

Aan: Directiebestuur CBS

Van: Bert Kroese

Auteur: 5.1.2.e

Onderwerp: Invoeren verPRINnen

Datum: 9 maart 2022

### **Inleiding**

In 2020 is binnen KERS het idee ontstaan om de huidige manier van pseudonimiseren van privacy gevoelige gegevens aan te passen. Deze nieuwe manier van pseudonimiseren zou tegemoet komen aan een aantal bezwaren die spelen bij de huidige manier van pseudonimiseren. Dit idee is met de rest van de CBS organisatie gedeeld en werd (bijna overal) positief ontvangen.

Het idee om het verRINnen te vervangen door het zgn. verPRINnen is in het voorjaar van 2021 in de Coördinatiegroep Continue Verbeteren (CCV) besproken. De leden van de CCV hebben in dat overleg aangegeven het eens te zijn met de in de nota geschetste richting en gaven aan dat het goed zou zijn om verdere stappen te zetten. In het afgelopen half jaar is binnen KERS verder aan dit idee gewerkt met de bouw van een prototype waarmee het verPRINnen plaats kan vinden.

Omdat dit nieuwe beleid voor het pseudonimiseren CBS brede impact heeft is het logischer om de verdere ontwikkeling, en mogelijk uitrol van het verPRINnen, niet binnen KERS te laten plaats vinden maar om het DB verdere besluiten te laten nemen over dit beleid en over de verdere ontwikkeling van benodigde tools.

In deze nota zal eerst kort het huidige verRIN beleid en het voorgestelde verPRIN beleid worden beschreven, inclusief de voordelen van het verPRINnen. In de daarop volgende paragraaf worden een aantal vragen beschreven waarbij de antwoorden op die vragen mede bepalend zijn voor de business case van dit nieuwe beleid. Voorgesteld wordt een werkgroep in het leven te roepen die deze vragen gaat beantwoorden waarna het DB een definitief besluit kan nemen over dit nieuwe beleid. De nota wordt afgesloten met vragen aan het DB.

### **Beschrijving verRINnen en verPRINnen**

Over verRINnen en verPRINnen zijn/kunnen dikke nota's worden geschreven. De poging om dit in enkele zinnen te beschrijven kan dus nooit compleet en volledig zijn. De volgende tekst is een hoog over beschrijving en heeft vooral als doel om een kort beeld te geven.

VerRINnen wordt al tientallen jaren binnen het CBS gebruikt als methode om identificerende gegevens te pseudonimiseren. Het gaat daarbij dus niet alleen om de BSN maar bijvoorbeeld ook over de exacte geboortedatum en het adres. De bestanden die zijn verRIND worden in een

afzonderlijke omgeving geplaatst (sec-omgeving) en medewerkers kunnen deze bestanden alleen met een apart account worden benaderd (sec account). Doordat iedere BSN wordt omgezet naar één specifieke RIN kunnen de bestanden eenvoudig worden gekoppeld. Bij het verPRINnen worden de BSN's niet naar één specifieke RIN omgezet maar wordt voor één bepaald proces (of groep processen of project) een aparte pseudoniem gebruikt. Verder kunnen, indien nodig, ook andere identificerende gegevens zoals de geboortedatum beschikbaar komen voor een proces. Bij het verRINnen worden of alle identificerende gegevens gepseudonimiseerd of juist allemaal niet.

In bijlage 1 zijn enkele procesplaten opgenomen waarin de huidige manier van verRINnen en de nieuwe manier van verPRINnen staan weergegeven. Hieruit blijkt dat de huidige manier van pseudonimiseren drie bezwaren kent. Al deze bezwaren zouden met het verPRINnen opgelost kunnen worden:

