



Gegevensbeschermings- effectbeoordeling (PIA): Mobile applicatie voor ODiN onderzoek

CENTRAAL BUREAU VOOR DE STATISTIEK (CBS)

DIRECTIE DATASERVICES, RESEARCH EN INNOVATIE (DRI)

Heerlen 10 november 2022

Maatregelen
nemen
Privacybewustwording
PIA
Doelbinding
Noodzaak
Effecten in kaart
Bescherming van
persoonsgegevens
Risico's
minimaliseren
Richtinggevend
Rechtsgrond
Met open vizier

Vaststelling verwerkersverantwoordelijke: 10 november 2022

Naam: 5.1.2.e

Advies functionaris voor gegevensbescherming: 10 november 2022

Naam: 5.1.2.e

Advies CIO: 10 november 2022

Naam: Wim van Nunspeet

Gegevensbeschermings- effectbeoordeling (PIA): ODiN mobile applicatie

CENTRAAL BUREAU VOOR DE STATISTIEK (CBS)
DIRECTIE DATASERVICES, RESEARCH EN INNOVATIE (DRI)

Contact:

Centraal bureau voor de statistiek

5.1.2.e

Email: 5.1.2.e@cbs.nl

Telefoonnummer (5.1.2.e): (5.1.2.e)

Versie: 1.0 10 november 2022

Inhoudsopgave

Inleiding	5
A. Beschrijving kenmerken gegevensverwerkingen	6
1. Voorstel 	6
2. Persoonsgegevens 	8
3. Gegevensverwerkingen 	8
4. Verwerkingsdoeleinden 	10
5. Betrokken partijen 	10
6. Belangen bij de gegevensverwerking 	10
7. Verwerkingslocaties 	10
8. Techniek en methode van gegevensverwerking 	11
9. Juridisch en beleidsmatig kader 	11
10. Bewaartermijnen 	11
B. Beoordeling rechtmatigheid gegevensverwerkingen	11
11. Rechtsgrond 	12
12. Bijzondere persoonsgegevens 	12
13. Doelbinding 	12
14. Noodzaak en evenredigheid 	12
15. Rechten van de betrokkene 	13
C. Beschrijving en beoordeling risico's voor de betrokkenen	13
16. Risico's 	13
D. Beschrijving voorgenomen maatregelen	15
17. Maatregelen 	15
E. BIJLAGEN	17
Bijlage 1: Variabelen gebruikt in Veldtest.....	17
Bijlage 2: Analyseplan CBS verplaatsingen app 2022 - 2023	19
Bijlage 3: Verplaatsingsonderzoek met de app: 1 versus 7 dagen waarnemen	29
Bijlage 4: Basisregels Informatiebeveiliging CBS.....	32
Bijlage 5: Persoonsgegevens, anonimiseren en pseudonimiseren.	33

Inleiding

Het Centraal bureau voor de statistiek (hierna: CBS) verwerkt gegevens voor het maken van statistieken voor de praktijk, beleid en wetenschap op basis van de CBS-wet. Hiervoor heeft het CBS een algemene Gegevensbeschermingseffectbeoordeling (ook wel Privacy Impact Assessment genoemd, hierna: PIA) opgesteld. Hierin is vastgelegd wat er met de gegevens gebeurt, op basis van welke grondslagen en hoe beveiliging van deze gegevens is geborgd binnen het CBS. Deze algemene PIA is bedoeld voor de reguliere processen van het CBS.

Voor afwijkende processen dienen afzonderlijke PIA's te worden opgesteld. Dit is onder meer het geval bij innovatieve processen waarbij nieuwe technieken en methoden worden toegepast zoals bij het onderzoek Onderweg in Nederland (ODiN) het geval is. Het betreft hier een veldtest met een smartphone app (ODiN app) om het reisgedrag van de respondent te kunnen meten. Er wordt aan de respondent gevraagd om van een aantal dagen alle verplaatsingen te registreren. In verband met het gebruiken van aanvullende privacy gevoelige gegevens en het afwijkende verwerkingsproces van dit onderzoek is deze aanvullende PIA nodig.

Dit onderzoek wordt uitgevoerd in opdracht van het Ministerie van Infrastructuur en Waterstaat (IenW).

A. Beschrijving kenmerken gegevensverwerkingen

Beschrijf op gestructureerde wijze de voorgenomen gegevensverwerkingen, de verwerkingsdoeleinden en de belangen bij de gegevensverwerkingen.

Onder A wordt de eerste stap beschreven van de PIA: een overzicht van de relevante feiten van de voorgenomen gegevensverwerkingen. Als de feiten onduidelijk zijn, werkt dit door in de beoordeling.

1. Voorstel



Beschrijf het voorstel waar de gegevensbeschermingseffectbeoordeling op ziet en context waarbinnen deze plaatsvindt op hoofdlijnen.

Algemeen: binnen het waarneemproces binnen het CBS worden verschillende modi (CAPI, CATI, CAWI, PAPI¹) ingezet. De inzet van deze modi is vastgelegd in de benaderstrategie die onderdeel uit maakt van het onderzoeksdesign van het betreffende onderzoek. Voor het onderzoek ODIN wordt momenteel alleen gebruik gemaakt van internetwaarneming (CAWI). Het onderzoek verschaft adequate informatie over de dagelijkse mobiliteit van de Nederlandse bevolking, beschreven naar plaats van herkomst, bestemming, tijdstip waarop het vervoer plaatsvindt, gebruikte vervoermiddelen en de reismotieven voor de verplaatsingen. De veldtest is bedoeld om te onderzoeken of en hoe een app onderdeel kan worden van de benaderstrategie van ODIN en welke instellingen en functionaliteiten deze app dan zou moeten hebben. De resultaten worden tevens gebruikt om de definitieve privacy borging in te richten en dataminimalisatie mogelijk te maken. Er wordt bijvoorbeeld gekeken naar het aantal te rapporteren/meten dagen door respondenten. Na de veldtest is tevens meer informatie beschikbaar om een definitieve PIA op te stellen, indien besloten wordt om de waarneming met behulp een app voort te zetten.

Taak: de ontwikkeling van de app valt binnen het team Vragenlijstontwikkeling en -bouw (DVL) van de sector Dataverzameling (DVZ). De analyse van de gegenereerde data en de uitspraken over de definitieve instellingen en functionaliteiten valt onder de teams van Methodologie Den Haag en Heerlen van de sector Research and Development (DRD) en de sector Verkeer en vervoer (SVV). Dit in samenwerking met de kennispartner: Universiteit van Utrecht, waarmee een raamovereenkomst waarbij o.a. de privacy van respondenten wordt geborgd is, afgesloten.

Uitvoeren (grondslag): de veldtest bestaat uit twee experimenten met aparte steekproeven. Het eerste experiment wordt uitgevoerd op een steekproef van voormalige ODIN 2022 respondenten die serieverplaatsingen hebben gemaakt en/of een of meer multi-modale verplaatsingen. Een serieverplaatsing is een aaneenschakeling van verplaatsingen met ertussen relatief korte, meestal beroepsmatige, stops. Een multi-modale verplaatsing is een verplaatsing waarbij meerdere vervoermiddelen worden gebruikt. Het tweede experiment is een verse steekproef die verder willekeurig wordt onderverdeeld naar een drietal experimentele condities. Voor de respondenten op het tweede experiment vindt ook een evaluatie plaats. Zie voor verdere details over de uitvoering van de veldtest de bijlage: "Analyseplan CBS verplaatsingen app 2022 – 2023".

Gebruik van de gegevens (doelbinding): voor de statistische gegevens, die vastgelegd worden via de vragenlijst in de app, geldt de centrale PIA van het CBS. Deze PIA heeft betrekking op het verzamelen van de data van de geolocatiegegevens (latitude, longitude en attitude coördinaten) op bepaalde tijdstippen en technische data van een toestel voor de veldtest zoals het model en het batterij verbruik van het toestel. Zie bijlage: "Variabelen gebruikt in Veldtest". De technische data van het toestel zijn van belang voor het verkrijgen van informatie over het functioneren van de app, en om de app te kunnen verbeteren. Het meten van locatiegegevens hangt samen met het batterijmanagement van de telefoon, en het aantal locatiepunten dat gemeten wordt, verschilt tevens per device en besturingssysteem. Als de app goed functioneert (dat wil zeggen: op een correcte manier locatiegegevens registreert) op het ene device, wil dat niet zeggen dat het op een ander device ook goed functioneert. Om de app in de toekomst in te kunnen zetten voor het verzamelen van data ten behoeve van statistiek, is het noodzakelijk dat de app op de meeste telefoons goed werkt, en dat er geen selectieve uitval is (bijvoorbeeld omdat bepaalde groepen mensen vaker een bepaald type device hebben, waarop de app niet functioneert). Om dat te kunnen analyseren, is technische data van de telefoon nodig.

Toegang tot de data: toegang, tot de door de app verzamelde data op de beveiligde CBS Cloud-server, is alleen mogelijk voor de specifieke ontwikkelaars binnen DVL die de data beheren en verbeteringen doorvoeren aan de functionaliteit van de ODIN app. De ontwikkelaars maken ook exports van de data naar een beveiligde CBS gebied. De toegang tot dit gebied is beperkt tot een specifieke groep medewerkers die analyses op de data gaat uitvoeren. De autorisatieprocedure van het CBS op toegang tot data is hierop van toepassing. CBS medewerkers zijn tevens vanuit de CBS gedragscode geïnstrueerd hoe met vertrouwelijke informatie om te gaan. Hiermee wordt geborgd dat gegevens uitsluitend voor de benoemde interne doeleinden worden gebruikt en de AVG wordt gevolgd. Zie ook bijlage: "Basisregels Informatiebeveiliging CBS". Toegang tot de data van de app op een smartphone is alleen mogelijk door de gebruiker van de Smartphone/deelnemer van de veldtest. De deelnemer, die op vrijwillige basis meedoet, krijgt een gebruikersnaam en password om mee te mogen doen aan de veldtest. Indien deze inloggegevens door een tweede persoon worden gebruikt en/of op een ander device, kan daarmee geen opgeslagen data van de eerste persoon en/of device ingezien worden. De app reageert dan alsof het de eerste keer is dat ingelogd wordt met deze inloggegevens.

Verder is de toegangsbeveiliging van de smartphone hier ook van toepassing. Na de veldtest wordt de deelnemer verzocht om de app te verwijderen. Daarmee wordt ook de data van de verplaatsingen verwijderd.

Beveiliging: de geolocatiegegevens en technische data van de smartphone wordt encrypted in realtime doorgestuurd naar een beveiligde CBS Cloud server en vervolgens encrypted naar een beveiligd CBS gebied. De persoonsgegevens worden bij binnenkomst volgens de bepalingen uit de algemene CBS PIA behandeld.

Geheimhouding: Na binnenkomst bij het CBS zijn de verzamelde persoonsgegevens zijn niet herleidbaar tot een identificeerbaar persoon. Persoonsgegevens worden geanonimiseerd. Hier gelden de bepalingen uit de algemene CBS PIA. Zie ook bijlage: Persoonsgegevens, anonimiseren en pseudonimiseren.

Toetsing: jaarlijks vinden op het CBS audits plaats op het gebied van: informatiebeveiliging (ISO 27001), kwaliteit (ISO 9001) en de privacybescherming (Privacy Control Framework van NOREA). De certificaten hiervan zijn te vinden op de website van CBS: [iso-en-privacycertificering](#).

