

# Data exchange rapport

marktscan  
& aanbeveling

15 april 2020



# Let's talk about

- **Executive summary** 3
- **Introductie** 5  
Doel en scope van het project.
- **Owners & Requestors** 13  
Wie zijn de gebruikers en wat is hun relatie?
- **Factoren voor delen** 17  
Welke factoren en voorwaarden wegen mee in het (niet) delen?
- **Data deel oplossingen** 22  
Welke oplossingen worden nu gebruikt?
- **Aanbeveling** 33
- **Bijlage** 36  
Bronnen en bijgevoegde modellen

# Executive summary

3 take-aways

1. De grootse uitdaging om tot datadelen te komen ligt bij de data aanbieders. De belangrijkste factoren in die bereidheid zijn Return-On-Investment (Effort / Gains) en Trust (Risk/Control). Het gewicht van deze factoren, waar techniek slechts een onderdeel van is, verschilt tussen de verschillende domeinen die we geïdentificeerd hebben.
2. Op dit moment bestaan er verschillende platforms en initiatieven die data exchange mogelijk maken met elk hun eigen focus en eigenschappen. Veel daarvan zijn veel nog in een vroege fase en het ecosysteem is nog volop in beweging.
3. Door het complexe en dynamische karakter van het datadelen is er niet een duidelijke market fit voor de Data Exchange aan te geven wat bevestigd werd in onze interviews de afgelopen maanden. Onze aanbeveling is om zoveel mogelijk in contact te komen met initiatieven om aansluiting te vinden met projecten die behoefte hebben aan de Data Exchange functionaliteit. Verdere product ontwikkeling zal dan ook moeten plaatsvinden binnen de context van deze projecten en/of initiatieven

/ Hoofdstuk 1

# Introductie

Doel en scope van het project

Het doel van het project was:

‘Identificeren van voorwaarden per markt segment voor het ontwikkelen van de data exchange’

## Conclusie op 31 januari, zoals gedeeld met de stuurgroep:

- We bevinden ons in een complex en dynamisch ecosysteem van vele initiatieven in verschillende phases van 'maturity' en oplossingsrichtingen en in verschillende deelgebieden met hun eigen uitdagingen;
- Elke manier van 'data sharing' is uniek, waar verschillende zaken (zoals motivatie, juridisch, data complexity, privacy, techniek van bestaande oplossingen) een rol spelen, en waarschijnlijk complexer zijn dan de technische oplossing zelf;
- Hierdoor is het oorspronkelijk doel van het project niet realistisch.

Op basis hiervan stelden we een pivot voor.

## Het doel na de pivot:

- Het komen tot een aanbeveling en het identificeren van voorwaarden van een pilot voor de Data Exchange (building blocks, advies gebied/market, mogelijke pilot partners) middels het houden van interviews en het organiseren van een co-creatie sessie.
- Als onderdeel daarvan wordt een market scan opgeleverd waar het volgende wordt onderzocht welke domeinen zijn er, wat speelt er in die domeinen, en wat voor oplossingen zijn er op dit moment

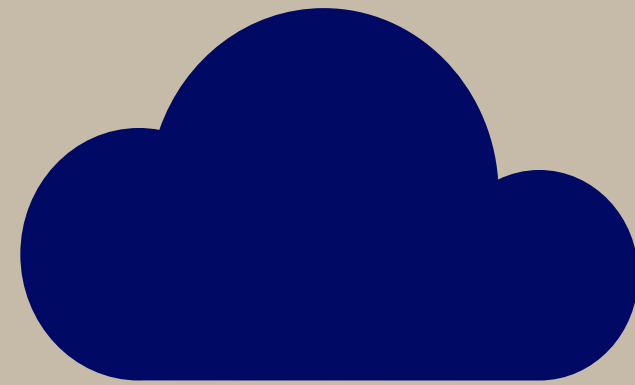


## De situatie na maart is veranderd:

- Co-creatie sessie niet mogelijk binnen korte tijd /  
Interesse voor sessie leek lauw. Daarom is het vinden van  
pilot partners nu niet realistisch
- Market scan bevestigt beeld van gefragmenteerd,  
complex ecosysteem in beweging

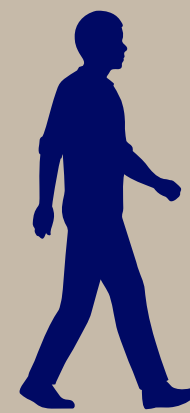
Op basis hiervan doen wij in dit rapport een suggestie voor  
vervolgstappen.

## Scope



# Onderwerp

Data die gedeeld kan worden onder bepaalde voorwaarden waar onderzoekers interesse in hebben.



# Doelgroep

(Externe) onderzoekers / onderzoeksgroepen die niet op locatie bij de data aanbieders werken.

Data aggregators, ziekenhuizen+ medische instellingen, onderzoekers + universiteiten en bedrijven die in het 'bezit' zijn van data.



# Regio

Dit onderzoek richt zich op Nederlands data-aanbieders en onderzoekers

# Data

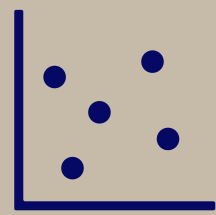
Schaal van open data  
tot gesloten data

Data 'alleen delen als' is  
in scope voor dit  
onderzoek.



# Soorten Data Owners

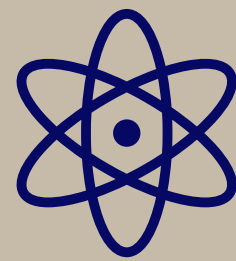
met voorbeelden die we tegengekomen zijn/mee gesproken hebben



Data  
aggregators

Zorgverlening  
(Palga, NZA)

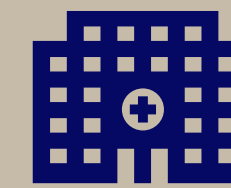
Sociaal-economisch  
(CBS, gemeente)



Ziekenhuizen  
+ medische instellingen

Ziekenhuis (AMC, vuMC,  
St. Antoniusziekenhuis)

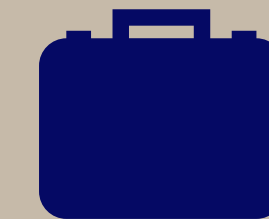
Verzekeringsmaatschappij  
en (Zilveren Kruis)



Onderzoekers  
+ universiteiten

Universiteiten  
(Twente, Wageningen,  
Groningen)

Onderzoekers



Bedrijven

Friesland-Campina,  
Elsevier

# Owners & Requestors

Wie zijn de gebruikers en wat is hun relatie?

