

Privacy in Practice Smart Pseudonymisation



Kristof Verslype
Cryptographer @ Smals Research

6 June 2024

Smals Research 2024



Innovation with
new technologies



Consultancy
& expertise



Internal & external
knowledge transfer



Support for
going live



**Physical
masks**

**Written
pseudonyms**

**Digital
pseudonyms**

Innovation @ Smals Research

Smart Pseudonymisation

Conversion from citizen identifiers to pseudonyms

Format-Preserving Pseudonymisation

Retroactive protection of personal data in TEST & ACC of legacy applications



eHealth Blind Pseudonymisation

Proactive protection of personal data in applications
Privacy by Design



Oblivious Join

Non-trivial join & pseudonymise projects for research purposes
Distributed & no integration



Format-Preserving Pseudonymisation

- Problem statement
- Concept & PoC
- Experimental service
- Conclusion



Format-Preserving Pseudonymisation

- **Problem statement**
- Concept & PoC
- Experimental service
- Conclusion



Widespread use of personal data in non-prod environments

“60% of organisations use raw production data in test environments”

World Quality Report, 2020

Security

Data breaches from non-prod environments

2016

UBER

Hacker exploited Uber's software development environments to break into the rideshare giant's cloud storage

2021

T Mobile™

Hacker leveraged an unprotected router to gain access to T-Mobile's production, staging, and development servers, which compromised over 48 million social security numbers and other details.

2022

LastPass...

The hacker targeted the home computer of a LastPass senior DevOps engineer

No negligible risk!

Compliance

Personal data in TEST/ACC subject to GDPR

❖ *Legitimate ground required*

- No informed and actively given consent
- Legitimate interest (gerechtvaardigd belang) questionable
- Special personal data (minors, medical data, sexual orientation, financial data, criminal data, ...)
- Other legitimate ground?

❖ *Appropriate measures*

- In general, TEST is less secured than PROD/ACC

Pseudonimisation

- ❖ Encouraged by GDPR to protect personal data
- ❖ Some rules by GDPR more relaxed
- ❖ Could help become more compliant
- ❖ Still Personal data

GDPR, Art 32.

*Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, **the controller and the processor shall implement appropriate technical and organisational measures** to ensure a level of security appropriate to the risk, **including inter alia as appropriate:***

- a) the **pseudonymisation** and encryption of personal data;*
- b) ...*

Guidelines on the protection of personal data in IT governance and IT management of EU institutions



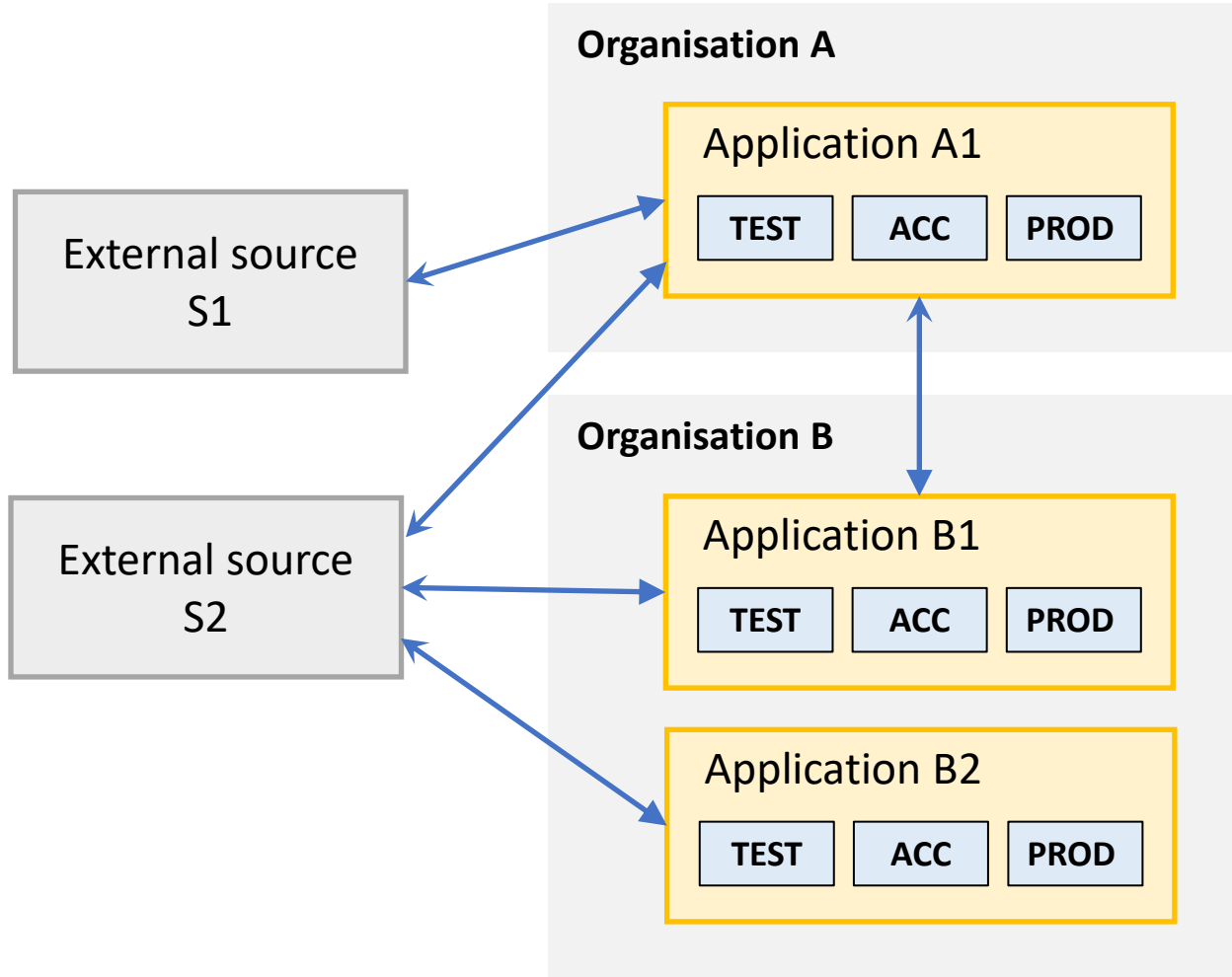
23 March 2018

80 In the testing phase, sampling of real personal data should be avoided, as such data cannot be used for purposes for which it was not collected and using it in testing environments may result in making personal data available to unauthorised individuals.

81 Where possible, artificially created test data should be used, or test data which is derived from real data so that its structure is preserved but no actual personal data is contained in it. Different such techniques have been applied successfully.

*82 Where thorough and cautious analysis shows that generated test data cannot provide sufficient assurance for the validity of the tests, a **comprehensive decision must be taken and documented**, which defines which real data shall be used in the test, **as limited as possible**, the **additional technical and organisational safeguards** which are established in the testing environment. Special categories of data can only be used in real data testing with the explicit consent of the individuals concerned.*

Theory meets reality



Question customer

**How to improve privacy
in TEST & ACC?**

Completely fictional data **not** an option

Because

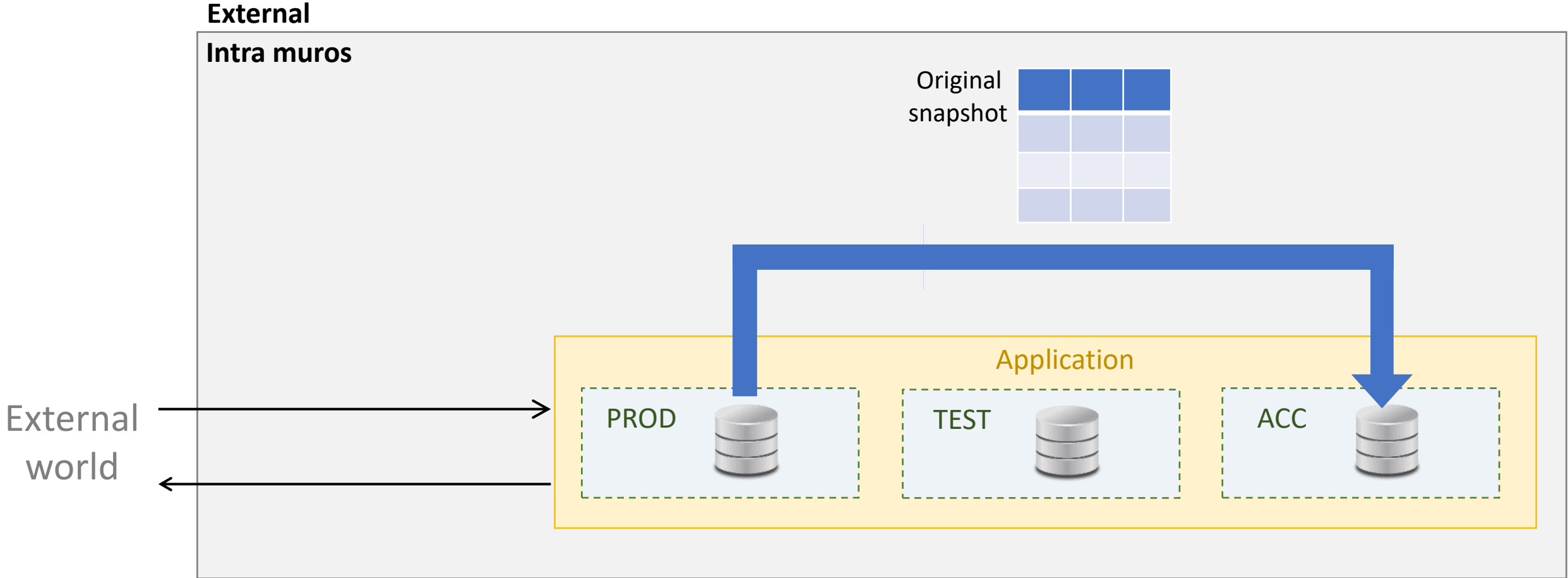
- Communication with external services
- We would miss edge cases
- Complex data models
- Labor intensive

Format-Preserving Pseudonymisation

- Problem statement
- **Concept & PoC**
- Experimental service
- Conclusion



Approach customer



Approach by member

Transforming batch of records with personal data copied to TEST or ACC

Structured identifiers	Unstructured identifiers		Domain-specific data		
Identifier	First name	Surname
18.32.08-903.41	Kasper	de Brouckère	A1	A2	A3
30.02.06-981.94	Melchior	Rogier	B1	B2	B3
72.43.27-109.21	Baltazar	Beernaert	C1	C2	C3

Identifier	First name	Surname
30.03.30-213.23	Melchior	Beernaert
66.08.15-286.27	Baltazar	de Brouckère
22.51.14-602.20	Kasper	Rogier

Transformed snapshot

Identifier	First name	Surname
30.03.30-213.23	Melchior	Beernaert	A1	A2	A3
66.08.15-286.27	Baltazar	de Brouckère	B1	B2	B3
22.51.14-602.20	Kasper	Rogier	C1	C2	C3

1. Pseudonymise

Replace structured identifier by format-preserving pseudonym

- Bidirectional
- By Smals Research

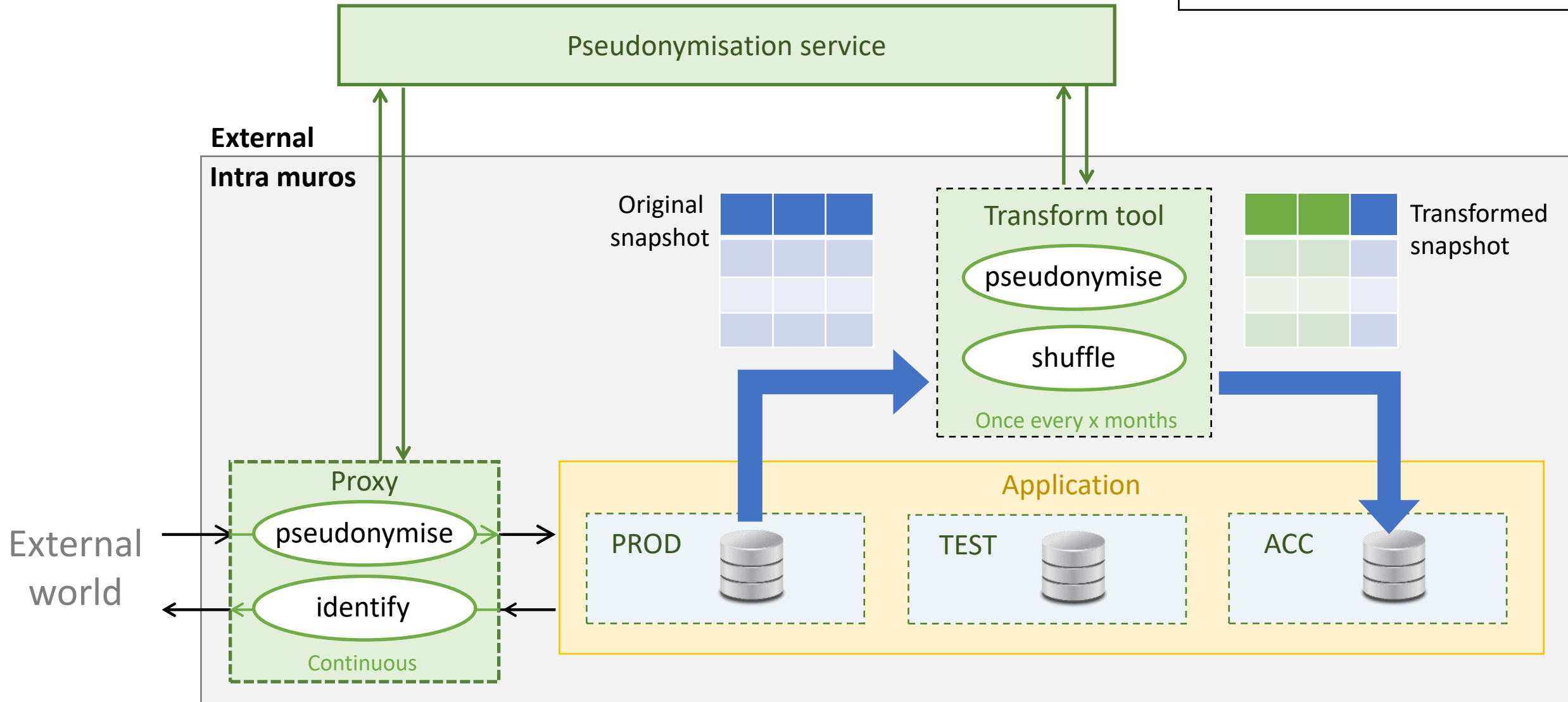
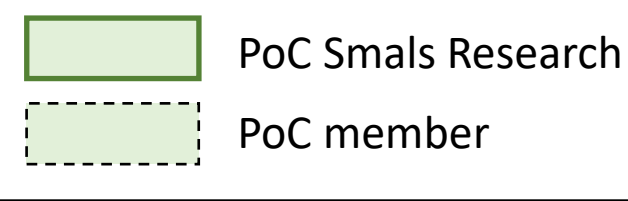
2. Shuffle

Column-wise permutation

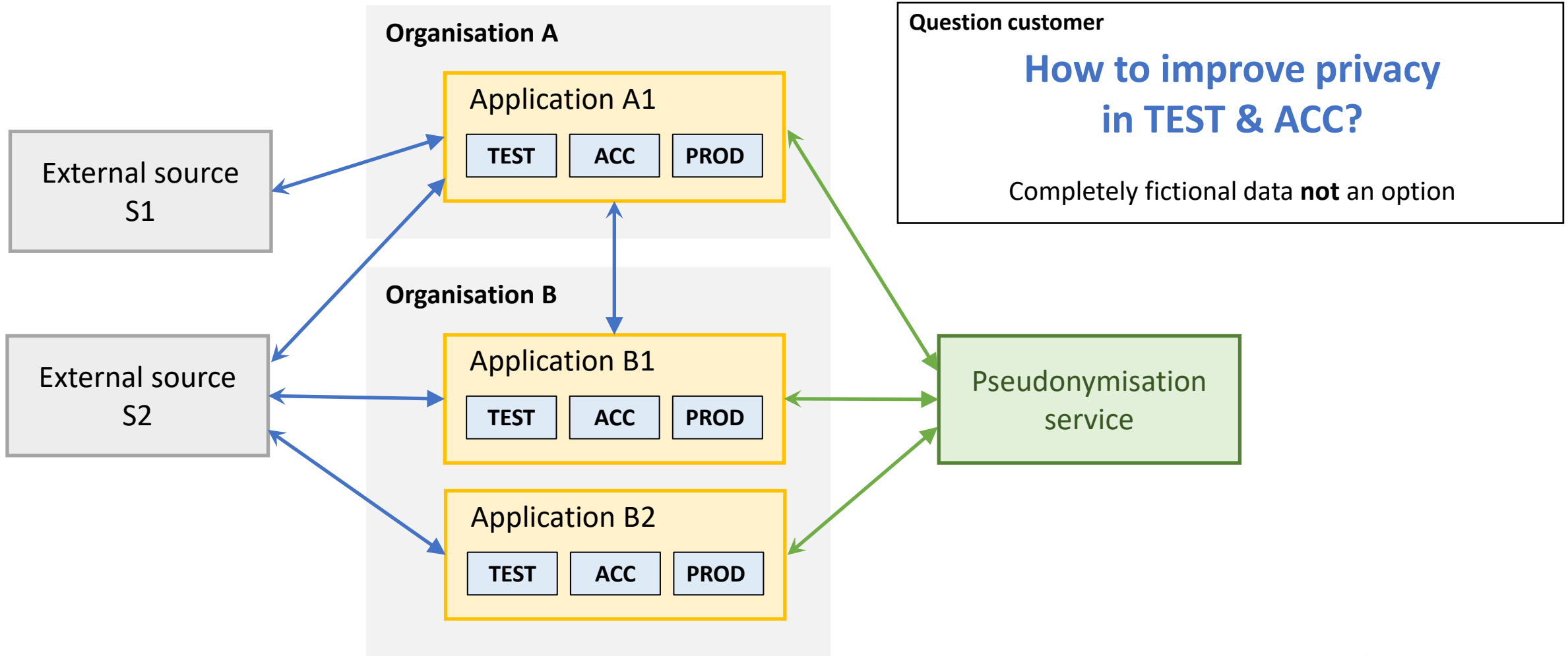
- Unidirectional
- By Customer

Records useful for TEST & ACC, while hard to identify!

