Pseudonimisatieservices voor wetenschappelijk onderzoek

Houten, 5 november 2019

ZorgTTP
Privacy & vertrouwen

Welkom!
Even voorstellen…

Simone van Wijngaarden

Adviseur, CIPP/e, Stichting ZorgTTP

Dienstverlening Stichting ZorgTTP

- Onomkeerbare pseudonimisatie d.m.v. ons pseudonimisatieplatform
- Omkeerbare pseudonimisatie d.m.v. onze encryptiedienst TRES (Trusted Reversible Encryption Service)
- Data Protection Impact Assessment (DPIA) / gegevensbeschermingseffectbeoordeling

Maar ook:

- Integratie van meerdere diensten
- Nieuw product/andere invulling opdracht met bestaande dienst als basis
    - Zoals de pseudonimisatievoorziening voor wetenschappelijk onderzoek waar we het vandaag over gaan hebben

ZorgTTP
Privacy & vertrouwen

Welkom!
Even voorstellen…

Jan Lucas van der Ploeg

Data Engineer, afdeling Informatiemanagement Onderzoek, UMCG

Robert Griffioen

Projectleider de-identificatie project, SURF

ZorgTTP
Privacy & vertrouwen

Waar gaan we het over hebben?

Pseudonimisatieservices voor wetenschappelijk onderzoek

- Introductie pseudonimisatievoorziening
- Noodzaak tot pseudonimisatievoorziening
- Toelichting op de eerste vier use cases
- Release 1
- Opschaling naar SURF
- Vragen en afronding

Vragen tijdens de presentatie? Stel ze gerust!

ZorgTTP
Privacy & vertrouwen

university of groningen

University Medical Center Groningen

Pseudonymisation Service for Research

Jan Lucas van der Ploeg and Francisco Romero Pastrana
Data Federation Hub - University Medical Center Groningen– University of Groningen

Pseudonymisation

Pseudonymisation means that direct identifiable identifiers such as a name, date of birth and address, are replaced with a pseudonym.

Pseudonymisation involves separating directly identifying personal data from *substantive* data, optionally maintaining a link through an arbitrary key. The GDPR explicitly mentions pseudonymisation as one approach for GDPR requirements compliance, increasing the privacy and security of personal data processing.

Pseudonymisation ≠Anonymization

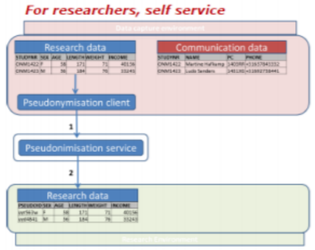
Pseudonymisation is one step in data process for research. Pseudonymised data is not necessarily anonymous. Re-identification is possible, because of indirect identifiers.

Pseudonymisation Service for research:

Pseudonymisation is not a trivial process. The UMCG and UG developed a pseudonymisation service to support researchers with pseudonymisation and linking datasets. The service provides software and a support desk for pseudonymisation, linking and anonymization.

Pseudonymisation: No data linkage possible

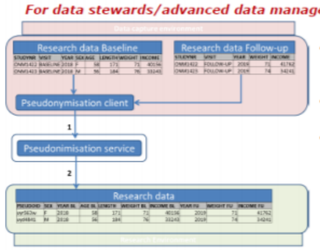
For researchers, self service



- Small studies
- Irreversible
- Different pseudonyms for different sessions
- Not possible to link published datasets
- Creating research pseudonyms

Pseudonymisation: Data linkage possible

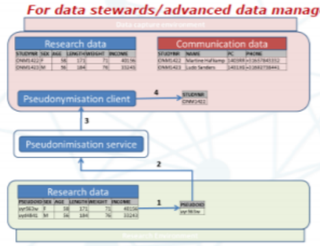
For data stewards/advanced data managers, no self service



- Longitudinal studies, large cohort studies and biobanks
- Subjects are traceable over time
- Pseudonym for subject is stable in project over different sessions

Process for reversible pseudonymisation

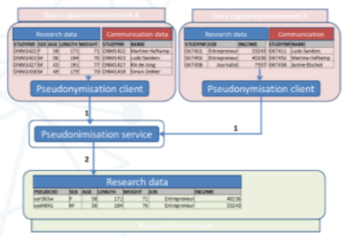
For data stewards/advanced data managers, no self service



- **Only when necessary**
- Not the default option, extra procedures and contracts when needed
- Incidental findings

Pre matching by Trusted Third Party

By Trusted Third Party (TTP), no self service



- Linking data sets from different sources
- Pre matching and linkage by TTP
- Only researcher knows overlap between sets

Pseudonymisation service for Research

First version of software available soon, with support desk for support for pseudonymisation, data linkage and anonymization

The service makes it as simple as possible for researchers to pseudonymise their data in a secure manner. The pseudonymisation service contributes to high-quality linking of health data for research that complies with legislation.