## Het delen van data

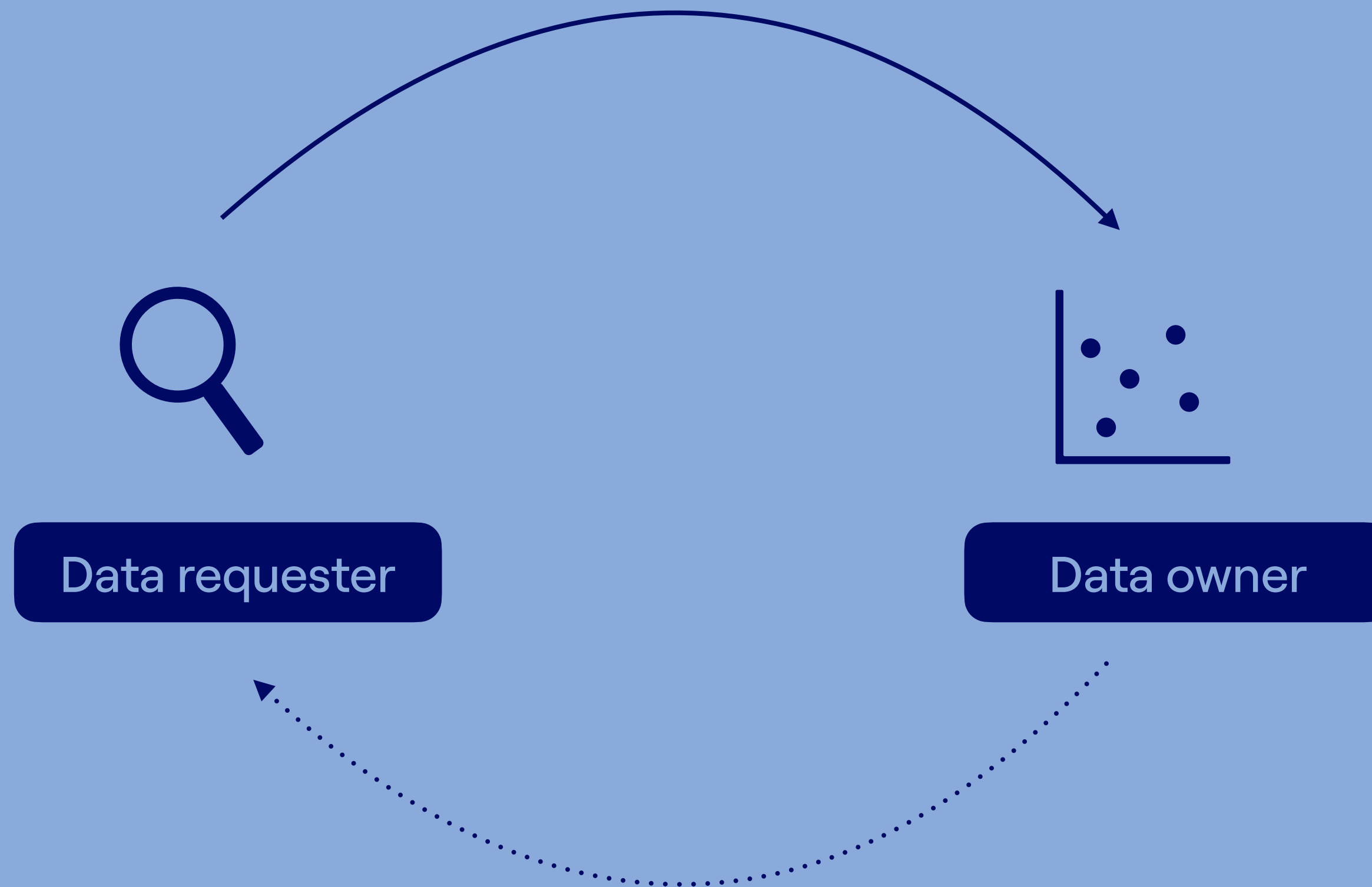
In het proces van data delen zullen we spreken over twee partijen:

**Data Owner.** Deze entiteit kan de data verzamelen en opslaan. Wettelijk gezien is er geen “eigenaar” van data er kan een verschil zijn tussen data beheerder (provider) en data rechthebbende.

**Data Requester.** In de scope van dit project gaat dit om een onderzoeker die voor zijn werk een databron wil gebruiken.

Er liggen meer voordelen bij de requester dan bij de owner van de data. Dit is een van de grootste struikelblokken voor het delen van data.

- + Data om onderzoek te onderbouwen.
- + Vooruitgang in wetenschap.
- Meer werk rondom verkrijgen, managen van data.



- ~ Mogelijk verkrijgen van nieuwe inzichten.
- Risico's rondom privacy en security.
- Extra werk zonder directe return on investments (ROI).

# Willingness to share data =

De bereidheid van owners om data te delen met requestors waar geen relatie bestaat wordt bepaald door de wisselwerking tussen return on investment (ROI) en trust (vertrouwen).

ROI

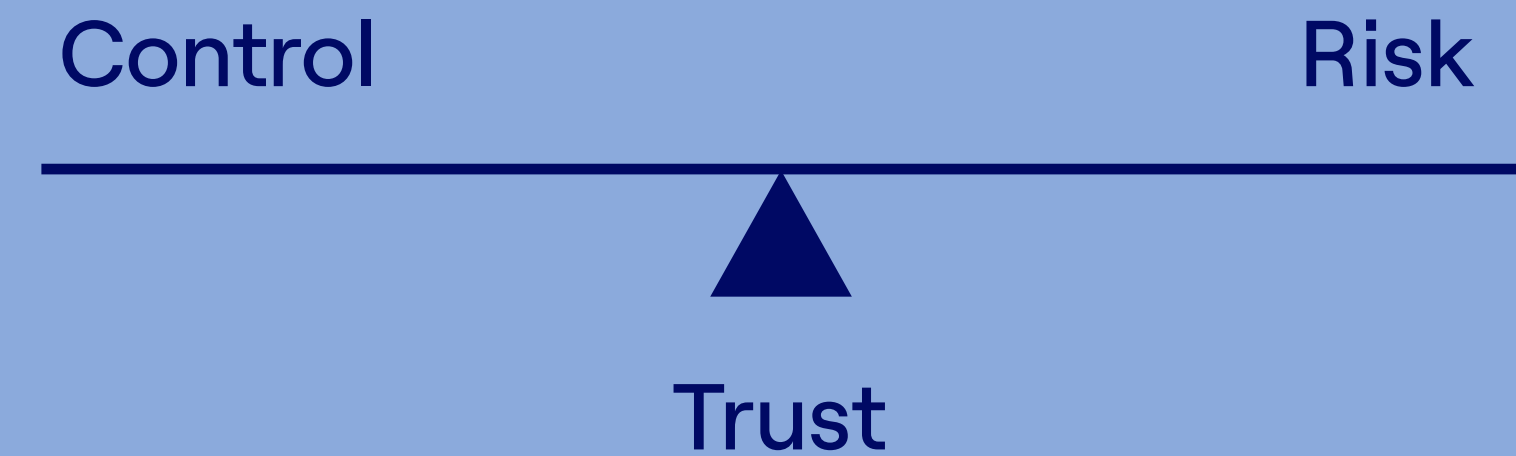
+

Trust



ROI wordt bepaald in de wisselwerking tussen hoeveel moeite het kost om de data te delen en hoeveel waarde de uitwisseling oplevert.