PoC in collaboration with customer



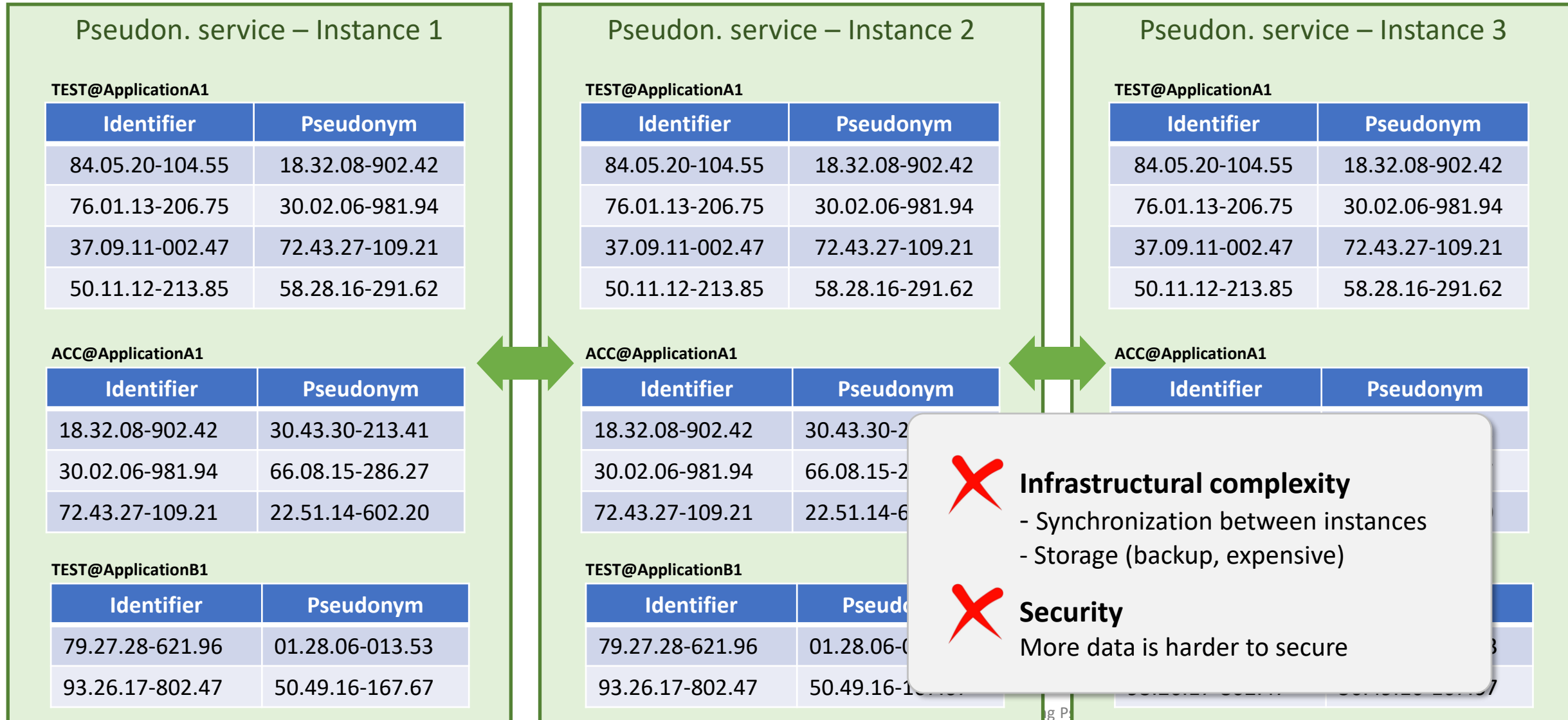
The reality in our sector



Question customer
How to improve privacy in TEST & ACC?
Completely fictional data **not** an option

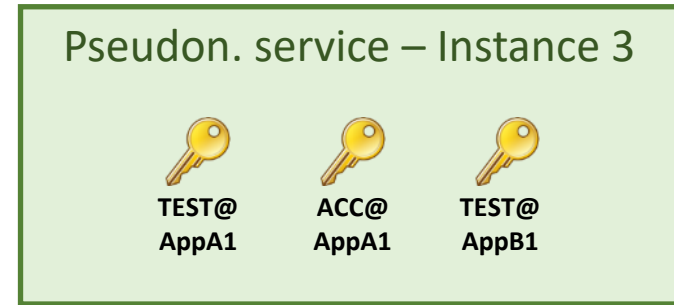
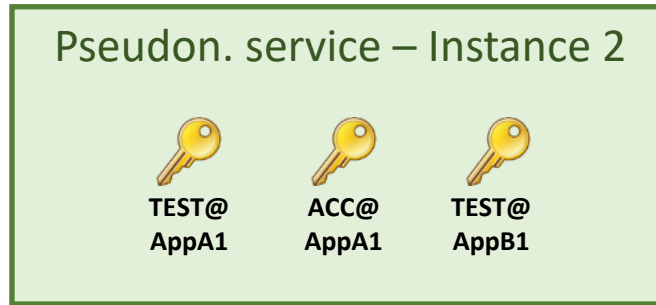
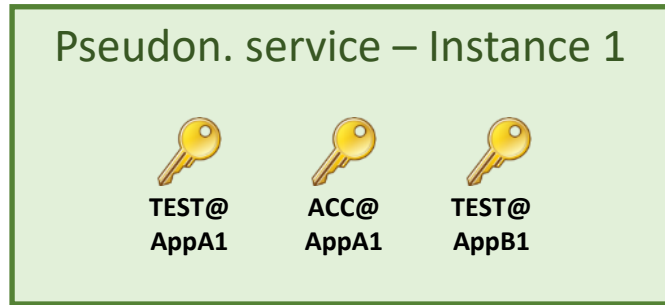
Classical approach

Pseudo service maintains **one table** per environment



Proposed approach

Pseudonymisation service maintains **single key** per environment



Key size is 32 bytes

Example

```
a4 71 c3 e0 9f 79 b3 64
3f 89 42 24 16 a1 9d 1e
6f f0 f6 4e 87 ea 34 03
68 a4 4e ee c0 14 dd 2d
```



Security

- Small keys more easy to secure (e.g. HSM)



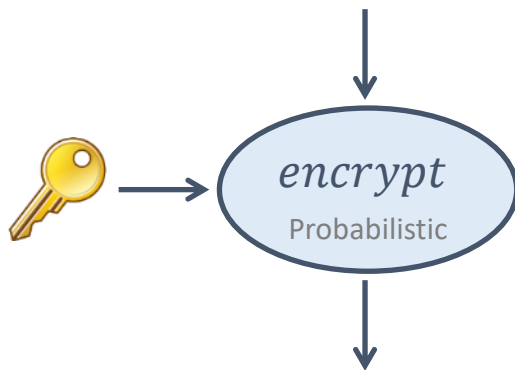
Infrastructural simplicity

- Minimal storage required
- Synchronization hugely simplified
- All keys derivable from single master key

Format-Preserving Encryption (FPE)

TRADITIONAL ENCRYPTION

83.06.21-123.62

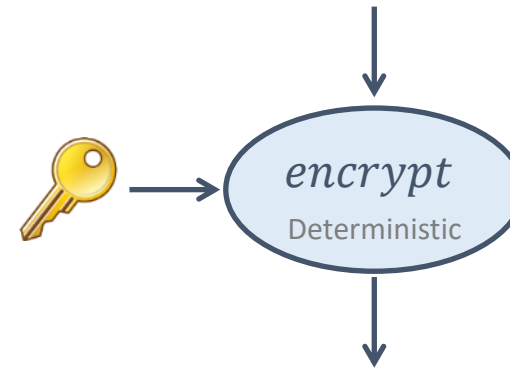


20 58 4d 87 3b 9a 97 dd
00 a0 50 83 20 00 08 b8
da 40 ab 08 04 06 07 2e
5b 08 7d 19 d8 44 40 a8
34 69 45 d3 3e 74 99 1f
0d fb 0a 50 3a 67 70 b4
a0 30 ba e0 bf 01 52 ac
13 40 02 58 7a 38 e2 09

Regular pseudonym

FORMAT-PRESERVING ENCRYPTION

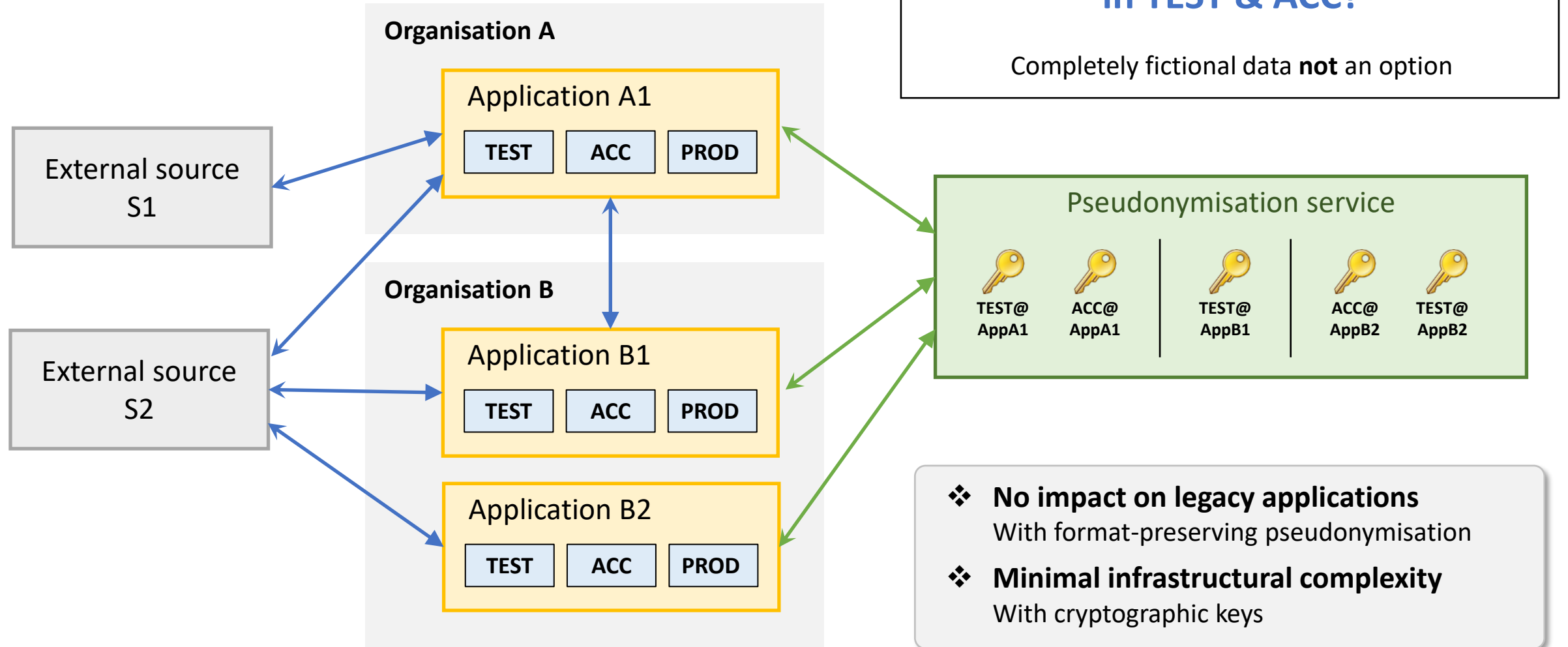
83.06.21-123.62



67.11.14-522.33
Format-preserving pseudonym

- ❖ Conversions happen on-the-fly
- ❖ Structure preserved, including valid checksum
- ❖ More details on blogpost Smals Research
- ❖ Described in NIST SP 800-38G Revision. 1 (2019)

The reality in our sector



Format-Preserving Pseudonymisation

- Problem statement
- Concept & PoC
- **Experimental service**
- Conclusion



Experimental REST service

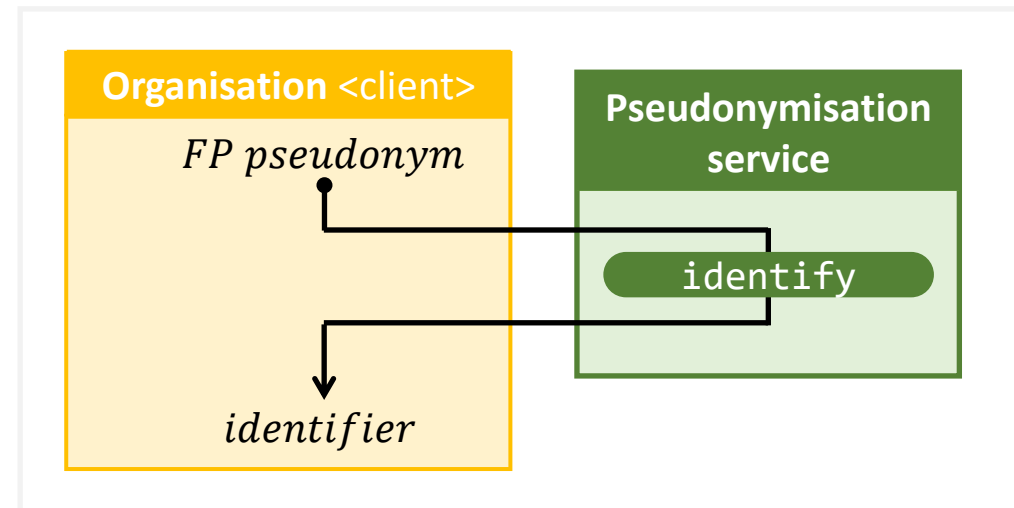
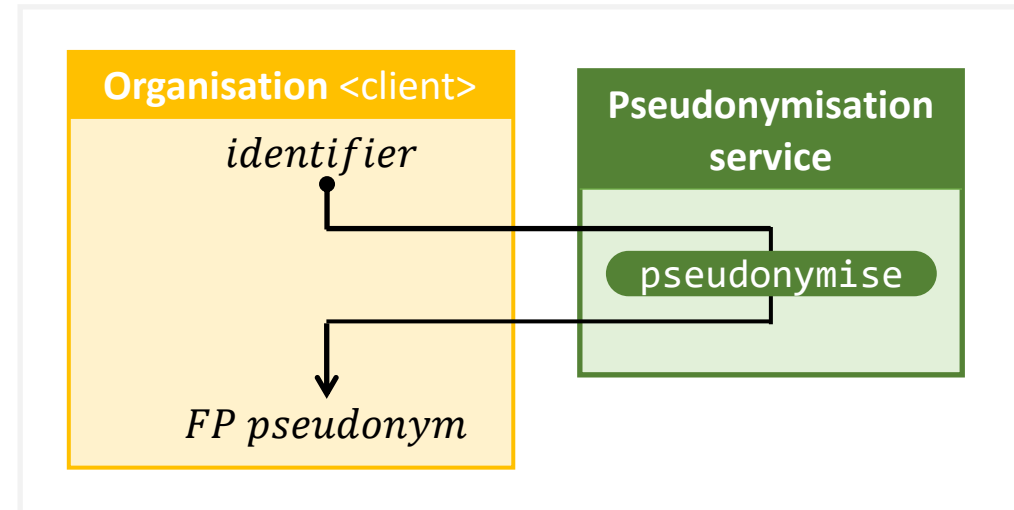
Built by Smals Research

