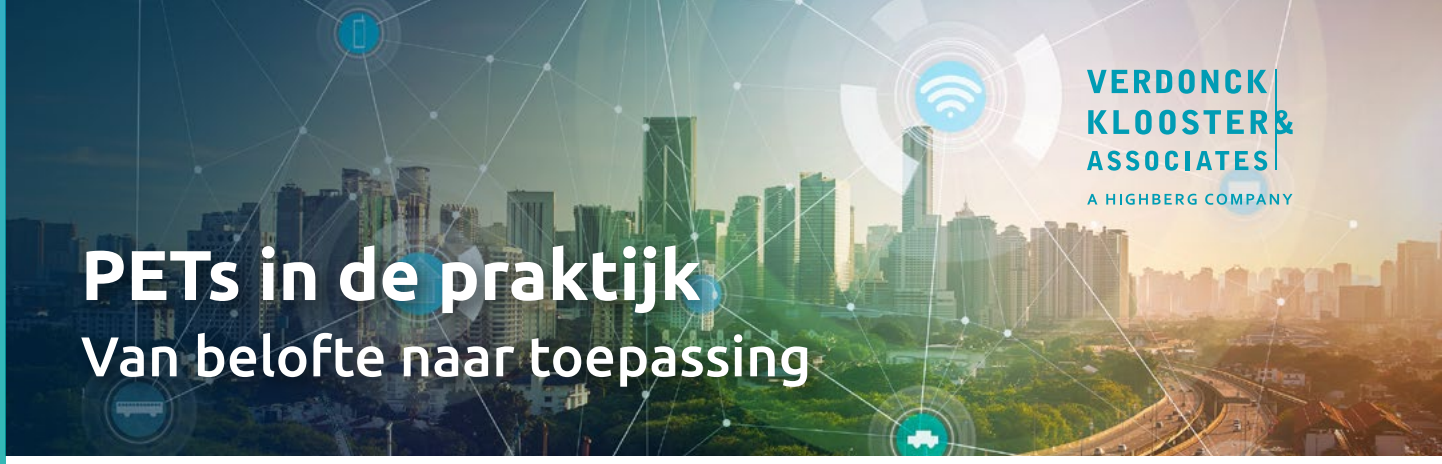

VERDONCK KLOOSTER & ASSOCIATES
A HIGHBERG COMPANY

PETs in de praktijk
Van belofte naar toepassing

Privacy Enhancing Technologies - demystified

- PET is een verzamelnaam van verschillende technieken voor vergaande bescherming van persoonsgegevens in informatiesystemen.
- Het concept PET bestaat al tientallen jaren. Historisch zijn er nog steeds maar weinig manieren om veilig, privacy vriendelijk samen te werken en gegevens uit te wisselen. Hoewel...
- Steeds vaker ontstijgen PETs de academische mogelijkheden, en zien we concrete toepassingen. De PET-beloftes worden langzaam maar zeker waargemaakt.
- Het wordt hoog tijd dat een breder publiek kennis maakt met de meest relevante PETs, hun mogelijkheden en beperkingen.
- Verdonck, Klooster & Associates (VKA) heeft daarom dit praktische overzicht voor jou gemaakt.
- In dit overzicht vind je de status van PETs die de wetenschappelijke belofte voorbij zijn en in de praktijk worden gebruikt.
- We leggen uit hoe deze PETs werken en waar je op moet letten. Dit illustreren we met concrete voorbeelden.

Voordat je met PETs aan de slag gaat, stel de volgende 4 vragen:

1. Wat is het toepassingsdoel?

Matching	Exploratie	Analyses

Wil je een specifiek individu kunnen identificeren (matchen), of wil je vrij kunnen grasduinen in details (exploreren)? Het kan ook zijn dat je gestructureerd te werk wilt gaan om iets te onderzoeken (analyse).