2. Persoonsgegevens i

Som alle categorieën van persoonsgegevens op die worden verwerkt. Geef per categorie van persoonsgegevens tevens aan op wie die betrekking hebben. Deel deze persoonsgegevens in onder de typen: gewoon, bijzonder, strafrechtelijk en wettelijk identificerend.

Er worden voor de veldtest de nodige data gebruikt. Het gaat om antwoorden op de startvragenlijst, waarbij geboortedatum en het geslacht van de respondent met meest identificerend zijn. De overige vragen in de startvragenlijst hebben met name betrekking op bezit en gebruik van vervoermiddelen. Daarnaast worden de door sensoren automatisch verzamelde informatie over de geolocatie en data van de door de respondent gebruikte smartphone zoals het batterijverbruik gebruikt. Zie bijlage voor de volledige lijst: "Variabelen gebruikt in Veldtest".

Er is geen sprake van bijzondere persoonsgegevens. De locatiegegevens vergroten wel de kans op onthulling van individuele respondenten en zijn in die zin dus privacygevoelig.

3. Gegevensverwerkingen i

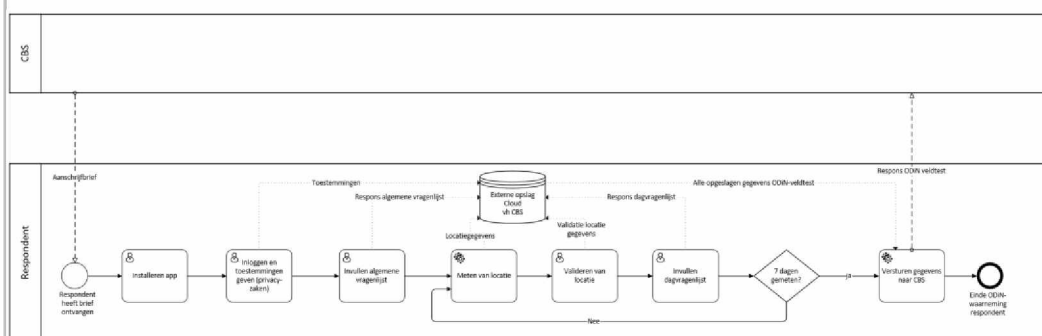
Geef alle voorgenomen gegevensverwerkingen weer.

¹ CAPI = Computer Assisted Personal Interview; CATI = Computer Assisted Telephone Interview; CAWI = Computer Assisted Web Interview; PAPI = Pen-and-Paper Personal Interview

Het dataverloop van de ODin app wordt hieronder beschreven (zie ook figuur 1).

De respondent krijgt een brief met inloggegevens (gebruikersnaam en password) en vult na het installeren van de ODin app deze gegevens in en accepteert de voorwaarden van het gebruik van de app en het ter beschikking stellen van zijn gegevens aan het CBS. De gegevens van de respondent worden geverifieerd op de CBS Cloud-server. Op deze Cloud-server hebben alleen de 3 app-ontwikkelaars van het CBS toegang. De respondent vult, voordat het reisgedrag wordt gemeten, eerst de algemene vragenlijst in. Deze gegevens worden encrypted verstuurd naar de CBS Cloud-server en daar opgeslagen. De algemene vragenlijst is leeg en bevat dus geen gegevens van de desbetreffende respondent. Hierdoor is er geen privacy-risico mochten de inloggegevens in verkeerde handen vallen. De ODin app monitort vervolgens 7 dagen lang het reisgedrag van de respondent (zie bijlage: Verplaatsingsonderzoek met de app: 1 versus 7 dagen waarnemen).

De geolocatiegegevens en gegevens van de smartphone worden in realtime opgeslagen op de CBS Cloud-server. De gegevens worden ook encrypted verstuurd naar de CBS Cloud-server. Na elke gemonitorde dag vult de respondent de gegevens aan die betrekking hebben op zijn reisgedrag van die dag. Ook deze gegevens worden encrypted verstuurd en op de CBS Cloud-server opgeslagen.



Figuur 1: Dataverloop ODin Veldtest

Na 7 dagen houdt de veldtest op en er kan niks meer ingevuld worden en de app stopt met monitoren van het reisgedrag. Na deze periode wordt aan de respondent verzocht de ODin app te verwijderen. De privacygevoelige gegevens zijn daarmee van de smartphone verwijderd.

Na de veldtest wordt alle respons encrypted verstuurd naar het CBS. De data wordt op een beveiligd gebied van het CBS geplaatst. Op dit gebied krijgen alleen degenen toegangsrechten die bij de analyse van de data betrokken zijn. De autorisatieprocedure van het CBS op toegang tot data is hierop van toepassing. De data wordt ontdaan van persoonlijke gegevens voordat het op enige manier buiten dit gebied komt. De algemene PIA van het CBS is hierop van toepassing.

4. Verwerkingsdoeleinden i

Beschrijf de doeleinden van de voorgenomen gegevensverwerkingen.

Algemeen: Het CBS heeft tot taak het van overheidswege verrichten van statistisch onderzoek ten behoeve van praktijk, beleid en wetenschap en het openbaar maken van de resultaten van deze statistische onderzoeken (artikel 3 CBS-wet) en is verantwoordelijk voor het uitvoeren van statistische Europese verordeningen (artikel 4 CBS-wet). Het CBS verwerkt daartoe gegevens voor statistische en wetenschappelijke doeleinden, voor het verbeteren van de kwaliteit van de statistiek en voor het ontwikkelen van nieuwe statistische informatie. Tevens kan het CBS in incidentele gevallen statistische werkzaamheden voor derden verrichten (artikel 5 CBS-wet).

Specifiek: De data die in de veldtest verzameld wordt, gaat niet gebruikt worden voor statistische doeleinden, maar wordt gebruikt voor: 1) verbetering van de app; 2) het verbeteren van de benaderstrategie van de toekomstige respondent en 3) voor het bepalen welke data nodig en overbodig is om te gebruiken voor het reguliere ODIN onderzoek.

Door de in de veldtest verzamelde data te analyseren is het mogelijk om inzicht te krijgen en keuzes te maken welke data minimaal nodig zijn voor het optimaal laten functioneren van de app en welke data nodig is om uiteindelijk het reguliere ODIN onderzoek te kunnen doen. Op deze manier wordt dataminimalisatie bereikt, kan de privacy borging optimaal worden ingericht, en ook heel belangrijk, kan de responsdruk omlaag worden gebracht waardoor de respondent minder hoeft in te vullen (Zie voor verdere detaillering bijlage: Analyseplan CBS verplaatsingen app 2022-2023).

5. Betrokken partijen i

Benoem welke organisaties betrokken zijn bij welke gegevensverwerkingen. Deel deze organisaties per gegevensverwerking in onder de rollen: verwerkingsverantwoordelijke, verwerker, verstrekker en ontvanger. Benoem tevens welke functionarissen binnen deze organisaties toegang krijgen tot welke persoonsgegevens.

Dit traject is uitgevoerd door het team Vragenlijstontwikkeling en -bouw van de sector DVZ, de teams van Methodologie Den Haag en Heerlen van de sector DRD, de sector Verkeer en vervoer, allen onderdeel van het CBS. Het CBS is verwerkingsverantwoordelijke. De Universiteit van Utrecht is de kennispartner van het CBS en speelt een rol in de analyse van de data van de veldtest. Met de Universiteit van Utrecht is een raamovereenkomst gemaakt die afspraken bevatten omtrent het gebruik van privacy gevoelige data.

6. Belangen bij de gegevensverwerking i

Beschrijf alle belangen die de verwerkingsverantwoordelijke en anderen hebben bij de voorgenomen gegevensverwerkingen.

Het CBS en de betrokken medewerkers willen de processen zo effectief en efficiënt mogelijk uitvoeren met het bijhouden/verbeteren van de kwaliteit t.b.v. het ODIN onderzoek.

De voorgenomen verwerking is noodzakelijk om zowel het algoritme en daarmee de functionaliteit als de gebruiksvriendelijkheid van de app te verbeteren. Dat is zowel in het belang van het CBS als die van de respondenten. De app zal de lastendruk op de respondent verlagen ten op zichte van de huidige waarneemmethode (de internetvragenlijst).

7. Verwerkingslocaties i

Benoem in welke landen de voorgenomen gegevensverwerkingen plaatsvinden.

Nederland (CBS in Den Haag en CBS in Heerlen).

8. Techniek en methode van gegevensverwerking



Beschrijf op welke wijze en met gebruikmaking van welke (technische) middelen en methoden de persoonsgegevens worden verwerkt. Benoem of sprake is van (semi-)geautomatiseerde besluitvorming, profilering of big data-verwerkingen en, zo ja, beschrijf waaruit een en ander bestaat.

Er is geen sprake van (semi-)geautomatiseerde besluitvorming of profilering en er worden geen gegevens gepubliceerd.

Zie bijlage Analyseplan CBS verplaatsingen app 20220-2023 voor de te gebruiken analysetechnieken op de data.

9. Juridisch en beleidsmatig kader



Benoem de wet- en regelgeving, met uitzondering van de AVG en de Richtlijn, en het beleid met mogelijke gevolgen voor de gegevensverwerkingen.

Buiten de AVG is hier geen specifieke wetgeving van toepassing. De data komt op de CBS Cloud server en werkgebieden van het CBS. Verder is er een raamovereenkomst met de Universiteit van Utrecht. De AVG benoemt real time Geolocatiegegevens niet met naam en toenaam maar vallen wel onder locatiegegevens. Daarom is er voor gekozen om de Geolocatiegegevens net zo te behandelen als alle andere privacygevoelige data.

10. Bewaartermijnen



Bepaal en motiveer de bewaartermijnen van de persoonsgegevens aan de hand van de verwerkingsdoeleinden.

Volgens de CBS Selectielijst² valt de ODiN Veldtest onder het Proces: "Het ontwikkelen en opstellen van methodieken voor (het uitvoeren van) statistisch onderzoek" Voor dit proces is de afspraak gemaakt dat de data 2,5 jaar bewaard mag worden na vaststelling of publicatie van de definitieve versies van de beschrijvingen.

B. Beoordeling rechtmatigheid gegevensverwerkingen

Beoordeel aan de hand van de feiten zoals vastgesteld in onderdeel A of de voorgenomen gegevensverwerkingen rechtmatig zijn. Het gaat hier om de beoordeling van de juridische rechtsgrond, noodzaak en doelbinding van de gegevensverwerkingen. Beoordeel tevens de wijze waarop invulling wordt

²De selectielijst is organisatie-georiënteerd en primair bedoeld om de archivistische neerslag van het CBS als archiefvormer te kunnen waarderen. Onder archivistische neerslag wordt verstaan: alle documenten en andere gegevens (incl. data) die worden ontvangen, gegenereerd of opgesteld in het kader van de taakuitvoering, ongeacht hun vorm.

gegeven aan de rechten van de betrokkenen. Voor dit onderdeel van de PIA is in het bijzonder juridische expertise nodig.

11. Rechtsgrond



Bepaal op welke rechtsgronden de gegevensverwerkingen worden gebaseerd.