- Only direct identifiers are pseudonymised.
- Only for quantitative data

For next releases:

- Pseudonymisation of subjects in research data set (e.g. names of doctors or nurses)
- Pseudonymisation of keys (maintaining data integrity)
- Other types of data (qualitative, video, audio, genetic)
- Anonymization

The Data Federation Hub / Human Data

The University Medical Center Groningen and the University of Groningen joined efforts to set up an integrated research support platform: the Data Federation Hub/Human Data. The DFH/Human Data support the whole research data lifecycle while ensuring data security and efficiently protecting the privacy of participants.

www.rug.nl/dfh dfh@rug.nl f.romero.pastrana@rug.nl j.l.van.der.ploeg@umcg.nl

ZorgTTP
Privacy & vertrouwen

Pseudonimisatieservices voor wetenschappelijk onderzoek

2016 – 2018
Uitvoerige gesprekken over use cases RUG-UMCG

Midden 2018
Eerste contact UMCG-ZorgTTP

Q3-Q4 2018
Vooronderzoek

Eind 2019
Demo voor eerste use case
- Pseudonimisatieservice voor kleine studies

Begin 2019
Start requirements en ontwikkeling tweede use case
- Pseudonimisatieservice voor longitudinale studies met EPD data

Midden 2019
Vervolg van de oplossingen via SURF dankzij gewonnen aanbestedingstraject

Eind september 2019
Oplevering eerste release
- Pseudonimisatieservice voor longitudinale studies

Q4 2019 – Q1 2020
Start requirements en ontwikkeling volgende use case
- Pseudonimisatieservice voor kleine studies

Pseudonimisatieservices en Use Cases

Jan Lucas van der Ploeg
Data Engineer, afdeling Informatiemanagement Onderzoek, UMCG

ZorgTTP
Privacy & vertrouwen

Pseudonimisatieservice Groningen

Samenwerking UMCG en RUG

- Onderdeel project Human Subject Research van UMCG en RUG

Onderscheid maken (in proces/stap en techniek)

- Pseudonimiseren
- Anonimiseren
- Koppelen

Zowel techniek als organisatie/proces

- Service met software plus support/consultancy

ZorgTTP
Privacy & vertrouwen

Scope en fasering

Scope bepalen

- Welke data pseudonimiseren?
- Direct identifiers, ook indirect identifiers, quasi identifiers, keys?
- Alleen research subjecten of ook andere personen (zorgverleners)?
- Kwantitatief, ook kwalitatief?
- Gestructureerde data, images, video, audio, geo-data, genetische data?

Gefaseerd

- In releases

ZorgTTP
Privacy & vertrouwen

Context en prioritering

Kleine studies

- Veel (± 1000 per jaar)
- Geen budget
- Weinig kennis pseudonimiseren & anonimiseren
- Scheiden directe persoonsgegevens van onderzoeksdata
- Grootste risico

Grote studies/projecten

- Wel budget
- Kennis en personeel aanwezig
- Maatwerkoplossingen beschikbaar
- Vaak koppelen aan andere bronnen

ZorgTTP
Privacy & vertrouwen

Eerste release van de service

Scope

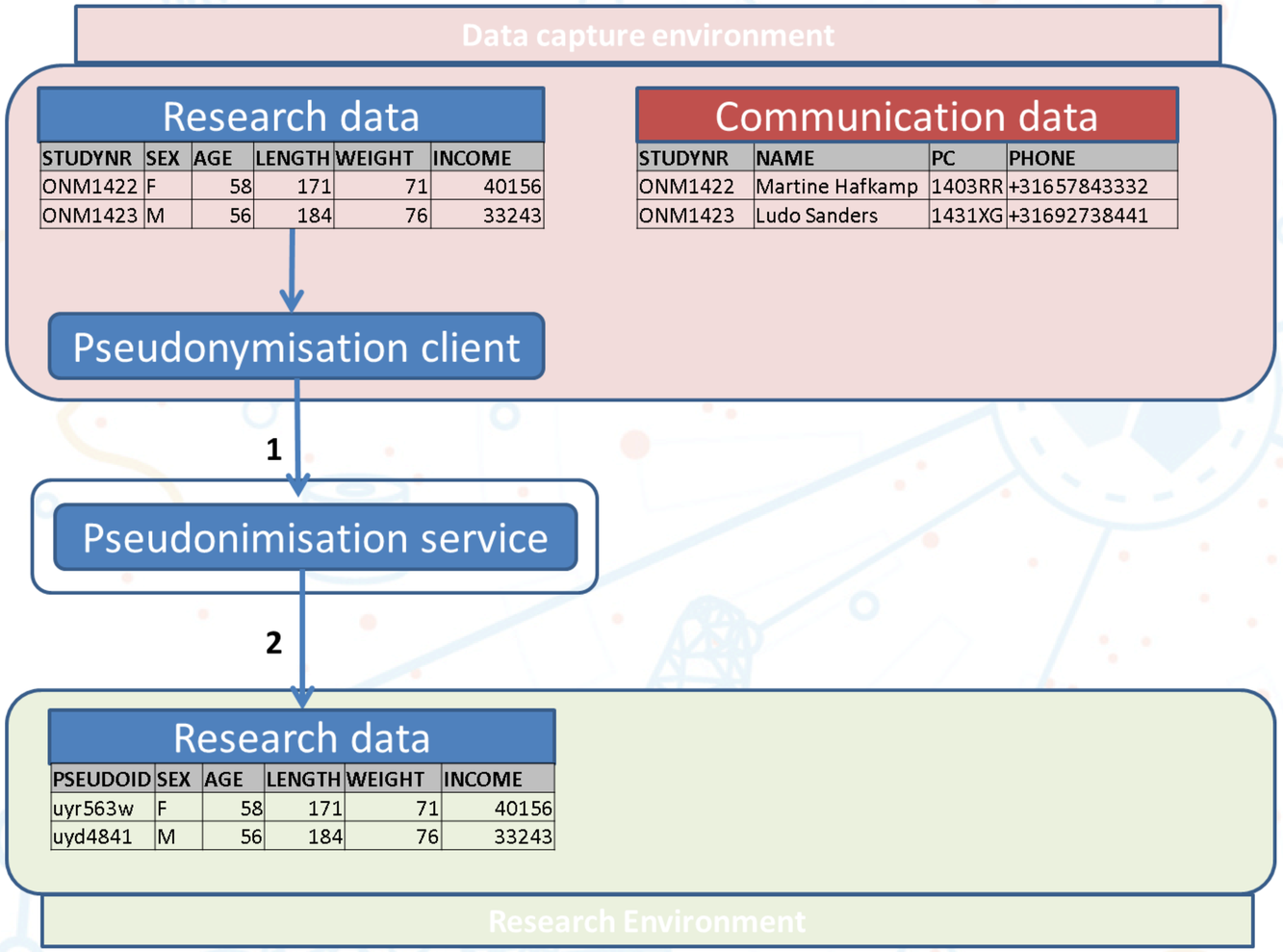
- Software for pseudonymisation
  - Only direct identifiers are pseudonymised
  - Only for quantitative data
- Support for other data and for anonymization and data linkage