Vertrouwen wordt bepaald in de wisselwerking tussen hoe groot de risico's zijn (vanwege privacy and competitie) en hoe veilig en controleerbaar de uitwisseling en het gebruik is.





# Factoren voor delen

Welke factoren en voorwaarden wegen mee in het (niet) delen

## Factoren voor delen

Data is nu veelal of open of gesloten. In het gesloten gebied zit echter een grijs vlak waarin er data gedeeld kan worden onder bepaalde voorwaarden.

Alhoewel deze voorwaarden lastig te generaliseren zijn en voor iedere case en domein anders lijken te zijn, zijn er factoren te identificeren die een rol spelen in het wel of niet delen van data.

# Factoren voor delen

## Besluitvorming

Hoe complex is het besluitproces? Welke wettelijke limitaties en afspraken moeten er worden vastgelegd?

## Doeleinde

Waar zal de data voor worden gebruikt?

## Data exposure requesters

Welke toegang geef je aan de requestor en wat kan die inzien (data, model, uitkomst)?

## Operationele effort

Hoeveel werk is het om de afspraken te maken en datasets effectief te delen?

## Privacy/GDPR

In hoeverre is de data onderhevig aan privacy/AVG wetgeving en/of gevoeligheden?

## Veiligheid

Hoe veilig is de omgeving waarin de data uitwisseling plaatsvindt?

## Waarde data inzicht

Welk voordeel kan de data aanbieder uit de analyses halen?

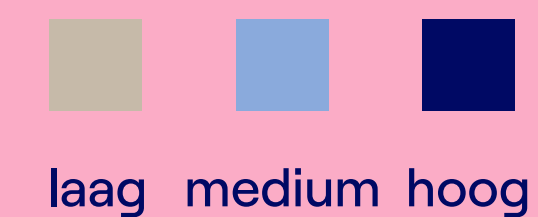
## Concurrentie-gevoelig

Welke risico's spelen er rondom concurrentie?

## Data controle

Wat willen aanbieders controleren tijdens en na het uitwisselen van de data?

# Belang in factoren per data owner groep

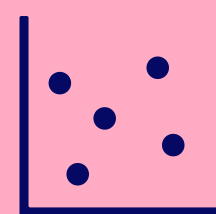


\*relatief

		Data aggregators	Ziekenhuizen + medische instellingen	Onderzoekers + universiteiten	Bedrijven
Effort	Besluitvorming	hoog	hoog	laag	medium
	Operationele effort	medium	hoog	medium	medium
Gains	Waarde inzicht voor provider	laag	hoog	medium	hoog
	Doeleinde	medium	medium	medium	hoog
Risk	Privacy / AVG	hoog	hoog	medium	medium
	Concurrentie-gevoelig	laag	laag	hoog	hoog
	Data exposure onderzoekers	medium	hoog	medium	hoog
Control	Veiligheid	hoog	hoog	laag	medium
	Data controle	hoog	hoog	laag	medium

## Beperkende factor in delen: privacy vs concurrentie

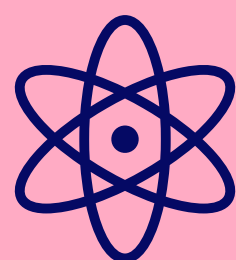
### *Privacy gevoelig*



Data  
aggregators

Zorgverlening  
(Palga, NZA)

Sociaal-economisch  
(CBS, gemeente)

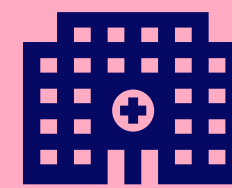


Ziekenhuizen  
+ medische instellingen

Ziekenhuis (AMC, vuMC,  
St. Antoniusziekenhuis)

Verzekeringsmaatschappij  
en (Zilveren Kruis)

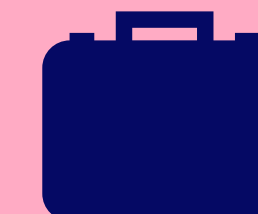
### *Concurrentie gevoelig*



Onderzoekers  
+ universiteiten

Universiteiten  
(Twente, Wageningen,  
Groningen)

Onderzoekers



Bedrijven

Friesland-Campina,  
Elsevier

# Data deel oplossingen

Welke oplossingen worden nu gebruikt

# Overzicht van bestaande oplossingen

## Ad hoc data uitwisseling

Via mail of online cloud service de data versturen.

## Ad hoc technische oplossingen

Oplossing speciaal gebouwd voor een specifieke vraag.

## Consortia/infrastructuur oplossingen

Vaak een meerjarig plan met grote visie, vind ook veel overleg in plaats over langere tijd.

## Commerciële oplossingen

Oplossingen gemaakt door commerciële partijen

## Specifieke tools

Gemaakt voor een klein onderdeel in het proces van data delen.

# Wat gebeurt er al?

## Ad hoc data uitwisseling

### 1. Ad hoc data uitwisseling

Data wordt als het ware op een USB stick gedeeld. Hebben gezien bij zowel aggregators (Palga) als bedrijven

#### + Voordelen

- Lage kosten, laagdrempelig

#### - Nadelen

- Onveilig
- Geen waarborgen privacy/AVG



# Wat gebeurt er al?

Ad hoc technische oplossingen

## 1. Ad hoc technische oplossingen

Deze worden zelf gebouwd voor een specifiek doeleinde. Zie je bij universiteiten als ook bedrijven

### + Voordelen

- Adresseert specifieke uitdagingen voor het project (e.g. data structuur, veiligheid)

### - Nadelen

- Vaak duur
- Omslachtig

# Wat gebeurt er al?

Consortia / infrastructuur oplossingen

## Consortium /infrastructuur oplossingen

Grote, compleet projecten met meerdere stakeholders die voor een bepaald domein een infrastructurele oplossing zoeken

### 1. Personal Health Train

(<https://www.dtls.nl/fair-data/personal-health-train/>)

- Created by Dutch Techcentre for Life Science (large consortium)
- ‘The PHT is primarily a FAIR data infrastructure approach so that a research question (including a specific type of data analysis) can be posed to the data without the need for data to leave the station (hospital, PHR, cohort, etc.)’

### 2. Enabling Personalised Interventions

(<https://delaat.net/epi/>)

A consortium of medical professionals, data scientists, IT-infrastructure experts, machine learning researchers and legal experts have designed Enabling Patient Interventions to liberate, analyse, and action medical data in a trustworthy way

# Wat gebeurt er al?

Consortia / infrastructuur oplossingen

## 3. ODISSEI ([www.odissei-data.nl/en/](http://www.odissei-data.nl/en/))

Ontwikkelt een duurzame wetenschappelijke infrastructuur voor de sociale wetenschappen in Nederland. Via ODISSEI krijgen onderzoekers in de sociale wetenschappen toegang tot grootschalige en longitudinale dataverzamelingen die gekoppeld zijn aan CBS-registraties.

