



DPO vs Data
Lost in Translation?

Privacy Café
Antwerpen, 4 Okt 2023

`/imagine` A movie poster with a man sitting on the bed in a hotel room, wearing only a robe and slippers, facing the camera, with incandescent lighting in the room and a dark city skyline in the background. Use the style of the poster of the movie "Lost in Translation".

Yours truly



bol.com



 **Ahold
Delhaize**

“Bij bol.com en Ahold Delhaize hebben we geleerd hoe we het managen van (privacy) risico’s in vertrouwen kunnen vertalen.

Privacy en security waren bijvoorbeeld de kern van ons aanbevelingen platform, waarvoor we gevoelige persoonsdata konden gebruiken zonder dat het onze snelheid of gebruikerservaring negatief beïnvloedde.”

[Pim](#), was Head of Data & ML/AI at bol.com

[Bart](#), was Head of R&D at Ahold Delhaize (Global)

Er zit een groot verschil tussen de werelden van data & data protection

[Abonneer](#)

(bij > 25 nieuwe inschrijvingen gaan we weer wekelijks schrijven!)





Translating privacy + technology

[Our Expertise](#)

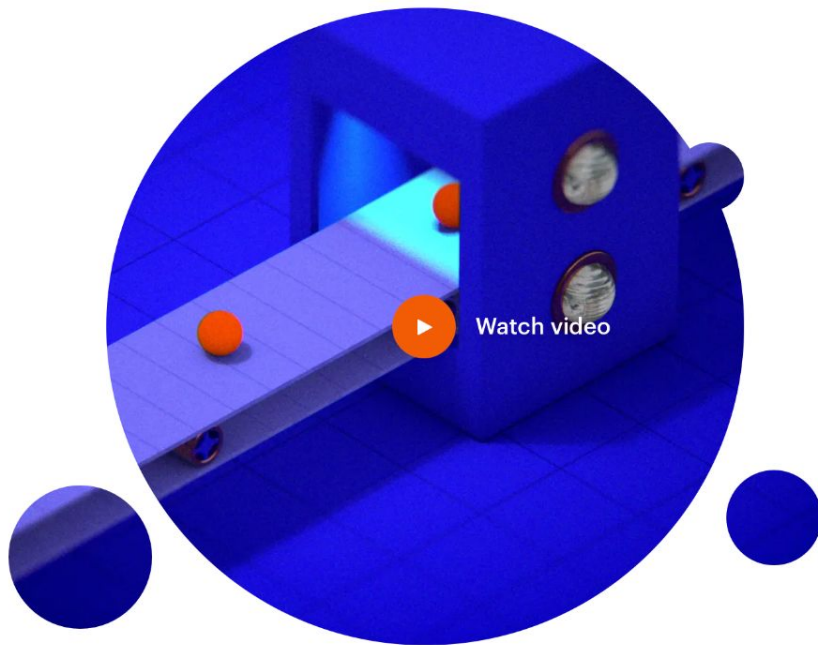


Privial

[Product](#)[Solutions](#)[Developers](#)[About](#)[Sign in](#)[Sign Up](#)[Request a Demo](#)

Accelerate data value.

Data is value. But can you use it? With STRM, you ensure sensitive data can be used securely and safely. So you can finally unlock the value in your (sensitive) data.

[Book a demo](#)[Learn more ▶](#)

Do you prefer compliance or usability? How about

Safe not slow

Het Plan voor vanavond?

- **Energie bijtanken** - pubquiz! 🍺🎓
- **Data vs DPO** - hoe kun je data & data protection samenbrengen?

Pubquiz: art.1

1

Ga allemaal staan, svp.

2

Fout = zitten, svp.



Wanneer werd GDPR ingevoerd?

***mei**

mei

2016

2018



Welk artikel GDPR sluit anonieme data uit?

***26**

29

 **Voor wie was de laatste boete van de gegevensautoriteit (BE), en hoe hoog was deze?**

**Orde van
Huisartsen,
€50k**

***Orde van
Apothekers,
€30k**

**Nu even wat
lastiger 🤦**

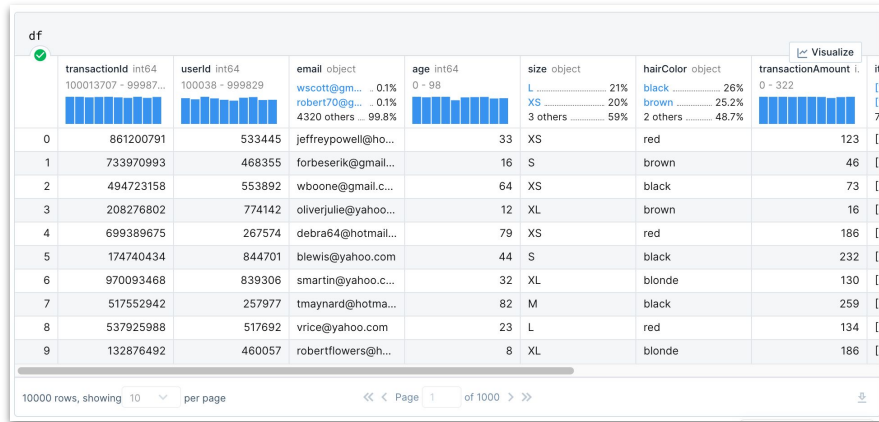


Wie is de betere wielrenner?





Wie van U heeft weleens de data zélf gezien?

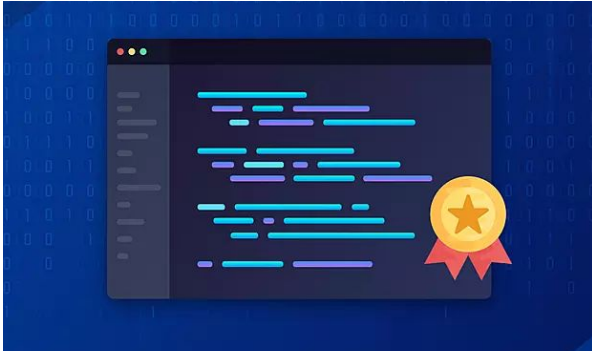


Echte data, van dichtbij.

Ik houd het bij het wetboek
en vragenlijsten



Wie van u kan code schrijven?



Ik schrijf ook code



Alleen inkt



Werelden te overbruggen

/imagine a bridge spanning different worlds in a sixties
artwork style.

Privacy ingewikkeld? Meet Data.



Matt Turck 
@mattturck · [Follow](#)



It's easy to succeed in data/AI. All you need is to put data into a DWH for OLAP using ETL, or ELT with DBT. Spot any issues on your DAG, run SQL for some BI, reverse ETL into SaaS or use a CDP, then leverage for ML/AI with CNNs, RNNs, RLs, GANs, and maybe LLMs or LDMs for GenAI.