4 Use cases

- Pseudonymisation: No data linkage possible
- Pseudonymisation: Data linkage possible
- Process for reversible pseudonymisation
- Pre-matching (and linking) by Trusted Third Party

ZorgTTP
Privacy & vertrouwen

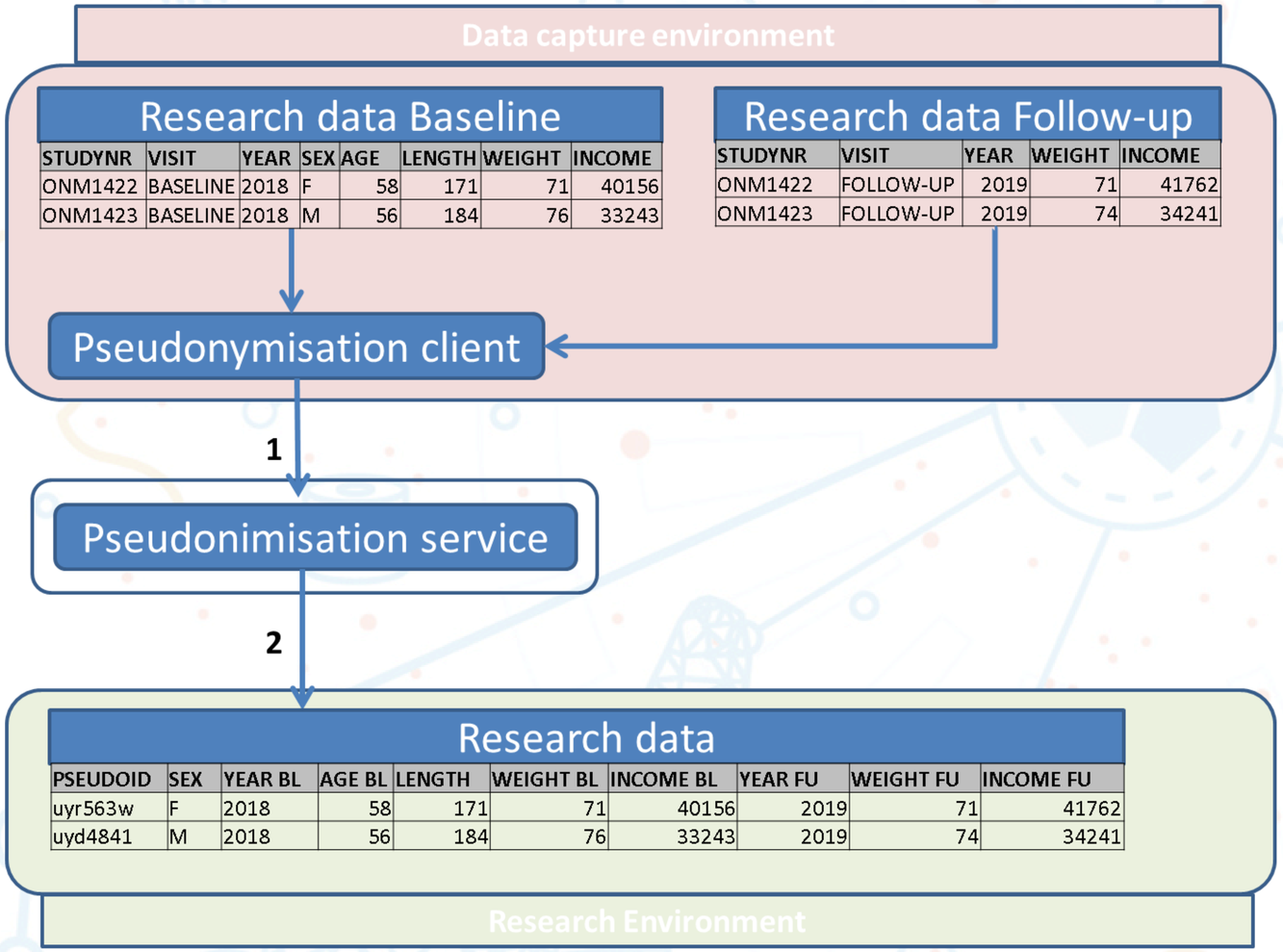
Pseudonimisatie: geen data linkage



- For researchers, self service
- Small studies
- Irreversible (default)
- Different pseudonyms for different sessions (default)
- Not possible to link published datasets
- Creating research pseudonyms

