/rfc/rfc1034.txt.
4. Wet bescherming persoonsgegevens, wetten.overheid.nl/BWBR0011468/.
5. Vrijstellingsbesluit Wbp, wetten.overheid.nl/BWBR0012461/.
6. 'Enhanced Top-level domain Resilience Through Advanced Data Analysis', CPB-melding 1591862, https://www.collegebeschermingpersoonsgegevens.nl/asp/ORDetail.asp?moid=8585898d8b88, mei 2015.
7. M. Müller, 'SIDekICk: SuspIcious DomaIn Classification in the .nl Zone' (masterthesis Enschede), 2015, www.sidnlabs.nl/downloads/publications/Muller_Master_Thesis_EIT_SP.pdf.
8. M. Davids, 'ENTRADA-koppeling met AbuseHUB', blogpost, september 2015, www.sidnlabs.nl/a/weblog/entrada-koppeling-met-abusehub.
9. 'De publieke kern van het internet. Naar een buitenlands internetbeleid', WRR-rapport nr. 94, maart 2015, www.wrr.nl/publicaties/publicatie/article/de-publieke-kern-van-het-internet-1/.
10. M. Wullink, G. Moura & C. Hesselman, 'ENTRADA: a High-Performance Network Traffic Data Streaming Warehouse', IEEE/IFIP Network Operations and Management Symposium (NOMS16), Istanbul, Turkije, april 2016, www.sidnlabs.nl/downloads/sidn-noms2016_EN.pdf.
11. G. Moura e.a., 'nDEWS: a New Domains Early Warning System for TLDs', IEEE/IFIP International Workshop on Analytics for Network and Service Management (AnNet 2016), Istanbul, Turkije, april 2016, www.sidnlabs.nl/downloads/presentations/sidn-annet2016.pdf.
12. M. Wullink e.a., 'DNS Big Data Applications with ENTRADA', Symposium on Electronic Crime Research (eCrime'16), Toronto, Canada, juni 2016, www.sidnlabs.nl/downloads/whitepapers/sidn-ecrime.pdf.
13. M. Wullink, 'ENTRADA: The Impact of a TTL Change at the TLD Level', presentatie, DNS-OARC Spring 2016 workshop, Buenos Aires, Argentinië, april 2016, www.sidnlabs.nl/downloads/presentations/DNS-OARC-Buenos-Aires-final.pdf.
14. S. Bortzmeyer, 'RFC7816: DNS Query Name Minimisation to Improve Privacy', DNS Query Name Minimisation to Improve Privacy, maart 2016, tools.ietf.org/html/rfc7816.
15. Autoriteit Persoonsgegevens, 'Melden verwerking persoonsgegevens', autoriteitpersoonsgegevens.nl/nl/melden/melden-verwerking-persoonsgegevens.

Appendix A Voorbeeldpolicy

Naam onderzoek/applicatie	ENTRADA Algemeen
Doel van de applicatie of het onderzoek	Het ENTRADA-platform (ENhanced Top-level domain Resilience through Advanced Data Analysis) is een platform waarmee DNS-querydata bewaard en geanalyseerd kunnen worden. Het doel van dit platform is de ontwikkeling van nieuwe diensten en applicaties waarmee we de veiligheid en stabiliteit van .nl verder kunnen vergroten, alsmede losse onderzoeken naar incidenten die de stabiliteit van .nl in gevaar zouden kunnen brengen. Deze policy betreft het platform zelf, prototypes en onderzoeken. Productiesystemen, applicaties en onderzoeken die niet binnen deze policy passen krijgen een eigen policy.
Persoonsgegevens	Omdat het hier gaat om het algemene platform dat gebruikt wordt voor onderzoek en ontwikkeling, kunnen we niet weten welke data al dan niet nodig zijn. Daarom worden DNS-querydata opgeslagen. Zoals beschreven in de policy paper 'Een privacyraamwerk voor DNS big data' houdt dat wat persoonsgegevens betreft in dat er IP-adressen en gezochte domeinnamen worden opgeslagen en verwerkt.
Grondslag	Het doel van het platform is onderzoek naar en ontwikkeling van applicaties om de veiligheid van .nl, en die van haar gebruikers, te verbeteren. Daarmee is gerechtvaardigd belang de grondslag van het gebruik van de data.
Filters	Er worden geen filters toegepast.
Retentie	De gegevens worden 18 maanden bewaard. Deze duur is gekozen om voldoende tijd te bieden om onderzoeken te doen die een jaar aan data gebruiken.
Toegang	De medewerkers van SIDN Labs hebben toegang tot de data, net als de DNS-beheerders van SIDN. De data kunnen alleen benaderd worden vanaf het interne SIDN Labs-netwerk, middels https, beveiligd met een wachtwoord dan wel via Kerberos-authenticatie. De DNS-beheerders van SIDN en SIDN Labs hebben instructie gekregen over het verantwoord omgaan met de data.
Publicatie/delen	De data worden niet gedeeld; als er onderzoeksresultaten gepubliceerd worden dan zullen deze geen specifieke persoonsgegevens bevatten. Data die onder deze policy verwerkt worden, worden niet gedeeld met derden. Voor projecten en diensten die wel data delen met derden zal een specifieke privacypolicy gemaakt worden.
Type	R&D-onderzoek
Andere beveiligingsmaatregelen	n.v.t.