## 4. Farm data train ([www.dtls.nl/fair-data/farm-data-train/](http://www.dtls.nl/fair-data/farm-data-train/))

Aims to connect agricultural data to make them more usable (zie Personal Health Train)

### + Voordelen

- In potentie een betrouwbare, volledige infrastructuur voor data uitwisseling.

### - Nadelen

- In vroege fase. Complex dus daarom niet snel operationeel.

# Wat gebeurt er al?

## Commerciële oplossingen

Veelal internationale oplossingen vooral gericht op het bedrijfsleven.

### 1. Data Republic ([www.datarepublic.com](http://www.datarepublic.com))

Cloud-based Data Exchange Technology: 'Data Republic's Senate solves the complex legal, data security and privacy barriers to data sharing with partners, making it simpler and more secure for your organisation to share and license data'

Meer gericht op 'secure data exchange' tussen bedrijven

- Connecting the dots - informatie van consumenten aanvullen met gegevens van andere bedrijven
- Data-analyse bedrijven die bedrijfsinformatie analyseren
- Marktplaats functie

Voornaamste elementen:

- Legal & licensing framework
- Managing control: who has access, what use of data & outputs are allowed,
- Anonymising PPI
- Information Security

Werkt als 'derde partij' - data aanbieders laden data op Data Republic platform waar de analyses plaatsvinden.

# Wat gebeurt er al?

## Commerciële oplossingen

### 2. Dawex ([www.dawex.com](http://www.dawex.com))

- ‘Sell-buy-share data’: sourcing, monetizing and exchanging data securely and in full compliance with regulations to ensure the integrity of licensing contracts.
- Blockchain as option
- Focus on marketplace aspect
- Claims to be market leader

### 3. Materials Zone ([www.materials.zone](http://www.materials.zone))

- ‘The global network for Connected Research Data: a platform combining computational, experimental and data tools creating a research network where researchers can trade knowledge, samples, and services to innovate faster’
- More for collaboration than for researchers to analyze third party data.

#### + Voordelen

- Geavanceerde, veilige omgeving met vele tools (volgens website informatie).

#### – Nadelen

- Commerciële platforms (kost geld)
- Overengineered
- Focus op marktplaats
- Niet toegespitst op onderzoek
- Vaak in vroege fase

## Wat gebeurt er al?

Specifieke tools voor specifieke momenten in het proces

Gemaakt voor een klein onderdeel in het proces van data delen.

### 1. ZorgTTP

- 'Biedt pseudonimisatie diensten voor health data

### 2. Podium

- Administratie tool voor aanvragers gebruik Bio and Databanking

#### + Voordelen

- Toegepaste oplossingen voor specifieke problemen in specifieke gebieden

#### - Nadelen

- Niet flexibel
- Lossen maar deel van uitdaging op

# Eigenschappen bestaande oplossingen



Negatief



Positief



NVT

		Ad-hoc data uitwisseling	Ad hoc technische oplossingen	Consortia / infrastruct. oplossingen	Commerciële oplossingen	Podium	ZorgTTP	Data Exchange
ROI	Snel in te zetten	Hoog	Laag	Laag	Hoog	Hoog	Hoog	Hoog
	Kosten van gebruik	Laag	Hoog	Hoog	Hoog	Laag	Laag	Laag
Trust: risk	Privacy / GDPR waarborgen	Laag	Depends	Hoog	Hoog	NVT	Hoog	Hoog
	Data exposure onderzoekers	Depends	Depends	Laag	Laag	NVT	Laag	Laag
Trust: control	Security	Laag	Depends	Hoog	Hoog	Hoog	Hoog	Hoog
	Data control	Laag	Depends	Hoog	Hoog	NVT	Hoog	Hoog
Other	Maturity	NVT	Laag	Laag	Medium	Medium	Medium	Laag
	Kosten van ontwikkeling	Laag	Hoog	Hoog	Hoog	Laag	Laag	Depends
	Flexibel	Hoog	Hoog	Depends	Hoog	Laag	Laag	Depends
	Herbruikbaarheid	Laag	Laag	Hoog	Hoog	Hoog	Hoog	Hoog
	Functionliteit	NVT	NVT	Breed	Breed	Beperkt	Beperkt	Beperkt

# Functionaliteiten in bestaande oplossingen

- nee   
 + ja   
 ~ nvt/soms

	Ad-hoc data uitwisseling	Personal health train	CBS microdata facility	Dexes	Farm data train	Data Republic	Dawex	Materials Zone	Podium	ZorgTTP	Data exchange
Data processing tools	-	+	~	?	?	+	+	+	-	-	-
Pseudonimiseren	+	+	+	?	?	+	+	-	-	+	-
Secure containers	-	+	+	-	?	+	-	-	-	-	+
Secure computing	-	-	-	-	-	-	-	-	-	-	-
Juridische overeenkomsten	+	+	+	+	+	+	-	-	?	+	~
Registratie	~	?	+	-	?	+	+	+	+	~	-
Data marktplaats	-	~	-	+	-	+	+	+	-	-	-



/ Hoofdstuk 5

# Conclusie & aanbevelingen

Stappen om te gaan nemen

## Conclusie & aanbevelingen voor volgende stappen

- Data delen gebeurt nu minder vaak dan zou kunnen. Oorzaken zijn het ontbreken van ‘mature solutions’ en het gebrek aan incentives vanuit de data owner kant. Ook is de vraag bij wetenschappers vaak latent vanwege het ontbreken van oplossingen om het data delen te bewerkstelligen.
- We bevinden ons op dit moment in een complex ecosysteem met vele initiatieven om data oplossingen te bieden. Deze zitten in verschillende niveaus van ‘maturity’ en oplossingsrichtingen, en in verschillende deelgebieden met hun eigen uitdagingen.
- Elke manier van ‘data sharing’ is uniek, waar verschillende zaken (motivatie, juridisch, data complexity, privacy, techniek van bestaande systemen) een rol spelen, en complexer zijn dan de technische oplossing zelf.
- The Data Exchange heeft een redelijk unieke functionaliteit die echter alleen succesvol zal zijn als die aansluiting als (pilot) partner vindt bij een initiatief dat juist op dat moment behoefte heeft aan een dergelijke oplossing. Het moet gezien worden als een potentieel puzzelstukje in een complexe en dynamische omgeving.
- Taak is daarom over zoveel mogelijk in contact te komen en te spreken met bestaande initiatieven om de juiste match te vinden (wordt reeds gedaan door SURF). In de context van die match moet gekeken worden naar specifieke functionaliteiten en de vorm van de samenwerking (e.g. pilot of full product development) en de product features
- Ook raden wij aan om zoveel mogelijk Surf te positioneren als een actieve partij op dit gebied. Als onderdeel daarvan adviseren wij het organiseren van een webinar met als thema ‘de staat van data delen in Nederland’. Deze webinar zal ook fungeren als afsluiting van dit project voor de mensen die geïnterviewd waren, en als acquisitie tool voor (pilot) partners for the Data Exchange.
- Onze aanbeveling is om een periode af te spreken waarin gezocht wordt naar (pilot) partners waarna er een go-no go volgt (bijvoorbeeld eind 2020).