ZorgTTP
Privacy & vertrouwen

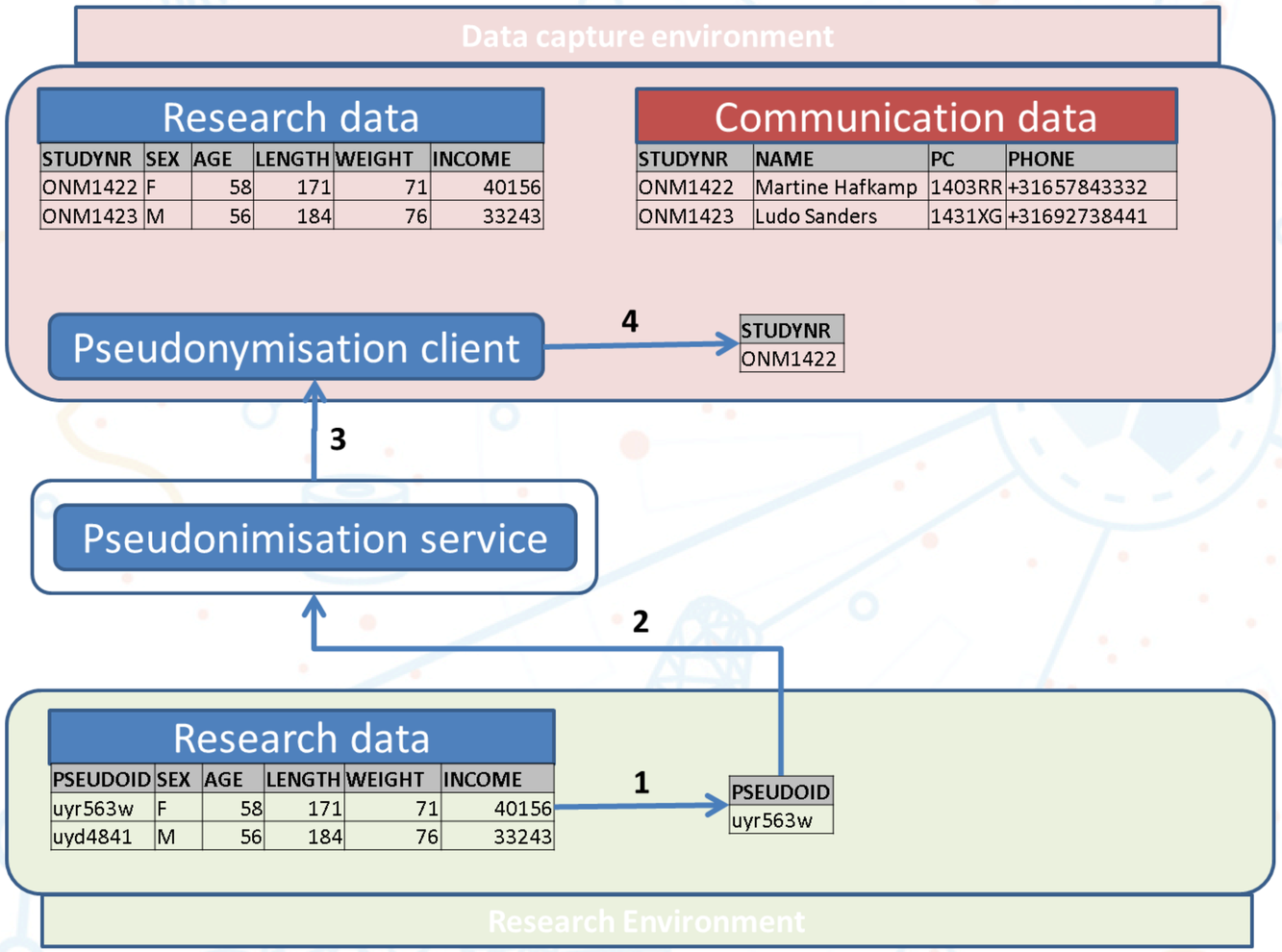
Pseudonimisatie: data linkage mogelijk



- For data stewards/advanced data managers, no self service
- Longitudinal studies, large cohort studies and biobanks
- Subjects are traceable over time
- Pseudonym for subject is stable in project over different sessions