De verwerking in de Veldtest beroept zich op de wettelijke taak en wettelijke verplichtingen van het CBS. Specifiek: "Het optimaliseren van processen om de invulling van de taak van algemeen belang zo goed mogelijk in te kunnen invullen".

12. Bijzondere persoonsgegevens



Indien bijzondere of strafrechtelijke persoonsgegevens worden verwerkt, beoordeel of één van de wettelijke uitzonderingen op het verwerkingsverbod van toepassing is. Bij verwerking van een wettelijk identificatienummer beoordeel of dat is toegestaan.

Niet van toepassing.

De extra procedurele beveiligingsmaatregelen, die het CBS heeft genomen om misbruik van bijzondere persoonsgegevens te voorkomen, zijn in deze PIA benoemd.

13. Doelbinding



Indien de persoonsgegevens voor een ander doel worden verwerkt dan oorspronkelijk verzameld, beoordeel of deze verdere verwerking verenigbaar is met het doel waarvoor de persoonsgegevens oorspronkelijk zijn verzameld.

Niet van toepassing, omdat gegevens worden verzameld voor het oorspronkelijke doel, al is dat niet voor het maken van statistiek.

14. Noodzaak en evenredigheid



Beoordeel of de voorgenomen gegevensverwerkingen noodzakelijk zijn voor het verwezenlijken van de verwerkingsdoeleinden. Ga hierbij in ieder geval in op proportionaliteit en subsidiariteit.

- a. Proportionaliteit: staat de inbreuk op de persoonlijke levenssfeer en de bescherming van de persoonsgegevens van de betrokkenen in evenredige verhouding tot de verwerkingsdoeleinden?**
- b. Subsidiariteit: kunnen de verwerkingsdoeleinden in redelijkheid niet op een andere, voor de betrokkene minder nadelige wijze, worden verwezenlijkt?**

De veldtest is de meest directe manier om informatie te verzamelen om te komen tot een optimale structurele oplossing, met minimale belasting en minimale privacy risico's in vergelijking tot het reguliere en ODin onderzoek dat via CAWI wordt uitgevoerd. Deze cawi-uitvraag is belastend voor de respondent, omdat er veel detail wordt uitgevraagd. Met behulp van de app wordt deze informatie automatisch geregistreerd en hoeft de respondent deze alleen maar aan te vullen en te bevestigen. Daarnaast is bekend dat de cawi-uitvraag foutgevoelig is: een verplaatsing of stop wordt snel vergeten of verkeerd geclassificeerd door de respondent. Om een goede classificatie te kunnen maken, is in deze fase van de veldtest alle locatiedata nodig. Achteraf kan dan geanalyseerd worden of een andere classifieerder beter zou hebben gewerkt. De data is dus nodig om de app te verbeteren, en om de kwaliteit van de output, die in de toekomst met de app moet

worden gemaakt, te borgen. De door de respondent gebruikte route voor de verplaatsing is belangrijke output van het reguliere ODIN. Hier wordt in de huidige internetvragenlijst echter niet naar gevraagd, omdat dit te complex is. De route wordt daarom achteraf geschat. Met hulp van de data die met de app wordt verkregen, kan wel de exacte route worden bepaald, terwijl de respondent minder belast wordt. De veldtest is nodig om de proportionaliteit en subsidiariteit van de toekomstige respondent van het nieuwe ODIN onderzoek te borgen. Het CBS ziet geen andere, voor de betrokkene minder nadelige, wijze om de verwerkingsdoeleinden te verwezenlijken.

15. Rechten van de betrokkene



Geef aan hoe invulling wordt gegeven aan de rechten van betrokkenen. Indien de rechten van de betrokkene worden beperkt, bepaal op grond van welke wettelijke uitzonderingen dat is toegestaan.

De Respondenten hebben toestemming gegeven voor deze aanpak en zijn (via de instructie op de app en de aanschrijfbrief) op de hoogte van de werking van de ODIN app. De respondenten zijn zelf actief betrokken bij het downloaden en aanzetten van de ODIN app. Risico's zijn, naast de genomen maatregelen, afdoende geadresseerd.

C. Beschrijving en beoordeling risico's voor de betrokkenen

Beschrijf en beoordeel de risico's van de voorgenomen gegevensverwerkingen voor de rechten en vrijheden van de betrokkenen. Houd hierbij rekening met de aard, omvang, context en doelen van de gegevensverwerking zoals in onderdeel A en B zijn beschreven en beoordeeld. Het gaat hierbij overigens niet om de risico's van de verwerkingsverantwoordelijke zelf.

16. Risico's



Beschrijf en beoordeel de risico's van de gegevensverwerkingen voor de rechten en vrijheden van betrokkenen. Ga hierbij in ieder geval in op:

- a. welke negatieve gevolgen de gegevensverwerkingen kunnen hebben voor de rechten en vrijheden van de betrokkenen;
- b. de oorsprong van deze gevolgen;
- c. de waarschijnlijkheid (kans) dat deze gevolgen zullen intreden;
- d. de ernst (impact) van deze gevolgen voor de betrokkenen wanneer deze intreden.

Negatieve gevolgen voor de rechten en vrijheden van de respondent

- Het op elk moment kunnen vaststellen waar een respondent zich bevindt.

Wat kan negatieve gevolgen veroorzaken (oorsprong risico's)

- de inloggegevens van de respondent kunnen door een derde worden onderschept en gebruikt worden. Hierdoor kan data worden gegenereerd die onbruikbaar is omdat deze niet hoort bij het reisgedrag van de geselecteerde persoon.
- de respondent kan zijn/haar smartphone verliezen, waarbij een derde toegang tot de toestel kan krijgen. Net zoals bij het eerste punt kunnen dan gegevens worden verzameld die niet bij de juiste persoon horen. Voor de respondent zit het risico in het feit dat, in geval van diefstal, door middel van de app inzicht kan worden verkregen in het reisgedrag van de respondent. Dit is echter ook het geval bij andere apps waarbij gebruik wordt gemaakt van Geolocatiegegevens.
- De Cloud server kan door onbevoegden benaderd worden.

De kans dat gegevens misbruikt worden

- Als de instructies correct worden gevolgd, is met de benoemde beveiligingsmaatregelen, de kans dat gegevens beschikbaar komen voor derden uiterst klein.
- Alleen geautoriseerde medewerkers die bij het project betrokken zijn, hebben toegang tot de data en hebben daarnaast een geheimhoudingsplicht met betrekking tot persoons- en bedrijfsinformatie waarmee men in aanraking komt.
Te weten: 3 developers, 4 helpdeskmedewerkers (hebben geen toegang tot Geolocatiegegevens), (max.) 10 Analisten.

De impact als (geolocatie)gegevens misbruikt worden

- Mocht het toch voorkomen dat (geolocatie)gegevens, vooral vanuit privé activiteiten, beschikbaar komen voor derden, dan kan de impact groot zijn voor de betrokken individuen. Op basis van de genomen maatregelen (zie punt 17 van deze PIA) resteert een minimaal restrisico.

D. Beschrijving voorgenomen maatregelen

In onderdeel D wordt gezien welke maatregelen kunnen worden getroffen om de in onderdeel C erkende risico's te voorkomen of te verminderen. Welke maatregelen in redelijkheid worden getroffen is een belangenafweging van de wetgever of verwerkingsverantwoordelijke. Voor dit onderdeel van de PIA is, als het gaat om beveiligingsmaatregelen, expertise over informatiebeveiliging belangrijk.

17. Maatregelen



Beoordeel welke technische, organisatorische en juridische maatregelen in redelijkheid kunnen worden getroffen om de hiervoor beschreven risico's te voorkomen of te verminderen. Beschrijf welke maatregel welk risico aanpakt en wat het restrisico is na het uitvoeren van de maatregel. Indien de maatregel het risico niet volledig afdekt, motiveer waarom het restrisico acceptabel is.

Voor het vaststellen van de noodzakelijke maatregelen hanteert het CBS de ISO 27001/2 norm en de BIR als uitgangspunt. In dat kader zijn in de (voorgestelde) verwerkingsprocessen reeds maatregelen opgenomen om privacy te beschermen. Het totaal overzicht aan maatregelen ziet er als volgt uit:

- Gebruik van de app is afgeschermd met inloggegevens die vooraf worden toegestuurd.
- Mocht de aanschrijfbrief met inloggegevens in verkeerde handen vallen leidt het inloggen met die gegevens niet tot het verkrijgen van persoonsgegevens over de respondent. In de vragenlijst zijn geen persoonsgegevens verwerkt.
- beveiliging van de app gebeurt met een gebruikersnaam en password.
- op bovenstaande beveiliging is een penetratietest uitgevoerd.
- encryptie van berichtenverkeer tussen de smartphone en de Cloud server en encryptie tussen de Cloud server en beveiligde omgeving van het CBS.
- De Cloud server is in ruime mate beveiligd tegen hacken.
- er wordt maximaal 7 dagen informatie verzameld. Na 8 dagen wordt de dataverzameling vanzelf gestopt door afbreken van de synchronisatie van de app met de server. De respondent wordt gevraagd de app te verwijderen nadat de data is verstuurd.
- de data is alleen toegankelijk voor geauthenticeerde personen.
- de app heeft geen optie om data met andere apps te delen. De GPS-data gegenereerd door de app kan niet voor andere doeleinden gebruikt worden.
- er zijn geen oplossingen vanuit het CBS voor het voorkomen van verlies van een toestel. De eigenaar van het toestel is daarvoor zelf verantwoordelijk. Producenten van mobiele telefoons bieden hierbij ondersteuning door de optie om apps in te schakelen: bij iOS de App "Find my iPhone", bij WindowsPhone "Find my Phone", of bij Android de dienst "Android Device Manager".
- In het verdere traject v.w.b. de verwerking van de data wordt, de data van de respondent geanonimiseerd of gepseudonimiseerd. Dit voorkomt verdere privacy-issues.

Met de genomen maatregelen zijn naar oordeel van het CBS de resterende risico's acceptabel. De maatregelen rondom verwerking van gegevens van de veldtest (na de uitvoering van de Veldtest, waarvoor deze PIA is opgesteld) vallen onder de algemene CBS PIA, omdat het verwerkingsproces van de veldtest niet afwijkt van reguliere CBS processen waarvoor al een PIA bestaat.

Tot Slot

Aanvullende maatregelen die voortkomen uit de evaluatie van de veldtest, worden gebruikt om de veldtest verder te optimaliseren, mits het CBS ervoor kiest om deze functionaliteit structureel te gaan toepassen. In dat geval wordt deze PIA verder geoptimaliseerd, zodat daarmee de privacy borging rondom dit aanvullende proces ook afdoende is geborgd. Dit in aanvulling op de CBS PIA, waarin dit proces niet is meegenomen.