Rest API

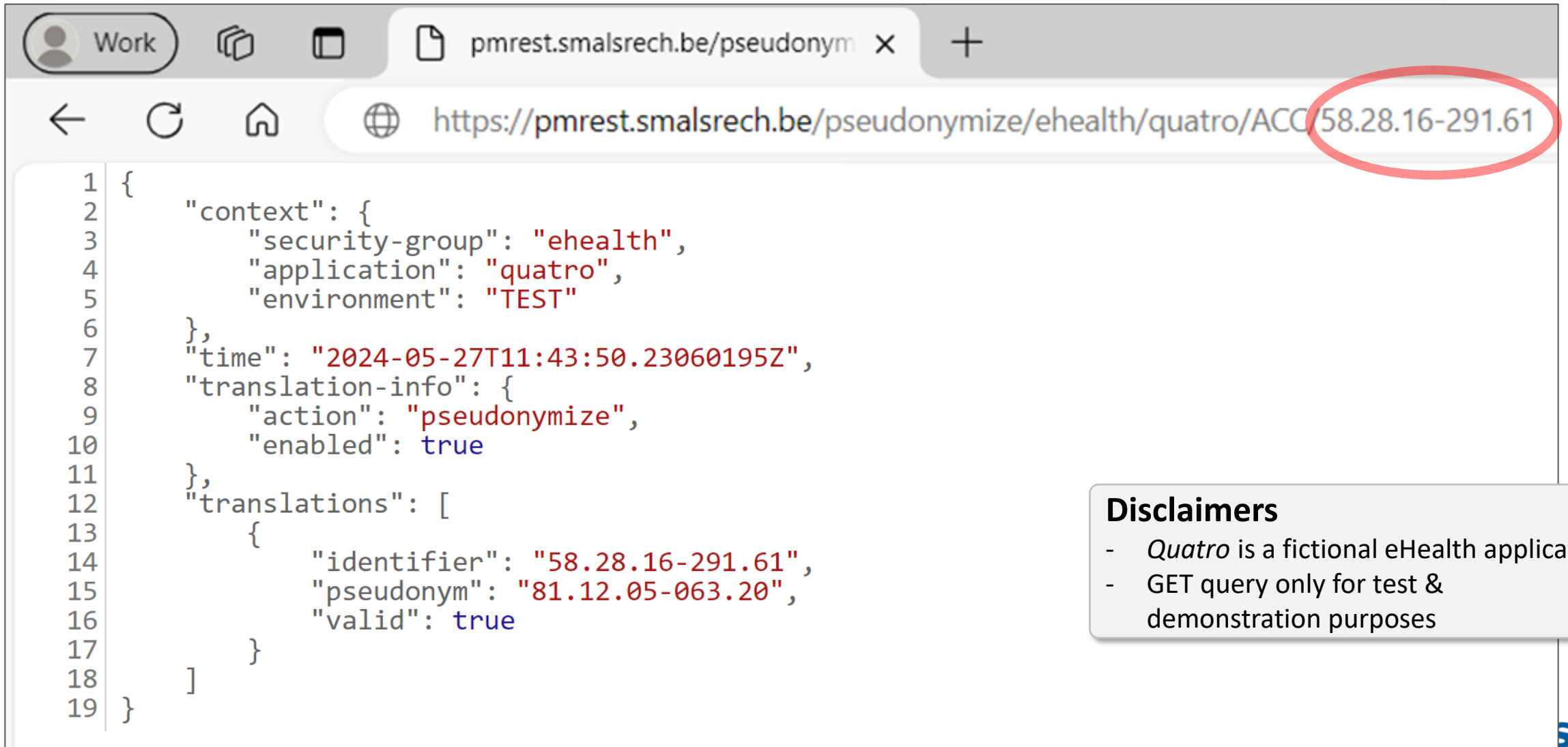
- ✓ Pseudonymise & Identify
- ✓ GET and POST
- ✓ Also batch (POST only)

Extensible

- ✓ Identifier types
 - Currently, only RRN, BIS, INSZ
 - KBO number, Bank account numbers, ...
- ✓ ...



GET query



```
1 {
2   "context": {
3     "security-group": "ehealth",
4     "application": "quatro",
5     "environment": "TEST"
6   },
7   "time": "2024-05-27T11:43:50.23060195Z",
8   "translation-info": {
9     "action": "pseudonymize",
10    "enabled": true
11  },
12  "translations": [
13    {
14      "identifier": "58.28.16-291.61",
15      "pseudonym": "81.12.05-063.20",
16      "valid": true
17    }
18  ]
19 }
```

Disclaimers

- *Quatro* is a fictional eHealth application
- GET query only for test & demonstration purposes

POST Request

```
1 {
2   "context": {
3     "security-group": "ehealth",
4     "application": "quatro",
5     "environment": "TEST"
6   },
7   "identifiers": [
8     "18.32.08-902.42",
9     "30.02.06-981.94",
10    "72.43.27-109.21",
11    "58.28.16-291.62",
12    "58.28.16-29X.61",
13    "58.28.16-291.90",
14    "79.27.28-621.96",
15    "30.43.04-205.53",
16    "93.26.17-802.47",
17    "33.24.16-568.07"
18  ]
19 }
```

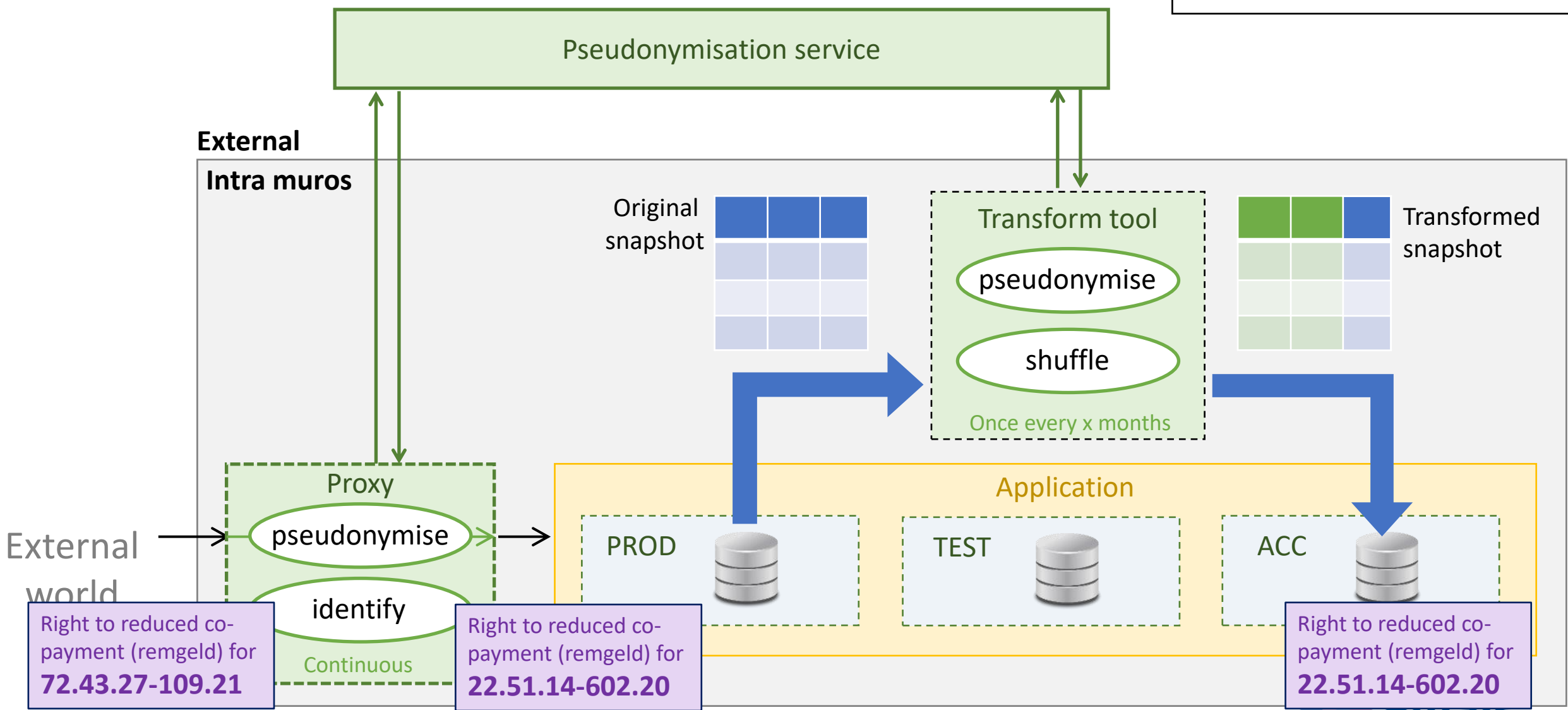
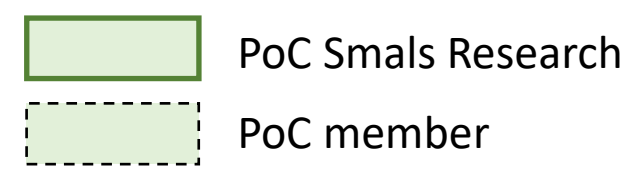


POST Response

```
1 {
2   "context": {
3     "security-group": "ehealth",
4     "application": "quatro",
5     "environment": "TEST"
6   },
7   "time": "2024-01-08T08:20:39.128207895Z",
8   "translation-info": {
9     "action": "pseudonymize",
10    "enabled": true
11  },
12  "translations": [
13    {
14      "identifier": "18.32.08-902.42",
15      "pseudonym": "30.43.30-213.41",
16      "valid": true
17    },
18    {
19      "identifier": "30.02.06-981.94",
20      "pseudonym": "66.08.15-286.27",
21      "valid": true
22    },
23    {
24      "identifier": "72.43.27-109.21",
25      "pseudonym": "22.51.14-602.20",
26      "valid": true
27    },
28    {
29      "identifier": "58.28.16-291.62",
30      "pseudonym": "null",
31      "valid": false,
32      "error": "checksum"
33    },
34  ]
35 }
```

- ✓ Easy to use
- ✓ Graceful error handling
- ✓ Efficient

PoC in collaboration with customer



External world

External Intra muros

Pseudonymisation service

Original snapshot

Transform tool
pseudonymise
shuffle
Once every x months

Transformed snapshot

Proxy
pseudonymise
identify
Continuous

Application
PROD
TEST
ACC

Right to reduced co-payment (remgeld) for 72.43.27-109.21

Right to reduced co-payment (remgeld) for 22.51.14-602.20

Right to reduced co-payment (remgeld) for 22.51.14-602.20

Format-Preserving Pseudonymisation

- Problem statement
- Concept & PoC
- Experimental service
- **Conclusion**



In summary

Concept

- **Building block to improve privacy in TEST and ACC environments**
- Citizen known under different pseudonym in each environment

As a service

- Reduce complexity side organisation
E.g. key management
- Separation of duties

Towards a generic service?

- Relatively simple service
- **Interest from members required**

Status

- Advanced PoC (Extensibility, unit tests, error handling, Smals standards, ...)
- **Experimental REST service** running
- No logging, access control, ...

Publication



Gegevensbescherming m.b.v. structuurbehoudende pseudonimisatie van rijksregisternummers

Protection des données par la pseudonymisation préservant la structure des numéros de registre national

www.smalsresearch.be/tag/pseudonymisation/

Innovation @ Smals Research

Smart Pseudonymisation

Conversion from citizen identifiers to pseudonyms

Format-Preserving Pseudonymisation

Retroactive protection of personal data in TEST & ACC of legacy applications



eHealth Blind Pseudonymisation

Proactive protection of personal data in applications
Privacy by Design



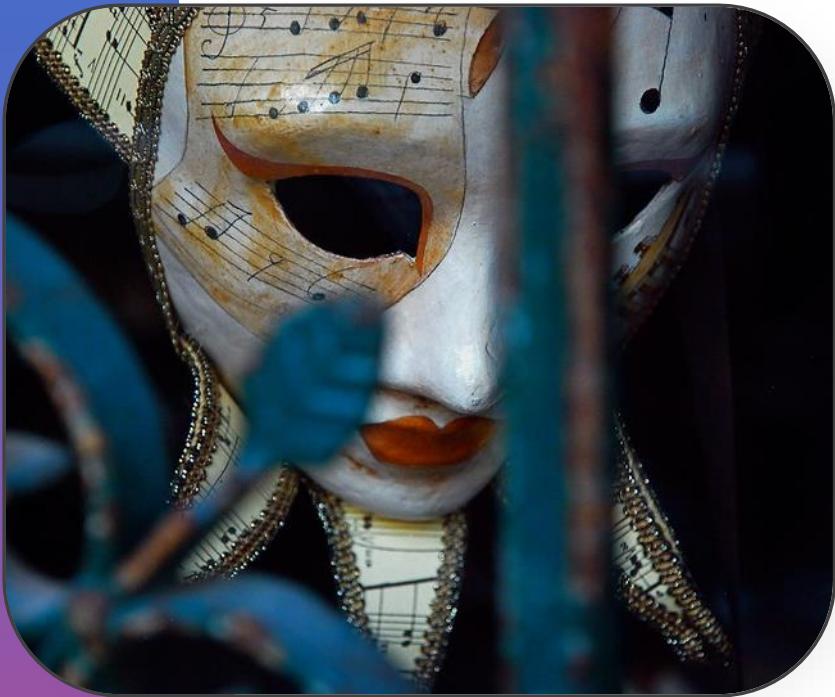
Oblivious Join

Non-trivial join & pseudonymise projects for research purposes
Distributed & no integration



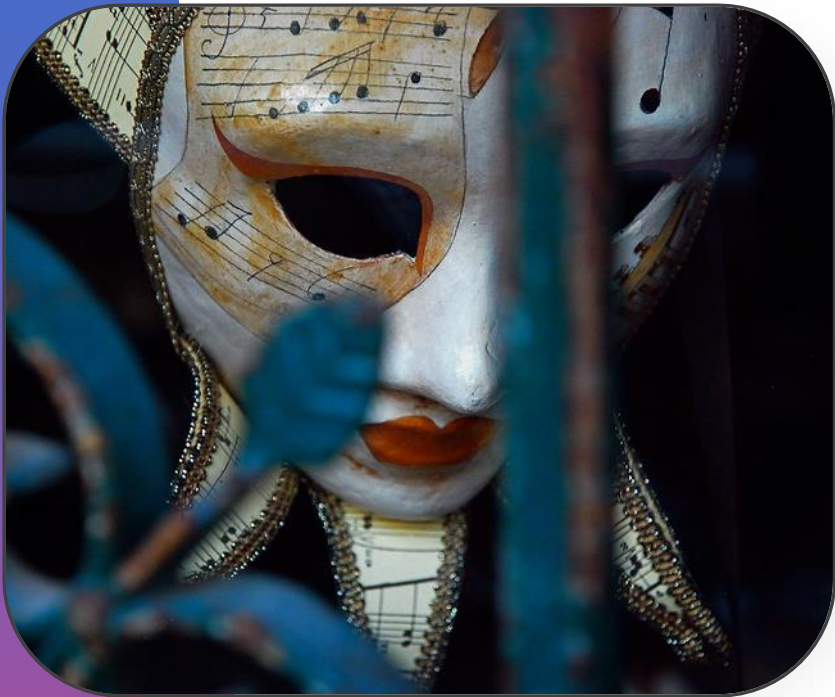
eHealth Blind Pseudonymisation

- Problem statement
- Secure records in live environments
- Join & pseudonymise for research
- Conclusion

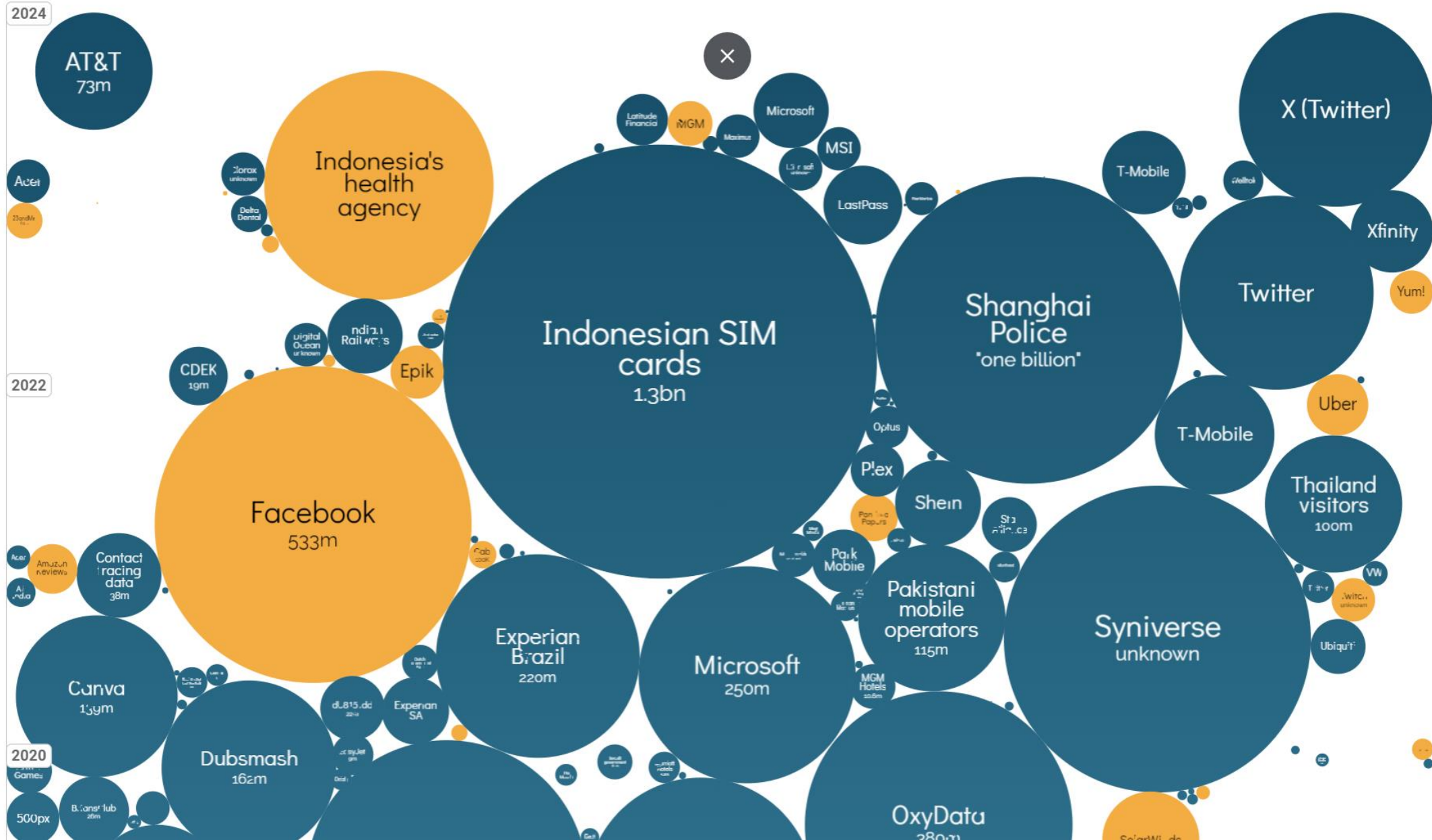


eHealth Blind Pseudonymisation

- **Problem statement**
- Secure records in live environments
- Join & pseudonymise for research
- Conclusion