- (Langdurige) toegang tot BSN. In een aantal processen wordt langdurig gewerkt met een BSN, bijvoorbeeld omdat een (bijna) exact adres nodig is. In de plannen van SER en EBN wordt verPRINnen dan ook als oplossing genoemd om de toegang tot BSN te beperken omdat bij het verPRINnen het nog wel mogelijk is om (als dat echt noodzakelijk is voor het proces) identificerende gegevens zoals adres te gebruiken.
- Inefficiënte uitwisseling BSN en RIN data. Het combineren van data met een BSN en data met een RIN is inefficiënt. In de praktijk komt deze situatie regelmatig voor, of omdat binnen processen de data op verschillende momenten worden gepseudonimiseerd (of helemaal niet worden gepseudonimiseerd) of omdat bijvoorbeeld wel een adres of geboortedatum nodig is waardoor er nu niet gewerkt kan worden met een RIN.
- Het combineren/koppelen van alle bestanden met een RIN is relatief eenvoudig en, zeker als er veel bestanden met RIN's op verschillende plekken binnen het CBS staan, brengt dit privacy risico's met zich mee.

Met de nieuwe manier van pseudonimiseren worden deze bezwaren in theorie weg genomen. In de praktijk zal één en ander nog wel uitgewerkt moeten worden en pas dan zal een goed zicht komen op de daadwerkelijk te realiseren baten en welke kosten hiermee gepaard gaan.

### **Vervolgstappen**

Ondanks eerste stappen die zijn gezet binnen KERS leven er nog een aantal vragen rond het verPRINnen. Het is goed om deze vragen in de komende periode/fase te beantwoorden. Het zijn niet zozeer beleidsvragen maar vragen die gaan over de wijze waarop implementeren kan plaats vinden en de inrichting van de governance. De antwoorden op deze vragen zijn (mede) bepalend voor het vaststellen van de benodigde functionaliteit van deze service, de wijze waarop deze service ingericht zal worden, investeringen die moeten worden gedaan bij bestaande tools, omgevingen etc. maar ook in de mate waarin voorziene baten kunnen worden gerealiseerd. De antwoorden op deze vragen helpen om een beter beeld te krijgen hoe deze nieuwe werkwijze in de praktijk gaat werken en wat de veranderingen zijn voor medewerkers, IT en de processen.

Het gaat daarbij in ieder geval over de volgende vragen:

1. Wat is de reikwijdte van de “P” in PRIN?  
De vraag is wat de reikwijdte zal zijn van de “P” , is dat een proces, een project, een team, een sector of een mix hiervan. Als de reikwijdte van een “P” klein wordt gedefinieerd dan is de privacy beter geborgd maar zal het systeem voor verPRINnen vaker aangeroepen moeten worden. Als het cluster erg groot wordt dan zal dit ten koste gaan van de privacy.
2. Wat betekent de PRIN voor de omgeving waarin gewerkt moet worden?  
Bij het huidige verRIN beleid worden bestanden met een RIN in een andere omgeving (de zgn. sec omgeving) verwerkt dan de omgeving waarin wordt gewerkt met bestanden met een BSN. In een eerder gevoerde gesprek met de (toenmalige) functionaris gegevensverwerking is aangegeven dat het onderscheid tussen de beide omgeving niet meer nodig is bij het gebruik van de PRIN. Het is goed om nog eens naar de omgevingen te kijken waar data met BSN, RIN en PRIN mogen worden verwerkt.
3. Kan de Remote Access (RA) omgeving ook werken met verPRINnen?  
Eerder is met medewerkers van de RA omgeving gesproken over het verPRINnen. Tijdens dat gesprek bleek dat het nog geen uitgemaakte zaak is dat verPRINnen op de RA omgeving (eenvoudig) te implementeren is. Dit zal nog uitgezocht moeten worden maar is wel een belangrijke vraag omdat dit één van de aanleidingen was om te komen met de PRIN.
4. Hoe gaat de beoogde volledig geautomatiseerde verPRIN service samen werken met de data abstractielaag of, als men nog niet is aangesloten op deze laag, buiten deze abstractielaag om?  
Binnen het CBS is afgesproken dat data uitwisseling tussen processen plaats zal vinden via de data abstractielaag met Denodo als tool. Nog uitgezocht moet worden hoe de verPRIN service gecombineerd kan worden met Denodo. Daarnaast zal het nog enige tijd duren voordat alle processen op deze abstractielaag zijn aangesloten en zal dus ook data uitwisseling buiten de abstractielaag om mogelijk moeten zijn. De vraag is hoe de beoogde volledig geautomatiseerde verPRIN service in deze situaties ingezet moet worden zodat in beide gevallen hetzelfde resultaat wordt bereikt.
5. Wat zijn de gevolgen voor het DSC en de huidige programmatuur voor verRINnen?  
Nog onduidelijk is of het introduceren van de verPRIN service gevolgen heeft voor het DSC en de programmatuur waarmee data in het DSC kunnen worden geplaatst en uit het DSC kunnen worden gehaald. Daarnaast zal in de situatie waarin we verPRINnen ook nog een RIN bestaan (denk aan de huidige bestanden in het DSC). Op dit moment is eveneens onduidelijk of de huidige programmatuur voor het verRINnen aangepast moet worden.