E. BIJLAGEN

Bijlage 1: Variabelen gebruikt in Veldtest

Voor de veldtest gaat het om de volgende gegevens:

- **CRUD, voor elke variabele** (Created at, Read at, Updated at, Deleted at)
 - o Timestamp met tijdzone
- **Automatisch gegenereerde primary keys** (voor elke variabele)
 - o Integer
- **Gebruiker** (privacy gevoelig)
 - o Gebruikersnaam – Tekst
 - o Wachtwoord – Tekst – Hash van het wachtwoord
- **Toestel** (Velden niet van toepassing op iOS of Android toestel worden leeg gelaten)
 - o Batterijniveau – big integer – batterijpercentage van 0% tot 100%
 - o Merk – Tekst - merk van het toestel
 - o Model – Tekst - model van het toestel
 - o Versie – Tekst – Versie van het besturingssysteem
 - o Productnaam – Tekst - specifieke productnaam van het toestel
 - o Android-ID – Tekst – Android ID van het toestel
 - o SDK-versie – Tekst – Versie van Android SDK
 - o Secure-ID – Tekst – Secure ID van de Android settings
 - o Width – Numeric – breedte van het toestel scherm
 - o Height – Numeric – hoogte van het toestel scherm
- **Verplaatsing**
 - A. Privacy gevoelig:
 - o Start_time – Bigint – tijdstip (uren en minuten) op welk een verplaatsing start
 - o End_time – Bigint – tijdstip (uren en minuten) op welk een verplaatsing eindigt
 - B. Gewoon:
 - o confirmed – Boolean – Bevestiging van een verplaatsing door respondent
 - o movement_category_id – Bigint - ID van de categorie waaronder de verplaatsing valt
- **Locatie** (via gebruikt de data vanuit 'sensor geolocatie')
 - A. Privacy gevoelig:
 - o Start_time – Bigint – tijdstip (uren en minuten) op welk een locatie start
 - o End_time – Bigint – tijdstip (uren en minuten) op welk een locatie eindigt
 - o lat – numeric - latitude coördinaten van het toestel
 - o lon – numeric - longitude coördinaten van het toestel
 - B. Gewoon:
 - o Reason_id – Bigint - ID van de categorie waaronder de bezoek-reden van de locatie valt
 - o confirmed - boolean - Bevestiging van een locatie door respondent
 - o name - tekst – adresnaam volgens de google-api bijbehorende latitude en longitude
- **Tracked LatLong** (via gebruikt de data vanuit 'sensor geolocatie')
 - B. Gewoon:
 - o Altitude – numeric – Altitude bijbehorend de latitude en longitude
 - o Mapped_date – bigint – datum van de geolocatie
 - Tracked Day (gewoon)
 - o Tracked_day_id – bigint – nummer van dag binnen survey van een gebruiker
 - o day – bigint – datum van de dag
 - o confirmed - boolean - Bevestiging van een locatie door respondent
 - o choice-id – bigint – keuze gemaakt in de dialogue bij het afsluiten
 - o validated – numeric – percentage of validated time

- o unvalidated – numeric - percentage of unvalidated time
- o missing – numeric – percentage of missing time
- o good_enough – boolean – om aan te geven of de dag compleet genoeg is (80%)

- **(Error)Logs** (gewoon)
 - o Message – tekst – error message
 - o Description – tekst – error beschrijving
 - o Type - tekst – type error

- **Google-API logs** (via gebruikt de data vanuit 'sensor geolocatie')
 - A. Privacy gevoelig:
 - o lat – numeric - latitude die wordt opgezocht
 - o lon – numeric – longitude die wordt opgezocht
 - B. Gewoon:
 - o Allowed – boolean – is user allowed to make this call
 - o Query – tekst – query die wordt gedaan bij api-call
 - o Request – tekst – type van google api request (geosearch/textsearch)

- **Sensor geolocatie** (via GPS sensor)
 - B. Gewoon:
 - o Sensor-type - tekst – type sensor gebruikt (fused, gps, cellular)
 - o Snelheid - numeric – snelheid gemeten door GPS satellieten in km/u
 - o Bearing – numeric – hoek van gps signaal
 - o Provider – tekst – type sensor provider (gps, fused, wifi, cellular)
 - o Accuracy – numeric – (nauwkeurigheid) afwijking van geolocatie in meters
 - o Date – bigint – datum en tijd van verkregen geolocatie

Bijlage 2: Analyseplan CBS verplaatsingen app 2022 - 2023

I.v.m. de privacywetgeving zijn de namen van de betrokkenen zwart gemaakt.

Versie: 06-11-2022

Auteurs: [REDACTED]

Samenvatting: Dit analyseplan geeft de onderzoeksvragen, analyses, beoogde resultaten en bijbehorende planning bij de veldtest AVA-EVA 2022 gekoppeld aan de app Onderweg in Nederland. We sluiten in de beoogde resultaten zoveel mogelijk aan op de belangrijkste projectonderdelen. We splitsen de analyse in twee stromen: een snelle analyse voor besluitvorming binnen project Innovatie ODIN2 en een uitgebreide analyse in het CBS-UU programma WIN. De snelle analyse leidt tot aanbevelingen en tussenrapportages tegen het einde van Q1 2023. De uitgebreide analyses lopen door tot na de zomer van 2023. De uitgebreide analyse dient ook uitsluitend te geven over hypothesen die van belang zijn voor de ODIN app PIA.

1. Achtergrond

Dit is niet de eerste versie van een analyseplan. In 2020 en 2021 zijn al verschillende plannen opgesteld. In het voorliggende plan zijn deze eerdere documenten verwerkt. Het is dus een op zichzelf staand document.

De analyses zijn gebaseerd op een grootschalige veldtest die plaatsvindt in Q4 2022 en Q1 2023. Het ontwerp hiervoor is beschreven in [REDACTED] (2022). De veldtest dient ter onderbouwing van keuzes in het ontwerp en de methodologie achter smart surveys die locatiemetingen inzetten, en in het bijzonder onderzoek Onderweg in Nederland (ODIN). Met RWS wordt samengewerkt binnen project Innovatie ODIN 2. De veldtest heeft ODIN als focus maar is daarvan geen exacte kopie.

Innovatie ODIN2 kent de volgende onderdelen/deelprojecten:

- App frontend UI: Algemene usability, introvragenlijst, techniek achter locatiemetingen en ingebouwde beslisregels;
- App frontend respondentacties: De implementatie en usability van toegestane acties die respondenten kunnen uitvoeren op voorgestelde stop-verplaatsing-rit beslisregels;
- App backend: De app database, server, interactie met Google API en queries voor monitoring;
- Logistiek: De aansluiting van de app backend op de DVZ logistiek inclusief monitoring en analyse;
- Benaderstrategie: De werving- en motivatiestrategie;
- Verwerkingsmethodologie: De keuze en methodologie achter beslisregels, gaafmaak procedures, imputatie van missende locatiedata, en afleiding van statistieken, inclusief inrichting van een eventueel extra rustpunt tussen de app backend en SVV;
- Juridisch-ethisch: De DPIA en eventuele afwegingen via de CBS ethische commissie;

Deze projectonderdelen zijn een uitwerking van de vier verschillende niveaus zoals die in ESSnet Smart Surveys zijn benoemd: IT en architectuur, methodologie, logistiek en juridisch. De IT en architectuur is hier gesplitst in app frontend UI en app backend. Methodologie is gesplitst in app frontend respondentacties, benaderstrategie en verwerking.

We bespreken eerst de doelen per onderdeel op hoofdlijnen, schetsen dan de data die worden verzameld in de veldtest, detailleren vervolgens naar analysevragen, en koppelen daaraan projectleden en een planning. In de analyses kiezen we voor een splitsing voor snelle en gedetailleerde analyses. De snelle analyses dienen voor besluitvorming rond ODIN en communicatie richting RWS. De uitgebreide analyses dienen de inrichting van de benaderstrategie en verwerkingsstraat in meer algemene zin.

2. Doelen op hoofdlijnen

De zeven projectonderdelen volgen elk een aantal mijlpalen en planning op hoofdlijnen en kennen een hoofdverantwoordelijke. Tabel 1 geeft de hoofdverantwoordelijke en de mijlpalen per onderdeel. [REDACTED] treedt op als projectleider overall.

Tabel 1: Mijlpalen en hoofdverantwoordelijke per projectonderdeel.

Onderdeel	Verantwoordelijke	Mijlpalen
App UI	[REDACTED]	<ol style="list-style-type: none"> 1. Baseline app (week ononderbroken meten plus tonen stop-verplaatsingen) 2. Baseline app aangevuld met respondentacties 3. App gereviseerd op basis van usability test 4. App met definitieve beslisregels en respondentacties 5. App gereviseerd op basis van usability tests;
App resp acties	[REDACTED]	<ol style="list-style-type: none"> 1. Aanbevelingen respondentacties usability test 2. Advies UI respondentacties op basis van veldtest 3. Aanbevelingen respondentacties usability test met definitieve beslisregels
App backend	[REDACTED]	<ol style="list-style-type: none"> 1. Baseline backend inclusief stress test 2. Standaard queries en exports database 3. Gereviseerde backend op basis van veldtest 4. Aansluiting backend op verwerking, inclusief rustpunten
Benaderstrategie	[REDACTED]	<ol style="list-style-type: none"> 1. Aanbevelingen wervingsstrategie via veldtest 2. Aanbevelingen communicatie- en motivatiestrategie via veldtest
Logistiek	[REDACTED]	<ol style="list-style-type: none"> 1. Voorlopige helpdeskprotocols 2. Voorlopig monitoring en analyse dashboard 3. Aanbevelingen logistiek en inregelen en aansluiting Phoenix 4. Aanbevelingen definitieve helpdeskprotocols 5. Definitief monitoring en analyse dashboard
Verwerking	[REDACTED]	<ol style="list-style-type: none"> 1. Veldtest analysebestand 2. Aanbeveling beslisregels stop-verplaatsing-ritten 3. Aanbeveling toegestane respondentacties 4. Aanbeveling steekproefontwerp (meetperiode) 5. Scripts voor microverwerking 6. Scripts voor macroverwerking
Juridisch	[REDACTED]	<ol style="list-style-type: none"> 1. Voorlopige DPIA voor veldtest 2. Aanbevelingen meetperiode en data minimalisatie 3. Definitieve DPIA

De veldtest is zeker niet het enige middel om de mijlpalen te bereiken; er worden parallel kleinere usability tests en functionele tests uitgevoerd. De veldtest is wel voor alle onderdelen onmisbaar in het behalen van de mijlpalen.

We geven per onderdeel de doelen van de veldtest en de opvolgende analyses.