11:30 PM · Jan 4, 2023



310



Reply

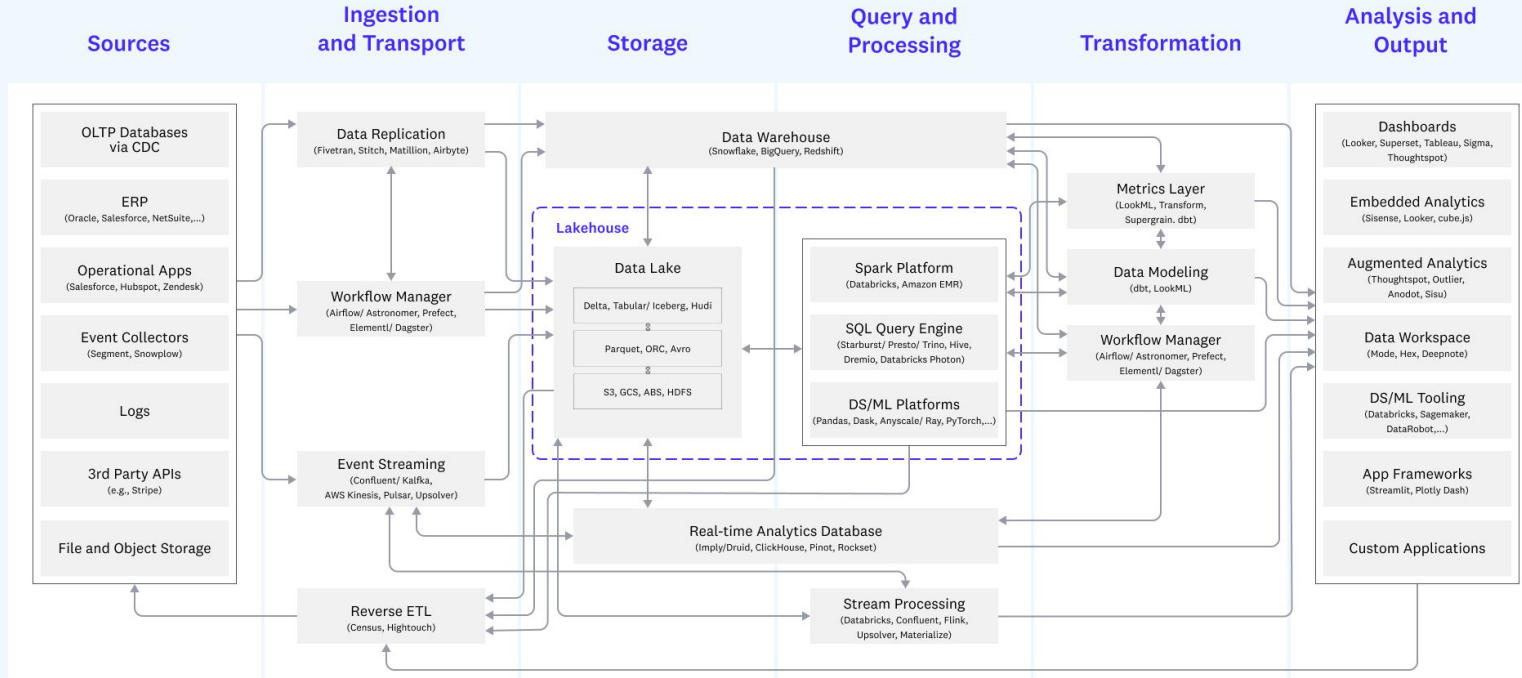


Copy link to post

[Read 29 replies](#)

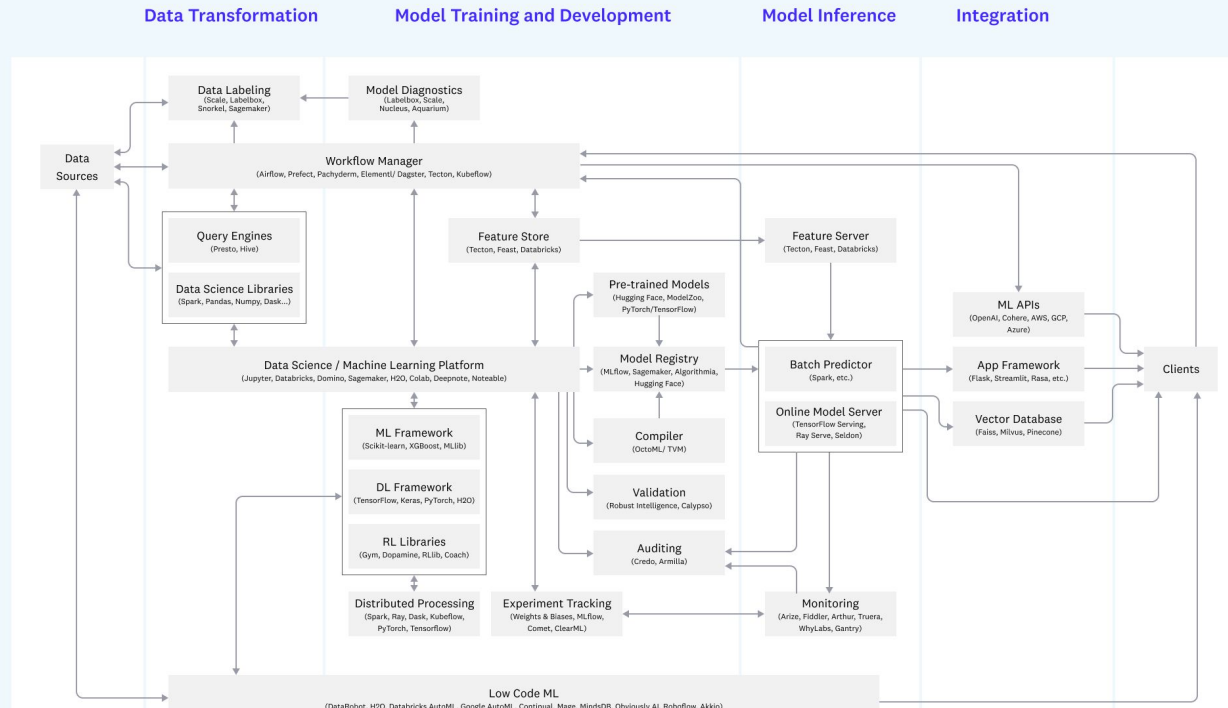
Privacy ingewikkeld? Meet Data.

Unified Data Infrastructure (2.0)



Privacy ingewikkeld? Meet machine learning.

Machine Learning Infrastructure (2.0)



"Kiest u maar" (een tooltje)

The landscape is divided into several main sections:

- INFRASTRUCTURE:** Includes STORAGE (AWS, Azure, Google Cloud), MPP DBs (Snowflake, Databricks), DATA LAKES / WAREHOUSES (Delta Lake, Amazon Redshift), DATA WAREHOUSES (Snowflake, Databricks), STREAMING / IN-MEMORY (Kafka, Flink), and DATA SCIENCE PLATFORMS (Databricks, AWS SageMaker).
- ANALYTICS:** Includes BI PLATFORMS (Tableau, Power BI), VISUALIZATION (Tableau, Power BI), DATA SCIENCE NOTEBOOKS (Databricks, Jupyter), ENTERPRISE ML PLATFORMS (Databricks, AWS SageMaker), and CUSTOMER DATA PLATFORMS (Salesforce, Oracle).
- MACHINE LEARNING & ARTIFICIAL INTELLIGENCE:** Includes DATA SCIENCE PLATFORMS (Databricks, AWS SageMaker), ENTERPRISE ML PLATFORMS (Databricks, AWS SageMaker), and various AI/ML tools like H2O, HuggingFace, and OpenAI.
- APPLICATIONS - ENTERPRISE:** Includes SALES (Salesforce), MARKETING (HubSpot), CUSTOMER EXPERIENCE (Salesforce), HUMAN CAPITAL (Workday), AUTOMATION & OPERATIONS (UiPath), and DECISION & OPTIMIZATION (Palantir).
- APPLICATIONS - HORIZONTAL:** Includes CODE & DOCUMENTATION (GitHub), TEXT (OpenAI), AUDIO & VOICE (Amazon Alexa), IMAGE (OpenAI), VIDEO EDITING (Runway), and SEARCH (Elasticsearch).
- APPLICATIONS - INDUSTRY:** Includes FINANCE & INSURANCE (Kenshoo), HEALTHCARE (Tempus), LIFE SCIENCES (Moderna), TRANSPORTATION (Uber), AGRICULTURE (Bayer), INDUSTRIAL & LOGISTICS (DHL), and GOVT & INTELLIGENCE (Palantir).
- OPEN SOURCE INFRASTRUCTURE:** Includes FRAMEWORKS (PyTorch, TensorFlow), FORMAT (JSON, XML), QUERY / DATA FLOW (Apache Airflow), DATA ACCESS (Apache Spark), DATABASES (PostgreSQL, MySQL), OLAP (Snowflake), ORCHESTRATION (Apache Airflow), INFRA-STRUCTURE (Kubernetes), DATA OPS (DataHub), STREAMING & MESSAGING (Kafka), STAT TOOLS & LANGUAGES (Python, R), MLOPS & AI INFRA (MLflow), AI FRAMEWORKS & LIBRARIES (PyTorch, TensorFlow), AI MODELS & ARCHITECTURES (OpenAI, Google AI), SEARCH (Elasticsearch), LOGGING & MONITORING (Datadog), VISUALIZATION (Tableau), and COLLABORATION (Slack).
- DATA SOURCES & APIS:** Includes DATA MARKETPLACES & DISCOVERY (Bloomberg), FINANCIAL & MARKET DATA (Bloomberg), AIR / SPACE / SEA (Airbus), PEOPLE / ENTITIES (LinkedIn), and LOCATION INTELLIGENCE (Esri).
- DATA & AI CONSULTING:** Includes Deloitte, IBM, Accenture, Capgemini, and others.



Privacy + Data in de realiteit

The
Intercept_



FACEBOOK ENGINEERS: WE HAVE NO IDEA WHERE WE KEEP ALL YOUR PERSONAL DATA

In a discovery hearing, two veteran Facebook engineers told the court that the company doesn't keep track of all your personal data.

Sam Biddle

September 7 2022, 7:00 a.m.

DONATE →



protocol

SUBSCRIBE



NEWSLETTERS

ENTERPRISE

FINTECH

ENTERTAINMENT

WORKPLACE

POLICY

CHINA

CLIMATE

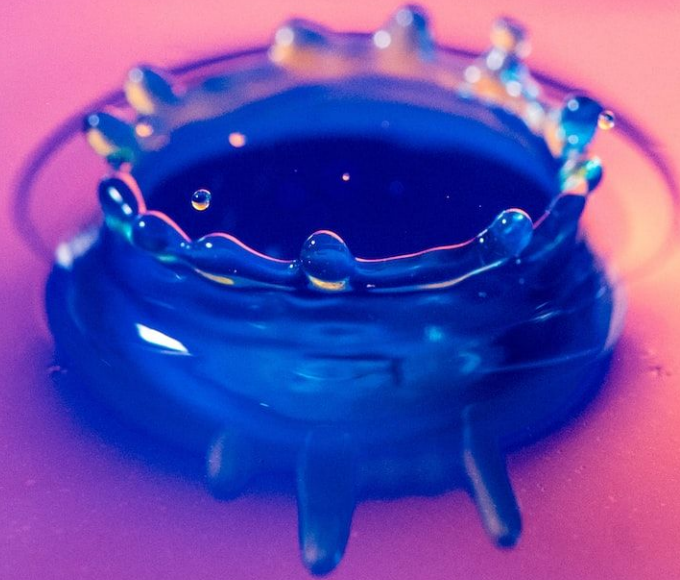
PRIVACY BY DESIGN

Privacy by Design laws will kill your data pipelines

The legislation could make old data pipelines more trouble than they're worth.



“Personal data is like a droplet in a data lake”



In de realiteit voldoet niemand volledig aan de wetgeving



**Privacy *kan*
eigenlijk helemaal
niet gewaarborgd
worden in de
digitale wereld.**

[uitleg](#)



Maar hoe dan wel?

(*Midjourney koos een man...)

`/imagine` a legal professional going crazy over an unsolvable puzzle in a dark room with demons

Complex (🕒, 💰, 🔥)

**Data Consumers
(Data Science)**

Bouw nieuwe
producten (€),
algoritmes, studies

**Platform
(IT engineering)**

Verzorg betrouwbare,
veilige data om te
gebruiken

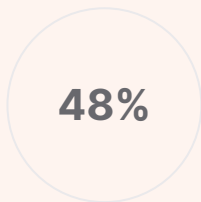
**Governance
(Privacy, Security
Legal)**

Draag zorg voor
compliance en het
verminderen van
risico

Datagebruik = kosten



coördinatiekosten rondom datagebruik kunnen oplopen tot **35K per verzoek** en *maanden* vertraging



van de professionals heeft niet genoeg tijd voor goede risico-assessments en een groeiend review-backlog



Veel data blijven **ongebruikt liggen**, terwijl ze zeer waardevol zijn (Syntho!)

  **Hoe lang duurt goedkeuring voor data toegang/gebruik in uw organisatie?**

< 2

> 2

wk

wk

Engineers: “ha een probleem”



**Data Consumers
(Data Science)**

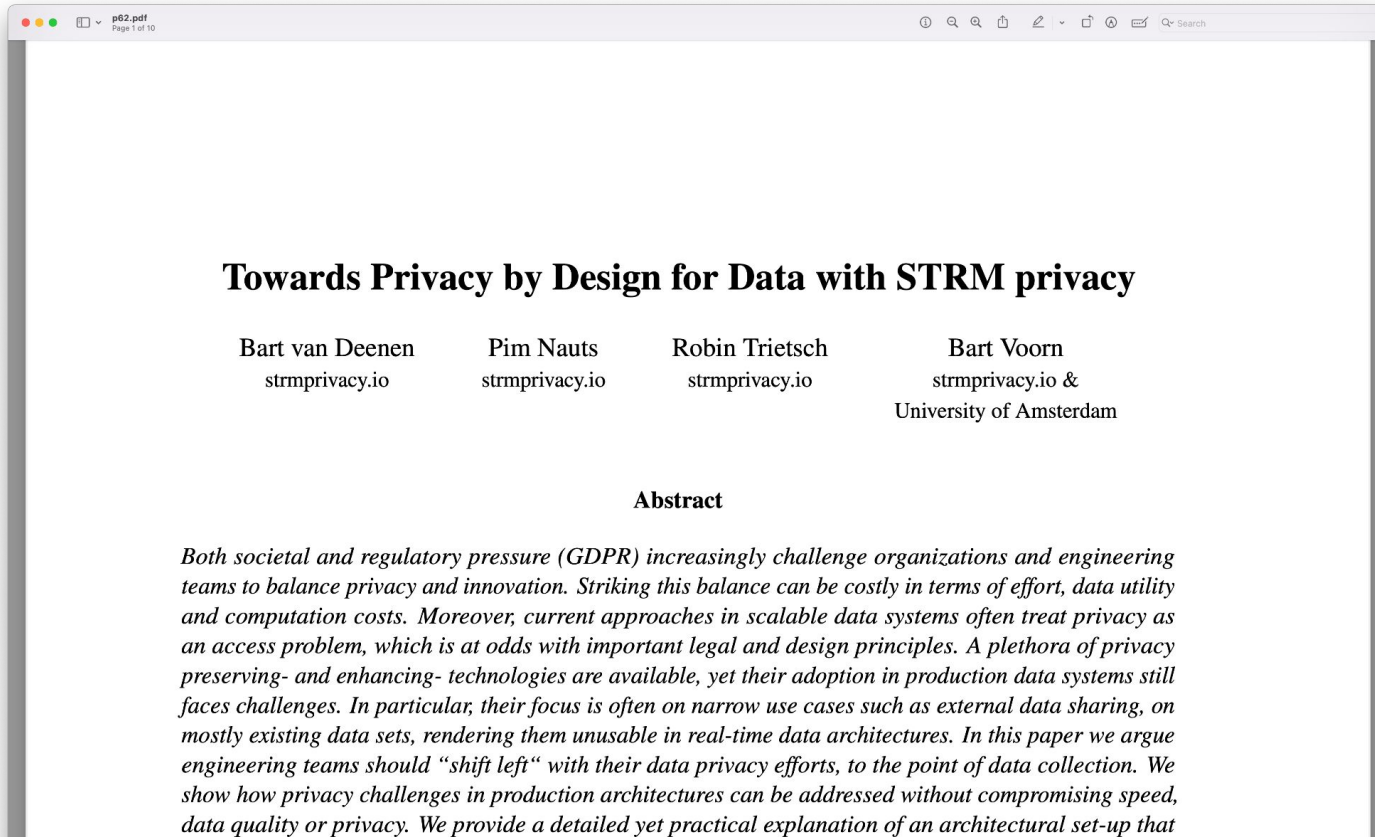
Bouw nieuwe
producten,
algoritmes, studies