6. Wie gaat de eigenaar worden van deze service/dit proces?

Het is belangrijk dat voor een dergelijke CBS brede dienst een eigenaar wordt aangewezen. De meest logisch kandidaat voor deze rol is de huidige groep Centrale Bestands Koppeling. Daar is op dit moment voldoende kennis aanwezig van pseudonimiseren en het gebruik van pseudoniemen in processen.

Het is niet zeker dat hiermee alle vragen ook echt in beeld zijn. Met name tijdens de bouw van de definitieve service kunnen aanvullende vragen opkomen. Het is daarom goed om gelijktijdig met het beantwoorden van bovenstaande vragen aan de slag te gaan met het verder uitwerken en specificeren van de tool die door team generiek gebouwd zal worden. Het gebouwde prototype kan daarbij dienen als eerste beeld.

Voorgesteld wordt om bovenstaande vragen te laten beantwoorden door een werkgroep bestaande uit de beoogd eigenaar (dus iemand van CBK), een business analist, een vertegenwoordiger van het SSB, een vertegenwoordiger uit KERS (indien anders dan de business analist), medewerkers van DRD (procesontwikkeling, projectleider) en vertegenwoordigers uit SER, EBN en DRI. Trekker van dit traject is Bert Kroese.

Gerealiseerd moet worden dat het nieuwe beleid niet van de ene op de andere dag geïmplementeerd kan worden. Er zal sprake zijn van een overgangperiode. Door te werken met specifieke PRIN's is een geleidelijke implementatie mogelijk. Zoals eerder gemeld zullen de antwoorden op de vragen een (verder) beeld geven van baten (op gebied van privacy en efficiëntie zoals het wegvallen van de sec omgeving) en lasten (omvang éénmalige investeringen en reguliere extra werkzaamheden).

**Aan het DB wordt gevraagd de volgende twee vragen te beantwoorden.**

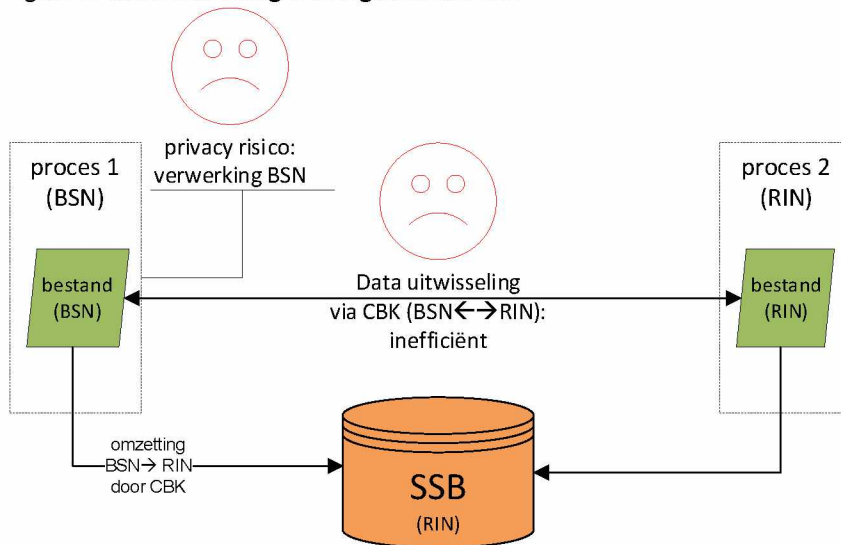
1. Gaat het DB akkoord met het verder uitwerken van het geschetste verPRIN beleid?
2. Is het DB akkoord met het instellen van een werkgroep om bestaande vragen te beantwoorden en daarnaast parallel te starten met de voorbereidingen voor de ontwikkeling van een verPRIN tool? De benodigde capaciteit zal daarbij vooral bestaan uit capaciteit van team generiek en de capaciteit van de genoemde leden van de werkgroep (per lid gemiddeld 100 uur).