- App frontend UI:
 - Leren welke UI elementen verbetering behoeven;
 - Leren hoe prestaties afhangen van type devices en waar grenzen liggen qua inzetbaarheid;
- App frontend respondentacties:
 - Leren of en hoe de verschillende respondentacties moeten worden toegepast;
 - Bepalen hoe noodzakelijke respondentacties het meest gebruiksvriendelijk ingericht kunnen worden;
- App backend:
 - Leren of de backend meer robuust gemaakt moet en kan worden;
 - Leren of en hoe het gekozen datamodel gebruiksvriendelijker gemaakt kan worden voor analyses;
 - Verbeteren van paradata logging van app gebruik;
 - Aanbevelingen doen voor een eventueel extra rustpunt tussen backend en verdere verwerking;
- Logistiek:
 - Vaststellen hoe helpdeskprotocols en -acties ingericht moeten worden;
 - Implementeren van een monitoring en analyse omgeving voor zowel algemene voortgang als helpdeskvragen;
 - Vaststellen welke benaderacties en -stappen automatisering behoeven via Phoenix;
- Benaderstrategie:

- Bepalen van de meest effectieve combinatie van een app en online ODIN vanuit het perspectief van gebalanceerde wervingskansen en responskansen;
- Vaststellen of en welke gedifferentieerde (doelgroepen) strategie de voorkeur heeft;
- Evalueren welke communicatie vooraf en gedurende veldwerk het meest effectief is in termen van data kwaliteit;
- Vaststellen welke duur van meten optimaal is met het oog op wervingskansen en voortijdig stoppen;
- Verwerkingsmethodologie:
 - Bepalen welke beslisregels de meest accurate stop-verplaatsing-rit onderverdeling geven;
 - Bepalen welke respondentacties noodzakelijk zijn;
 - Voorstellen hoe om te gaan met missende data en met uitbijters, oa via beschikbare POI data en opentripplanner;
 - Duiden van verschillen tussen de reguliere ODIN en een app-gebaseerde ODIN;
 - Inrichten van verwerkingsstappen en bijbehorende scripts op microniveau;
- Juridisch-ethisch:
 - Onderbouwen van een definitieve data protection impact assessment (DPIA) voor een app-gebaseerde ODIN;
 - Legitimeren van de keuze in het aantal te rapporteren/meten dagen door respondenten;
 - Onderbouwen van keuzes in in-device verwerking versus in-huis verwerking;

3. Veldtestontwerp en data

Het ontwerp van de veldtest is in detail beschreven in [REDACTED] (2022). We geven hier enkel de belangrijkste details. De veldtest heet AVA22.

De veldtest bestaat uit twee experimenten met aparte steekproeven. Het eerste experiment wordt uitgevoerd op een steekproef van voormalige ODIN 2022 respondenten die een serieverplaatsingen hebben gemaakt en/of een of meer multi-modale verplaatsingen. Een serieverplaatsing is een aaneenschakeling van verplaatsingen met ertussen relatief korte, meestal beroepsmatige, stops. Een multi-modale verplaatsing is een verplaatsing waarbij meerdere vervoermiddelen worden gebruikt. Het tweede experiment is een verse steekproef die verder willekeurig wordt onderverdeeld naar een drietal experimentele condities. Voor de respondenten op het tweede experiment vindt ook een evaluatie plaats.

Het kleinere, eerste experiment vraagt respondenten om zowel de app te gebruiken als een dag de ODIN online vragenlijst in te vullen. Respondenten wordt gemeld dat de app-gebaseerde ODIN een nieuwe, experimentele optie is die met dit onderzoek wordt vergeleken. Respondenten wordt daarom gevraagd om voor het invullen van de online vragenlijst niet eerst in de app te kijken. Dit experiment is dus vooral gericht op datakwaliteit en evaluatie van antwoorden en antwoordgedrag. Het leent zich om verschillen in rapportage te schatten. Daarbij moet wel bedacht worden dat de respondent de app als experimenteel zal zien, gegeven de instructie.

Het tweede, grotere experiment nodigt een steekproef uit om mee te doen aan ODIN en voert de app op als een volwaardige optie. Het kent drie experimentele variaties:

1. Online vragenlijst als keuze: Het moment waarop de bestaande online ODIN vragenlijst als alternatief wordt aangeboden wordt willekeurig gevarieerd tussen direct bij uitnodiging of tijdens het eerste of tweede rappel;
2. Respondentacties: De mogelijke acties die respondenten kunnen uitvoeren op aan hen voorgelegde stop-verplaatsingen met ingebouwde beslisregels varieert willekeurig. Een helft krijgt alle mogelijke acties en een helft kan alleen stops en verplaatsingen verwijderen. De beslisregels worden zo ingesteld dat er gemiddeld genomen net wat te veel stops worden gevonden;
3. Meetperiode: De steekproef wordt willekeurig verdeeld in twee porties waarbij één deel enkel gevraagd wordt ten minste een etmaal mee te doen en een ander deel een week. Deze variatie is enkel in de brief genoemd (en daar gekoppeld aan beloning) maar in de app verder niet expliciet zichtbaar;

De drie condities worden willekeurig met elkaar gekruist. Omdat de twee varianten met toegestane respondentacties (op dit moment) om aparte apps vragen in app stores, vinden deze achter elkaar plaats in plaats van naast elkaar. De twee andere condities worden binnen de twee batches wel willekeurig verdeeld. Dit experiment richt zich naast datakwaliteit op wervingskansen en afbreekkansen.

Alle respondenten uit beide experimenten worden na afloop uitgenodigd voor een online evaluatievragenlijst. Deze uitnodiging is gelijktijdig met het versturen van de beloning.

In het ontwerp van AVA22 is het eerste experiment fase 1 en het tweede experiment omvat fasen 2 t/m 4. Hier noemen we het voor het gemak AVA22-1 en AVA22-2. De volgende data wordt verzameld en gekoppeld voor zowel AVA22-1 als AVA22-2:

- Steekproefbestand gekoppeld aan een verzameling registerkenmerken (via het SAL doelpopulatiebestand). De selectie aan registerkenmerken moet nog bepaald worden maar omvat in ieder geval standaard demografische en socio-economische variabelen, geregistreerd voertuigbezit en geregistreerd rijbewijsbezit;
- Paradata over verzending van brieven en rappels;
- Type device gebruikt voor deelname en eventuele devicewissels;
- Informatie CCI in geval van contact respondent met CBS;
- Bij respons:
 - Indicator of persoon de online evaluatievragenlijst heeft ingevuld;
 - Bij deelname evaluatie: Antwoorden op de online evaluatievragenlijst

Aanvullend is er voor AVA22-1:

- Selectie van antwoorden respondent op reguliere ODIN. De selectie moet nog gemaakt worden maar in ieder geval de relevante afgeleide ODIN statistieken;
- Responsindicator: enkel registratie in app, enkel vragenlijst, beide, geen van beide;
- In-app data (tenzij respondent dit onderdeel niet doet):
 - Locatiemetingen, ruw en verwerkt tot stop-verplaatsingen;
 - Antwoorden in-app vragenlijst;
 - Paradata over in-app antwoordgedrag;
 - Logging van inputwaardes voor sensorkeuze na eerste uur;
 - Eventuele informatie uit helpdesk emails;
- Online vragenlijstdata (tenzij respondent vragenlijst niet invult):
 - Alle data zoals gebruikelijk in de reguliere ODIN;
 - Paradata over afbreken en responstijden;

En voor AVA22-2:

- Experimentele condities waaraan steekproefpersoon is toegewezen;
- Responsindicator: deelname via app, deelname via vragenlijst, geen deelname;
- Bij deelname in-app:
 - Locatiemetingen, ruw en verwerkt tot stop-verplaatsingen;
 - Antwoorden in-app vragenlijst;
 - Paradata over in-app antwoordgedrag;
 - Logging van inputwaardes voor sensorkeuze na eerste uur;
 - Eventuele informatie uit helpdesk emails;
- Bij deelname vragenlijst:
 - Alle data zoals gebruikelijk in de reguliere ODIN;
 - Paradata over afbreken en responstijden;

4. Analyse vragen en output

We onderscheiden de eerder genoemde zeven projectonderdelen. Tabel 2 geeft aan per onderdeel of AVA22-1 en AVA22-2 worden ingezet.

Tabel 2: Inzet experimenten per projectonderdeel.

Onderdeel	AVA22-1	AVA22-2
App UI	X	
App resp acties	X	X
App backend		X
Benaderstrategie		X
Logistiek		X
Verwerking	X	X
Juridisch		X

Per onderdeel geven we de analysevragen en beoogde resultaten. Deze worden telkens gesplitst in een snelle en een uitgebreide analyse.

De analysevragen kunnen door projectverantwoordelijken nog verder gedetailleerd en aangevuld worden in de aanloop naar de veldtest dan wel tijdens de analyse van de veldtest.

Voor onderdelen App respondentacties, Benaderstrategie, Verwerking en mogelijk ook Juridisch zullen congrespapers en tijdschriftartikelen worden gemaakt. Nu al is bekend dat onderdeel Verwerking een presentatie gaat geven op NTTS 2023. Deze resultaten worden niet apart benoemd.

4.1 App user interface

Snelle analyse:

- Hoe verliep het gebruik van de app in algemene zin?
- Hoeveel tijd besteedden respondenten in de app aan de verschillende onderdelen?
- Zijn er verschillen te zien in gebruik op basis van type en leeftijd van het device?
- Hoe hangt het overschakelen van hoog- naar laagfrequente locatiemetingen (dus van 'normal' naar 'fused/balanced') samen met kenmerken van het device en het devicegebruik van de respondent?

Uitgebreide analyse:

- Hangen eventuele problemen/issues samen met kenmerken van de respondent?
- Kunnen op basis van in-app paradata gebruikersprofielen worden onderscheiden?
- Wat is de optimale inrichting van de keuze van type locatiemetingen in termen van batterijgebruik?

Resultaten:

1. Verslag met bevindingen en leerpunten voor de app frontend algemeen
2. Notitie met aanbevelingen voor batterijgebruik rond locatiemetingen

4.2 App respondentacties

Snelle analyse:

- Hoe veel aanpassingen doen respondenten gemiddeld per type actie?
- Hoeveel tijd besteedden respondenten gemiddeld per type actie?
- Of en hoe hangt dit aantal/deze tijd af van het aantal dagen in app?
- Hangt het aantal stops wat verwijderd wordt af van de variant van de app?
- Hangt het aantal redenen en vervoermiddelen wat ingevuld wordt af van de variant van de app?
- Hangt het type en aantal uitgevoerde acties af van kenmerken van de respondent?
- Werden acties correct uitgevoerd (ofwel leidden tot daadwerkelijke verandering)?
- Is er een interactie in bovenstaande analyses met type en leeftijd van het device?
- Welke verschillen zijn zichtbaar tussen app en vragenlijstmetingen als functie van respondentacties?
- Welke verschillen zijn zichtbaar tussen de twee varianten van de app als functie van respondentacties?
- Zien we verschillen in de respondent-tevredenheid in de evaluatievragenlijst tussen de twee varianten van de app als functie van respondentacties?

Uitgebreide analyse:

- Hangt het type en aantal uitgevoerde acties af van reisprofielen?
- Kunnen profielen onderscheiden worden in in-app antwoordgedrag?
- Welke kenmerken komen voor in multivariate modellen voor in-app antwoordgedrag?
- Is er een interactie in die modellen met dagen in app?
- Is er een interactie in die modellen met device?

Resultaten:

1. Verslag met beschrijvende statistieken over uitgevoerde respondentacties
2. Rapport met karakteristieken van respondentacties als functie van experimentele condities en respondentkenmerken

4.3 App backend

Snelle analyse:

- Of en wat voor technische problemen of storingen traden op in de backend?
- Hoe snel groeide de backend database als functie van aantal respondenten en dagen in app?
- Welk aandeel werd ingenomen door de verschillende componenten (locatiemetingen, paradata, overig)?

Uitgebreide analyse:

- Welke verbeteringen zijn mogelijk aan de backend om statistische queries beter uit te voeren?

Resultaten:

1. Verslag met functioneren backend en leerpunten

4.4 Benaderstrategie

Snelle analyse:

- Wat zijn app en vragenlijst wervingskansen uitgesplitst naar de meest relevante kenmerken?
 - Voor een etmaal en voor een week;
 - Bij direct online alternatief, online alternatief in eerste rappel of online alternatief in tweede rappel;
- Wat zijn kansen op volledige afronding uitgesplitst naar de meest relevante kenmerken?
 - Voor een etmaal en voor een week;
 - Bij direct online alternatief, online alternatief in eerste rappel of online alternatief in tweede rappel;
 - Bij weinig of veel toegestane respondentacties;
- Welk aandeel van de respondenten neemt contact op met CBS via telefoon en/of email om zich af te melden en/of om meer uitleg te krijgen?