ZorgTTP
Privacy & vertrouwen

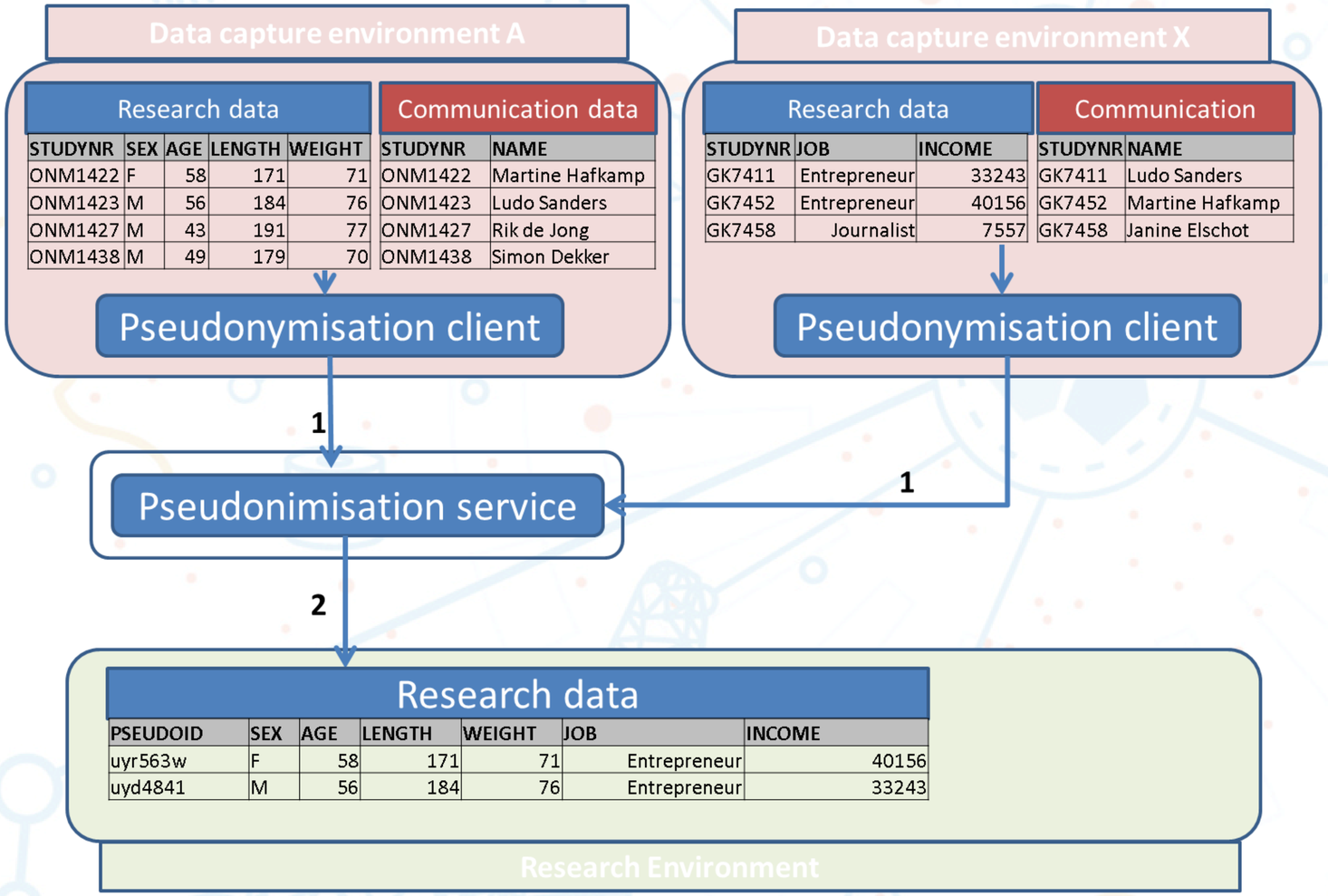
Proces voor omkeerbare pseudonimisatie



- For data stewards/advanced data managers, no self service
- **Only when necessary**
- Not the default option, extra procedures and contracts when needed
- Incidental findings