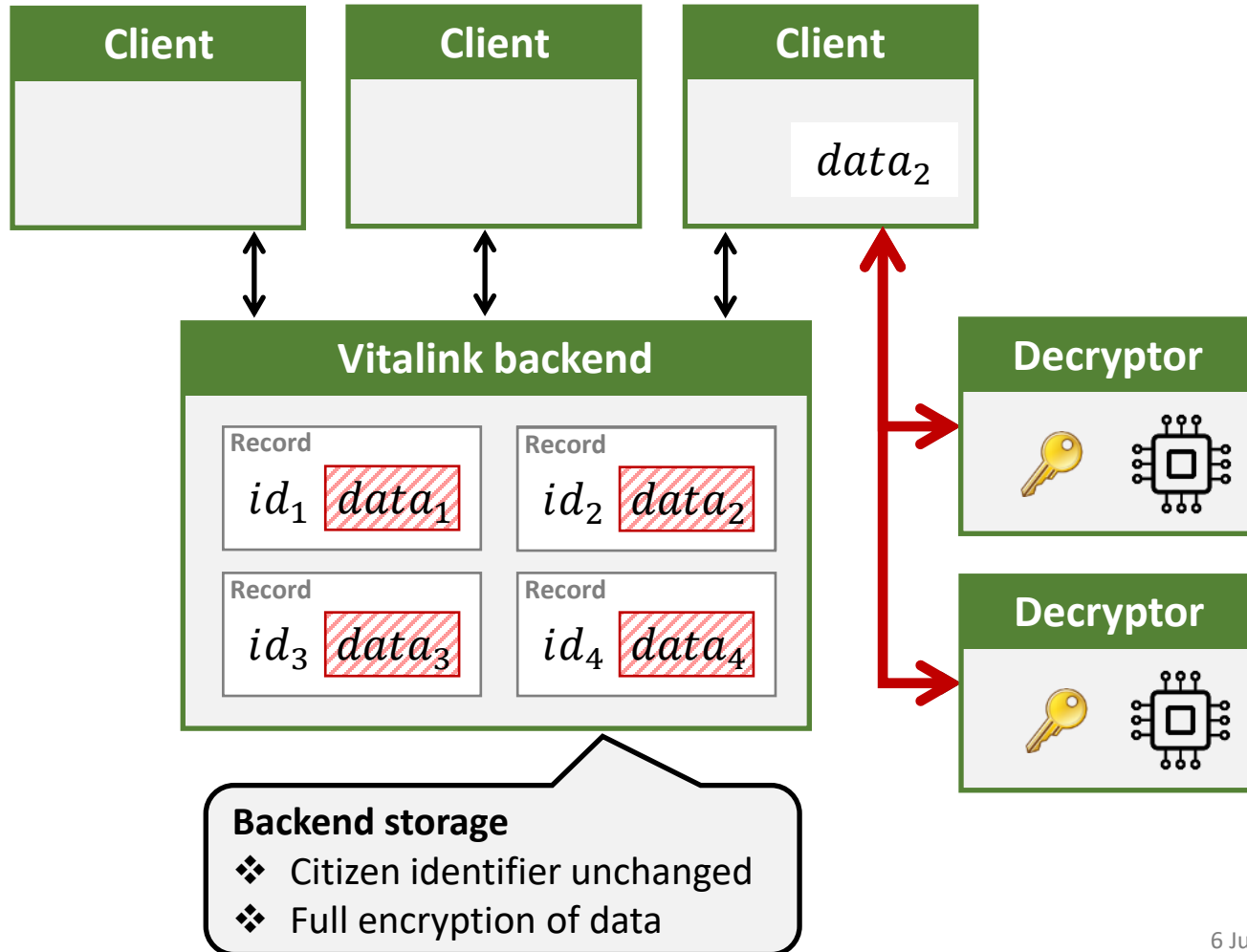


World's Biggest Data Breaches & Hacks





Enables healthcare providers to share digital data about their patients with each other



Design principles

- ❖ **Privacy by design**
Privacy should be taken into account when designing and building products and services
- ❖ **Separation of duties**
Entity managing encryption keys should not have access to protected data (and vice versa)

Privacy requirements

- ❖ Prevent backend from learning personal data
- ❖ Only authorized entities can access data
- ❖ Decryptors don't learn personal data

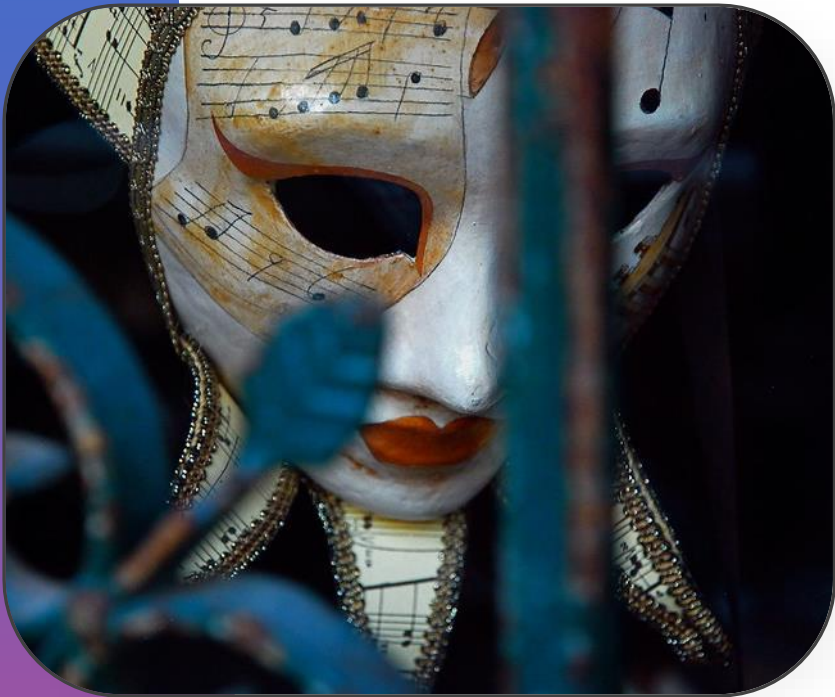
Evaluation

- ✓ High security
- ✗ Full encryption limits functionality
Input verification, statistics, analytics

Can we do better?

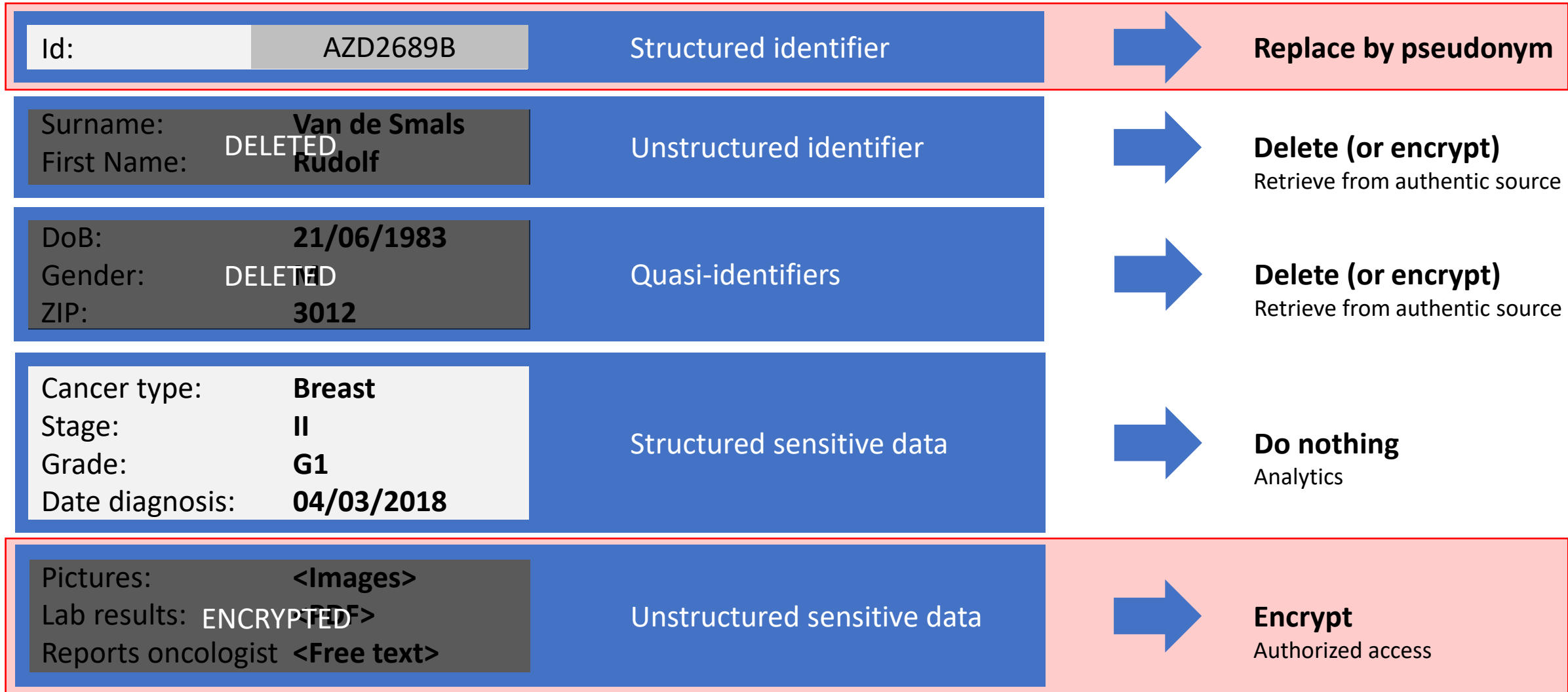
eHealth Blind Pseudonymisation

- Problem statement
- **Secure records in live environments**
- Join & pseudonymise for research
- Conclusion



Protection of personal data record

Fictional example



Records hard to identify while still preserving some functionality

Use case 1 - Live

UHMEP: Referral prescription = Verwijsvoorschrift / Prescription de renvoi

Unaddressed Health Message Exchange Platform

What?

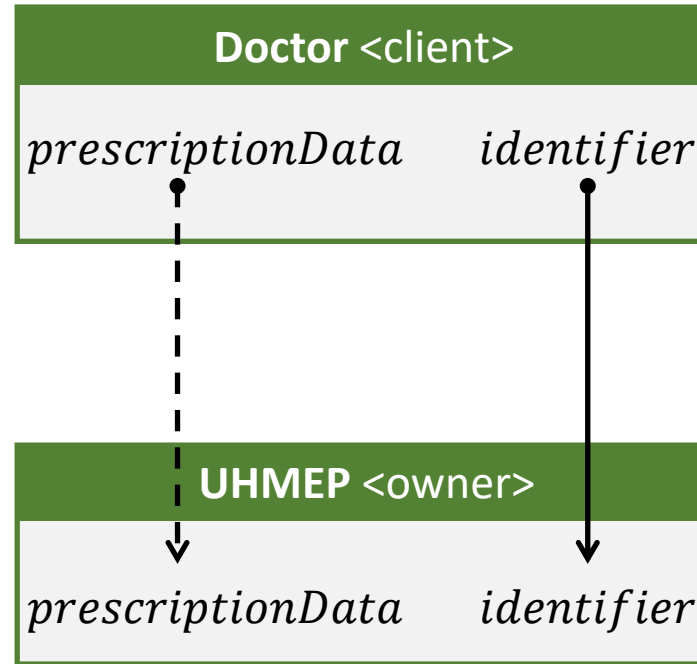
A certificate to start a certain treatment (e.g. physiotherapist, dieticians, speech therapists). Without a referral prescription issued by a doctor, the treatment may not be started.

Requirements

- ❖ No full encryption of data (maybe selective)
- ❖ UHMEP backend should never be able to link prescription data to a natural person

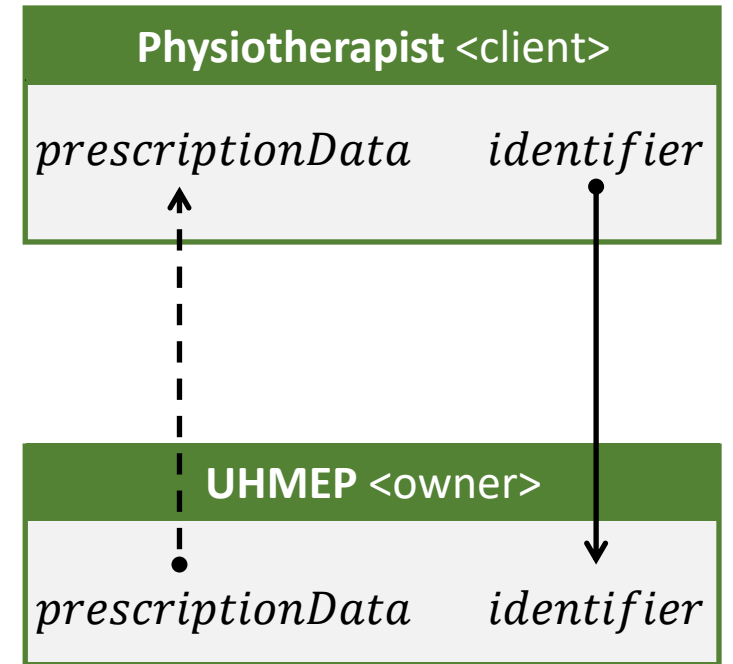
Scenario 1

Doctor (client) requests UHMEP (owner) to register prescription



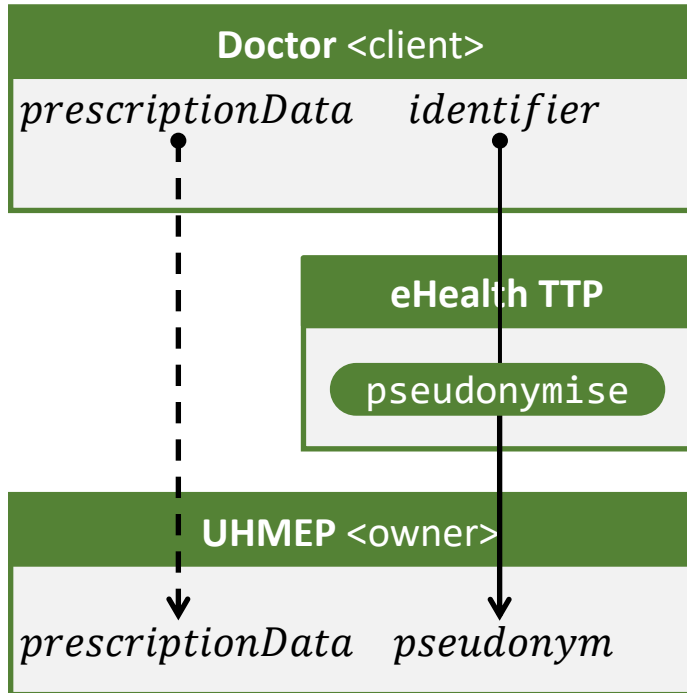
Scenario 2

Physiotherapist (client) requests access to prescription for a specific citizen from UHMEP (owner)



Why no reuse existing eHealth pseudonymisation service?