Uitgebreide analyse:

- Welke kenmerken komen in multivariate modellen op als voorspellers voor werving?
- Is er een interactie met meetperiode en/of mode-strategie in deze modellen?
- Welke kenmerken komen in multivariate modellen op als voorspellers voor afronding?
- Is er een interactie met meetperiode en/of mode-strategie in deze modellen?
- Komen geïdentificeerde kenmerken ook voor in modellen voor de belangrijkste ODin kernvariabelen (gemiddelde afstand per verplaatsing, gemiddeld aantal stops, wel/niet serieverplaatsing, gemiddelde aantal vervoermiddelen per verplaatsing)?
- Welke kenmerken in multivariate modellen voorspellen voortijdig afbreken?
- Is er een interactie in al deze modellen met type en/of leeftijd van het device?
- Of en hoe zou een gedifferentieerde benadering zinvol kunnen zijn?

Resultaten:

1. Notitie met resultaten en eerste aanbevelingen uit de snelle analyse
2. Rapport met uitgebreide analyse
3. Notitie met definitieve aanbevelingen

4.5 Logistiek

Snelle analyse:

- Hoe verliep het veldwerk in algemene zin?
- Welk aandeel van respondenten zoekt contact met het CBS voor assistentie?
- Wat zijn de meest voorkomende vragen/problemen vanuit respondenten?
- Welke aanwijzingen voor verbeteringen geeft de evaluatievragenlijst?

Uitgebreide analyse:

- Welke handmatige stappen in de veldtest kostten het meeste tijd?
- Welke frequentie van monitoring en analyse is het meest efficiënt?
- Hoe kan in-app paradata het meest effectief worden ingezet in monitoring en analyse?

Resultaten:

1. Een veldwerkverslag
2. Een notitie met aanbevelingen/leerpunten voor logistiek

4.6 Verwerking

De analysevragen voor dit onderdeel worden onderverdeeld in:

1. Methode-effecten: Duiding van verschillen tussen app en vragenlijst
2. Beslisregels: Keuze van beslisregels en parameters binnen de regels
3. Respondentacties: Onderbouwing voor keuze in welke respondentacties worden toegestaan
4. Vergelijkbaarheid: Duiding van potentiële risico's in vergelijkbaarheid door verschillen tussen type devices en/of telefoongebruik

Snelle analyse:

- Methode-effecten:
 - Hoe groot zijn verschillen tussen de app- en vragenlijstmetingen voor respondenten die beide invullen?
 - Welke verplaatsingsscenario's/reisprofielen laten de grootste verschillen zien?
 - Wat is het effect van meetperiode op kernvariabelen?
 - Wat is datakwaliteit als functie van het aantal dagen in-app?
 - Wat is het aandeel missende locatiemetingen?
- Beslisregels:
 - Welke duur-radius parameters sluiten het beste aan bij acties van respondenten?
 - Wat zijn de prestaties van verschillende beslisregels?
 - Dient in beslisregels rekening gehouden te worden met de nauwkeurigheid van verschillende sensoren (Wi-Fi, GPS en GSM)?
- Respondentacties:
 - Hoe vaak wordt elk van de in-app respondentacties uitgevoerd als functie van tijd in-app?
 - Welke effecten hebben respondentacties op kernvariabelen (afgelegde afstand per verplaatsing, aantal stops, etc)?
 - Hoe vaak zien respondenten acties over het hoofd?
 - Wat is het aandeel multimodale verplaatsingen voor en na respondentacties?
- Vergelijkbaarheid:
 - Zijn er verschillen naar type en leeftijd van het device?
 - Hoe hangt de daadwerkelijk frequentie van locatiemetingen af van type en leeftijd van devices?
 - Hoe vaak en onder welke omstandigheden schakelt het type sensormetingen na het eerste uur over van hoogfrequent ('normal') naar laagfrequent ('fused/balanced')?
 - Kan het overschakelen van hoog- naar laagfrequente metingen voorspeld worden met kenmerken van respondenten?

Uitgebreide analyse:

- Methode-effecten:
 - Welke verbeteringen in nauwkeurigheid kunnen worden bereikt met geavanceerde imputatiemethoden?
- Beslisregels:
 - Dienen duur-radius parameters gedifferentieerd te worden (bijvoorbeeld als functie van vervoermiddel, respondentkenmerken, woonomgeving)?
 - Hoe nauwkeurig kunnen vervoermiddelwissels gedetecteerd worden?
 - Welke beslisregels geven de hoogste nauwkeurigheid?
 - Wat zijn de verbeteringen in nauwkeurigheid als point-of-interest data en opentripplanner data worden toegevoegd?
 - Wat zijn de prestaties van ML modellen die vervoermiddelpredictie uitvoeren?
- Respondentacties:
 - Welke respondentacties zijn vanuit het oogpunt van nauwkeurigheid noodzakelijk?, ook bij geavanceerde beslisregels
- Vergelijkbaarheid:
 - Welke meetfrequentie is optimaal in termen van vergelijkbare nauwkeurigheid over devices heen?
 - Welke invloed heeft het overschakelen op laagfrequente metingen op vergelijkbaarheid tussen relevante deelpopulaties?

Resultaten:

1. Notitie met uitleg over verschil tussen gemiddeldes over respondenten of gemiddeldes over alle respondentdagen
2. Rapport over effecten van beslisregels en respondentacties
3. Rapport over geschatte verschillen tussen app- en vragenlijstmetingen, inclusief duiding van verschillen en inclusief effect van meetperiode (binnen versus tussen persoonvarianties)
4. Scripts in R/Python voor toepassing beslisregels (naast pre-processing e.d.)
5. Eindrapport over optimale beslisregels, meetperiode en meetfrequentie en gevolgen voor respondentacties en imputaties deden?

4.7 Juridisch

De AVA22 veldtest heeft meer vrijheden gekregen in ontwerp en dataverwerking dan het reguliere ODIN. De tussentijdse ODIN app DPIA erkent expliciet dat de AVA22 veldtest input moet geven voor een definitieve ODIN app DPIA. Het gaat daarbij om twee cruciale keuzes die de data kwaliteit en rijkheid beïnvloeden maar ook vanuit het data minimalisatie principe van de AVG extra goed gemotiveerd moeten worden. De eerste keuze is die van lengte van meten, namelijk een enkele dag tegen meerdere dagen (in AVA22 gezet op zeven dagen). Meerdere dagen meten maakt data rijker en makkelijker te corrigeren voor uitbijters en kleine gaten in locatiemetingen. Steekproeven kunnen ook verkleind worden. De tweede keuze is interne (en dus uitgestelde) verwerking van locatiemetingen tegen directe in-app afleidingen van ODIN statistieken. Zowel locatiemetingen zelf als respondent correcties zijn dermate feilbaar dat in-huis verwerken grote voordelen biedt in data kwaliteit. AVA22 dient dus twee hypothesen kracht bij te zetten:

1. Meerdere dagen meten betekent een meetbare verbetering van kwaliteit en verkleining van steekproeven terwijl respondenten geen grote bezwaren hebben;
2. In-huis verwerking betekent een meetbare verbetering van kwaliteit zonder dat zowel gepercipieerde als werkelijke kracht van data beveiliging afnemen;

Snelle analyse:

- Voor hypothese duur van meten:
 - Of en wat voor bezwaren kwamen er via telefoon of email vanuit steekproefpersonen tegen locatiemetingen, en in het bijzonder tegen een langere meetperiode?
 - Wat voor kanttekeningen plaatsten respondenten in de evaluatievragenlijst?
 - Wat was het in-app afbreekpercentage?
 - Wat zijn waarschijnlijke verklaringen voor afbreken (technisch, belasting in tijd, UI)?
- Voor hypothese in-huis verwerking:
 - Welke acties voeren respondenten uit op gepresenteerde stops-verplaatsingen?

Uitgebreide analyse:

- Voor hypothese duur van meten:
 - Wat is een optimale meetperiode in termen van nauwkeurigheid?
- Voor hypothese in-huis verwerking:
 - Welke verlies aan nauwkeurigheid treedt (globaal) op als verwerking lokaal in-device moet plaatsvinden?
 - Welke aanwijzingen zijn er dat respondenten acties hadden moeten toepassen maar dat niet deden?

Resultaten:

1. Een notitie ter onderbouwing van keuzes in een definitieve DPIA

5. Planning

Het onderscheid tussen snelle en uitgebreide analyses is bewust gemaakt. De snelle analyses dienen communicatie met belanghebbenden in/extern en geven de eerste belangrijke aanzet tot besluitvorming over de meest invloedrijke ontwerpkeuzes. De uitgebreide analyses geven onderbouwing voor meer subtiële ontwerpkeuzes en inrichting van het verwerkingsproces.

Veldwerk is verdeeld in drie periodes:

- Kleine veldtest en grote veldtest eerste batch: 6 okt – half nov (laatste rappel op 4 nov)
- Grote veldtest tweede batch: 22 nov – eind december (laatste rappel op 22 dec)
- Evaluatiesurvey: 19 jan – eind februari

De snelle analyses kunnen starten aan het einde van de eerste periode rond half november. De data zijn dan nog niet compleet voor de grote veldtest, maar scripts/analyses kunnen al wel worden opgezet. De uitgebreide analyses starten begin januari als alle veldtestdata compleet zijn. Ook hier zouden al wel eerder scripts kunnen worden gemaakt.

De resultaten voor de snelle analyse staan gepland voor 30 april 2023 en omvatten:

- App frontend algemeen:
 - Verslag met bevindingen en leerpunten voor de app frontend algemeen
- App frontend respondentacties:
 - Verslag met beschrijvende statistieken over uitgevoerde respondentacties
- App backend:
 - Verslag met functioneren backend en leerpunten
- Benaderstrategie:
 - Notitie met resultaten en eerste aanbevelingen
- Logistiek
 - Een veldwerkverslag
- Verwerking:
 - Notitie met uitleg over verschil tussen gemiddeldes over respondenten of gemiddeldes over alle respondentdagen
 - Rapport over effecten van beslisregels en respondentacties
 - Rapport over geschatte verschillen tussen app- en vragenlijstmetingen, inclusief duiding van verschillen en inclusief effect van meetperiode (binnen versus tussen persoonvarianties)
- Juridisch:
 - Geen

De resultaten voor de uitgebreide analyses worden geleidelijk opgeleverd vanaf september 2023 tot uiterlijk einde 2023. Deze omvatten alle overige resultaten.

De planning wordt per onderdeel in detail uitgewerkt. Voor onderdeel Verwerking bestaat al een gedetailleerder plan.

6. Projectinrichting

Tabel 3 geeft per onderdeel en type analyse de beoogde projectleden. In de uitgebreide analyses zijn ook onderzoekers van UU betrokken. [REDACTED] rondt in 2022 haar PhD af over correctie voor missende locatiemetingen. [REDACTED] start in zomer 2022 haar PhD naar push-to-app/smart benaderstrategieën

Tabel 3: Projectinrichting per projectonderdeel en type analyse.