ZorgTTP
Privacy & vertrouwen

Prematching en linkage door Trusted Third Party

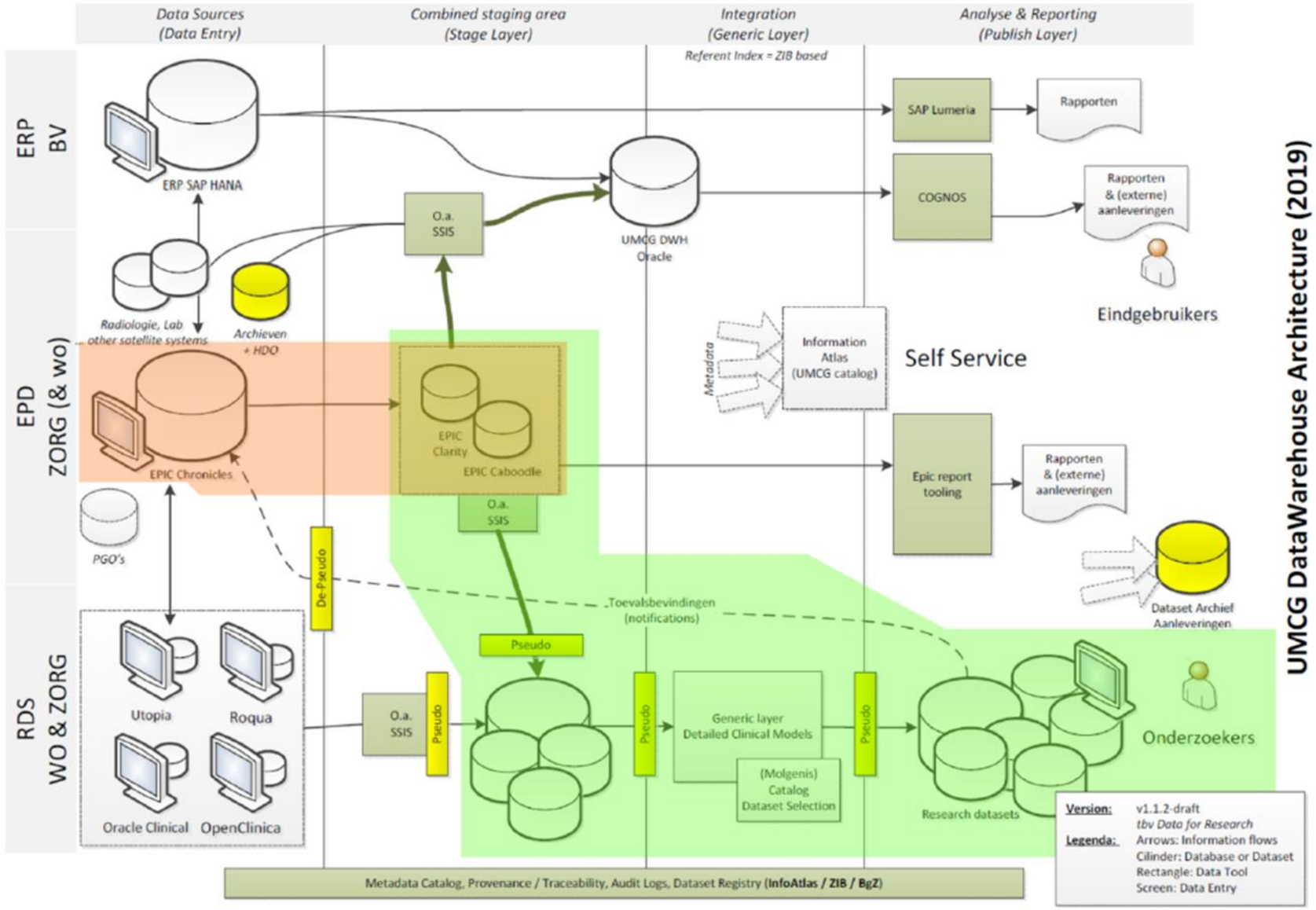


- By TTP, no self service
- Linking data sets from different sources
- Pre matching and linkage by TTP
- Only researcher knows overlap between sets

ZorgTTP
Privacy & vertrouwen

2019: Prioriteit EPD data voor onderzoek

- Pseudonimisatie service is voorwaarde voor EPD data voor onderzoek
- Prio 1ste release software
  - 2e use case; longitudinale studie met stabiele pseudoniemen



ZorgTTP
Privacy & vertrouwen

Release 1: Pseudonimisatieservice voor longitudinale studies

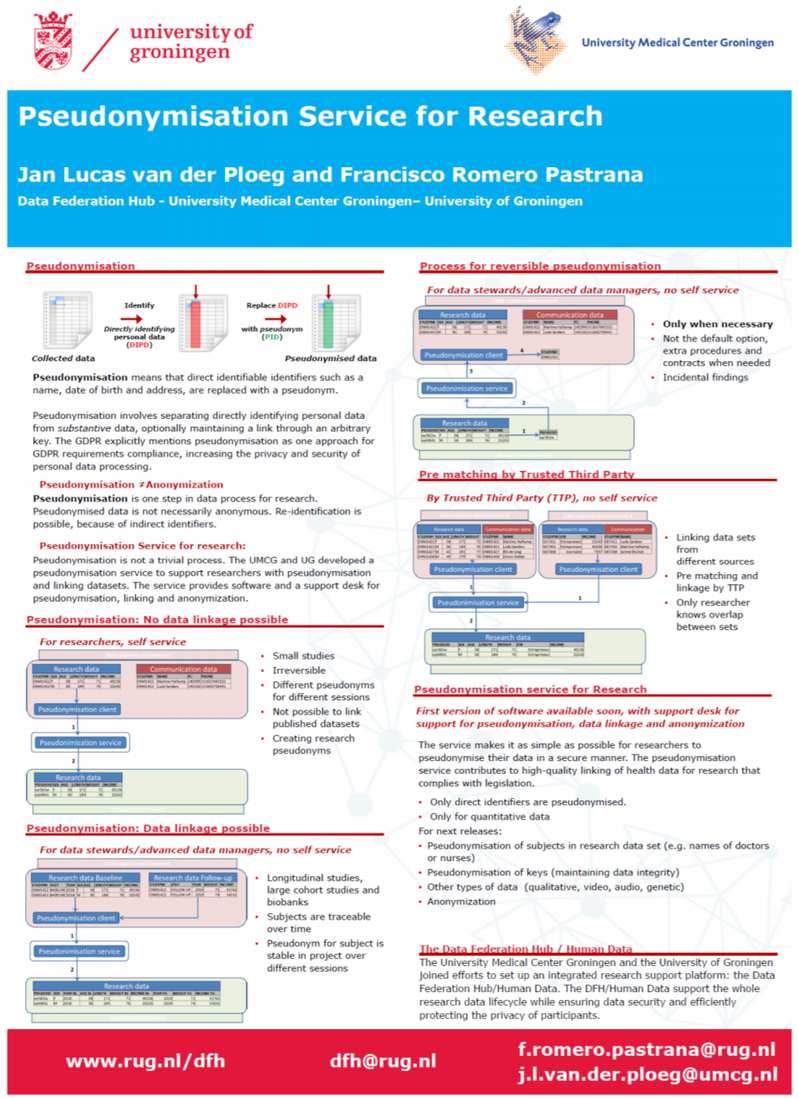
Simone van Wijngaarden
Projectleider, Stichting ZorgTTP

ZorgTTP
Privacy & vertrouwen

Vereisten pseudonimisatieservice longitudinale studies (use case 2)

- REST API
- FHIR bundle
- JSON
- Self service
- Onomkeerbare pseudoniemen
- Pseudoniemen
  - Zelfde input, andere studie = ander pseudoniem
    - Stabiele pseudoniemen binnen studie
    - Unieke pseudoniemen
- Limieten
- Responstijd
- Geteste oplevering
- Eindgebruikersdocumentatie