eHealth WS TTP

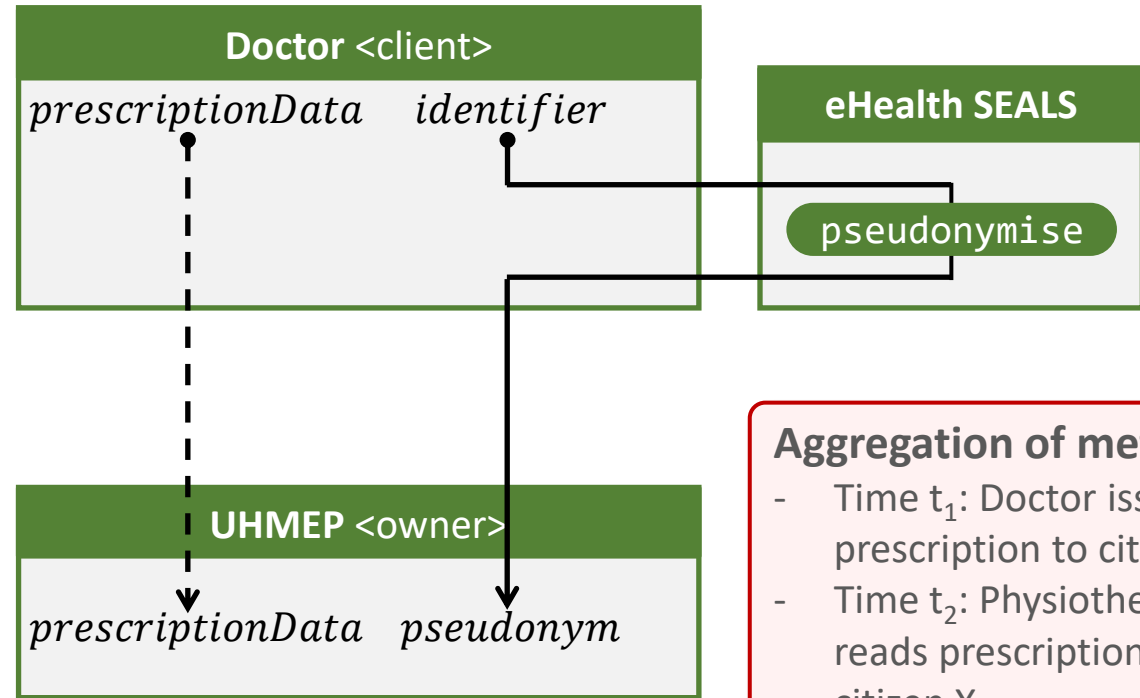


X Trust & identification risk
Data leaks to one entity

X Two communication flows
To be combined somehow

eHealth WS SEALS

Previously known as 'dienst codage'



X Trust & identification risk
Data leaks to two entities

Aggregation of metadata

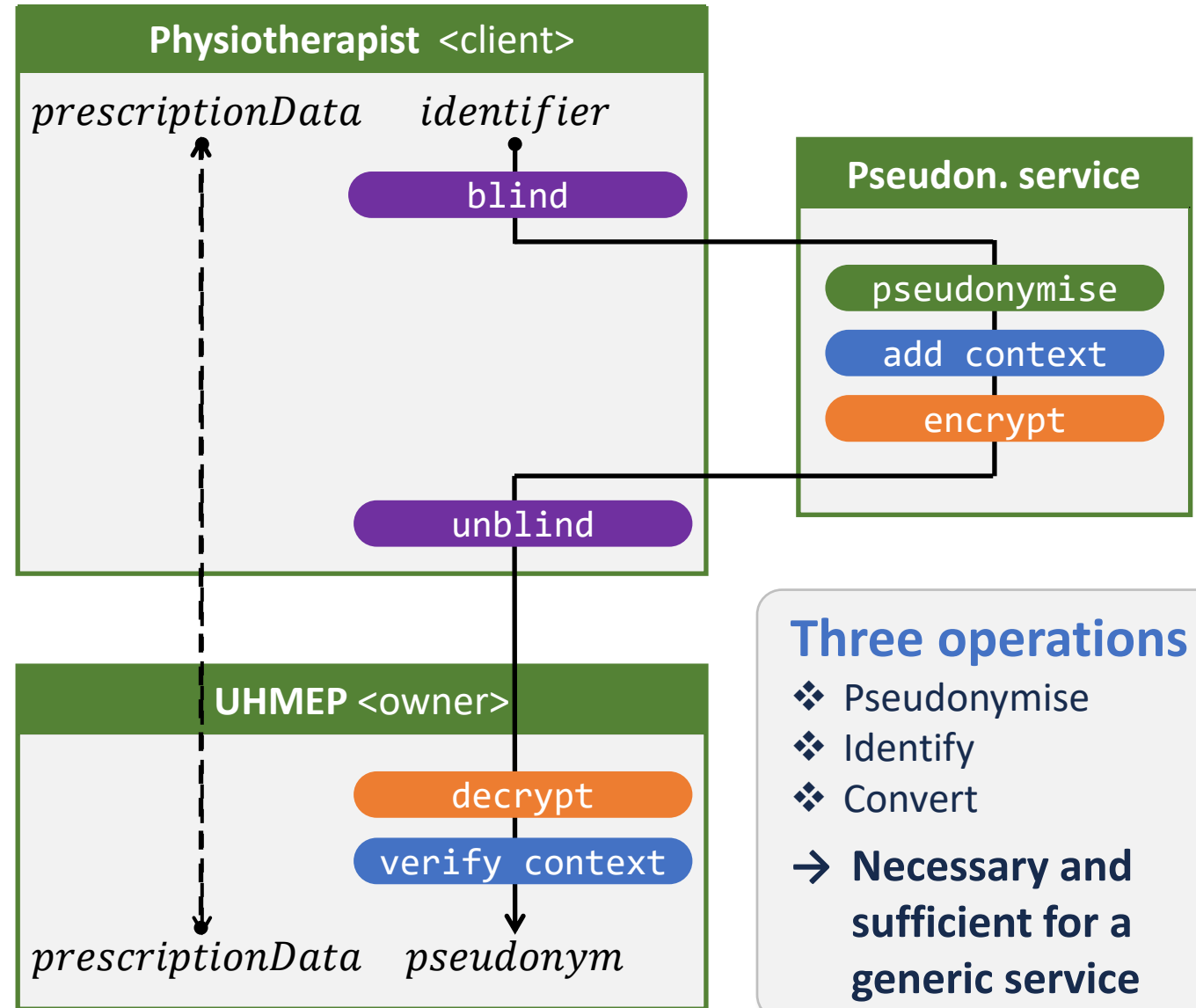
- Time t_1 : Doctor issues prescription to citizen X
- Time t_2 : Physiotherapist reads prescription of citizen X
- Time t_3 : Doctor issues prescription to citizen X
- Time t_4 : Data of citizen X involved in research project on diabetes
- ...

Blind Pseudo Service Pseudonymise

Structure blinded identifier, blinded pseudonym and final pseudonym

(AV+VXF9H5LdTe4b1 SSC7bHjp6b2enJmf pIC6a3/jCR5fUHxX RSaRniYR8h7ugNqa lGvP49cZnv6lf9B7 2RUG0rA/, eSmII52CEtsZzSseU DY3YKltSgqh1wLPm 9ncHBzGiv1wMlxmc1 jSmpW36GhTt/s1P5s hZGhG8ncoWKSgkJDy fw=)

- ✓ **Each party only sees what is absolutely necessary**
- ✓ **Business flow not interrupted**
 - Pseudo service on the sidelines
 - Only direct communication between healthcare professional and UHMEP backend
- ✓ **Low-intrusive client-side**
 - No extra keys requires
 - Relatively simple implementation



Three operations

- ❖ Pseudonymise
- ❖ Identify
- ❖ Convert

➔ **Necessary and sufficient for a generic service**

	Sender		Translator		Receiver	
	identifier	pseudonym	identifier	pseudonym	identifier	pseudonym

<i>Seals</i>	●	●	●	●	○	●
<i>TTP</i>	●	○	●	●	○	●
<i>Blind</i>	●	○	○	○	○	●

Blind Pseudo Service

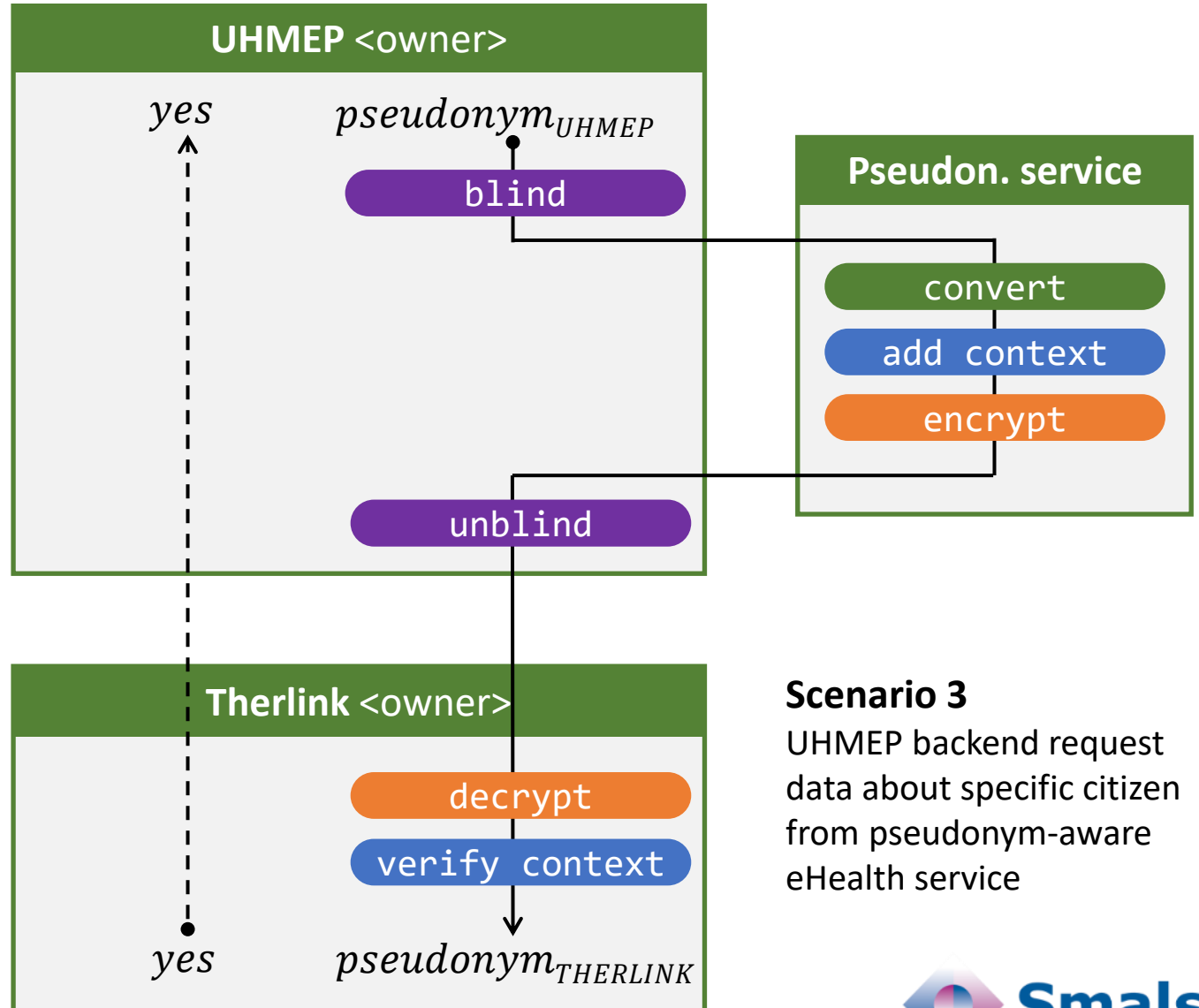
Convert

- ✓ Interaction possible with system that do (not) use pseudo service
- ✓ Each party only sees what is absolutely necessary

Three operations

- ❖ Pseudonymise
- ❖ Identify
- ❖ Convert

→ Necessary and sufficient conditions for a generic service

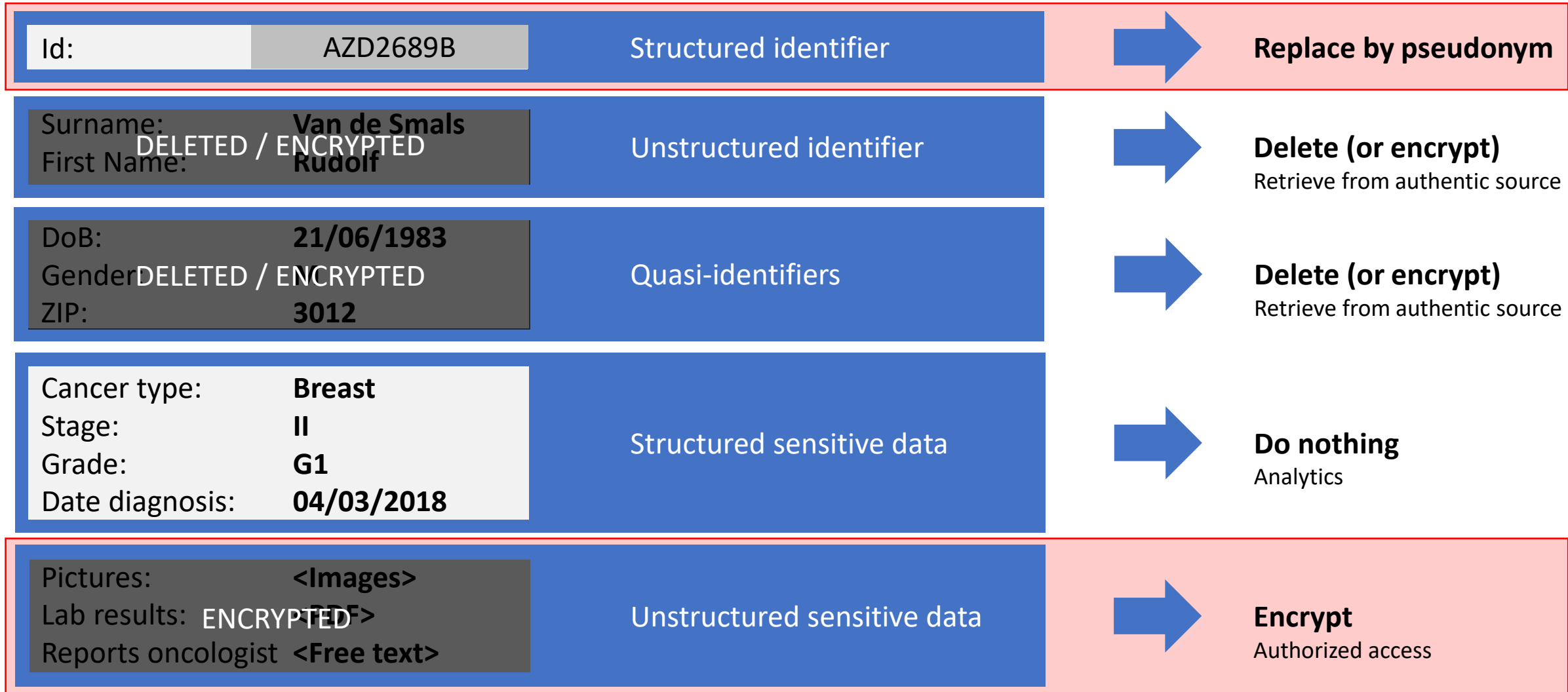


Scenario 3

UHMEP backend request data about specific citizen from pseudonym-aware eHealth service