That's it,  
let's chat

**bit** Prototyping  
futures

# Bijlagen

We'll keep it understandable

# Let's talk about

- **Overzicht elementen** 38  
Wat wegen we tegen wat af.
- **Overzicht geïnterviewden** 39  
Wie hebben we gesproken?
- **Modellen** 40  
Anderen modellen en figuren die zijn gemaakt.
- **Owners** 48  
Overzicht data owners
- **Functionaliteiten** 53  
Overzicht van technische oplossingen
- **Colofon** 59

# Overzicht van elementen

	Factoren in data delen (why)	Huidige oplossingen (what)	Functionaliteiten in huidige oplossingen (how)
Effort	Besluitvorming	Snel in te zetten	Data processing tools
	Operationele effort	Kosten van gebruik	
Gains	Waarde inzicht voor provider	Operationele effort	
Risk	Privacy / GDPR	Privacy / GDPR waarborgen	Pseudonimisering
	Concurrentie-gevoelig	Data exposure onderzoekers	
	Data exposure onderzoekers		
Control	Security	Security	Secure containers
	Data control	Data control	Secure computing
			Juridische overeenkomsten
			Registratie
Other		Maturity	
		Kosten voor ontwikkeling	
		Flexibel	
		Herbruikbaarheid	

# Lijst geïnterviewden:

Geïnterviewd	Bedrijf	Rol	Categorie
Mark Siebert	Elsevier	Director Research Collaborations	Bedrijf
Wim Sijstermans	NZA	CTO	Aggregator
Bahar Mehmani	Elsevier	Reviewer Experience Lead	Requester
Jaap van Ekris	Palga	Data protection officer	Aggregator
Rinus Voorham	Palga	Adviseur	Aggregator
Erik van der Bergh	Universiteit wageningen	Data Infrastructure Coordinator	Universiteit
Erik Flikkenschild	LUMC	Informatiemanager	Ziekenhuis
Erik Kentie	Surfsara	Communitymanager	Bedrijf
Rick Groenendijk	UVA	PHD student	Universiteit

Screening (telefonisch)	Bedrijf	Rol	Categorie
Hydide Eidhof	NovelT, Universiteit Twente	Jurist	Bedrijf + Universiteit
Leon Gommans	KLM	Science officer	Bedrijf
Kasia Karpinska	ODISSEI	Project manager	Onderzoeks-instituut
Jan-Jurjen Uitterdijk	UMC Groningen	Program Manager Research-IT	Universiteit
David de Koning	VUmc / AMC	Data Management Consultant	Ziekenhuis
Peter Vermeijs	MBO raad	Jurist	-
Menno Rasch	Universiteit Utrecht	Programmamanager Research IT	Universiteit
Irene Luijendijk	227 data	Business development manager	Bedrijf
Arnaud Wijnant	Center data, LISS panel	Senior wetenschappelijk programmeur	Onderzoeks-instituut
Esther Bouma	Universiteit Groningen	Onderzoeker sociale psychologie	Universiteit
Roland van Rijswijk-Deij	Onderzoeker DNS	Universiteit Twente, NLnetLabs	Universiteit
Marielle Berden	Zilveren Kruis	Business developer	Bedrijf