Onderdeel	Snelle analyse	Uitgebreide analyse
App UI	[REDACTED]	
App resp acties	[REDACTED]	[REDACTED]
App backend	[REDACTED]	
Benaderstrategie	[REDACTED]	[REDACTED]
Logistiek	[REDACTED]	[REDACTED]
Verwerking	[REDACTED]	[REDACTED]
Juridisch		[REDACTED]

Bijlage 3: Verplaatsingsonderzoek met de app: 1 versus 7 dagen waarnemen**Inleiding**

In het huidige ODIN, volledig via internet waargenomen, wordt een respondent gevraagd om alle verplaatsingen van 1 specifieke, vooraf vastgestelde dag ('peildag') door te geven, retrospectief dus. Dat is een tijdrovende klus voor een respondent. De app moet dit eenvoudiger maken voor een respondent, omdat verplaatsingen en stops automatisch worden gemeten. De respondent hoeft idealiter alleen nog neveninformatie te geven, zoals gebruikte vervoermiddelen en motief van de verplaatsing. De respondentbelasting is een stuk lager, waardoor meer dagen waarnemen overwogen kan worden. Achterliggende gedachte is dat, door een week waar te nemen bij een respondent, uiteindelijk de steekproef verlaagd kan worden. Privacy-issues spelen hierbij echter een rol, want door een week waar te nemen, is de impact op de privacy van de respondent groter (denk aan het herkennen van patronen).

De te beantwoorden vraag is: Waarom is een week waarnemen dan toch een wens vanuit statistisch oogpunt, als dat in het huidige ODIN ook niet gebeurt?

In dit document worden de argumenten op een rijtje gezet. Hierbij is het belangrijk om op te merken dat het verzoek om m7 dagen waar te nemen gericht is op de veldtest in het najaar van 2022. Uit de resultaten van die veldtest moet blijken of en hoe de app in de toekomst onderdeel kan worden van het ODIN design. Daarnaast kan aan de hand van de data-analyse bepaald worden of het meten van meerdere dagen te rechtvaardigen is tegen de meerwaarde in relatie tot het doel. Tot is het goed om op te merken dat hier dus niet gaat om een 'definitieve' app die gebruikt wordt voor het maken van output.

Waarom 7 dagen in plaats van 1 dag?

De veldtest in het najaar wordt gedaan met ODIN als basis, maar het CBS kan de kennis van waarneming via de app breder inzetten. De veldtest moet daarom niet enkel worden gezien als ODIN kopie, maar als een veldtest waarmee meerdere toekomstige (app)onderzoeken bediend kunnen worden.

Juist in deze testfase is het meten over meerdere dagen de meest efficiënte manier om te kijken hoe het app-onderzoek het beste ingericht kan worden. Ook met uiteindelijk data-minimalisatie in het achterhoofd. Door nu meerdere dagen te meten, wordt een beter inzicht verkregen in hoe eventuele hiaten in de data gerepareerd kunnen worden.

Maar ook voor de output van ODIN biedt meer dagen meten meer mogelijkheden. Door meer dan één dag te meten, wordt er voor gezorgd dat met zekerheid een volledige responsdag van een respondent wordt waargenomen. Bovendien is het vaststellen van een peildag, zoals nu in ODIN gebeurt, niet meer nodig wanneer zeven dagen worden waargenomen. Eén dag hoeft niet representatief te zijn voor een gemiddelde week. Na zeven dagen meten kan een nauwkeurig reisprofiel van personen worden gemaakt, wat een duidelijke meerwaarde is ten opzichte van het huidige ODIN. Het wordt bijvoorbeeld mogelijk om het verschil tussen een weekenddag en een werkdag voor een respondent te bekijken, en variatie in het verplaatsingsgedrag over de week te zien. Het feit dat dit nu niet kan, betekent niet dat die wens er niet ligt. De responslast wordt met de huidige waarneemmethode (internetvragenlijst) te groot als meerdere dagen worden waargenomen. Met de app is deze responslast een stuk lager.

Bijkomend voordeel is dat de steekproef uiteindelijk kleiner kan worden. Hoeveel kleiner moet blijken uit de veldtest, maar op basis van de veldtest van 2018 is hier wel al een inschatting van gemaakt, dit wordt verderop in dit stuk beschreven.

Ook legitimeert het waarnemen van meerdere dagen het gebruik van de app. Een app installeren op de telefoon voor slechts één dag is ongebruikelijk. De drempel om de app te installeren voor één dag dataverzameling is vrij hoog; bij meerdere dagen wordt die drempel lager. Waar het optimale punt ligt is onduidelijk, maar in de veldtest van 2018 werd na dag 1 slechts een zeer geleidelijk afbreken waargenomen. Het lijkt erop dat personen expliciet kiezen voor meedoen of niet en dat duur dan niet zoveel meer uitmaakt. Wat het optimale punt is, zou in de veldtest gemeten kunnen worden.

Waarom dan zeven dagen en niet drie of vijf?

Wanneer tijdens de veldtest zeven dagen worden waargenomen, is een compleet week- en weekendpatroon te zien. Weekenden en weekdays zijn voor veel mensen verschillend, en zaterdag en zondag verschillen vaak ook nog van elkaar. Alleen als vijf weekdays identiek zijn, zou volstaan kunnen worden met minder dagen, maar dat is op voorhand niet vast te stellen.

Op basis van de data van de veldtest van 2018 is gekeken naar het aantal unieke versus herhaalde trips. Uit die data blijkt dat ook na 7 dagen nog om en nabij 70% van alle trips op die dag uniek zijn. Dus voor de statistieken is het nuttig mensen langer te observeren. De inschatting is dat als respondenten één dag gevolgd

worden in plaats van zeven, de steekproef vier keer zo groot moet zijn om hetzelfde aantal unieke trips te halen. Met andere woorden: met een steekproef vier keer kleiner dan het huidige ODIN, kan eenzelfde precisie worden behaald als respondenten een week worden waargenomen.

Bij minder dan zeven dagen bestaat nog steeds het probleem van het vaststellen van de peildag (startdag). Voor de waarneming van ODIN is het van belang om alle dagen van de week evengoed vertegenwoordigd te hebben in de output, om zo iets te kunnen zeggen over de verplaatsingen gedurende het hele jaar. Om die reden wordt in het huidige internetonderzoek aan elke respondent een peildag toegewezen. Dat betekent dat er iets extra's gevraagd wordt van de respondent, namelijk precies op die dag verplaatsingen bijhouden. Dus niet een zelfgekozen dag. Dit heeft uitval tot gevolg, want de kans bestaat dat respondenten het op die dag vergeten, en dan niet meer meedoen, of verplaatsingen van een andere dag gaan invullen op die dag. Met zeven dagen waarnemen is een peildag niet meer nodig en kunnen respondenten zelf kiezen welke dag ze starten met de app.

Waarom en hoe dat gerechtvaardigd is

Of er nu één dag waargenomen wordt, of zeven: in alle gevallen moet het CBS haar zaken qua dataveiligheid en privacyrichtlijnen op orde hebben. Er moeten heldere afspraken zijn wie later wat mag doen met de data en hoe lang deze data na verzameling bewaard mag worden. De outputdatabase kan zo worden opgezet, dat er toestemming gegeven moet worden om dagen van dezelfde personen bij elkaar te brengen en onderliggende ruwe data kan worden gescheiden.

Wat van belang is, is heldere communicatie naar respondenten. Respondenten moeten kunnen zien wat er gebeurt, en controle houden. Uiteraard kunnen ze, net als bij andere surveys, weigeren om mee te doen, of tussentijds afhaken. Ook bij één dag meten staat het thuisadres van de respondent in de output, en ook dan vragen we naar doel van verplaatsing. Eén aangewezen peildag kan net zo gevoelige informatie bevatten als meerdere dagen.

Veldtest

Door in de veldtest zeven dagen waar te nemen, kan onderzocht worden bij hoeveel dagen respondenten uitvallen. Daarnaast is het plan om in de veldtest users in groepen te verdelen, waarbij de ene groep zeven dagen wordt waargenomen en de andere groep één dag (waarbij de app drie dagen gebruikt wordt: dag 1 = installatie van de app, dag 2 = waarneming, dag 3 = afsluiting). Op die manier kan inzicht worden verkregen in het verschil in kwaliteit van de output, maar ook welke groep beter respondeert en of er sprake is van selectiviteit. De steekproef van de veldtest is groot genoeg om statistische uitspraken te doen over de effecten van lengte van meten, uitgesplitst naar deelpopulaties.

Daarnaast is een onderzoeksvraag met hoeveel personen je de steekproef kunt verlagen, als je meerdere dagen waarneemt. De resultaten uit de veldtest van 2018 kunnen daarmee gevalideerd worden.

Wat doen wij bij andere smart surveys en wat doen anderen?

Nieuwe meetmethoden komen met nieuwe eisen en kansen. Het CBS heeft onderzoek gedaan naar beweegmeters, waarbij een week gemeten is, terwijl deze informatiebehoefte in huidig onderzoek bij het CBS wordt afgedaan met een paar vragen. Een experiment metindoormeters om het binnenklimaat te meten duurt 1 of 2 maanden. Zo zijn er meerdere voorbeelden te noemen waarbij gedurende langere periode informatie verzameld wordt van een respondent. In die zin is de verplaatsingenapp dus niet uniek.

In Nederland zijn nog twee verplaatsingsonderzoeken: het NVP (Nederlands Verplaatsingspanel) en het MPN (Mobiliteitspanel Nederland). Beide zijn panelonderzoeken, waarbij respondenten van het NVP continue hun locaties doorgeven en bij MPN wordt 3 dagen achtereen gemeten. Idealiter zou daar zeven dagen worden gemeten, echter de verwachting is dat de respondentlast dan te groot wordt met als gevolg grotere kans op uitval uit het panel. Drie dagen is daar als pragmatische oplossing gekozen.

Ook is gekeken naar buitenlandse studies. Zo wordt in de British National Travel Survey een week gemeten, en Oostenrijk doet twee dagen. Specifiek over verplaatsingsonderzoeken met appwaarneming en meerdere dagen meten is weinig gevonden. Wel is het zo dat gerapporteerde app studies vrijwel altijd langer zijn dan één dag.

Samengevat

Het voorstel is om in de veldtest die in het najaar van 2022 plaatsvindt, twee condities rondom lengte van waarnemen te testen: 1 dag waarnemen (3 dagen app: dag 1 installeren, dag 2 waarnemen, dag 3 afsluiten) versus 7 dagen waarnemen. Aan de hand van de resultaten van deze veldtest kan dan een gefundamenteerd besluit worden genomen over het te rechtvaardigen en ideale aantal dagen waarnemen in toekomstig onderzoek. Wanneer in de veldtest zeven dagen wordt waargenomen, kunnen we altijd terug naar kortere periodes van waarnemen, maar langer zal speculeren zijn.

Bijlage 4: Basisregels Informatiebeveiliging CBS

Bij het CBS wordt gewerkt met heel veel vertrouwelijke gegevens; gegevens over individuele organisaties, bedrijven, instellingen, overheden en vooral heel veel persoonsgegevens. Het is van groot belang dat de vertrouwelijkheid van die gegevens en daarmee de privacy van iedereen waarover we informatie hebben, wordt gegarandeerd.