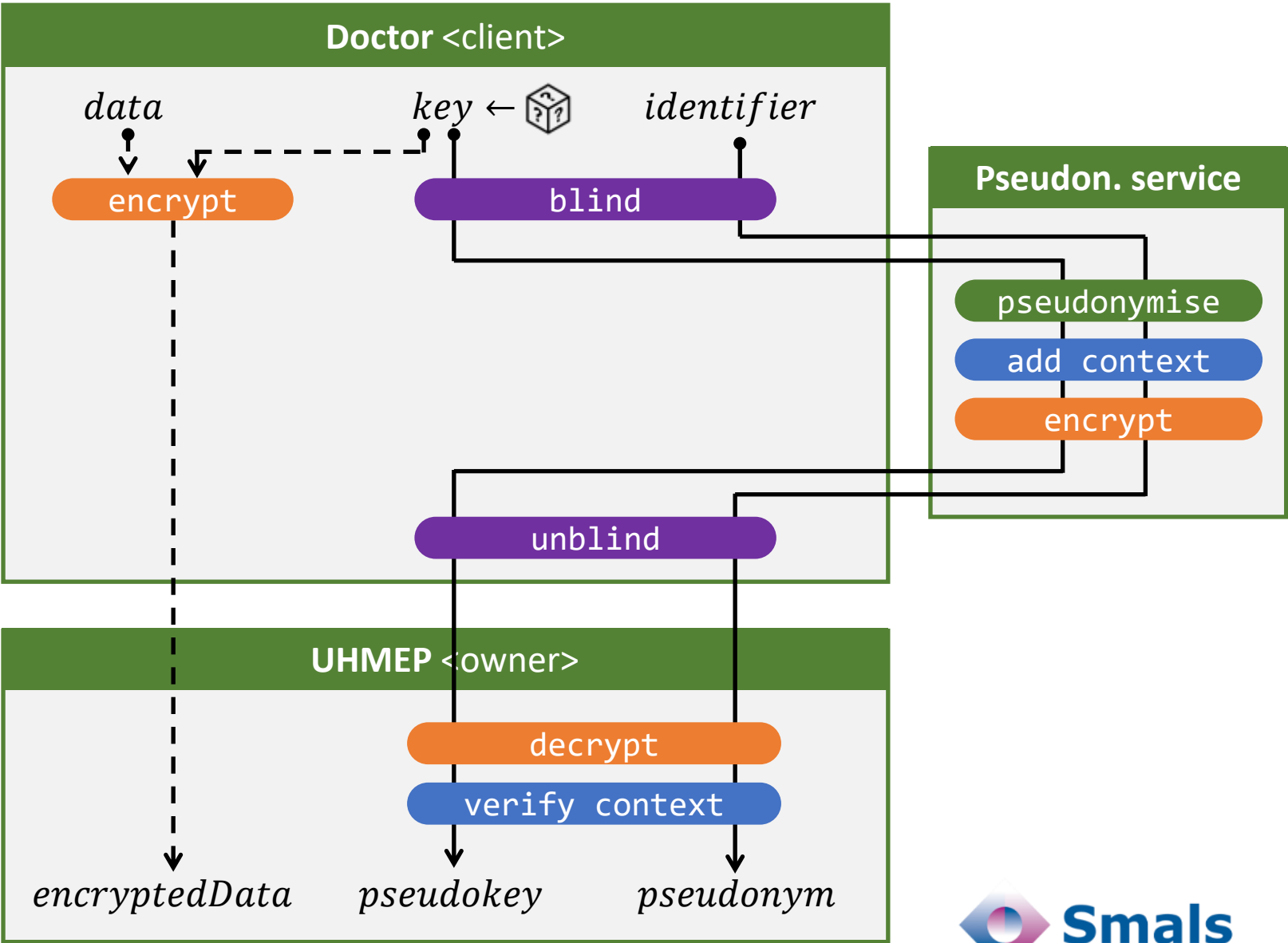
Protection of personal data record

Fictional example



Blind Pseudonymisation Service

Encrypt



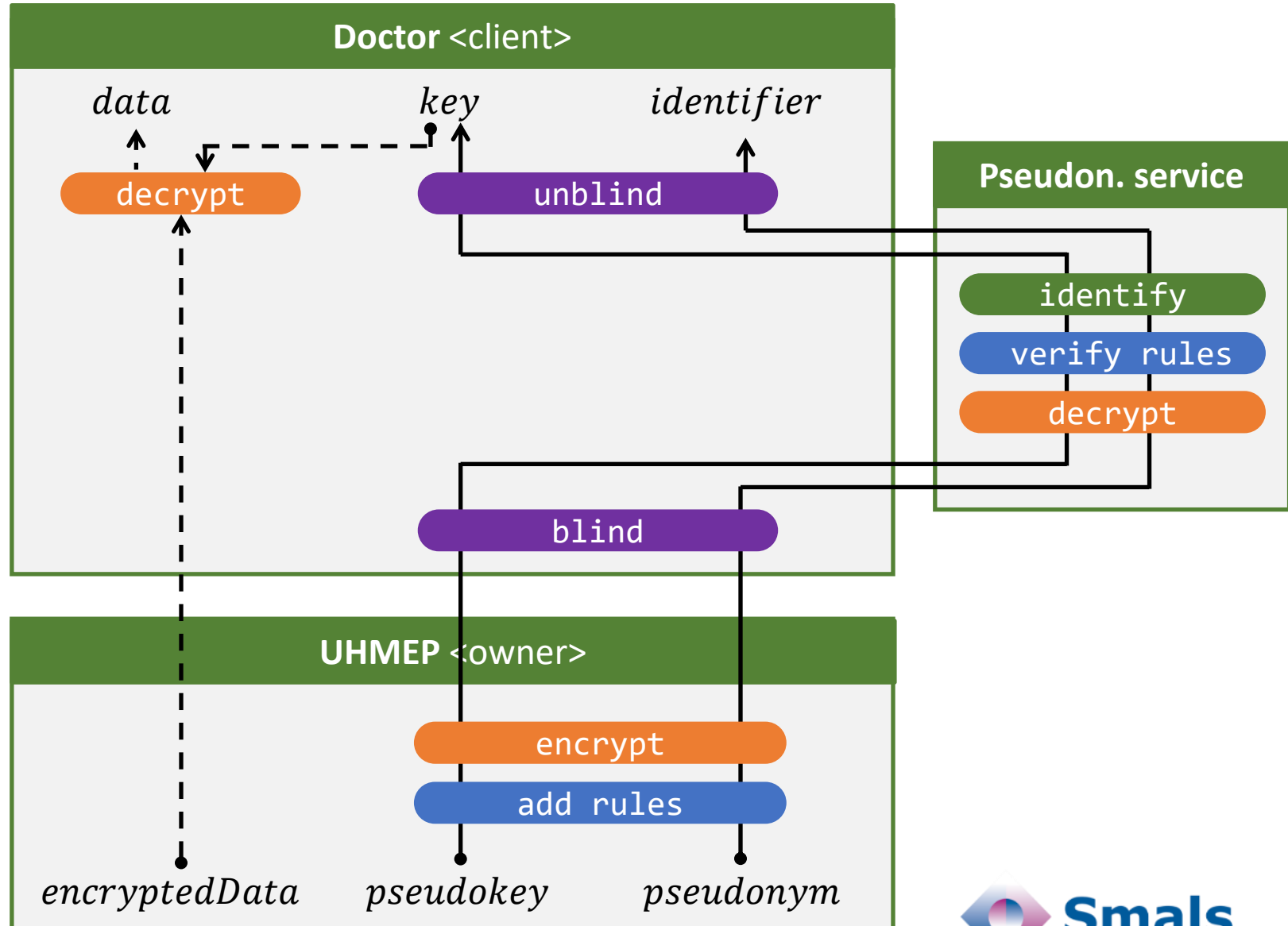
Blind Pseudonymisation Service

Decrypt

- ✓ UHMEP (or hackers) cannot access data
- ✓ Authorized healthcare professional can access data
- ✓ No new logic required

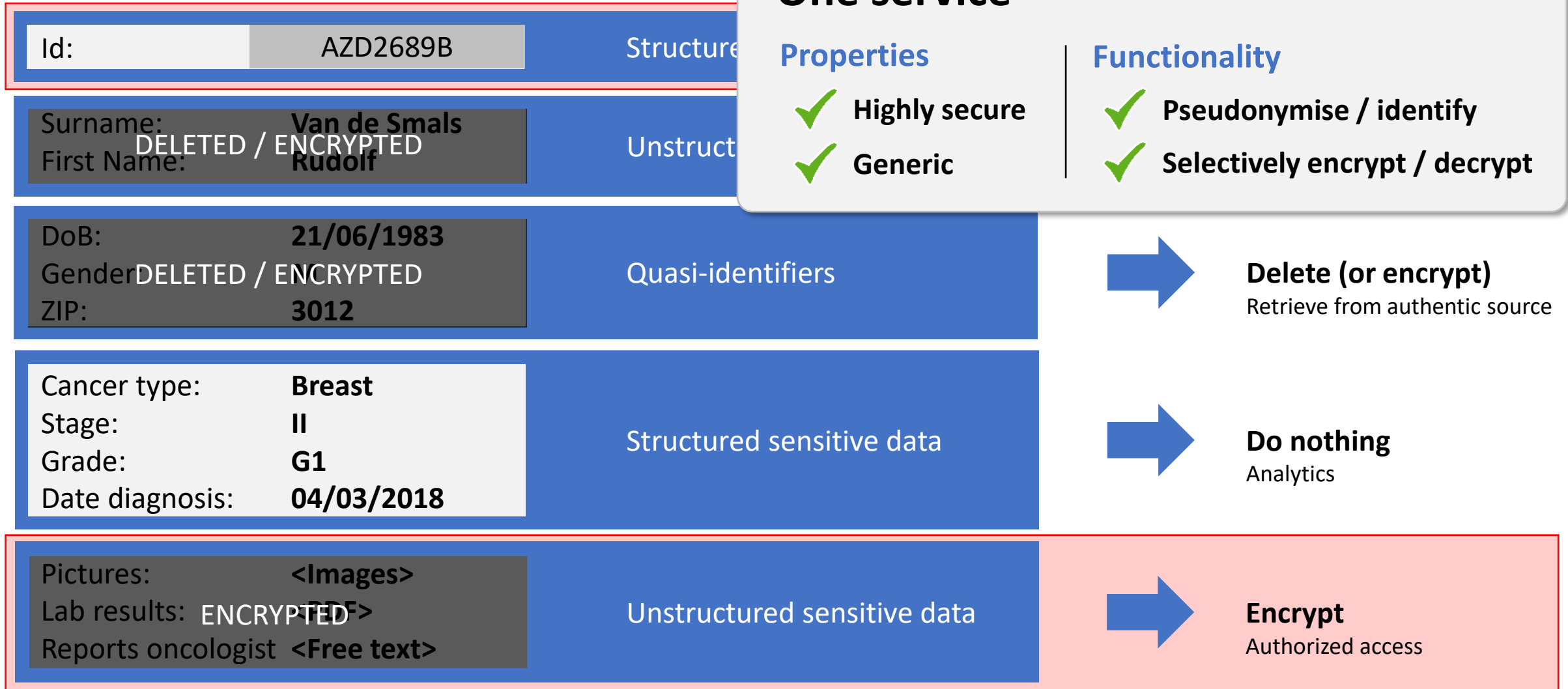
Crucial that pseudon. service

- ❖ is independent
- ❖ is well secured
- ❖ has proper procedures for authorization

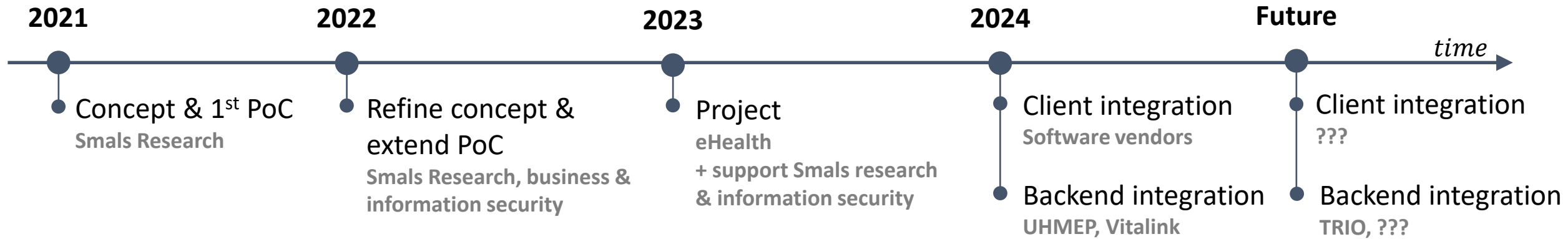


Protection of personal data record

Fictional example

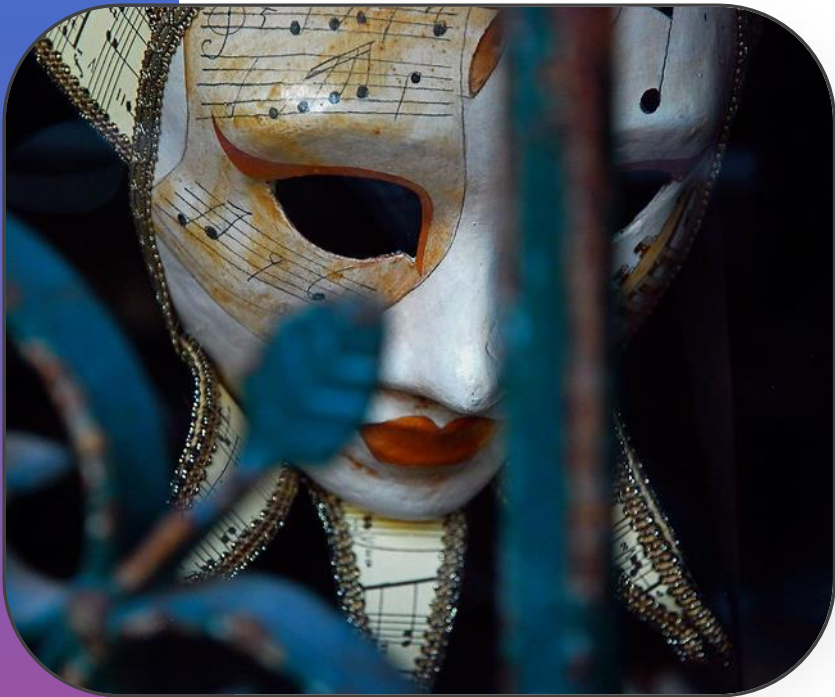


History

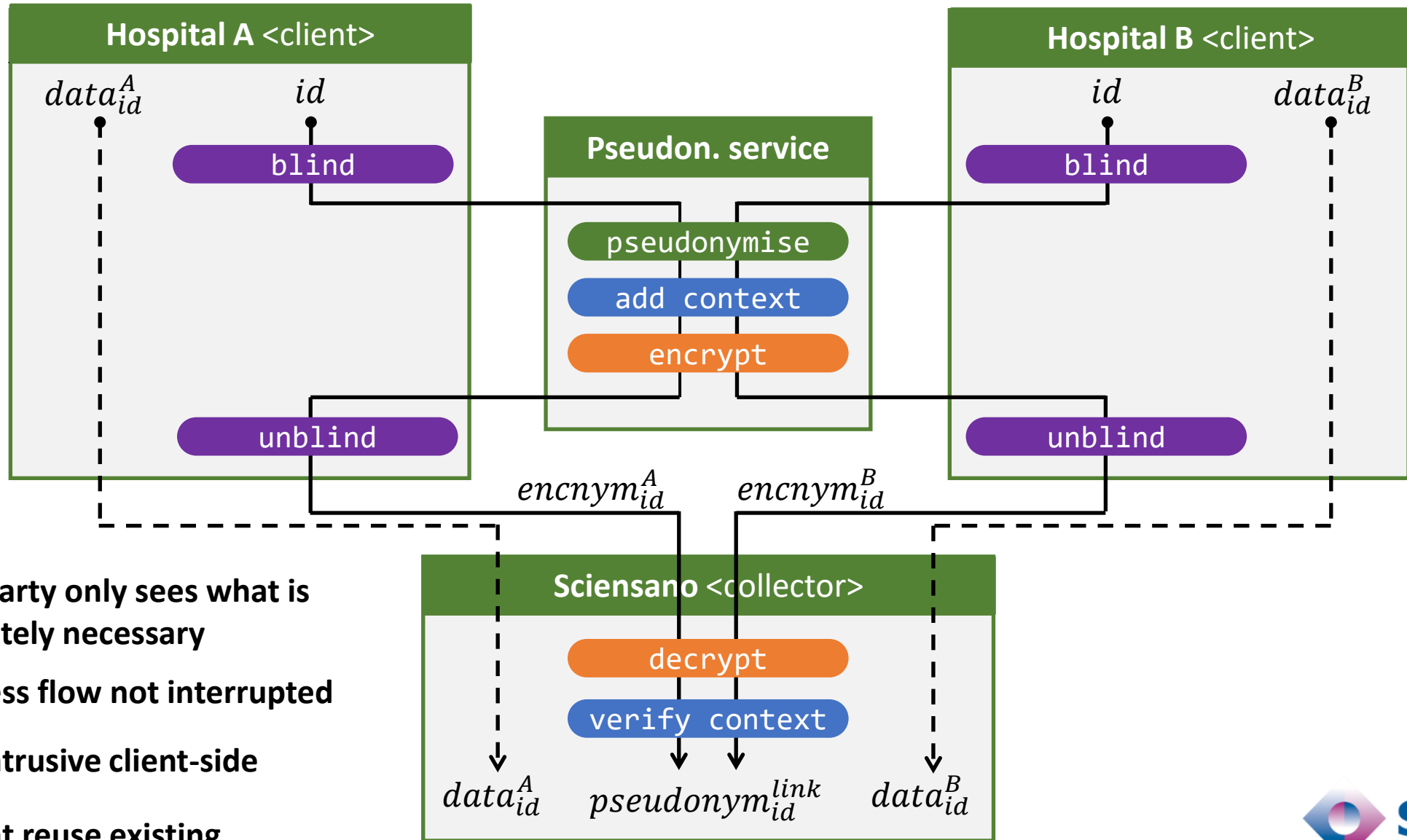


eHealth Blind Pseudonymisation

- Problem statement
- Secure records in live environments
- **Join & pseudonymise for research**
- Conclusion



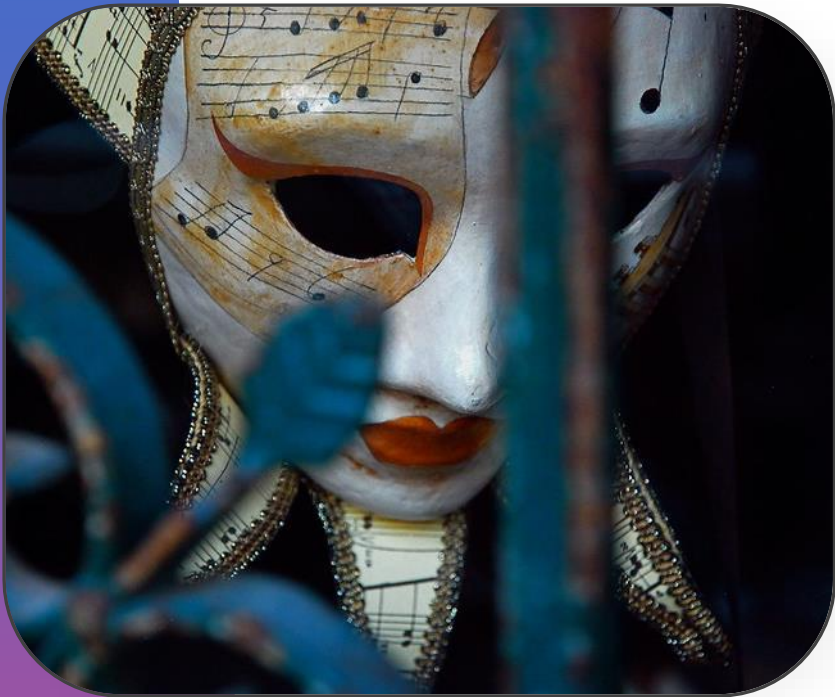
Linking & pseudonymizing data for research purposes