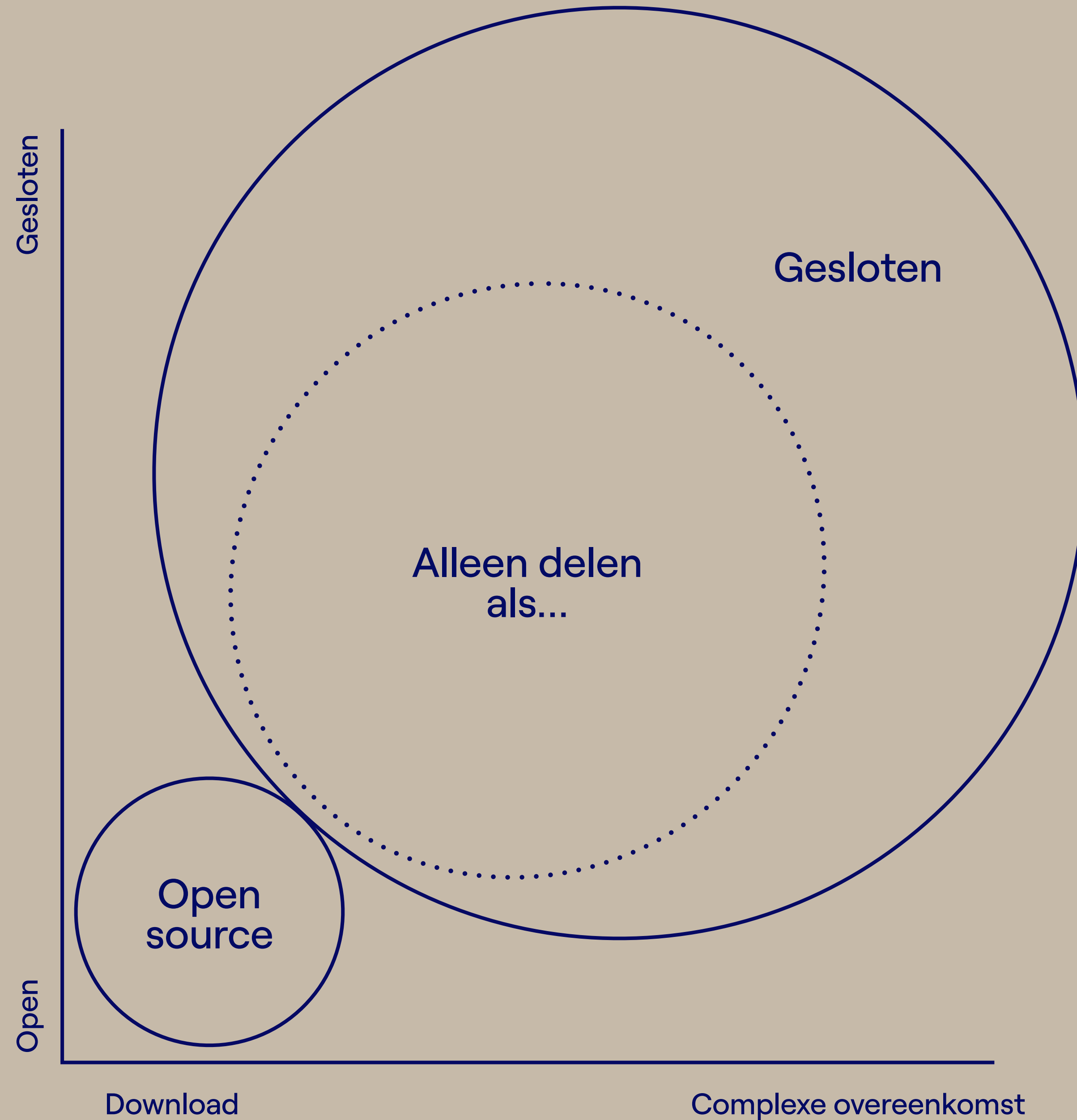
# Modellen

Overzicht van visuals en modellen die zijn gemaakt



# Data

Schaal van open data tot gesloten data

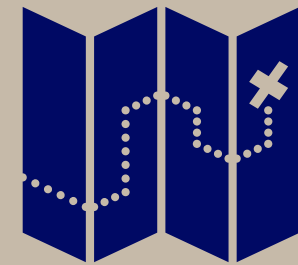


# Datadeel proces

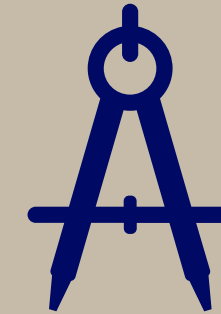
Voor zover bekend



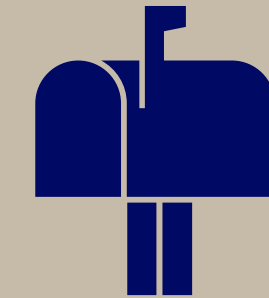
Onderzoeker  
maakt plan



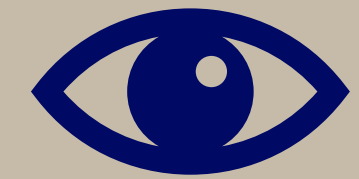
Data match  
vinden



Data match  
vormgeven



Data  
delen



Gebruik van  
data controleren

Requester

Search for data

Request data

Agree on terms

Sign contract

Receive data

Use data



Provider

Scan for relevant datasets

Respond

Propose contract

Send data

Control data use

# Shareable factors

1. **Relevance**
2. **Amount of data** **Data format**
3. **Data explainability** **Data sensitivity**
4. **Continued relevance (incl. updates)**

# Controlable factors

1. **Private** **Public** **Personal**
2. **Purpose** **Value** **Benefit**
3. **Amount of partners**

# Terms of access

## Who

Data owner

Data holder

Public

Private

## Whatfor

Commercial

Common good

## When

Single permission

Ongoing permission

Specific purpose

Partnership

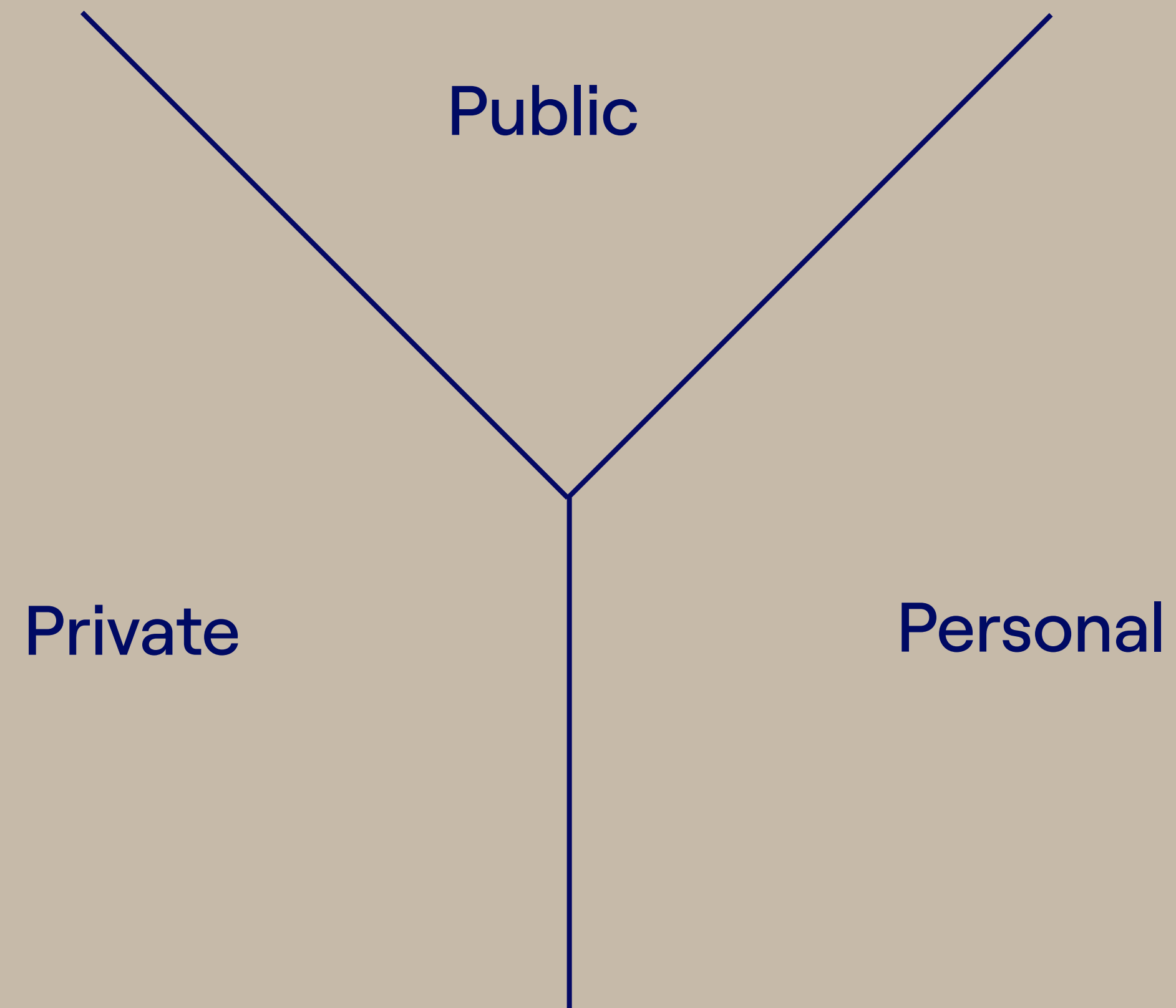
## How

'Er tegen rekenen'

