



Privacypolicy

SIDN-dataplatform

Datum

1 januari 2024

Classificatie

Publiek

Auteur

SIDN

Blad

1/5

SIDN BV

Contact

T 026 352 55 00

support@sidn.nl

www.sidn.nl

Bezoekadres

Meander 501

6825 MD Arnhem

Postadres

Postbus 5022

6802 EA Arnhem

KvK

Arnhem 88 77 28 96

Naam

onderzoek/applicatie

SIDN-dataplatform

Ingangsdatum policy

1 januari 2024

Projectbeschrijving

Het SIDN-dataplatform is een centrale opslagplaats waarin grote hoeveelheden gegevens in de oorspronkelijke vorm worden opgenomen en opgeslagen.

De gegevens kunnen worden verwerkt en gebruikt als basis voor verschillende analysebehoeften.

De gegevensbestanden worden opgeslagen in gefaseerde zones: raw, standardized, integrated en consumable, zodat verschillende typen gebruikers de gegevens in de verschillende vormen kunnen gebruiken om aan hun behoeften te voldoen.

Doel van de applicatie of het onderzoek

Het dataplatform heeft de volgende doelen:

1. Verbeteren van datakwaliteit en consistentie: door gebruik te maken van een data lake standaardiseren we het ophalen en structureren van gegevens uit elke bron. Hiermee voorkomen we het maken van foute aannames en besparen we tijd.
2. Eén versie van de waarheid: het data lake (house) is de centrale plek voor alle informatievraagstukken en zorgt ervoor dat iedereen met dezelfde data werkt. De

afhankelijkheden van het bronsysteem voor de informatievoorziening worden gereduceerd, waardoor onze informatievoorziening minder kwetsbaar is. Daarnaast heeft een eventuele vervanging van een bronsysteem geen invloed op de structuur van de informatievoorziening. Hiermee wordt de continuïteit geborgd.

3. Mogelijkheid tot het combineren van databronnen: door gebruik te maken van een data lake (house) kunnen we gegevens uit verschillende bronnen verzamelen. Door deze gegevens vervolgens met elkaar in relatie te brengen kunnen er interessante inzichten ontstaan. Dit geeft bijvoorbeeld de registrar meer inzicht in zijn portfolio.
4. Reduceren tijd beheer en onderhoud: het beheer en onderhoud van lokale servers kost steeds meer tijd. Ook het opzetten van nieuwe applicaties en/of rapportagemodules kost veel tijd. Als gevolg hiervan is er minder tijd voor innovatie. Door gebruik te maken van een centraal data lake in de cloud, verminderen we de werkdruk van productteams.
5. Betere beveiliging: een data lake (house) is erg geschikt voor het afschermen van data. Door op een centrale plek ervoor te zorgen dat gebruikers alleen toegang hebben tot de data die zij ook daadwerkelijk mogen zien. Omdat dit centraal wordt ingeregeld hoeft dit ook maar op één plek te worden beheerd.

De volgende databronnen worden ontsloten met daarbij behorende doelen van dataverwerking:

- Domein Registratie Systeem (DRS): voor interne rapportages over transacties van registrars, DRS reports om medewerkers van support te voorzien van juiste registrar- en domeinnaaminformatie en SIDN Insights om registrars inzicht te geven in hun eigen performance.
- DMAP: om registrars inzicht te geven in hoe zij scoren op de Registrar Scorecard (RSC) via SIDN Insights en interne rapportages om voortgang op technische standaarden en internettrends te meten.
- Valimon: om registrars inzicht te geven op het gebied van DNSSEC en intern de voortgang te kunnen meten.
- Google Analytics/PiwikPro: om inzicht te geven in wat de online conversie is van registrars in de domeinnaamsuggestietool.



- Netcraft: data waarmee we zien op welke domeinnamen abuse plaatsvindt, wat de verschillen tussen registrars zijn en wat de afhandelingsnelheid is.
- Business Central: om registrars inzicht te geven in hun facturen.

Het dataplatform maakt van de bovenstaande bronnen rapportages en dashboards voor medewerkers en klanten zoals registrars, ICANN en CENTR. Dit zijn al bestaande rapportages die nu uit meerdere applicaties komen. Het dataplatform neemt de functionaliteit van het genereren van rapportages over.

Deze policy betreft alleen het dataplatform. Voor toepassingen en onderzoeken, die gebruik maken van data uit het dataplatform en die niet binnen deze policy passen, moet een aparte privacy policy worden opgesteld.