- ✓ Each party only sees what is absolutely necessary
- ✓ Business flow not interrupted
- ✓ Low-intrusive client-side
- ✓ Efficient reuse existing infrastructure

eHealth Blind Pseudonymisation

- Problem statement
- Secure records in live environments
- Join & pseudonymise for research
- **Conclusion**



In summary

Versatile tool

- Pseudonymise identifiers, identify or convert pseudonyms
- Encrypt & decrypt data
- Linking & pseudonymizing data for research purposes

High security

- Each party only sees what is absolutely necessary
- Separation of duties
- Privacy by design
- HSMs

Acceptable complexity

- Manageable
- Especially client-side (integration software vendors)

Status: live



Publications Smals Research

Introductie tot de nieuwe eHealth pseudonimiseringsdienst

<https://www.smalsresearch.be/tag/pseudonymisation/>

Introduction au nouveau service de pseudonymisation eHealth

Technical documentation

Pseudonimiserings & Anonimiserings



<https://ehealth.fgov.be/ehealthplatform/nl/service-pseudonimiserings-anonimiserings>

Innovation @ Smals Research

Smart Pseudonymisation

Conversion from citizen identifiers to pseudonyms

Format-Preserving Pseudonymisation

Retroactive protection of personal data in TEST & ACC of legacy applications



eHealth Blind Pseudonymisation

Proactive protection of personal data in applications
Privacy by Design



Oblivious Join

Non-trivial join & pseudonymise projects for research purposes
Distributed & no integration



Oblivious Join

- Problem statement
- Concept
- In practice
- Conclusion



Oblivious Join

- **Problem statement**
- Concept
- In practice
- Conclusion



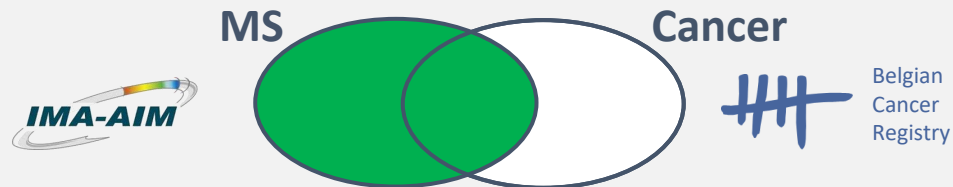
Use case

Deliberation nb. 20/020 – 14/01/2020

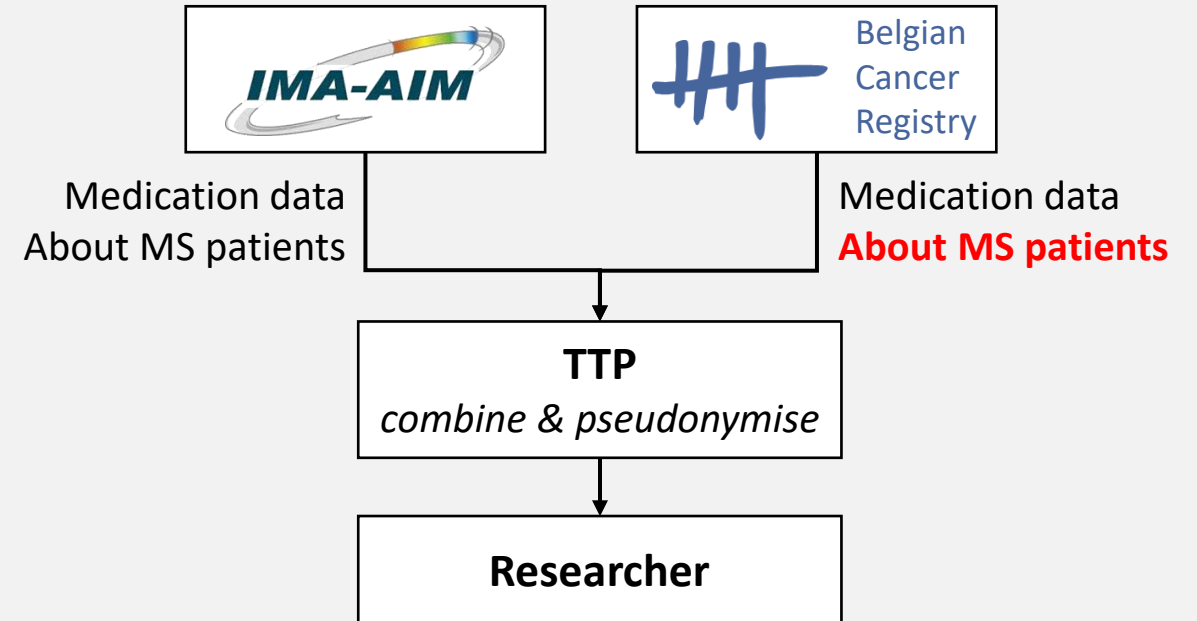
Research question

Do MS patients who take medications with the molecule teriflunomide or alemtuzumab have an increased cancer risk compared to MS patients treated with other medications?

Involved citizens



Naive flow



How can BCR deliver only records about MS patients without learning who has MS?

Current practice

Observations

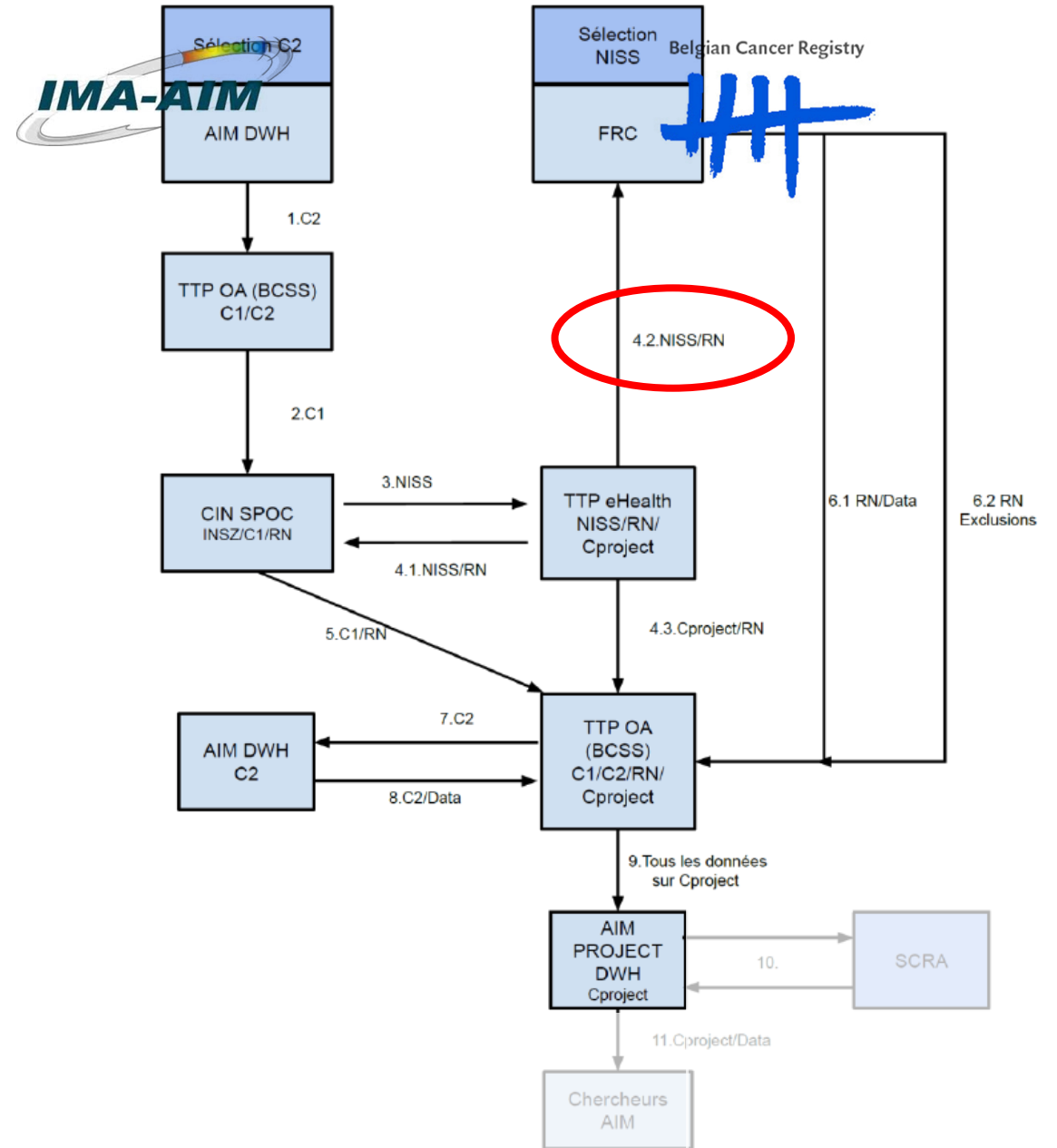
- ✗ Complex flow
- ✗ Expensive
- ✗ Bespoke
- ✗ Doesn't scale well
- ✗ Slow
- ✗ Security risk (data leakage)

Global issue: Heavy reliance on combination of trusted parties and strong legal regulations

Comments

"Lasts weeks, months, even years"

"Requires an exorbitant amount of resources"



Central question

Can we for specific research projects
combine and pseudonymise personal data
originating from different sources

Requirement 1

**Not all data sources able to independently
select pertinent records**

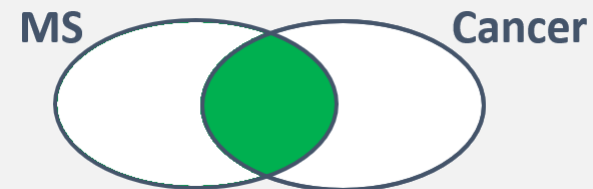
Requirement 2

**Cost-efficient, lightweight, non-intrusive way
(because each research question is different)?**

Focus: set intersection

How can we deliver pseudonymised
data of citizen that have MS and cancer

Extensible from there

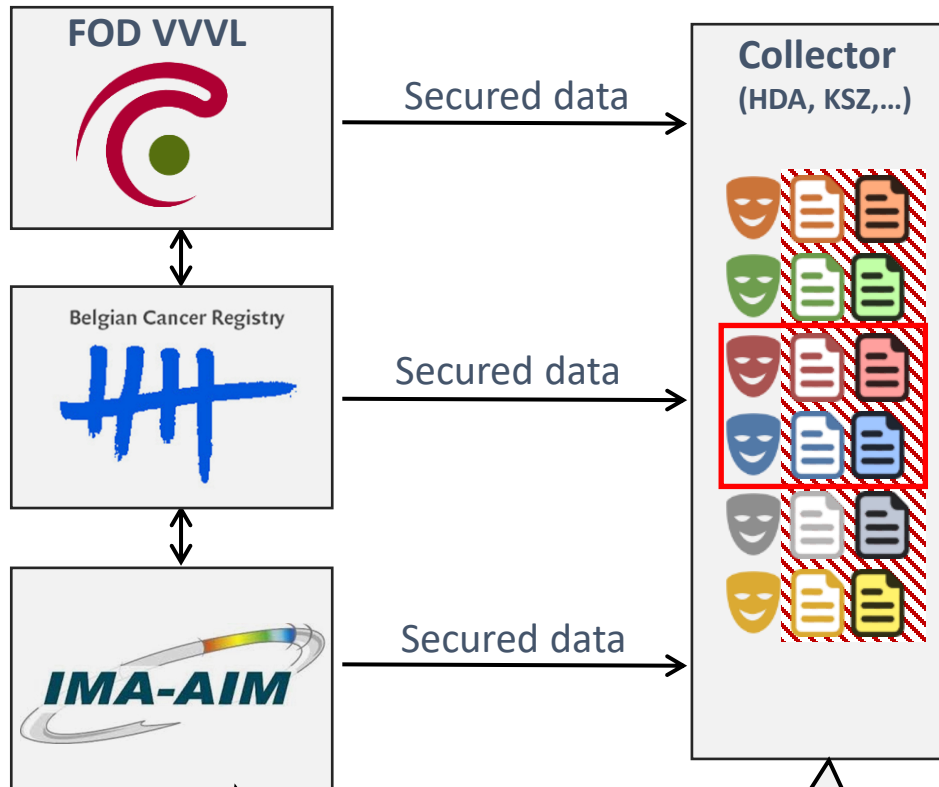


Oblivious Join

- Problem statement
- **Concept**
- In practice
- Conclusion



Concept



- ❖ Does not learn any new personal or statistical data
- ❖ Only see identifiers of their data

- ❖ Learns only minimum required pseudonymised personal data
- ❖ Learns high-level statistical data
E.g. number of citizens with cancer diagnosis
- ❖ Only sees pseudonyms

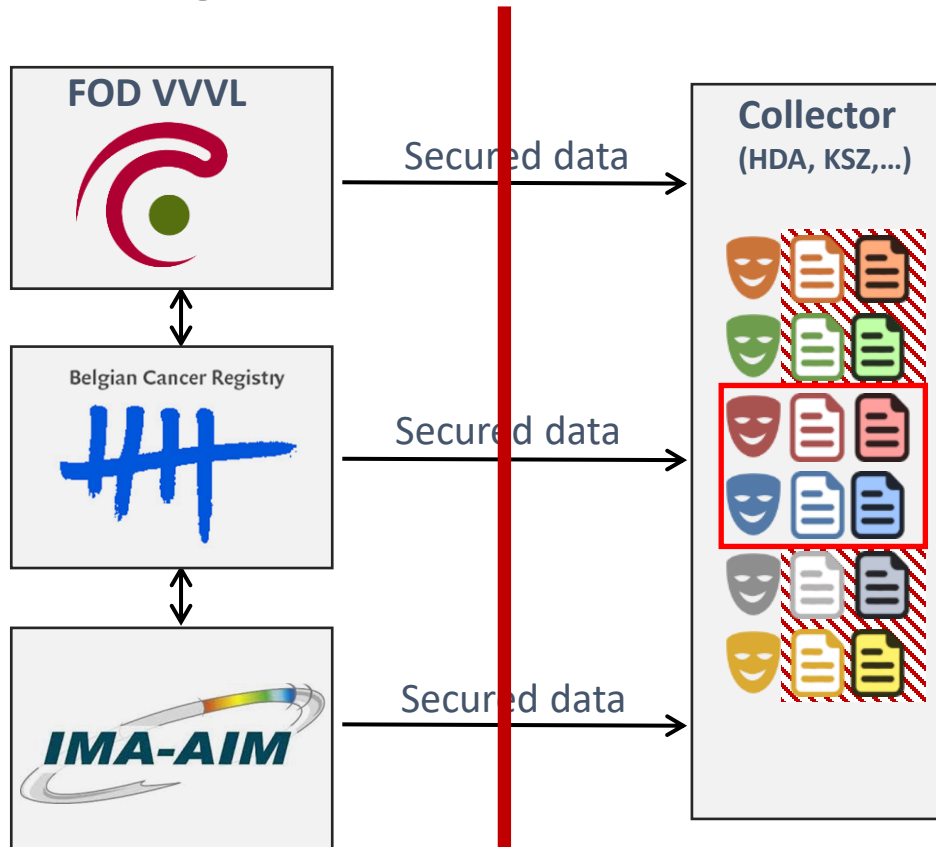
Properties

- ✓ Privacy-friendly & secure
- ✓ Distributed: no pseudon. service
- ✓ Harmonized & no integration
- ✓ Fast & cost-efficient

3 steps

1. Fully automated agreements between data sources (no human intervention)
2. Each data source sends all potentially relevant data encrypted & pseudonymised to collector
3. Thanks previous agreements (step 1) collector can only decrypt & combine pertinent records