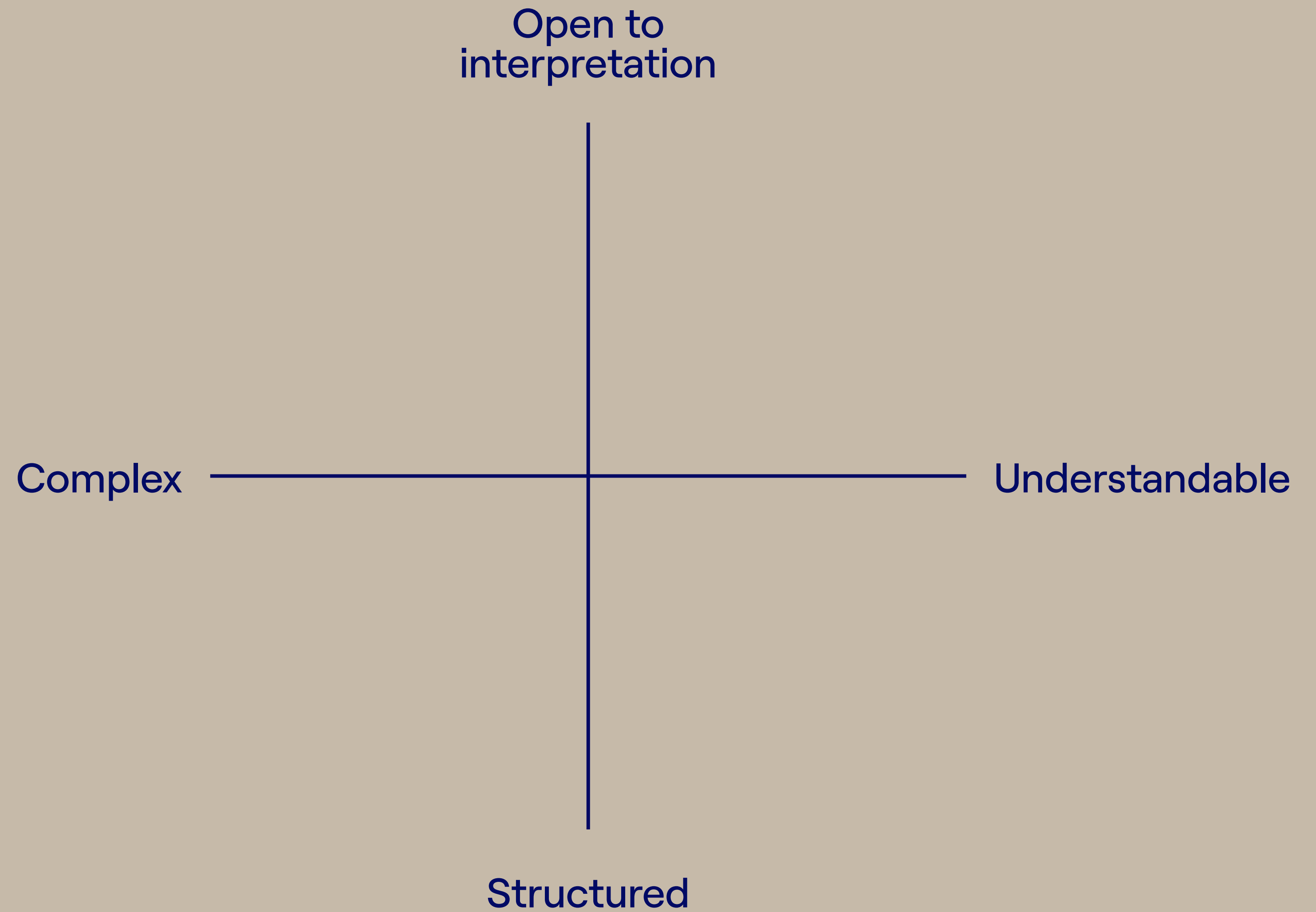
Data inzien

# Data “eigenaren”

Vallen in een van de  
volgende categorieën.



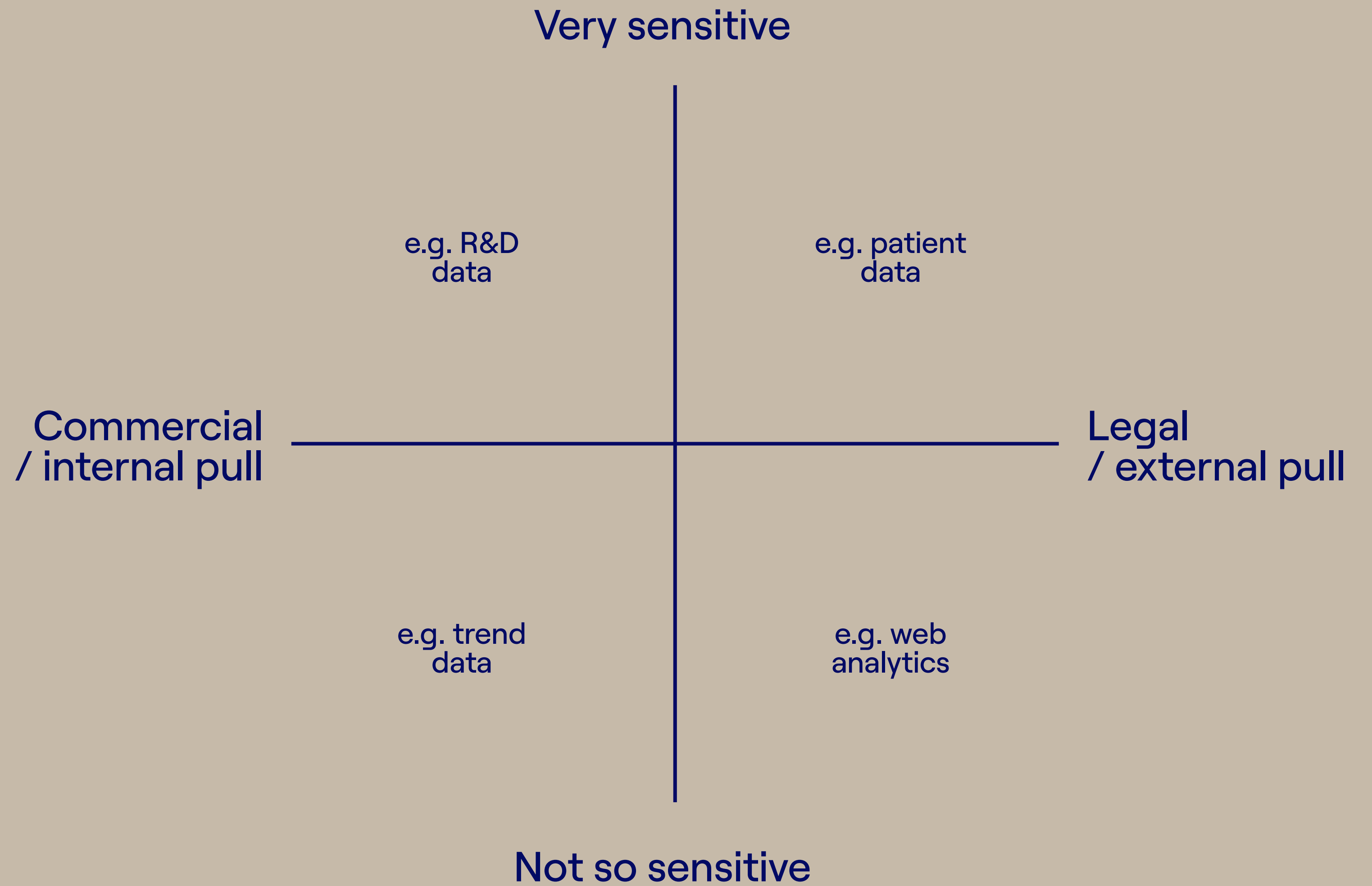
# Data explain- ability



# Data sensitivity

Each factor has subsets.