2. Waar staan de gevoelige gegevens?

Eén bron	Meerdere bronnen

Staan de gevoelige gegevens op één plek of verspreid over meerdere bronnen? Veilig combineren van gegevens uit meerdere bronnen vormt een extra uitdaging.

3. Welke beveiligingsfuncties moet de PET bieden?

Voorkom identificatie	Grip op gebruik	Waarborg vertrouwelijkheid

Wil je voorkomen dat gevoelige informatie over een individu wordt herkend? Strikter controleren welke analyse wel of niet wordt gedaan? Of de vertrouwelijkheid van alle gegevens, verwerkingsacties en (tussentijdse) uitkomsten beschermen.

4. Welk budget is beschikbaar voor implementatie en gebruik?

€	€€	€€€

Sommige technieken zijn relatief eenvoudig in te zetten en goedkoop, terwijl andere technieken extra tijd en resources (en rekenkracht) vergen in gebruik.

Welke PET's zijn er voor welke situaties?

Matching	(Data-)Exploratie	Analyses (op meerdere bronnen)
Ik wil checken of een individu voorkomt in een gegevens-verzameling, zonder toegang tot de gegevens zelf.	Ik wil vrij kunnen grasduinen in details in een gegevens-verzameling, maar een exacte match op individu is niet nodig.	Ik wil diverse inzichten opdoen, waarvoor een exacte koppeling van gegevens per individu onder strikte voorwaarden ook mogelijk moet zijn.
Zero-knowledge proof **Bloom filter**	**Differential Privacy** **Synthetische data**	**Trusted Execution Environments** **Federated Learning** **Multi-party computation** **Fully Homomorphic Encryption**

Kies de PET die past bij jouw context en toepassing

	Zero-knowledge proof	Bloom filters	Differential privacy	Synthetische data	Trusted Execution Environments	Federated learning	Multi Party Computation	Fully Homomorphic Encryption
Wat is het toepassingsdoel?								
Matching ('staat naam X op deze lijst?')	■	■			■	■	■	■
Exploratie (vrije toegang details, maar géén matching)			■	■	■	■	■	■
Analyses (inclusief matching, onder voorwaarden)					■	■	■	■
Waar staan de gevoelige gegevens?								
Eén bron	■	■	■	■	■	■	■	■
Meerdere bronnen					■	■	■	■
Welke beveiligingsfuncties moet de PET bieden?								
Voorkom identificatie (herleidbaarheid)	■	■	■	■	■		■	■
Grip op gebruik (doelbinding in de techniek verankerd)	■				■	■	■	
Waarborg vertrouwelijkheid	■						■	■
Welk budget is beschikbaar?								
Kosten voor implementatie	€€	€	€€	€	€€	€€	€€	€€
Kosten in gebruik	€€	€	€	€	€	€€	€€	€€€

PETs voor matching

Zero knowledge proof

- Deze techniek levert bewijs voor specifieke claims, zoals “ik ben ouder dan 18 jaar.”
- Een partij kan de claim bewijzen zonder gevoelige informatie te delen. De ontvanger leert niets over de informatie behalve of het waar/onwaar is.
- Beschermt privacy door een wiskundig bewijs te geven van de claim, zonder iets over de onderliggende gegevens te onthullen.

Let op!
De veiligheid van deze techniek komt het beste tot zijn recht in enkelvoudige toepassingen met één bewijs voor één verificatie.

Matching	Exploratie	Analyses
Eén bron	Meerdere bronnen	
Voorkom identificatie	Grip op gebruik	Waarborg vertrouwelijkheid
Implementatie	€	€€	€€€
Gebruik	€	€€	€€€

Voorbeeld: Crypto currency ZeroCash

Een inherente eigenschap van cryptocoins zoals Bitcoin is dat de transactiegegevens van elke gebruiker publiekelijk zijn opgeslagen. Deze gegevens zijn vastgelegd in een gedistribueerd grootboek, de blockchain.
ZeroCash biedt een privacy-beschermend alternatief: de Zerocoin. Transacties in deze munteenheid zijn mogelijk zonder de verzender, ontvanger of het bedrag openbaar te maken.
Met behulp van zero-knowledge proof wordt aangetoond dat transacties juist zijn en voldoen aan de voorwaarden.