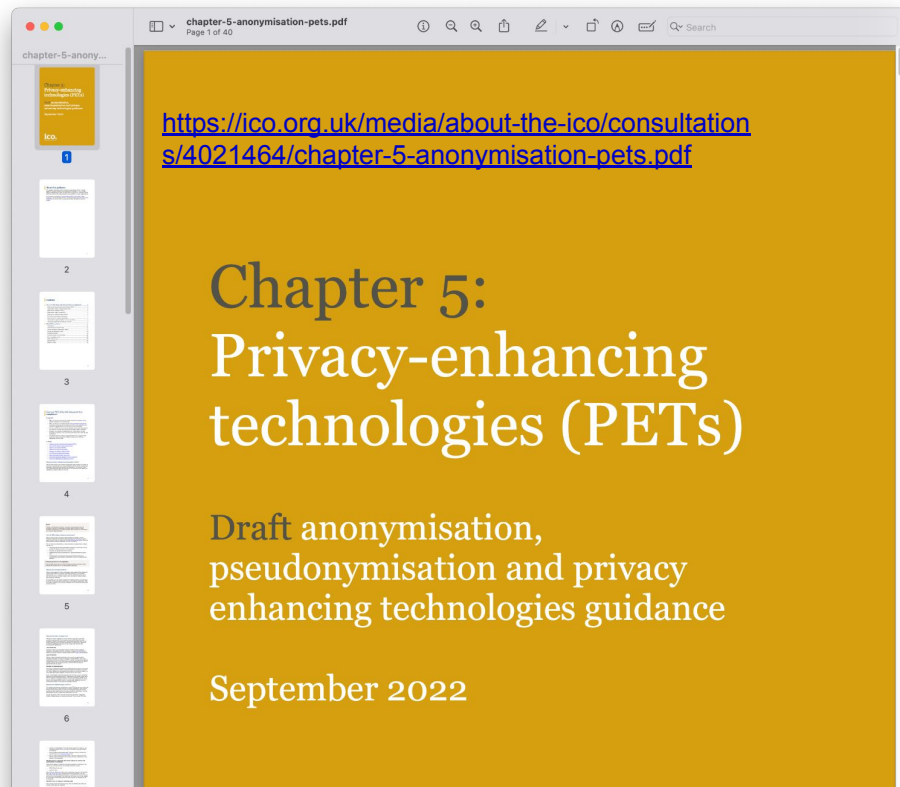
**Platform
(IT engineering)**

Verzorg betrouwbare,
veilige data om te
gebruiken

Enter *Privacy by design*



Het is Dierendag: PETs!



Simple Privacy-by-design

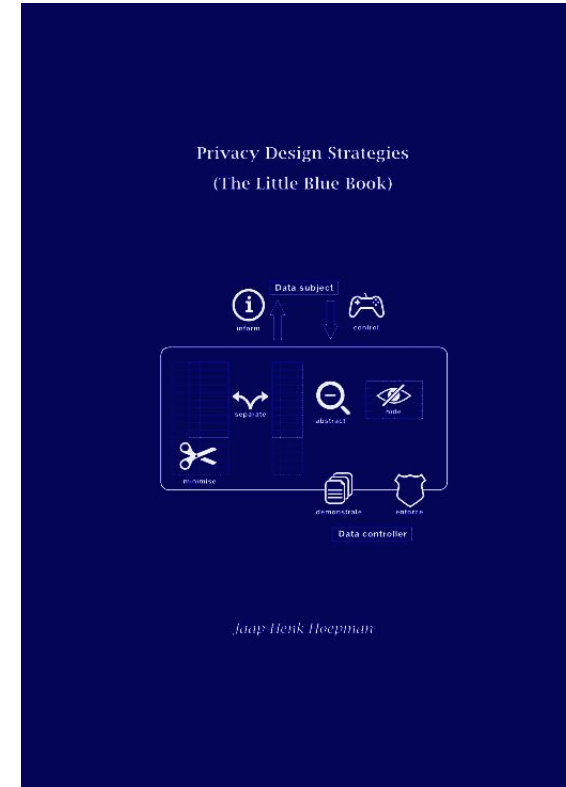
<https://www.cs.ru.nl/~jhh/blue-book.html>

Data-georiënteerd

1. Minimaliseer - gebruik niet wat niet nodig is
2. Scheid - koppel persoonsdata los van niet-persoonsdata
3. Abstraheer - laat niet onbewerkt, wat ook als categorie kan
4. Bescherm, maak onherleidbaar - PET's!

Proces-georiënteerd

- ~~5. Informeer~~
- ~~6. Geef controle~~
- ~~7. Dwing af~~
- ~~8. Toon aan~~



Privacy Enhancing Technologies →



Encryption, tokenization, permutation



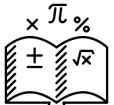
Pseudo/anon



Differential Privacy



Synthetiseren

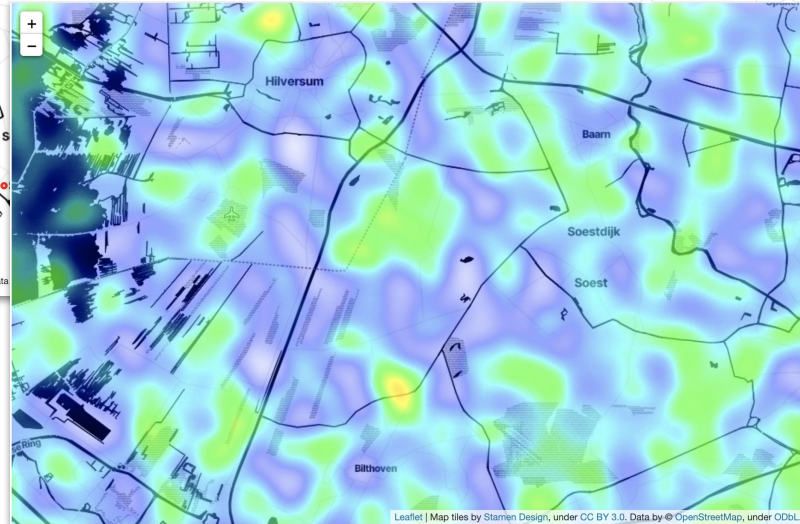


MPC/FHE

Uitdaging: balanceren van *veiligheid* met *bruikbaarheid*.