ZorgTTP
Privacy & vertrouwen



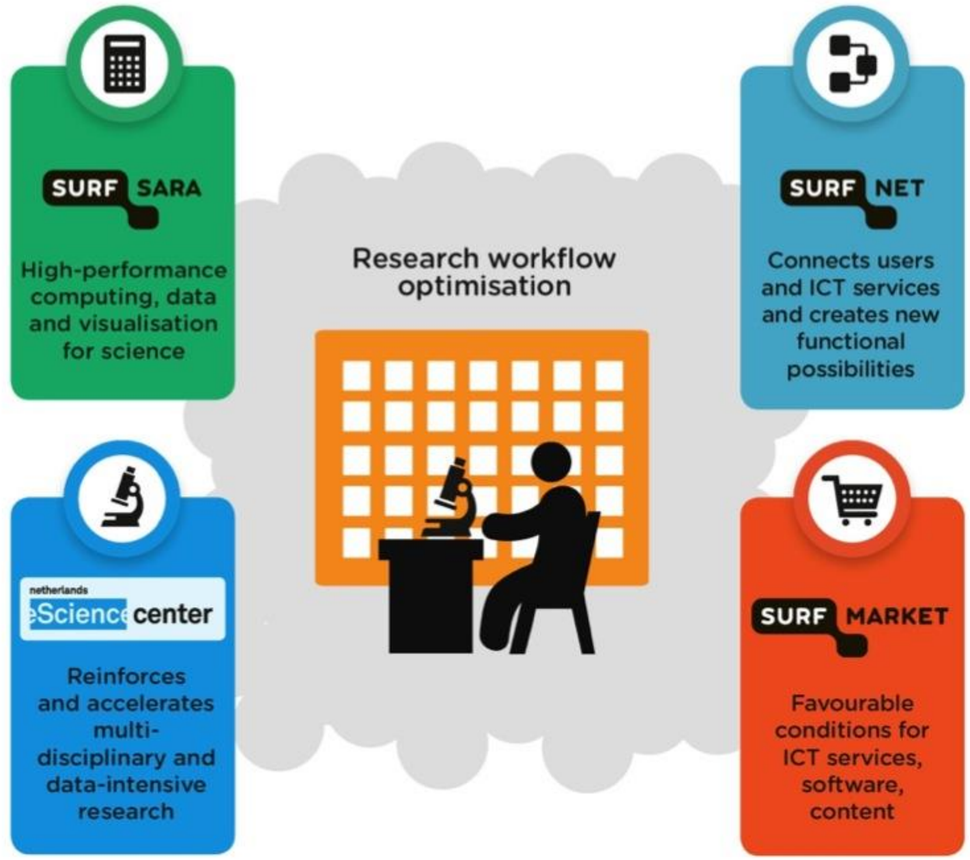
Opschaling pseudonimisatieservices

Robert Griffioen
Projectleider de-identificatie project, SURF

ZorgTTP
Privacy & vertrouwen

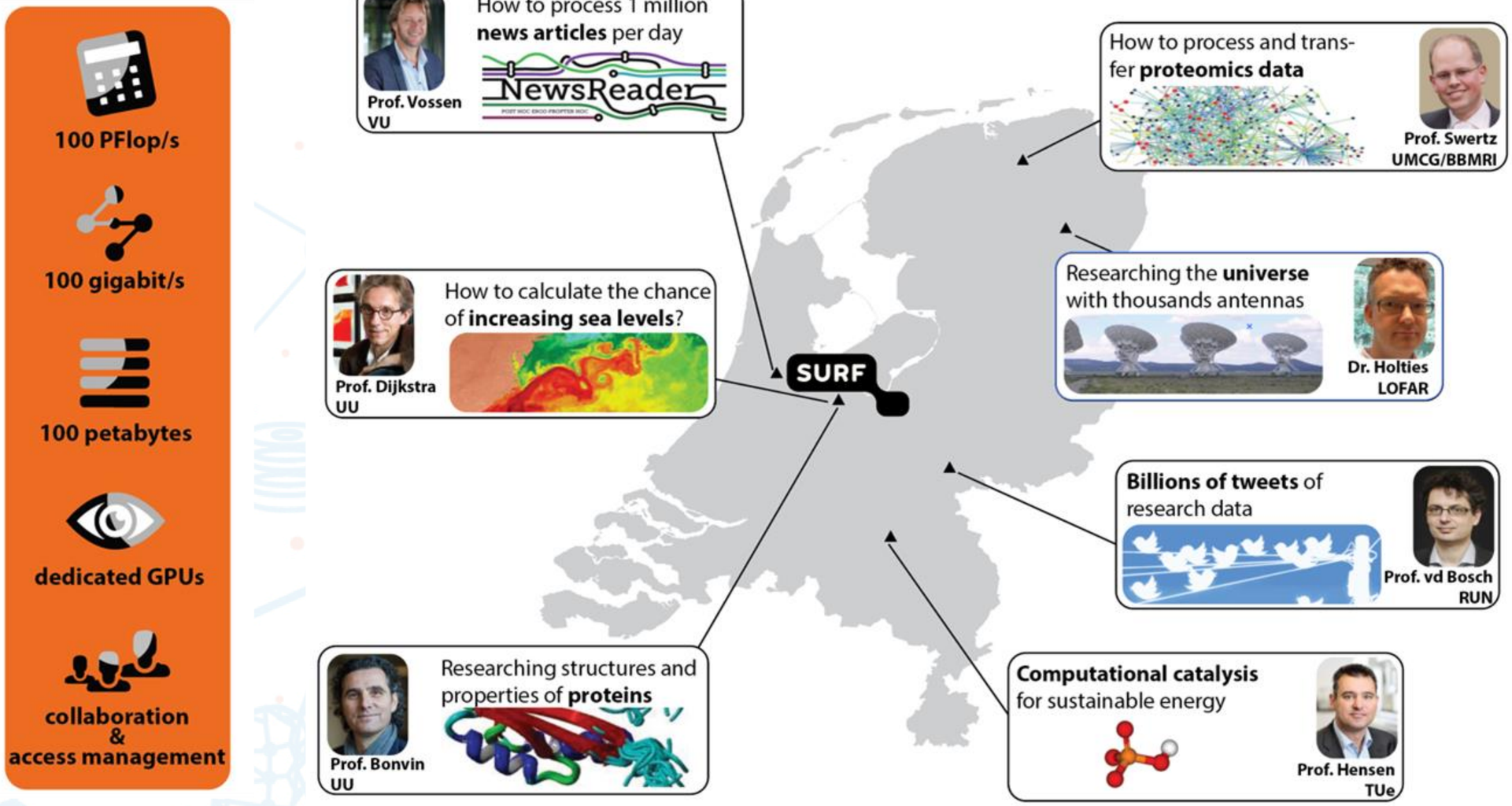
SURF coöperatie voor onderzoek en onderwijs