Lees meer: http://zerocash-project.org/how_zerocash_works

PETs voor matching

Bloom filter

- Deze techniek is ontwikkeld om efficiënt te toetsen of een gegeven (bijv. een naam) voorkomt in een gegevensverzameling, ook wel 'matching' genaamd.
- De toets zegt ofwel dat een item zeker niet voorkomt, ofwel dat een waarde mogelijk wel voorkomt.
- Bloom filters werken op een heel beknopte samenvatting van de gehele lijst, waarin afzonderlijke oorspronkelijke objecten niet herkenbaar zijn.
- Door gebruik van kansberekening kost toepassing op grote datasets weinig opslagruimte en tijd.

Let op!

- Een juiste configuratie van de test is nodig om (te) hoge ratio van *false positives* ('mogelijk wel') te voorkomen
- Vooral geschikt voor gebruik in gesloten en vertrouwde gebruikersgroepen

Matching	**Exploratie**	**Analyses**	
Eén bron	**Meerdere bronnen**		
Voorkom identificatie	**Grip op gebruik**	**Waarborg vertrouwelijkheid**	
Implementatie	**€**	**€€**	**€€€**
Gebruik	**€**	**€€**	**€€€**

Voorbeeld: Ma3tch

Ma3tch is een techniek die specifiek geschikt is om te checken of een andere partij gegevens heeft over een persoon of rechtssubject waar jij bewust naar op zoek bent, zonder dat je inzage krijgt in de brongegevens van die andere partij.

Lees meer: https://magazines.rijksoverheid.nl/jenv/jenvmagazine/2021/13/reportage-ma3tch

PETs voor data-exploratie

Differential Privacy

- Deze techniek voegt een statistische 'ruis' (random data) toe aan de individuele gegevens voordat deze worden gebruikt.
- De toegevoegde data is zodanig gekozen dat statistische eigenschappen van de totale dataset in stand blijven. Het resultaat van een analyse op de data blijft daarmee relevant.
- Door de ruis zijn resultaten niet meer terug te voeren tot op individueel gevalsniveau.

Let op!

- Het 'privacy budget' weegt de hoeveelheid ruis af tegen de hoeveelheid echte informatie. Dit 'privacy budget' moet goed worden afgestemd.
- Minder geschikt voor analyse op details, omdat toegevoegde ruis de uitkomsten beïnvloedt.

Matching	Exploratie	Analyses
Eén bron		**Meerdere bronnen**
Voorkom identificatie	**Grip op gebruik**	**Waarborg vertrouwelijkheid**

Implementatie	€	€€	€€€
Gebruik	€	€€	€€€

Voorbeeld: United States Census Bureau

Sinds 2020 past de US Census differential privacy toe om ongewenste onthulling te voorkomen. Zij gebruiken hiervoor het "TopDown Algorithm".
Dit algoritme maakt een aantal dwarsdoorsneden van de gegevens, en voegt vervolgens ruis aan elke celwaarde op basis van het berekende en toegestane onthullingsrisico.
De ruis wordt eerst verdeeld over de grootste dwarsdoorsneden (met gegevens op nationaal niveau), en vervolgens uitgesplitst naar kleinere schaal (states, counties, enzovoort).

Lees meer: Case study 17 in de 2023 UN Guide to PETs https://unstats.un.org/bigdata/task-teams/privacy/guide/

PETs voor data-exploratie

Synthetische data

- Dit zijn kunstmatig gegenereerde data die lijkt op echte data, maar geen persoonsgegevens bevatten.
- Kunstmatige intelligentie (AI) zorgt dat eigenschappen uit de originele data worden gereproduceerd in de synthetische data.
- Synthetische data kan worden gebruikt als testdata voor softwareontwikkelaars, of voor het trainen van AI modellen.