Abstraheer: waar zijn **onvoldoende** andere punten om de exacte locatie te verstoppen → blur.



Combineer tech en proces: fraude detectie



Rekening	Tegenrekening	Bedrag	Tijdstip
f6c61d552c58224456 c60ab2d8a61660ee64 6d255499d678ec608a 052bb3c163	93f031a92138891e07 3eea53929e61c6f364 e42bff5735cc893a6e 0e83894e62	€ 10,00	11:00 GMT+1
153bee3d966ed04766 10fc14dc8fd8555b80 ca2687b8e1e5fef2a6 6a85a6f0ab	63ab505cf0d6eb5c71 ce6405d519c3add1a4 aad14c15c4b6ad8bc2 65155a1353	€ 11,14	15:19 GMT+1
1e8e1bce1f50e67961 94cd4d64a7be25b6df 40c8d0702271b2b14a d614fb7ba4	205d585c07c6f23b06 1c7b7c729837b18644 1829e60782863066a2 236432f68a	€ 32000,00	04:19 GMT+1
c029f41706feca35ab dae77708233397421b 954396c28c66697d8e 389546d786	2b06474df528f59422 efb73ed7d782b992e1 bfd5361a649278cd4f a2404e16d6	€ 4822,21	19:59 GMT+1

Minimaliseer + classificeer transacties als anonieme gegevens. **Ontsluitel** pas als er een redelijke verdenking is (*proportionaliteit*).



→ **NL34 INGB 0467 5209 11**

Bepaal zelf de grens: K-anonymity



 code-alert

Draai zelf diagnostics:

<https://github.com/strmprivacy/strm-privacy-diagnostics>

Voorbeeld code:

<https://deepnote.com/workspace/STRM-demos-2614c69d-1aae-4c75-a0b8-ee631006da30/project/Data-team-in-a-day-with-STRM-eb9f78ee-b796-48e5-b1ff-b77815a3952a/notebook/Anonymisation%20pipelines%20with%20STRM%20Privacy-681be7708cf844589c24db36e0a5d2d9#a11f3f0fd8c4d9ebf6b7a15e7ed5b5d>

```
d = PrivacyDiagnostics("/datasets/faker-demo-bucket/output/decrypted-0-kmma.csv")

# calculate the statistics
d.calculate_stats(qi=['age', 'transactionAmount', 'size', 'hairColor'])

print(f'Your data has a k-anonymity of: {min(d.k)}')

with open('/datasets/faker-demo-bucket/output/decrypted-0-kmma_metrics.txt') as f:
    for line in f.readlines():
        print(line, end='')
```

Your data has a k-anonymity of: 3

Applied Mona_0.5 to 6657 records in 211 ms

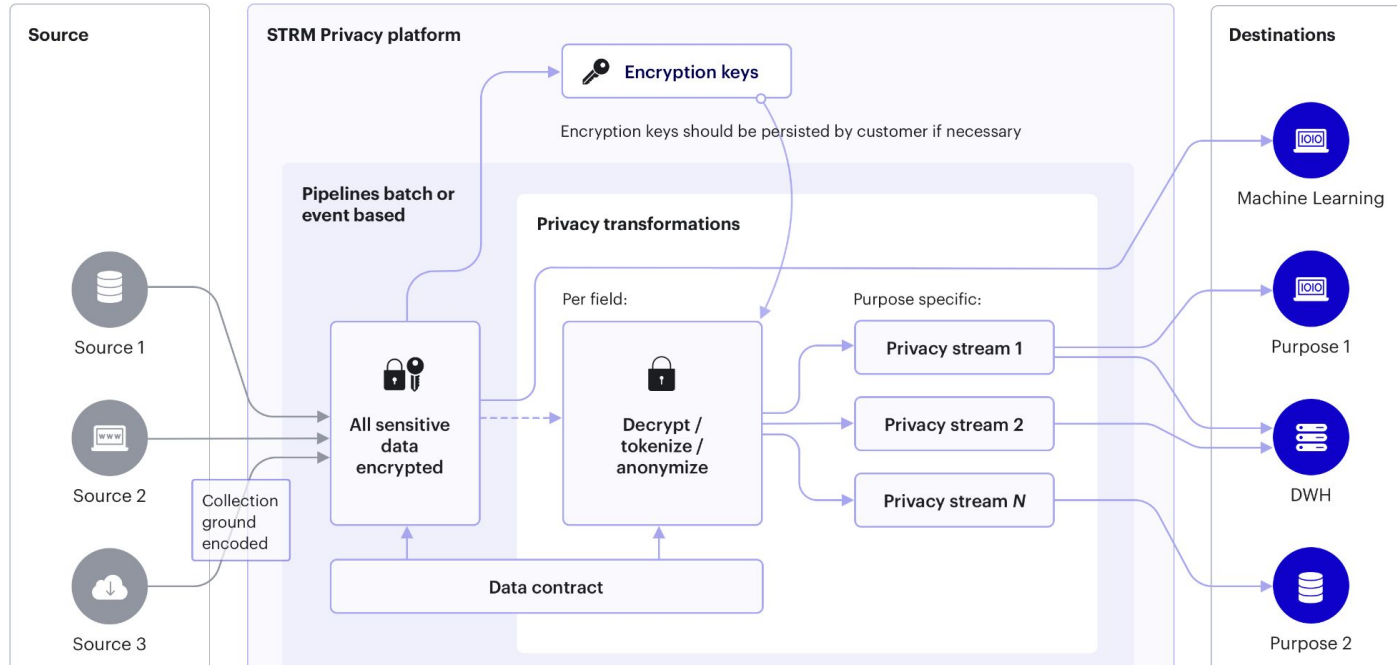
Iterations: 1, clusters: 1921

Diversity: 6.278350633598083E7, totalCost: 625562.4999999999, informationLoss: 0.99 %

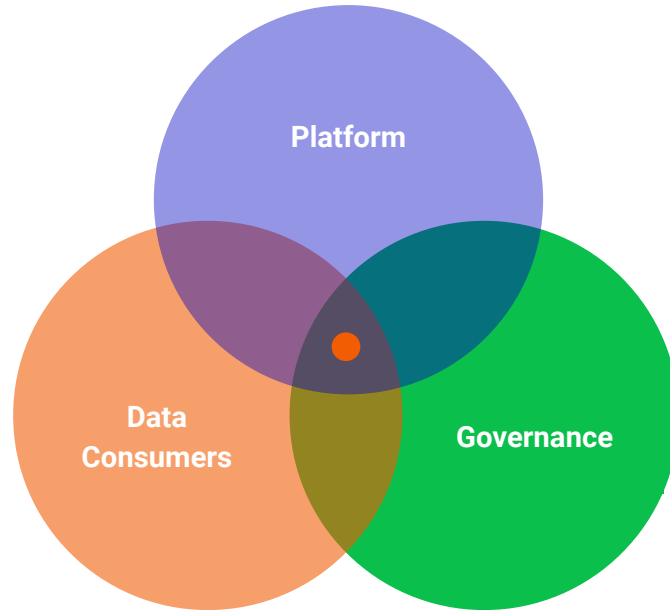
Cluster size	Count
3	1224
4	502
5	193
6	2