Concept



**No collusion
between data
source and
collector**

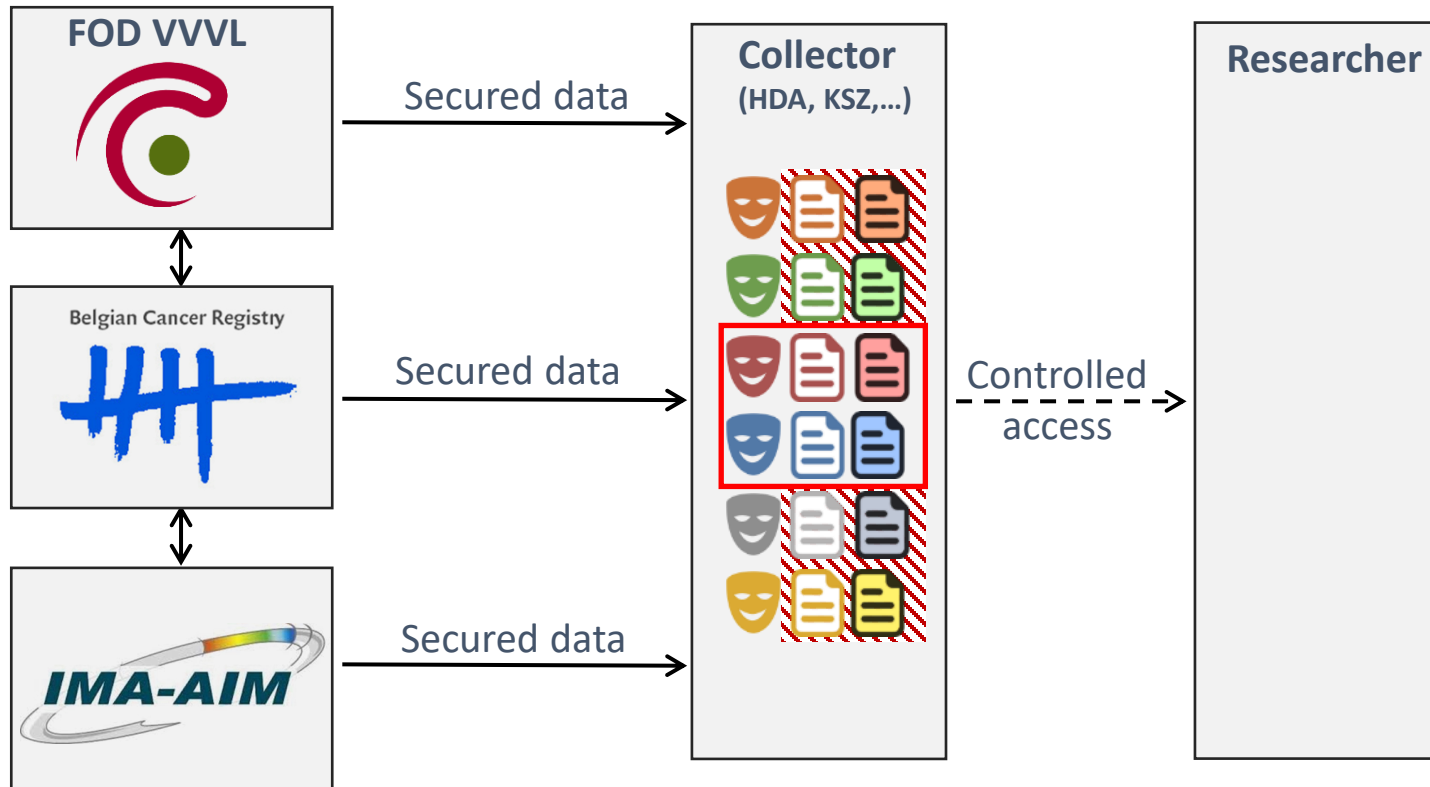
Properties

- ✓ Privacy-friendly & secure
- ✓ Distributed: no pseudon. service
- ✓ Harmonized & no integration
- ✓ Fast & cost-efficient

3 steps

1. Fully automated agreements between data sources (no human intervention)
2. Each data source sends all potentially relevant data encrypted & pseudonymised to collector
3. Thanks previous agreements (step 1) collector can only decrypt & combine pertinent records

Concept



Collector

TTP or shielded environment

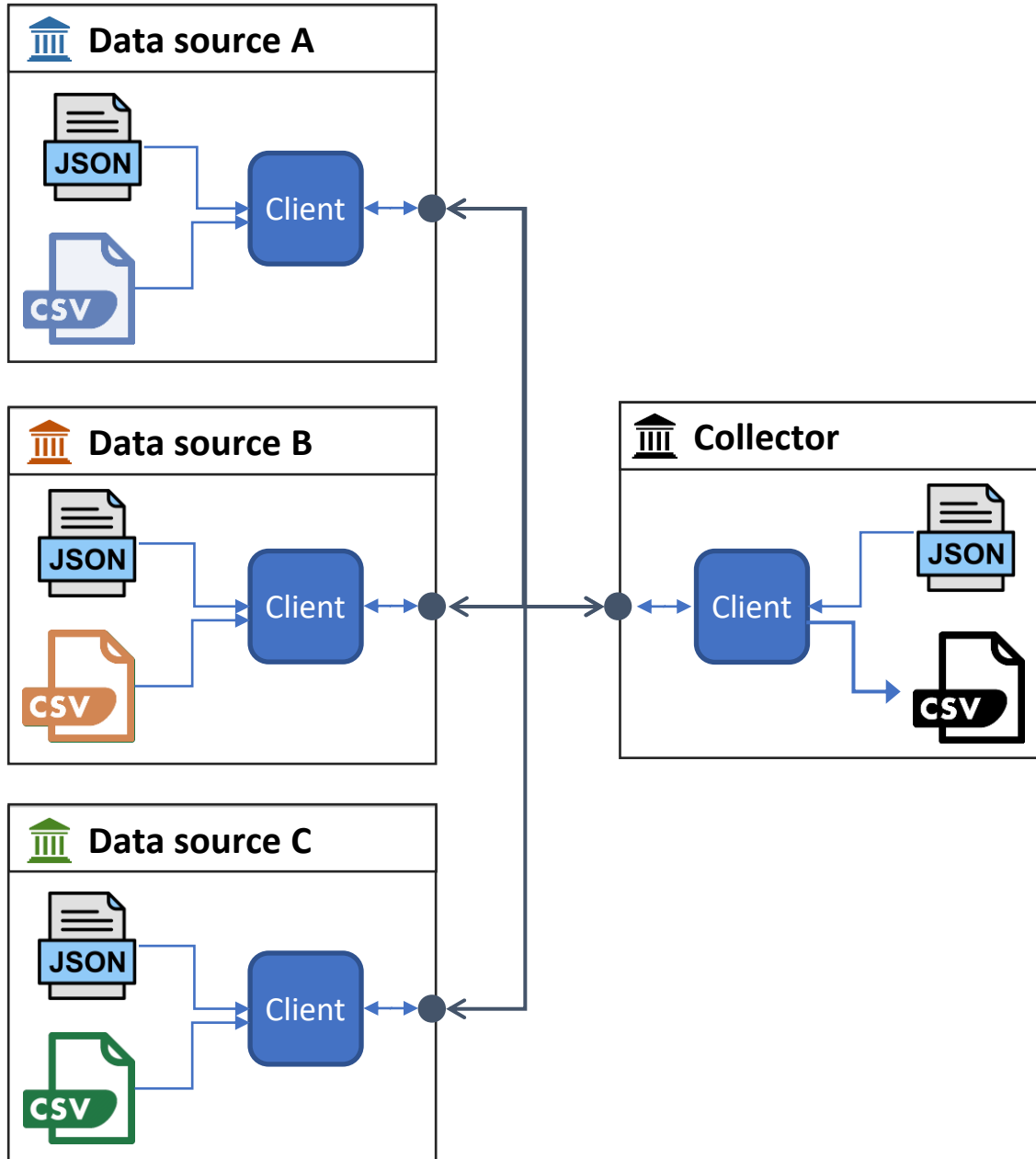
1. Deletes asap irrelevant ciphertexts
 2. Can do additional checks on the data
 3. Control access by researcher
- 'Trust' in collector limited

Oblivious Join

- Problem statement
- Concept
- **In practice**
- Conclusion



In practice



Client

- Java jar
- No integration required → non-intrusive, flexible
- All parties use same client (software)
- Command-line interface



Project description

- JSON file
- Created by coordinating party
- Contains all info required to execute protocol
- All parties use same project description



Input files

- CSV file
- Created by individual data source (out of scope)
- Contains all, potentially relevant, identified personal data



Output file

- CSV file
- Collector's output after protocol execution
- Contains minimal required combined & pseudonymised personal data

Test with fictional data

Extract input CSV

Data source 1 (IMA-AIM)

60.01.03-231.73	Teriflunomide
60.01.03-562.33	Alemtuzumab
60.01.03-697.92	Glatiramer acetate
60.01.04-606.56	Interferon beta
60.01.04-681.78	Dimethyl fumarate
60.01.05-045.05	Teriflunomide
60.01.05-186.58	Tysabri
60.01.05-617.15	Ocrelizumab
60.01.05-715.14	Alemtuzumab

200 000 records

E.g. Citizens with MS

Extract input CSV

Data source 2 (BCR)

60.01.03-782.07	Melanoma	3	G1
60.01.04-124.53	Colorectal	1	G3
60.01.04-345.26	Prostate	2	G2
60.01.04-562.03	Breast	2	G1
60.01.05-045.05	Lung	1	G3
60.01.05-893.30	Pancreas	4	G2
60.01.06-401.07	Breast	3	G1
60.01.06-696.03	Stomach	2	G1
60.01.07-203.78	Thyroid	1	G3

500 000 records

E.g. Citizens with cancer

Extract input CSV

Data source 3 (VVVL)

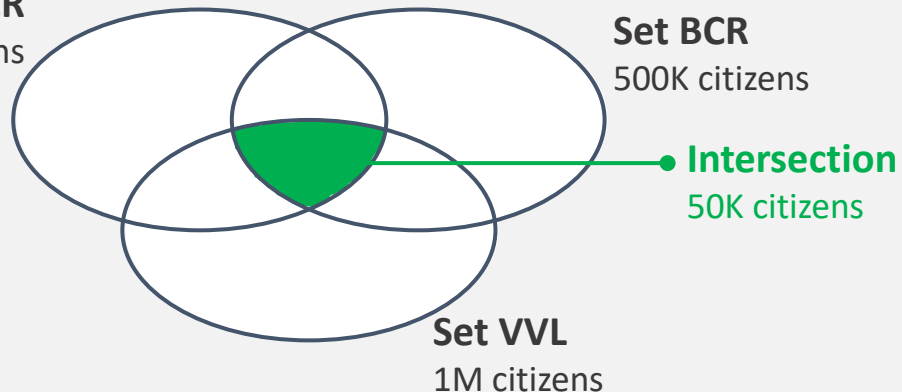
60.01.03-542.53	C
60.01.03-559.36	G
60.01.03-606.86	D
60.01.03-697.92	A
60.01.04-697.62	G
60.01.04-816.40	B
60.01.05-045.05	D
60.01.06-701.95	B
60.01.06-886.07	F

1 000 000 records

E.g. Citizens with high-risk profile

Set IMA-BCR

200K citizens



Performance test

Parameters

- MinNbRecords: 10
- 128 bit security

Infrastructure

- Data sources: 4 i9-7940x cores @ 3.10 GHz, 16GB RAM
- Collector: 2 i9-7940x cores @ 3.10 GHz , 16GB RAM

Results

- < 2 min calculations
- Excl. a few hundred MBs data transfer

Test with fictional data

Extract input CSV

Data source 1 (IMA-AIM)

60.01.03-231.73	Teriflunomide
60.01.03-562.33	Alemtuzumab
60.01.03-697.92	Glatiramer acetate
60.01.04-606.56	Interferon beta
60.01.04-681.78	Dimethyl fumarate
60.01.05-045.05	Teriflunomide
60.01.05-186.58	Tysabri
60.01.05-617.15	Ocrelizumab
60.01.05-715.14	Alemtuzumab

200 000 records

E.g. Citizens with MS

Extract input CSV

Data source 2 (BCR)

60.01.03-782.07	Melanoma	3	G1
60.01.04-124.53	Colorectal	1	G3
60.01.04-345.26	Prostate	2	G2
60.01.04-562.03	Breast	2	G1
60.01.05-045.05	Lung	1	G3
60.01.05-893.30	Pancreas	4	G2
60.01.06-401.07	Breast	3	G1
60.01.06-696.03	Stomach	2	G1
60.01.07-203.78	Thyroid	1	G3

500 000 records

E.g. Citizens with cancer

Extract input CSV

Data source 3 (VVVL)

60.01.03-542.53	C
60.01.03-559.36	G
60.01.03-606.86	D
60.01.03-697.92	A
60.01.04-697.62	G
60.01.04-816.40	B
60.01.05-045.05	D
60.01.06-701.95	B
60.01.06-886.07	F

1 000 000 records

E.g. Citizens with high-risk profile

Extract output CSV

Collector (KSZ)

50 000 records

99338454821...	Teriflunomide	Lung	3	G1	F
12056965607...	Alemtuzumab	Cervix uteri	2	G2	B
15380767762...	Daclizumab	Pancreas	1	G2	A
15380767762...	Teriflunomide	Lung	1	G3	D
31309444464...	Ocrelizumab	Stomach	3	G1	C
99921347021...	Dimethyl fumarate	Breast	2	G2	H
69025938558...	Ofatumumab	Prostate	3	G3	A
38469942453...	Alemtuzumab	Melanoma	4	G1	E
18048091119...	Aubagio	Prostate	3	G3	D

Who sees what?

- ❖ Data sources only see identifiers
- ❖ Collector only sees pseudonyms
- ❖ No pseudonymisation service

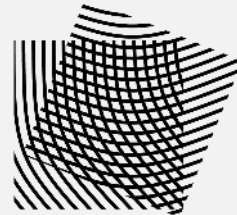
Oblivious Join

- Problem statement
- Concept
- In practice
- **Conclusion**



Collaboration universities

Expert paper



COSIC

KU LEUVEN

KU LEUVEN

UNIVERSITY OF
WATERLOO



*Goal is an A-tier
conference*

Multidisciplinary paper



KU LEUVEN

*Public Governance and
Emerging Technologies –
Values, Trust, and
Compliance by Design*



Utrecht
University

Evaluation

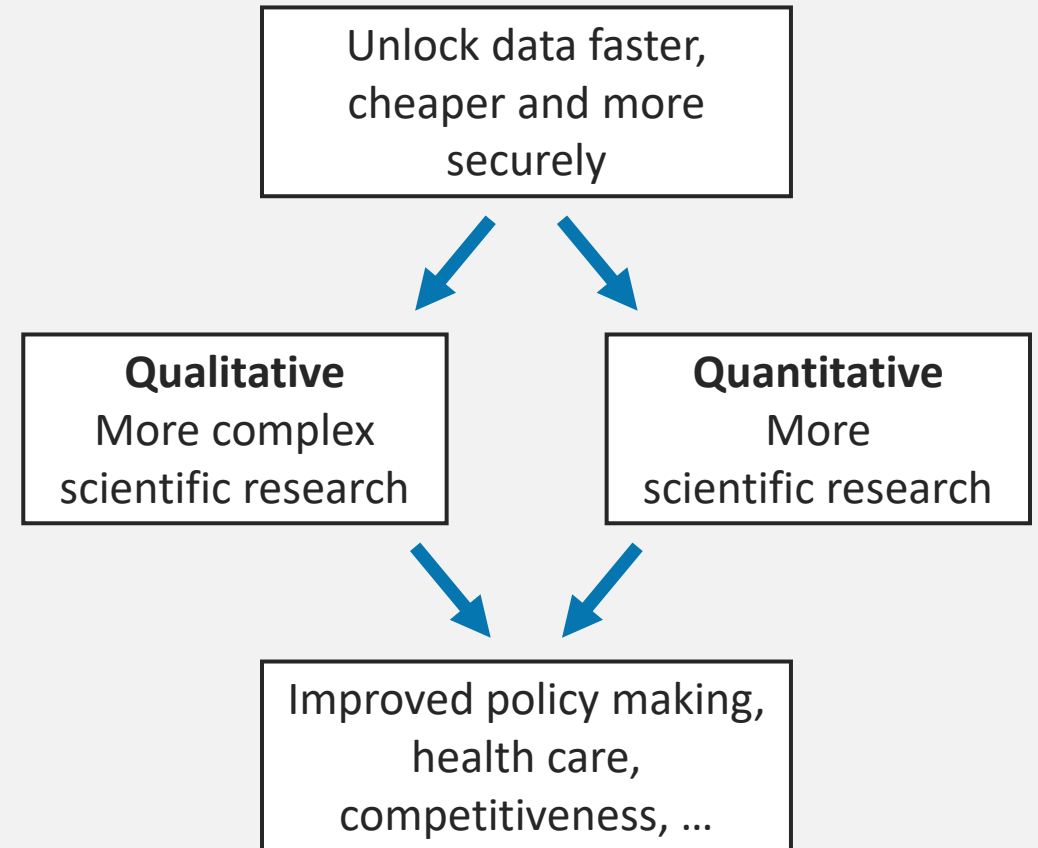
Advantages

- ✓ Answer on business need
- ✓ Privacy-friendly & secure
- ✓ Distributed (no pseudonymisation service)
- ✓ Harmonized & no integration
- ✓ Fast & cost-efficient
- ✓ Formal academic validation

Challenges

- ⚠ Only passive interest
- ⚠ Still in research phase
- ⚠ Higher development complexity (but lower infra)
- ⚠ Extensions required

Opportunities



Wrapping up

Innovation @ Smals Research

Smart Pseudonymisation

Conversion from citizen identifiers to pseudonyms

Format-Preserving Pseudonymisation

Retroactive protection of personal data in TEST & ACC of legacy applications



eHealth Blind Pseudonymisation

Proactive protection of personal data in applications
Privacy by Design



Oblivious Join

Non-trivial join & pseudonymise projects for research purposes
Distributed & no integration



No one-fits-all solution for pseudonymisation

Do you see use cases where **pseudonymisation** seems promising?

Format-Preserving

Ps

eHealth Blind

Ps

Proa

Oblivious

Unknown

Privacy technology

Perfect fit your use case



Thanks for your attention

If you have any questions, do not hesitate to contact us!

 kristof.verslype@smals.be

 +32(0)2 7875376

 [linkedin.com/in/verslype](https://www.linkedin.com/in/verslype)



www.smals.be

www.smalsresearch.be

www.cryptov.net

Images



Judy Dean
Creative Commons
<https://flickr.com/photos/peterscherub/53152339550/>



estorde
Creative commons
<https://flickr.com/photos/estorde/4572006561>



Pixabay
Pixabay License
<https://pixabay.com/fr/photos/femme-les-yeux-masquer-carnaval-411494/>



Aris Gionis
Creative Commons
Flickr



Daniel Bruce
Creative Commons
<https://iconscout.com/free-icon/mask-126>