Let op!

- Synthetische datasets kun je niet aan elkaar koppelen, een exacte match is niet mogelijk.
- Vooral geschikt voor gebruik in gesloten en vertrouwde gebruikersgroepen, waarbinnen eigenschappen over de dataset open gedeeld kunnen worden.*

Matching	**Exploratie**		**Analyses**
Eén bron		**Meerdere bronnen**	
Voorkom identificatie		**Grip op gebruik**	**Waarborg vertrouwelijkheid**
Implementatie	**€**	**€€**	**€€€**
Gebruik	**€**	**€€**	**€€€**

Voorbeeld: Churn predictie modellen

SAS en Syntho werken samen om data-toegang voor AI-ontwikkeling te verbeteren. Voor een churn-predicte model hebben zij uitkomsten vergeleken van modellen getraind op echte data, synthetische data en “klassiek” geanonimiseerde data. De uitkomsten:

- Modellen getraind op synthetische data laten vergelijkbare performance zien ten opzichte van modellen getraind op echte data
- Modellen getraind op “klassiek” geanonimiseerde data laten beduidend mindere performance zien ten opzichte van modellen getraind op echte data
- Synthetische data werkt eenvoudig en snel omdat het toepasbaar is op elke vorm van data en er geen domein kennis nodig is en het voor de eindgebruiker werkt alsof je met echte data werkt.

Lees meer: https://blogs.sas.com/content/hiddeninsights/2022/07/07/ai-generated-synthetic-data-easy-and-fast-access-to-high-quality-data/

* Voor een wetenschappelijke evaluatie van de privacyeigenschappen, zie https://arxiv.org/abs/2011.07018

PETs voor analyses op meerdere bronnen

Trusted Execution Environment (TEE)

- Dit is afgeschermde omgeving op een computersysteem waarbinnen gevoelige gegevens worden verwerkt.
- Systeembeheerders en analisten hebben geen toegang tot deze omgeving. Alleen de uitkomsten van rekenopdrachten worden vrijgegeven.
- Dit is een laagdrempelige manier om samen op gegevens te werken, die wordt aangeboden door grote leveranciers (Intel, Microsoft, e.a.)

Let op!

- Er zijn veiligheidsrisico's aangetoond in deze manier van beveiligen. Gevoelige informatie kan onder meer lekken via het rekenproces.*
- De veiligheidsspecificaties van TEE aanbieders vallen onder intellectueel eigendom en zijn niet onafhankelijk te verifiëren.

Matching		Exploratie		Analyses	
Eén bron		Meerdere bronnen			
Voorkom identificatie		Grip op gebruik		Waarborg vertrouwelijkheid	
Implementatie	€		€€		€€€
Gebruik		€		€€	€€€

Voorbeeld: AI voor de zorg

DarkCovidNet is een AI-model getraind om COVID-19 infectie op te sporen in röntgenfoto's van longen. Een groot volume aan gevoelige data moet bijeen gebracht worden om dit model te trainen. Confidential computing specialist Fortanix demonstreert hoe dit mogelijk is binnen een Trusted Execution Environment van IntelSGX:
Zorgpartijen controleren de integriteit van het systeem voordat ze gegevens aanleveren. Zij leveren gegevens vervolgens versleuteld aan de veilige omgeving aan. De gegevens worden binnen de veilige omgeving ontsleuteld zodat het AI algoritme kan worden getraind.
De TEE zorgt ervoor dat gegevens en berekening niet toegankelijk zijn.

Lees meer: https://www.fortanix.com/blog/2020/12/securing-healthcare-ai-with-confidential-computing

* Zie onder meer: D_2_116_An_Overview_of_Vulnerabilities_and_Mitigations_of_Intel_SGX_Applications_c1282b1505.pdf (cyber.ee)

PETs voor analyses op meerdere bronnen

Federated learning

- Een rekenopdracht wordt rondgestuurd langs verschillende bronnen (partijen). De uitkomsten worden gedeeld en op één plek samengevoegd tot resultaat.
- Deze methode wordt gebruikt voor o.m. het trainen van modellen over een grote populatie (van gegevens cq. Big data).
- Vereenvoudigd samenwerking doordat brongegevens de organisatie niet verlaten.