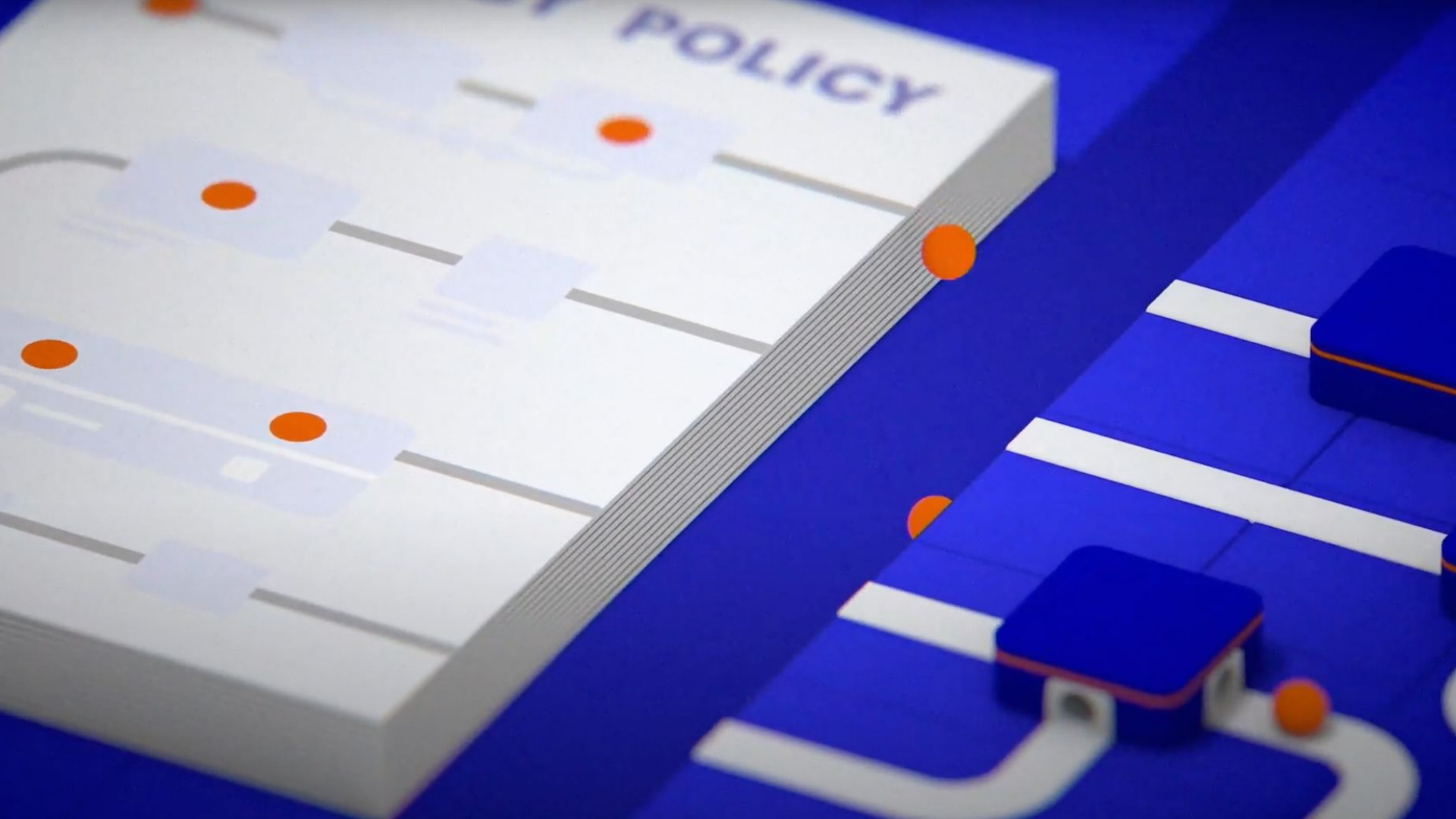
By design: datapipelines met een purpose

Deel in op **gebruiksdoel** → bepaal **contract** → **pas toe** → data kan altijd gebruikt worden!



Maar: veilig datagebruik is ul-teamwork





Verzekeraar, €20 miljard omzet

“We handle over 450 data requests per year, with a full team of data stewards, lawyers, and data people.



Purchasing a cotton picker?



“ We had our biggest HR project delayed for a full year because of the privacy team. During meetings, our VP screamed, yelled and left the room. ”



PREMIUM



Jay Averitt

Privacy @ Microsoft | Privacy Engineer | Privacy Evangelist | Writer/Speaker

[View full profile](#)



Jay Averitt · 1st

Privacy @ Microsoft | Privacy Engineer | Privacy Evangelist | Writer/Speaker
13h ·

What makes technical privacy reviews so uniquely challenging?

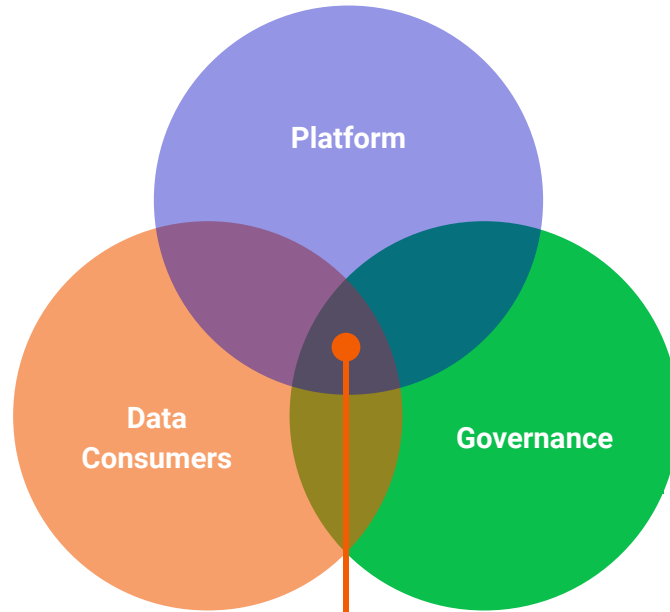
Every Wednesday I complete 6-8 technical privacy reviews in a single day, I'm mentally drained, but yet energized when my Wednesday ends. What about the process makes my brain work so hard?

1) To be really good, you need to be thinking about many things outside of privacy. First and foremost, I see it as our jobs to develop trust with our users. Ensuring that we have proper privacy controls in place is only one way to help develop that trust.

Since Privacy has shifted so far left in the SDLC in orgs with good Privacy by Design principles in place, there is a real opportunity for Privacy Engineers to develop trust in other ways too.

For example, are we doing something that isn't a real privacy concern, but will otherwise upset our users? Raise that with the team.

HOE OPTIMALISEER JE DATAGEBRUIK?



**ST
TS
MR**

DATA MANAGEMENT;

What	Know Your Data Ensure you have an overview of your most valuable data assets cross data systems	Control Your Data Enable fast data usage whilst extra securing your crown jewels with a variety of security and privacy controls, automatically put in place.		Watch Your Data Enforce compliance by monitoring usage and risks, update data automatically
How	Locating Classification Connect catalogues Data Quality	Data Access Controls Data Contracts Data Approvals	Mask Encrypt Tokenize Synthesize	Logging Report and Alert Update

Beschrijf data projecten of studies

ST
RM



Department ▾

Upgrade Help DA

Your project

Edit

This project plan serves as the central, standardized starting point for your team.