- Coöperatie van Nederlandse instellingen hoger onderwijs en onderzoek (>160), bijv. universiteiten, academische medische centra, hoge scholen en mbo's.
- Leverancier diensten of makelaar van diensten
- Innovatie partner en coördinator van dienstontwikkeling
- Kenniscentrum
- We worden 1 SURF per 1 juli 2020: verschillende doelgroepen en culturen
- SURFsara van oudsher gericht op onderzoek



ZorgTTP
Privacy & vertrouwen

SURFsara: voorbeeld projecten



ZorgTTP
Privacy & vertrouwen

TTP-pseudonimisatie project

- Coördinerende SURF Contactpersonen – WO project
- Brede scope: koppelen van data centraal
- Analyse fase
- Meervoudige onderhandse aanbesteding:

Hierbij kwamen project ZorgTTP-Groningen en dit project samen

1	Opslag of archivering van anonieme persoon-herleidbare data of persoonsgegevens
2	Pseudonimisatie van persoonsgegevens
3	Sleutelmanagement (sleutel-generatie, -opslag, -beheer en –vernietiging)
4	Beheer van versleutelde data
5	Data koppeling op basis van versleutelde data
6	Supportloket ter ondersteuning van onderzoekers
7	Informed consent

ZorgTTP
Privacy & vertrouwen

De-identificatie pilot

- Nadruk van de pilot ligt of de-identificeren, pseudonimiseren of anonimiseren, van persoonsdata.

- Nut/Noodzaak: allerlei lokale oplossingen binnen instellingen waarvan men op centraal bestuurlijk niveau niet op de hoogte is, laat staan over de kwaliteit. Vanuit governance oogpunt is het daarom wenselijk een uniforme wijze van de-identificeren te introduceren die voldoet aan de privacy en security richtlijnen van de verschillende instituten.

- Tool is vereiste, maar voor een complete oplossing zijn bijvoorbeeld ook organisatorische aspecten een eis.

ZorgTTP
Privacy & vertrouwen

De pilot: werkwijze

- Deelnemers: Rijksuniversiteit Groningen, UMC Groningen, Universiteit Leiden en Vrije Universiteit

- Doel: ontwikkeling van pilot- tot Nationale SURF productiedienst/*SURF*-dienst. Verder uitwerken van reeds ontwikkelde diensten van 'Groningen' en ZorgTTP (zie presentatie Jan Lucas).
  - Use case 1: Kleinschalige studies (onomkeerbare pseudonimisatie)
  - Use case 2: Studies die periodiek worden bijgewerkt.

- Agile ontwikkeling: ontwikkeling en evaluatie tussenproducten

- Requirements en evaluatie door:
  - Onderzoekers
  - Data Steward
  - Functionaris Gegevensbescherming
  - IT

ZorgTTP
Privacy & vertrouwen

De pilot: verder..

- Voordelen:
  - Werken met een ervaren en erkende partij op privacy gebied
  - Inrichten van organisatie samen met partner instellingen

- Deelname aan de productiedienst door instellingen
  - Reeds kenbaar maken van interesse om op de hoogte te worden gehouden:

  Simone van Wijngaarden (simone.van.wijnaarden@zorgttp.nl) en
  Robert Griffioen (robert.griffioen@surfsara.nl).

ZorgTTP
Privacy & vertrouwen

De toekomst: ontwikkeling nieuwe privacy diensten; van pilot tot dienst

- Ontwikkeling nieuwe privacy diensten:
  - Use case 3: Omkeerbare pseudonimisatie
  - Use case 4: Anonimiseren van data
  - Koppelen van data
  - Beelddata, etc.
- Binnen dit project, maar mogelijk ook: andere deelnemers, andere leverancier, ander project.

ZorgTTP
Privacy & vertrouwen

Vragen?

Dank!

Jan Lucas van der Ploeg

Data Engineer, Afdeling Informatiemanagement Onderzoek, UMCG

Robert Griffioen

Projectleider de-identificatie project, SURF

Simone van Wijngaarden

Adviseur, CIPP/e, Stichting ZorgTTP

Simone.van.wijngaarden@zorgttp.nl

06-13631500

ZorgTTP
Privacy & vertrouwen