Hieronder tien basisregels voor het veilig werken met vertrouwelijke gegevens:

1. Je hebt toegang tot de data op onze IT -systemen, althans het deel dat voor jouw werk nodig is, via je CBS account. Dat account is persoonsgebonden. Geef je inloggegevens nooit aan anderen, ook niet aan je collega's. Kies een veilig wachtwoord en hou het geheim.
2. Gebruik je CBS wachtwoord nooit voor andere toepassingen. Het is in het algemeen aan te bevelen om voor elke toepassing een verschillend wachtwoord te gebruiken.
3. Vergrendel je PC bij het verlaten van je werkplek, ook als je maar 'even' weg bent. (bijvoorbeeld via de snelkoppeling "Computer Vergrendelen" (het hangslotje op de taakbalk).
4. Krijg je vertrouwelijke gegevens (bijvoorbeeld informatie van berichtgevers of voorlopige resultaten) op papier of heb je gegevens geprint, bewaar die documenten dan in afgesloten kasten of laden. Dat geldt ook voor gegevensdragers en mobiele apparatuur. Laat deze nooit onbeheerd achter. Het CBS hanteert een clean desk beleid.
5. Vernietig vertrouwelijke informatie op papier of datadragers wanneer deze niet meer nodig is. Lever opslagmedia in bij de IT-servicedesk, vernietig documenten in de papierversnipperaar.
6. Als je thuis of onderweg wilt werken, maak dan gebruik van de telewerkfaciliteiten of een door het CBS verstrekte laptop. Het is niet toegestaan om vertrouwelijke CBS data buiten het CBS te verspreiden. Je mag deze data dus ook niet naar je privé mailadres sturen of bewaren op je op je privé computer of laptop.
7. Als je thuis of onderweg werkt, zorg er dan voor dat anderen niet kunnen meelezen op je scherm. Wees voorzichtig met vertrouwelijke documenten; zorg dat namen, titels of teksten niet zichtbaar of door derden leesbaar zijn. Wees voorzichtig met (telefoon) gesprekken in openbare gelegenheden of in het openbaar vervoer.
8. Hou de computer waarop je telewerkenfaciliteiten gebruikt up-to-date en vrij van virussen en malware. Installeer altijd de laatste updates van het besturingssysteem en de virusscanner.
9. Vertrouwelijke CBS data op datadragers of mobiele apparatuur moet altijd versleuteld worden opgeslagen. De harddisk van CBS Laptops wordt standaard versleuteld, apps voor CBS-mail op telefoon of tablet (Blackberry Work e.d.) maken gebruik van een versleutelde container.
10. Meld beveiligingsproblemen en -incidenten, waaronder vermissing van vertrouwelijke documenten, opslagmedia of apparatuur altijd direct bij je leidinggevende en via TopDesk.

Bijlage 5: Persoonsgegevens, anonimiseren en pseudonimiseren.**Wat is het en wat is het belang voor privacybescherming en informatiebeveiliging?**

Het CBS ontvangt een veelheid van gegevens uit een grote verscheidenheid van bronnen. Veel van deze bestanden bevatten persoonsgegevens. Een bestand met gegevens over individuele personen is vanuit beveiligingsperspectief kwetsbaar. Persoonsgegevens kunnen, in handen van ongeautoriseerde personen, worden misbruikt c.q. gebruikt voor doelen waarvoor ze niet bedoeld zijn. Door het anonimiseren of pseudonimiseren van zo'n bestand wordt die kwetsbaarheid verminderd. Ook andere maatregelen worden ingezet om misbruik te voorkomen. Denk hierbij onder andere aan versleuteling of encryptie en fysiek en virtueel toegangsbeheer. Het CBS is wettelijk verplicht om de gegevens bij ontvangst, verwerking en verstrekking met de geldende beveiligingsstandaarden tegen misbruik en onbedoeld gebruikt passend te beveiligen.

Wat zijn geanonimiseerde gegevens?

Anonimiseren of anoniem maken wil zeggen het verwijderen van gegevens die de identiteit van een persoon kunnen onthullen. Voorbeelden van direct identificerende gegevens zijn iemands naam, het burgerservicenummer (BSN), de geboortedatum, het adres, het IP-adres, het e-mailadres, etc. Deze direct identificerende gegevens moeten in elk geval volledig worden verwijderd. Wat er nodig is om een bestand volledig te anonimiseren is mede afhankelijk van de inhoud van dat bestand: een vrouwelijke notaris gevestigd te Urk is vermoedelijk nog steeds herkenbaar. De variabelen geslacht, beroep en vestigingsplaats zijn geen van alle direct identificerend, maar in combinatie met elkaar kan wel identificatie optreden. Om identificatie op basis van een combinatie van variabelen minder gemakkelijk of zelfs geheel onmogelijk te maken kunnen scores op variabelen uit deze combinatie worden verwijderd of samengenomen met andere scores (dat laatste wordt ook wel indikken genoemd).

Anonimiseren is een onomkeerbaar proces. Het is dus niet mogelijk om identificerende gegevens in een later stadium (opnieuw) toe te voegen. Deze koppeling is verbroken. Hoewel geanonimiseerde gegevens wel betrekking hebben op personen zijn het in de zin van de privacywetgeving geen persoonsgegevens meer omdat deze niet meer te herleiden zijn tot identificeerbare natuurlijke personen.

Definitie van Artikel 29 werkgroep³: *Anonimiseren is het verwisselen⁴ van persoonsgegevens in gegevens die niet langer gebruikt kunnen worden om een natuurlijk persoon te identificeren, daarbij in ogenschouw nemende 'alle middelen die hiervoor redelijkerwijs gebruikt kunnen worden' door zowel een verantwoordelijke als een derde partij. Een belangrijke factor hierbij is dat de verwerking onomkeerbaar moet zijn.*

Wat zijn gepseudonimiseerde persoonsgegevens?

Bij het pseudonimiseren van gegevens worden direct identificerende gegevens verwijderd en vervangen door een betekenisloze code: het record identificatie nummer (RIN). Bestanden die met eenzelfde set van RIN's zijn gepseudonimiseerd, kunnen eenvoudig met elkaar worden gekoppeld. Dat is dan ook de reden om te pseudonimiseren in plaats van te anonimiseren. Om statistiek te maken worden veel verschillende bronnen ook over de jaren heen aan elkaar gekoppeld.

Definitie uit AVG⁵: *Pseudonimiseren is het verwerken van persoonsgegevens op zodanige wijze dat de persoonsgegevens niet meer aan een specifieke betrokkene kunnen worden gekoppeld zonder dat er aanvullende gegevens worden gebruikt, mits deze aanvullende gegevens apart worden bewaard en technische en organisatorische maatregelen worden genomen om ervoor te zorgen dat de persoonsgegevens niet⁶ aan een geïdentificeerde of identificeerbare natuurlijke persoon worden gekoppeld.*

Pseudonimiseren doorloopt hetzelfde proces als anonimiseren, met twee uitzonderingen: de relatie tussen de identificerende kenmerken en het RIN ligt vast. Daarmee is het proces, in tegenstelling tot het anonimiseren, in principe wél omkeerbaar. Daarnaast worden bij pseudonimiseren binnen het CBS de indirect identificeerbare gegevens niet ingedikt of verwijderd zoals bij anonimiseren het geval is. Er is daarmee geen garantie dat onze gepseudonimiseerde gegevens niet meer identificeerbaar zijn voor derden ook al beschikken deze niet over de koppeling tussen RIN en bijvoorbeeld BSN. Dit is ook de reden dat het CBS geen gepseudonimiseerde bestanden aan derden levert. Gepseudonimiseerde gegevens worden daarom in de privacywetgeving wel als persoonsgegevens beschouwd, ook al zijn de gegevens niet *direct* te herleiden tot identificeerbare personen. Bij het samenstellen van tabellen wordt er op toegezien dat ook daar de vrouwelijke notaris in Urk niet herkenbaar

³ De Artikel 29-werkgroep is het onafhankelijke advies -en overlegorgaan van Europese privacytoezichthouders.

⁴ Aanvulling CBS: Hieronder wordt verstaan verwijderen, veranderen, indikken etc.

⁵ Algemene Verordening Gegevensbescherming (opvolger van de Wet Bescherming Persoonsgegevens (Wbpg)).

⁶ Aanvulling CBS: door iedereen

is. Voor derden zijn daarmee tabellen en alle overige open data welke het CBS ter beschikking stelt, geanonimiseerd.

Wat zijn persoonsgegevens?

We hebben het in deze notitie over persoonsgegevens; gegevens waarop de privacywetgeving van toepassing is. Volgens die wetgeving is elk gegeven over een geïdentificeerde of *identificeerbare* natuurlijke persoon een persoonsgegeven. Het woordje identificeerbare bepaalt het verschil tussen geanonimiseerde en gepseudonimiseerde gegevens. Het feit dat pseudonimiseren in principe omkeerbaar is, maakt dat de gegevens daarna nog steeds als persoonsgegevens moeten worden beschouwd. Ondanks het feit dat de kwetsbaarheid van de gegevens door het pseudonimiseren aanzienlijk wordt verminderd, is de privacywetgeving onverminderd van kracht. Geanonimiseerde gegevens zijn per definitie géén persoonsgegevens.

Definitie uit AVG: *Persoonsgegevens zijn alle informatie over een geïdentificeerde of identificeerbare natuurlijke persoon („de betrokkene”); als identificeerbaar wordt beschouwd een natuurlijke persoon die direct of indirect kan worden geïdentificeerd, met name aan de hand van een identifier zoals een naam, een identificatienummer, locatiegegevens, een online identifier of van een of meer elementen die kenmerkend zijn voor de fysieke, fysiologische, genetische, psychische, economische, culturele of sociale identiteit van die natuurlijke persoon;*

In hoeverre dragen anonimiseren of pseudonimiseren bij aan de beveiliging?

Het CBS is op grond van de CBS-wet verplicht om maatregelen te nemen “*ter beveiliging van zijn gegevens tegen verlies of aantasting en tegen onbevoegde kennisneming, wijziging en verstrekking van die gegevens.*” Deze verplichting geldt voor alle vertrouwelijke informatie c.q. voor alle informatie over individuele bedrijven, instellingen, personen, huishoudens, etc. Specifiek voor personen gelden daarnaast de beperkingen en verplichtingen die de privacywetgeving oplegt.

Om invulling te geven aan deze verplichtingen heeft het CBS een groot aantal beveiligingsmaatregelen genomen op tal van verschillende terreinen. Denk aan restrictieve maatregelen zoals beperking van de fysieke toegang tot gebouwen, beperking van de toegang tot bestandsmappen, bescherming van de IT-infrastructuur tegen ongeautoriseerde toegang van buitenaf, encryptie van gegevensdragers, etc. Maar denk ook aan de minder ‘harde’ maatregelen zoals de geheimhoudingsverklaringen, awareness programma’s, clean-desk-beleid, etc. Alle maatregelen zijn erop gericht om het risico op verlies of oneigenlijk gebruik van vertrouwelijke gegevens te verkleinen. Ook het anonimiseren en pseudonimiseren zijn beveiligingsmaatregelen. De kans dat de gegevens in verkeerde handen vallen wordt er niet door verkleind, maar de kans dat onbevoegden misbruik kunnen maken van die gegevens wordt wél een stuk kleiner.