Let op!

- Toepassing vereist dat elke bron een gelijke structuur heeft.
- Rekenopdrachten kunnen onbedoeld informatie over een individu of kleine groep vrijgegeven.
- Geschikt voor vertrouwde gebruikersgroepen, waarbinnen deel-uitkomsten vanuit elke bron of partij open gedeeld kunnen worden.

Matching	Exploratie	Analyses
Eén bron	Meerdere bronnen	
Voorkom identificatie	Grip op gebruik	Waarborg vertrouwelijkheid
Implementatie	€	€€	€€€
Gebruik	€	€€	€€€

Voorbeeld: analyse overleving mondholtekanker Nederland & Taiwan

In 2020 hebben onderzoekers uit Nederland (IKNL, Maastricht University) en Taiwan samen de overlevingskansen bij mondholtekanker te analyseren, zonder gegevens over individuele patiënten te delen.
Vanwege beperkende regelgeving zijn analyses op patiëntniveau of combineren van data uit Nederland en Taiwan niet uitvoerbaar.

Lees het persbericht van IKNL: https://iknl.nl/nieuws/2020/analyse-overleving-mondholtekanker-nederland-taiwa

Wetenschappelijke publicatie: https://pubmed.ncbi.nlm.nih.gov/33239719/

PETs voor analyses op meerdere bronnen

Multi-Party Computation (MPC)

- Met deze techniek worden gegevens aan de bron versleuteld en opgedeeld in verschillende sets die onder meerdere partijen worden verspreid.
- Koppeling en rekenopdrachten worden decentraal uitgevoerd zonder de gegevens te ontsleutelen. Analisten zien alleen de uitkomsten van analyses.
- Vooral geschikt voor koppelen en analyseren van gegevens die niet onderling gedeeld kunnen worden. Training en toepassing van machine-learning modellen is ook mogelijk.
- Vertrouwelijkheid is gewaarborgd met wiskundige zekerheid (door toepassing van cryptografie).

Let op!

- Kost meer rekenkracht dan niet-versleuteld rekenen, vanwege de benodige communicatie tussen partijen.
- Geschikt voor verwerking van grote datasets tot ca. 100M rijen.

Matching	Exploratie	Analyses
Eén bron		Meerdere bronnen
Voorkom identificatie	Grip op gebruik	Waarborg vertrouwelijkheid
Implementatie	€	€€	€€€
Gebruik	€	€€	€€€

Voorbeeld: samenwerken voor een digitaal veilig Nederland met NCSC/SecureNed

Het Nationaal Cybersecurity Centrum (NCSC) werkt samen met overheden en bedrijven in SecureNed. In dit samenwerkingsverband wordt meldingen over cyberdreigingen en incidenten gedeeld, zodat partijen gezamenlijk maatregelen kunnen nemen.
SecureNed verwerkt gevoelige informatie via MPC. Het NCSC creëert hiermee een breed en gemeenschappelijke beeld van acute cyberdreiging, en informeert deelnemers inzichten waarin aanleverende partijen niet herleidbaar zijn.
De gevoelige gegevens blijven te allen tijde verspreid en versleuteld. Het systeem kan een compromittering van één van de privacy servers weerstaan.

Lees meer: https://emagazine.one-conference.nl/2021/secure-net-ncscs-partnership-for-rapid-and-safe-information-sharing/ en https://www.ncsc.nl/onderwerpen/securened

PETs voor analyses op meerdere bronnen

Fully Homomorphic Encryption (FHE)

- Met deze vorm van versleuteling kun je gevoelige gegevens samenbrengen en verwerken op één plek (door één partij).
- Berekeningen worden uitgevoerd zonder de gegevens te ontsleutelen.
- Vertrouwelijkheid is gewaarborgd met wiskundige zekerheid (door toepassing van cryptografie).
- Vooral geschikt voor eenvoudige bewerkingen zoals optellen en gemiddelden berekenen.