Persoonsgegevens

In het kader van het project worden de volgende persoonsgegevens verwerkt:

- Alle transacties van een domeinnaam.
- Alle transacties van een registrar (naam, e-mail, telefoon, adres).
- Classificaties van registrars (segmenten).
- Alle gegevens van een houder (naam, e-mail, telefoon, adres).
- Domeinnamen met abuse.
- Classificatie van abuse.
- Tijdstip van abuse.
- IP-adres van domeinnaam met abuse.
- Geraadpleegde domeinnaam WhoIS.
- Gevalideerde domeinnamen door Valimon (domeinnaam, eventueel foutmelding).
- KvK-nummer (DMAP).
- Facturatiegegevens van registrars (e-mail, inhoud van factuur).

Deze gegevens kunnen onder omstandigheden persoonsgegevens zijn.

Grondslag

Door data centraal beschikbaar te stellen, voorkom je dat er veel onnodige kopieën worden gemaakt. Bovendien is het makkelijker om rollen en rechten te bepalen, is het eenvoudiger te beheren en geeft het ons inzicht in gebruik van data en toegang tot data. Dit draagt bij aan de veiligheid, transparantie en beheersbaarheid van de organisatie.



Datum
1 januari 2024

Classificatie
Publiek

Blad
4/5

Door data uit diverse bronnen te combineren kun je klanten meer inzichten verschaffen in hun portfolio en dragen we o.a. bij aan de groei van .nl en kunnen we beter sturen op de veiligheid van de zone (abuse, RSC)

Filters

Er worden geen filters toegepast.

Retentie

Er wordt per bron of businessobject een bewaartermijn opgegeven. De standaard bewaartermijn is 2 jaar voor ruwe data en 5 jaar voor geaggregeerde data.

Toegang

Het dataplatform kent verschillende lagen. Er worden per laag autorisaties ingesteld.

Raw data: deze laag bevat ruwe data zoals door de bron aangeleverd (met toevoeging van metadata) en is alleen toegankelijk voor geautomatiseerde processen (pipelines) om data door te zetten naar de volgende laag.

Standardized data: deze laag bevat data in file of tabelstructuur. De historische volgorde is bepaald en er hebben technische transformaties plaatsgevonden zoals compressie en uniformering. De data is query-baar en is alleen toegankelijk voor geautoriseerde gebruikers van het datateam en data scientists van SIDN Labs in een beveiligde omgeving met een username, wachtwoord en 2FA.

Integrated data: in deze laag vindt de technische implementatie van businessobjecten plaats. De data is query-baar en is alleen toegankelijk voor geautoriseerde gebruikers van het datateam in een beveiligde omgeving met een username, wachtwoord en 2FA.

Consumable data: in deze laag is de data functioneel geoptimaliseerd door middel van filtering, aggregatie en specifieke business rules. De uitkomsten van deze laag worden beschikbaar gesteld in zogeheten datamarts en/of -sets. Deze marts zijn beschikbaar voor geautoriseerde medewerkers van SIDN om rapportages te bouwen. Deze autorisaties kunnen per dataset of mart verschillen.

Export laag/api: specifieke laag voor klanten waar rapportages of datasets beschikbaar worden gesteld (bijv. aan CENTR, ICANN).



Datum
1 januari 2024

Classificatie
Publiek

Blad
5/5

Publicatie/delen

Er worden gegevens vanuit de consumable laag gedeeld met medewerkers van de afdelingen Business & Support, Finance, Labs en het MT. Het uitgangspunt is dat er geen persoonsgegevens gedeeld worden, tenzij dat noodzakelijk is voor uitvoering van de functie.

Via de consumable laag worden ook rapportages of datasets beschikbaar gesteld aan externe gebruikers zoals registrars of ICANN.

Deze gegevens kunnen persoonsgegevens bevatten wanneer er behoefte is aan informatie over een specifieke domeinnaam of houder. Een voorbeeld hiervan is een medewerker van Support die een klant helpt met een vraagstuk of een registrar die op domeinniveau een vraag heeft.

Er worden geaggregeerde gegevens (statistieken) gebruikt voor diverse publicaties. Deze geaggregeerde gegevens bevatten geen persoonsgegevens.

Type

Productie

Andere beveiligingsmaatregelen

Via de datamanagementmodule is er inzicht in het datagebruik voor audits.

Samenvatting projectbeschrijving

Het dataplatform is een centrale opslagplaats waarin grote hoeveelheden gegevens in de oorspronkelijke vorm worden opgenomen en opgeslagen.

De gegevens kunnen worden verwerkt en gebruikt als basis voor verschillende analysebehoeften.

De gegevensbestanden worden opgeslagen in gefaseerde zones, raw, standardized, integrated en consumable, zodat verschillende typen gebruikers de gegevens in de verschillende vormen kunnen gebruiken om aan hun behoeften te voldoen.