Activity log

Apr 1, 2023 Darren Adams checked **Business needs, team and roles** task as done

Apr 2, 2023 Brijamohan Mallick created **Risk Mitigation** in Assessments

Apr 3, 2023 Darren Adams added a file **Meeting 12 Apr 2023 Team.docx**

[See all updates](#)

Tasks

- Business needs, team and roles
- Functional and technical requirements
- Functional and technical design
- Process risk assessments
- Register sensitive data (data contracts)
- Enforce data contracts

+ Add task...

Risk assessments

Project risk:

Sensitive ▾

What's the status?

● Under review ▾

Set due date

Brijamohan Mallick joined
2 days ago

Project status is changed to under review
Marie Labrone 4 days ago

Kevin McCallister joined
1 month ago

Marie Labrone joined
27 Jan, 2023

You joined
27 Jan, 2023

Project created
Darren Adams 27 Jan, 2023

Stel een data contract op (met consent)



Encode data contracts

Your benefits

Confidence

Understand where to begin your data privacy journey, eliminate risky blindspots by creating clear structure.

The screenshot shows a web interface for configuring a data contract. The breadcrumb trail is 'Projects > Project > Layer 3'. The interface has a navigation bar with steps: 01 Fields (active), 02 Privacy, 03 Annotations, 04 Description, and 05 Review. Below the navigation, there are buttons for 'View code' and 'Upload file'. The main area contains a table of fields:

FIELD	TYPE	REQUIRED	REPEATED	DESCRIPTION	
device_id	string	<input checked="" type="checkbox"/>		the id of the scooter	[edit] [delete]
ride_id	string	<input checked="" type="checkbox"/>			[edit] [delete]
GPS	string			the GPS coordinates	[edit] [delete]
charge	int			the charge left in the battery	[edit] [delete]
current_speed	nodes				[edit] [delete]
distance	float		<input checked="" type="checkbox"/>		[down] [edit] [delete]
time	float		<input checked="" type="checkbox"/>		[up] [edit] [delete]
brake_pressure	float				[edit] [delete]
producer_id	string	<input checked="" type="checkbox"/>			[edit] [delete]

Below the table, there is a configuration panel for the selected 'producer_id' field:

- Field: producer_id
- Type: string
- Required: Yes
- Repeated: Yes
- Description: Max. 50 chars
- Save button

At the bottom, there is a 'New field' input area and a 'Next' button.

01 Fields →

02 Privacy →

03 Annotate →

04 Describe →

05 Review

Review your data contract

25% done



Fields ^

FIELDS	DATATYPE	REQUIRED	REPEATED	DESCRIPTION
device_id	string	✓		the id of the scooter
rider_id	string	✓		
GPS	string			the GPS coordinates
charge	int			the charge left in the battery
current_speed	nodes			
— distance	float		✓	
— Time	float		✓	
brake_pressure	float			
producer_id	string	✓		

Edit

Approve

Privacy ▾

Annotations ▾

Description ▾

Info panel >|

 Information Activity Comments Kevin

@Marie please review this data contract and let me know if it need changes.

5 minutes ago

 Marie

@Kevin could you add a description to the keyfield? Thank you!

Just now

Write a comment

Laat een data contract goedkeuren



Apply contracts to data

Your benefits

Scalability

Create the contract once and we scale effortlessly with you. Giving your teams valuable time back.

Projects > Project > Layer 3

Upgrade Help

01 Fields → 02 Privacy → 03 Annotate → 04 Describe → 05 Review

Review your data contract 25% done

FIELDS	DATATYPE	REQUIRED	REPEATED	DESCRIPTION
device_id	string	✓		the id of the scooter
rider_id	string	✓		
GPS	string			the GPS coordinates
charge	int			the charge left in the battery
current_speed	nodes			
— distance	float		✓	
— Time	float		✓	
brake_pressure	float			
producer_id	string	✓		

Info panel

- Information
- Activity
- Comments

Kevin
@Marie please review this data contract and let me know if it need changes.
5 minutes ago