Let op!

- De techniek maakt berekeningen zeer complex en is nog niet rijp voor toepassing op grote datasets.
- Vereist zeer veel rekenkracht en opslagcapaciteit, geschikt voor datasets tot ca. 10,000 rijen.

Matching	Exploratie	Analyses
Eén bron		Meerdere bronnen
Voorkom identificatie	Grip op gebruik	Waarborg vertrouwelijkheid
Implementatie	€	€€	€€€
Gebruik	€	€€	€€€

Voorbeeld: NHS Digital gebruikt een vorm van homomorfe encryptie om gegevens veiliger te koppelen

Om patiëntgegevens veilig te verwerken en toegankelijk te maken in het zorgproces, combineert de NHS pseudonimisatie met homomorfe encryptie.
De uitdaging met traditionele pseudonimisatie is dat gegevens moeilijk of onmogelijk te verbinden zijn wanneer ongelijke “tokens” worden gebruikt om een uniek patiëntnummer te vervangen. Het gebruik van één standaardtoken als alternatief levert echter teveel risico op.
Privitar combineert tokenisation met een partiële vorm van homomorfe encryptie (PHE), om ongelijke tokens te verbinden zonder de oorspronkelijke waarde te onthullen.

Lees meer: https://www.privitar.com/wp-content/uploads/2022/07/P1026-CS-NHS-Digital-DP.pdf en In het Royal Society Rapport “Protecting Privacy in Practice”, p.34

Verantwoording en copyright

- Dit whitepaper is geschreven door Frank van Vonderen van Verdonck, Klooster & Associates (VKA), onderdeel van de Highberg group. Bij de totstandkoming heeft VKA veel hulp gehad van Roseman Labs.
- Over de individuele PETs is inhoudelijk veel te vertellen. We wilden de PETs echter zo kernachtig mogelijk beschrijven. De korte samenvattingen zijn opgeschreven om begrijpelijk te zijn. Dan ontbreken er ook ongetwijfeld nuances, we hebben hier een afweging gemaakt tussen volledigheid en leesbaarheid.
- De beschrijving van de eigenschappen is gebaseerd op diverse openbare bronnen. Vanuit VKA hebben we geprobeerd om alle kenmerken en eigenschappen objectief te beschrijven. We hebben geen enkele voorkeur voor een PET of een leverancier ervan.
- De techniek staat niet stil. Door nieuwe ontwikkelingen kunnen de overzichten over een tijdje niet meer actueel zijn.
- Met dank aan Syntho, Martijn Hunsche, Christian Verhagen en Laura Natrop voor hun review.
- Speciale dank aan Toon Segers en vooral aan Freya de Mink van Roseman Labs. Zonder jullie input was dit whitepaper nooit uit de kraamkamer gekomen. We begonnen als volslagen onbekenden aan dit avontuur en met de passie voor ons vak hebben we elkaar geïnspireerd. Veel dank daarvoor.

Over de auteur

- **Frank van Vonderen** maakt graag moeilijke dingen begrijpelijk. Als bedrijfskundige denkt hij bij alle nieuwe technologie maar aan één ding: wat kan ik er mee. Hij maakt dingen graag praktisch.
- Eerder schreef hij handzame boekjes over Auditing (*Help! Een audit*), over Privacy by Design, over DPIA's (*Hoera, een DPIA!*) en over algoritmes. Daarnaast blogt hij regelmatig en deelt hij kennis via Linkedin.
- Volg Frank via Linkedin.

Wil je aan de slag met PETs? VKA kan je helpen: door een case te vinden die geschikt is voor PET's, met de techniek en met de organisatie.

Neem contact op met Frank van Vonderen: **frank.vanvonderen@vka.nl**.