That can be divided in:

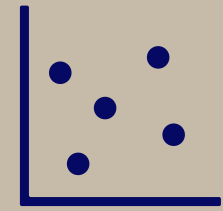


# Owners

Overzicht van data owners



# Data aggregators



Privacy  
gevoelig

Zorgverlening  
(Palga, NZA)

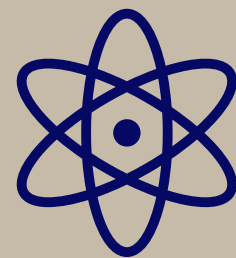
Sociaal-economisch  
(CBS, gemeente)

Data aggregators zijn partijen die data verzamelen vanuit het veld bij verschillende partijen. Ze hebben vaak gestructureerde, grote datasets.

De aard van de data is vaak social en hierdoor is het grootste probleem dat ze kennen privacy gevoeligheid. Dit wordt nu vaak opgelost door de data te pseudoniemiseren of geen persoonsgegevens te delen met andere partijen.

In sommige gevallen is dit echter wel gewild zoals een postcode hiervoor worden dan oplossingen gezocht die de data aggregator tijd kost.

# Ziekenhuizen + medische instellingen



Privacy  
gevoelig

Ziekenhuis (AMC, vuMC,  
St. Antoniusziekenhuis)

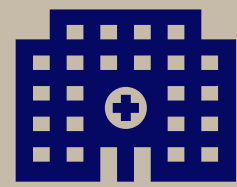
Verzekeringsmaatschappij  
en (Zilveren Kruis)

De data die ziekenhuizen en medische instellingen hebben zijn gevoelig omdat hier speciale persoonsgegevens in staan. Net als bij de aggregators zien we dat pseudoniemisering vaak gebruikt wordt als oplossing.

Verder zijn ze binnen dit veld bezig met “Personal health train” en FAIR data.

Ze willen graag samenwerken maar het verschil tussen de aggregators en deze instellingen is, is dat hun data niet op dezelfde manier wordt gestructureerd. Hierdoor kost het opschonen en samenvoegen van de data meer moeite.

# Onderzoekers + universiteiten



Concurrentie  
gevoelig

Universiteiten  
(Twente, Groningen)

Onderzoekers

Binnen universiteiten wordt er een beleid van open data gevoerd. Echter komt dit niet altijd uit door twee redenen.

Ten eerste omdat de data soms verkregen is onder een contract met een bedrijf of via een verklaring die alleen het benoemde doel als gebruik toestemt.

Ten tweede zijn onderzoekers soms bang voor het door een andere partij eerder publiceren van een soortgelijk onderzoek.



Concurrentie  
gevoelig

Friesland-Campina,  
Elsevier

# Bedrijven

Bedrijven die data bezitten weten dat er veel waarde zit in hun data die concurrentie gevoelig is.

Wat nu gebeurd is dat een onderzoeker onder contract staat bij het bedrijf waardoor er een makkelijke samenwerking is (op eigen terrein) en een NDA.

# Functionaliteiten

Overzicht van technische oplossingen

## Overeenkomsten

Het risico wordt verkleind door het implementeren van procedures, afspraken of wettelijke kaders.

Risk is reduced by implementing procedures or legal frameworks.

- Stipulation what can or cannot be done with the data.
- Signing of a contract or NDA between data provider and data requester
- Implementation of a dispute resolution proces, or an assignment of an arbiter for when data provider and data requester don't agree that the data is processed according to the agreed conditions.

Collaborations are often used when the parties have a symmetrical relation, i.e. when data is shared between two or more equal parties.

## Registratie

Registratie van de aanvragen over de verschillende datasets. Zorgt voor controle over het gebruik.

Data requests may be complex, especially if data needs to be obtained from different sources (e.g. for correlation of different datasets, or when similar data is distributed among different organisations). A registration system keeps track of requests, and authenticates (the affiliation of) requesters.

## Pseudoniemisering

Om privacy te kunnen bewaren moeten we ervoor zorgen dat records niet te herleiden zijn naar een individu.

Pseudonymization ensures that records in a data set are not trivially traceable to individual persons.

Some common methods are:

- Filtering, only showing relevant records
- Pruning, only showing relevant properties
- Aggregation, only showing combined records
- Made more coarse grained, replace detailed information with a coarse grained variable (e.g. "low", "medium" or "high", or only reporting age instead of birthdate)
- Slightly altered (e.g. replace 12.31 with a value between 12.21 and 12.41 so that it can't be matched exactly with other records)
- One-way hashing of records, so that it can be matched with other data sources, but not easily identified
- Replace identifiers (like name or BSN number) with one-time id (which allows matching over time with the same record, but no matching with other data sources)



## Secure containers

De data wordt verwerkt in een container waardoor deze nooit bij de requester komt.

Data is not released to the data requester, but processing of the data is done at the premise of the data provider or a trusted third party. The data requester provides an algorithm and only gets to see the result, but not the actual data. Computation is done in a secure environment which prevents leaking of data to the outside world.

## Secure computing

Gebruik maken van cryptografische technieken kan het mogelijk maken gegevens te berekenen zonder die gegevens te onthullen.

Use of complex cryptographic techniques may allow computing on data without revealing that data. It is an alternative without the need for a trusted third party.

- Secure multiparty computation, allows a group of data owners that do not trust each other to jointly compute a result that depends on all of their private inputs.
- Homomorphic encryption, performs computation on encrypted data, so that only the data owner can decrypt the returned results to obtain the output. No scalable, fully homomorphic solution exists today.

# Got Q's? We've got A's

**Freek Dijkstra**

Design Expert, Surfsara

[freek.dijkstra@surfsara.nl](mailto:freek.dijkstra@surfsara.nl)

**Joris van Rossum**

Business developer

[jpvanrossum@gmail.com](mailto:jpvanrossum@gmail.com)

**Nadia Piet**

Researcher, Bit

[nadia@wearebit.com](mailto:nadia@wearebit.com)

**Ocean Conijn**

Researcher, Bit

[ocean@wearebit.com](mailto:ocean@wearebit.com)