**Op basis van de resultaten van deze fase (afronding van deze fase is medio 2022 voorzien) zal aan het DB een definitieve vraag voor besluit worden voorgelegd.**

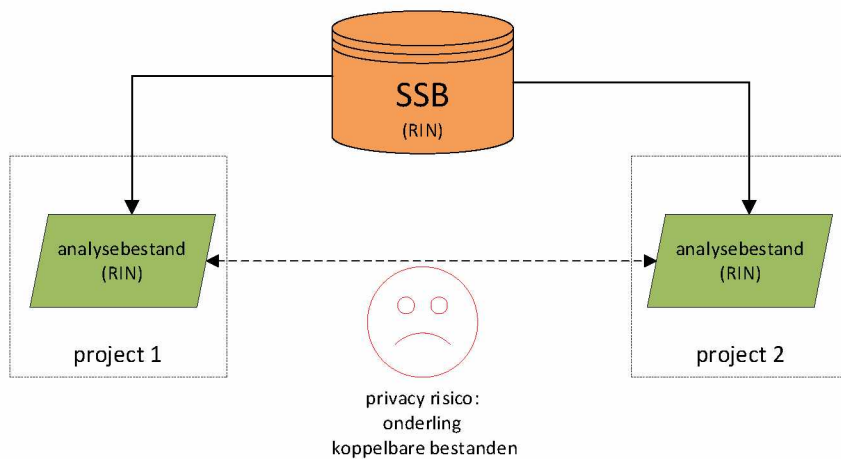
## Bijlage 1: korte beschrijving huidige situatie en toekomstige situatie van pseudonimiseren.

### Huidige situatie

Figuur 1: dataverwerking voorafgaand aan SSB

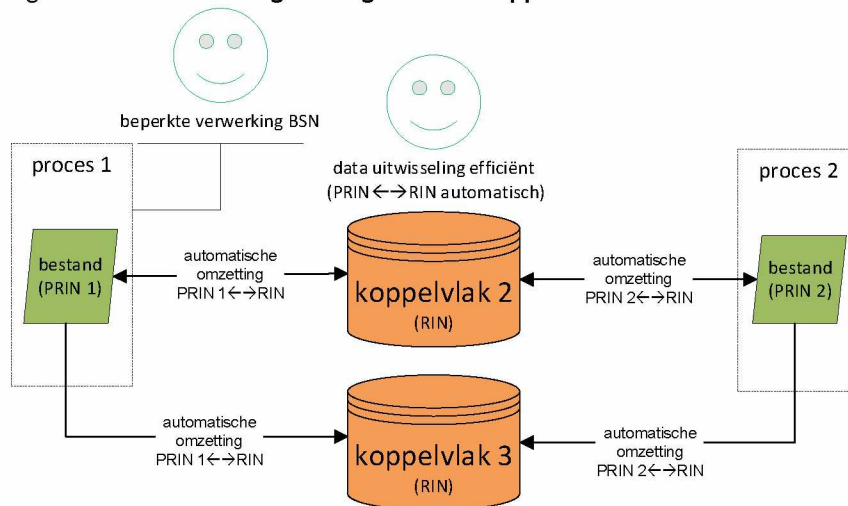


Figuur 2: data delen via het SSB



## Toekomstige situatie

Figuur 1: dataverwerking voorafgaand aan koppelvlaak 3



Figuur 2: data delen via koppelvlaak 3