Appendix B Voorbeeldbeoordeling

Policy		
	Naam van policy	ENTRADA Algemeen
	Datum van beoordeling	5 januari 2016
Doelbinding		
	Wbp van toepassing?	Worden er persoonsgegevens verwerkt? *De privacyboard is van mening dat de onderzochte querydata tot een ip-adres herleid kunnen worden en dat ip-adres in een beperkt aantal gevallen weer tot een natuurlijke persoon. Daarmee kan een ip-adres een gegeven betreffende een identificeerbare natuurlijke persoon zijn (art. 1 sub a Wbp). De privacyboard adviseert dan ook de data te behandelen alsof deze persoonsgegevens bevatten. Hetzelfde geldt voor de gezochte domeinnamen.* Is er sprake van een geheel of gedeeltelijk geautomatiseerde verwerking van persoonsgegevens of een niet-geautomatiseerde verwerking van persoonsgegevens die in een bestand zijn opgenomen? *Ja en i.v.m. art. 2 lid 1 Wbp kan geconcludeerd worden dat de Wbp van toepassing is.*
	Doel	Is het doel welbepaald, uitdrukkelijk omschreven en gerechtvaardigd? *Ja, de privacyboard is van mening dat in de privacypolicy conform art. 7 Wbp het doel welbepaald en uitdrukkelijk omschreven is. Tevens vindt de privacyboard het verder vergroten van de veiligheid en stabiliteit van .nl een gerechtvaardigd doel.*
	Grondslag	Is er een legitieme grondslag/verwerkingsgrond voor de verwerking? *De privacyboard is van mening dat een gerechtvaardigd belang van zowel SIDN zelf (i.v.m. ons doel de betrouwbaarheid en veiligheid van .nl en van het internet in de breedte verder te verhogen) als derden (namelijk de gebruikers van .nl) wordt gediend. Hiermee is er sprake van een legitieme grondslag o.b.v. art. 8 sub f Wbp.*

Borging en maatregelen		
	Borging doelbinding	Is geborgd dat de persoonsgegevens niet verder worden verwerkt op een wijze die onverenigbaar is met de doeleinden waarvoor ze zijn verkregen? *Ja, de privacyboard is van mening dat voldaan wordt aan art. 9 Wbp doordat ENTRADA slechts wordt gebruikt voor intern onderzoek en vervolgens afdoende maatregelen zijn genomen om de toegang ertoe te beveiligen.* *Op het moment dat de data met anderen worden gedeeld volgt een privacypolicy met een bijbehorende beoordeling door de privacyboard.*
	Bewaartermijn	Is geborgd dat de gegevens niet langer bewaard worden dan nodig voor het doel? *Ja, de privacyboard is van mening dat conform art. 10 Wbp een bewaartermijn van 18 maanden reëel is i.v.m. het gedurende een half jaar onderzoeken van een jaar aan data.*
	Minimale set gegevens	Is geborgd dat de gegevens slechts worden verwerkt voor zover ze voor het doel toereikend, ter zake dienend en niet bovenmatig zijn? *Ja, de privacyboard is van mening dat conform art. 11 Wbp wordt gewerkt met de minimale dataset die benodigd is om het doel van de verwerking te bereiken. Voor het uit te voeren onderzoek is een volledige dataset nodig.*
	Correctheid gegevens	Is geborgd dat de verzamelde gegevens correct zijn? *Ja, de privacyboard is van mening dat conform art. 11 Wbp wordt gewerkt, omdat met gegevens o.b.v. een eigen meting door SIDN op haar eigen systemen wordt gewerkt. Door de toegang daartoe vervolgens te beveiligen wordt voorkomen dat deze door derden worden aangepast.*
	Gegevensverwerkers	Is geborgd dat de gegevens slechts worden verwerkt in opdracht van de verantwoordelijke? *Ja, de gegevens worden conform art. 12 Wbp verwerkt door medewerkers van SIDN Labs en de DNS-beheerders van SIDN, d.w.z. die medewerkers van Verantwoordelijke die toegang nodig hebben om de werkzaamheden uit te kunnen voeren.*
	Beveiliging gegevens	Zijn er passende technische en organisatorische maatregelen genomen om de gegevens te beveiligen? *Ja, de privacyboard is van mening dat conform art. 13 Wbp de beveiliging van de toegang afdoende is.*

Overig		
	Bijzondere persoonsgegevens	Worden er bijzondere persoonsgegevens verwerkt? *Nee, mede o.b.v. een extern advies is de privacyboard van mening dat er geen bijzondere persoonsgegevens conform art. 16 Wbp worden verwerkt.*
	Melding bij AP	Is voldaan aan de meldingsplicht uit art. 27 Wbp? *Ja, de verwerking valt onder meldingsnummer 1591862.*
	Rechten betrokkene	Wordt voldaan aan de informatieverplichtingen uit artikel 33 en 34 Wbp? *Ja, de privacyboard is van mening dat art. 34 van toepassing is.*
	Buiten EU	Worden de gegevens doorgegeven naar landen buiten de EU waardoor voldaan moet worden aan artikel 76 Wbp? *Nee, de gegevens binnen ENTRADA worden slechts verwerkt door medewerkers van SIDN.*
Conclusie		
	Beoordeling	*De privacyboard gaat akkoord met de privacypolicy ENTRADA Algemeen*