Marie
@Kevin could you add a description to the keyfield? Thank you!
Just now

Write a comment

Transformeer en “de-risk” gegevens

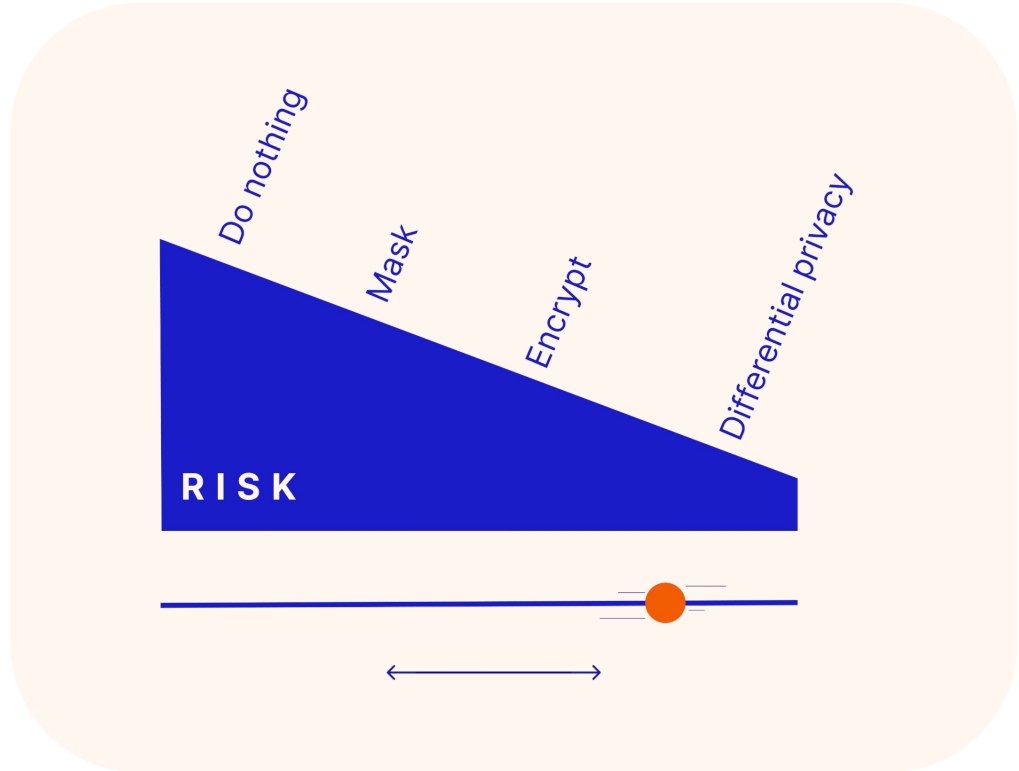


Purpose based transformations

Your benefits

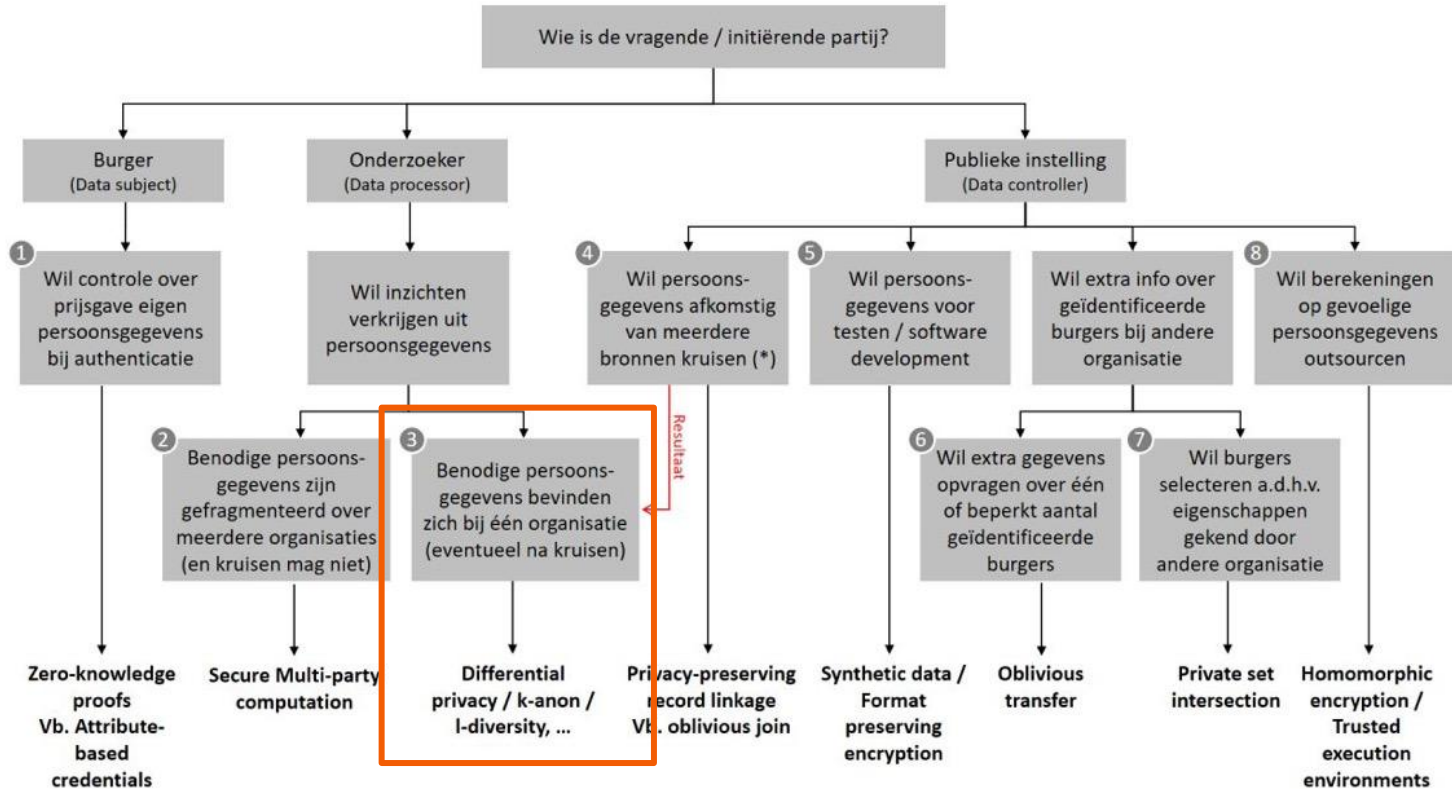
Compliance

Choose the degree of privacy you require based on your business needs and risk-appetite.

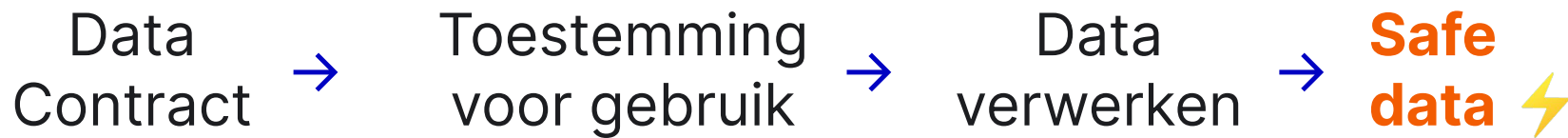


Voorbeeld

PRIVACY TECHNOLOGIE IN ACTIE



STRM.in



1. Data contract (bepaal welke data)

Please define purpose and classify the privacy of this data contract

FIELD	PURPOSE	SENSITIVE	PERSONAL
device_id (string)			the id of the scooter
rider_id (string)	ID	✓	
GPS (string)	Location	✓	the GPS coordinates
charge (int)			
current_speed (nodes)			
distance (float)			
time (float)			
brake_pressure (int)			
producer_id (string)		✓	

Back

```
1 name: exampleSimpleSchema
2 nodes:
3   - type: STRING
4     name: directIdentifier
5     repeated: false
6     required: true
7     doc: 'a direct identifier such as a name or postal code'
8     nodes: []
9   - type: STRING
10    name: indirectIdentifier
11    repeated: false
12    required: true
13    doc: 'a direct identifier such as a customer id'
```

2. Vraag toestemming voor gebruik

Review your data contract 25% done

FIELDS	DATATYPE	REQUIRED	REPEATED	DESCRIPTION
device_id	string	✓		the id of the scooter
rider_id	string	✓		
GPS	string			the GPS coordinates
charge	int			the charge left in the battery
current_speed	nodes			
— distance	float		✓	
— Time	float		✓	
brake_pressure	float			
producer_id	string	✓		

Review your data contract 100% done

- Fields ✓
- Privacy ✓
- Annotations ✓
- Description ✓

Info panel

- Information
- Activity
- Comments

Comments

- Kevin McCallister**
@Marie please review this data contract and let me know if it need changes.
5 minutes ago
- Marie Labrone**
@Kevin could you add a description to the keyfield? Thank you!
Just now
- Kevin McCallister**
@Marie please review this data contract and let me know if it need changes.
5 minutes ago
- Marie Labrone**
@Kevin could you add a description to the keyfield? Thank you!
Just now

Approve

Write a comment

3. Transformeer de data

The image displays three overlapping screenshots of the STRM (Stream Transformation and Routing Manager) dashboard, illustrating the workflow for data transformation:

- Top Screenshot:** Shows the 'My Project' overview with a 'Create pipeline' modal open. The modal prompts the user to 'Create a new pipeline and start sending events.' It includes input fields for 'Name' and 'Description', and a 'Data policy' dropdown set to 'Select'.
- Middle Screenshot:** Shows the 'My Project' overview with a 'Create Data connector' modal open. The modal prompts the user to 'Create a new data connector to load data from and/or export data to.' It includes input fields for 'Name' and 'Type' (set to 'Select').
- Bottom Screenshot:** Shows the 'My Project' overview with a 'Name of exporter' modal open. This modal is for configuring a Kafka exporter and includes fields for:
 - Pipeline:** Pipeline x
 - Kafka cluster:** shared-export
 - Bootstrap servers:** export-kafka-bootstrap.kafka-export-9099
 - Auth mechanism:** SASL_PLAIN
 - Token URI:** https://auth.strm.services/token

The STRM logo is visible in the bottom left corner of the overall image.

Vragen

Aan U!

DPO:
meer of minder
technisch?

Wat wint: Privacy of Security?

**Wie heeft er al
PETs ingezet?**



Translating privacy
+ technology

[Our Expertise](#)



<https://privial.io>

Privial





<https://getstrm.com>

[mail: go@getstrm.com](mailto:go@getstrm.com)

Privacy Café

Antwerpen, 4 Okt 2023

/imagine A movie poster with a man sitting on the bed in a hotel room, wearing only a robe and slippers, facing the camera, with incandescent lighting in the room and a dark city skyline in the background. Use the style of the poster of the movie "Lost in Translation".

STRM