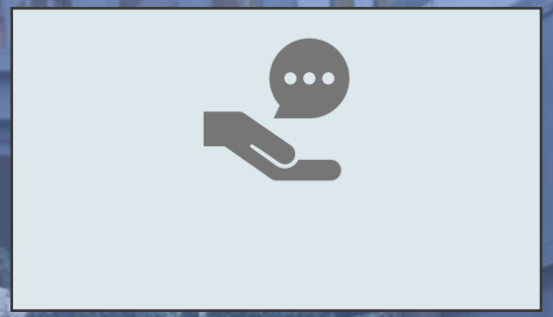
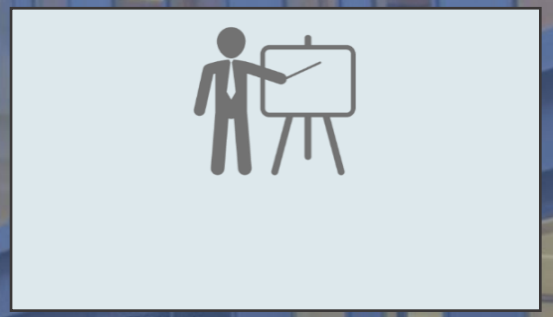
Privacy in Practice

Kristof Verslype
Cryptographer @ Smals Research

6 June 2024

Smals
ICT for society





Vector Databases

Causal AI

Customer Support Optimization

AI & Cybersecurity

Securing Kubernetes Containers

Semantic Search

Crypto-agility

Data Ingestion

ATMS

Workplace “co-pilot”

HSM as a Service

PII Filtering

Event Driven Architecture in practice

Graph ML

Rules and Code Generation

ReUse Format-preserving encryption

Generative AI Agents

Verifiable Credentials & EU Digital Wallet

Smals
ICT for society

Smals
ICT for society

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

ICT for society

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

Smals ICT for society

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

Smals ICT for society

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

World Quality Report, 2020

Smals ICT for society

Data breaches from non-prod environments

UBER

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

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.

LastPass

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

Smals ICT for society

❖ ***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

❖ Encouraged by GDPR to protect personal data

❖ Some rules by GDPR more relaxed

❖ Could help become more compliant

❖ Still Personal data

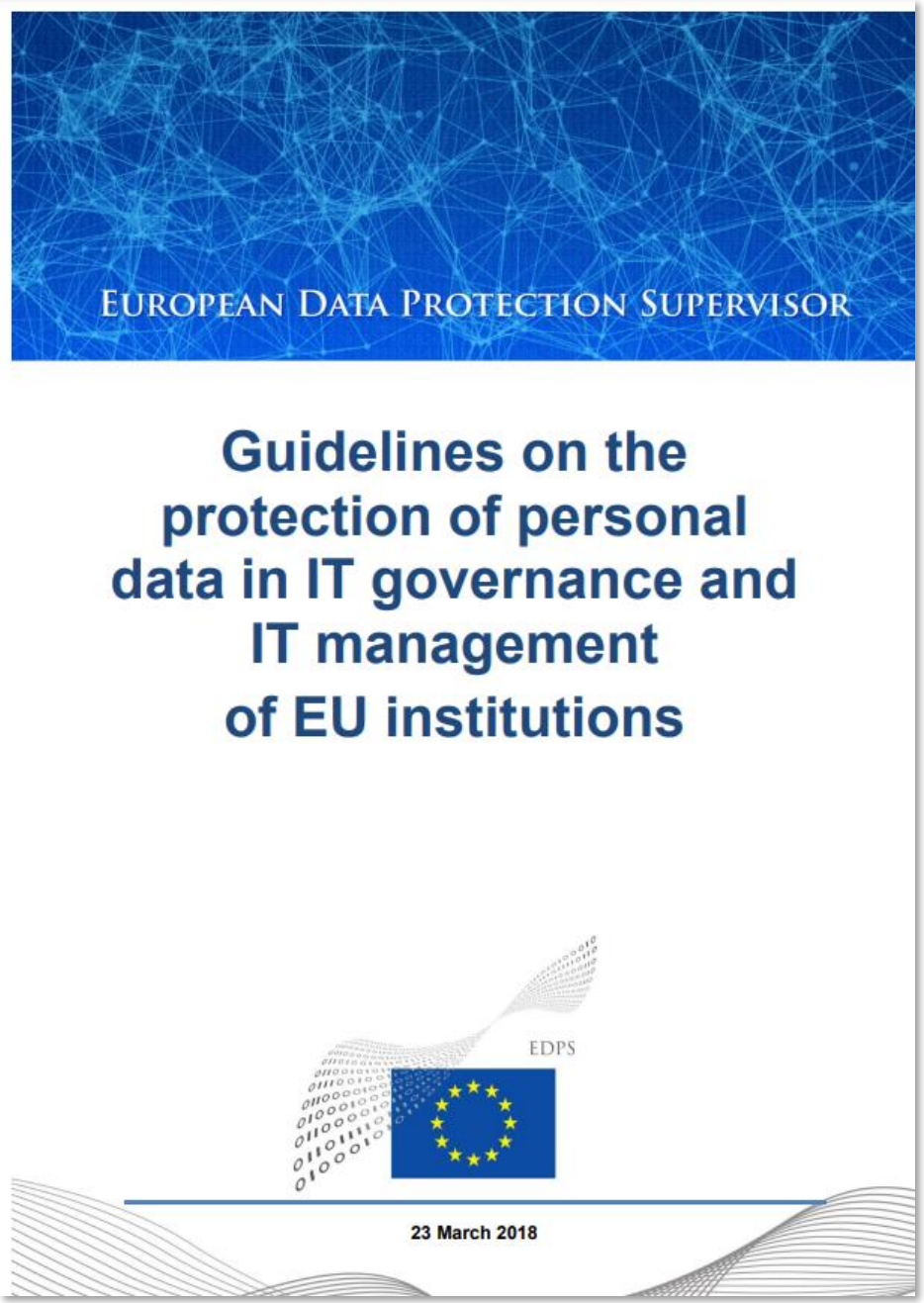
GDPR, Art 32.

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) …

Smals ICT for society



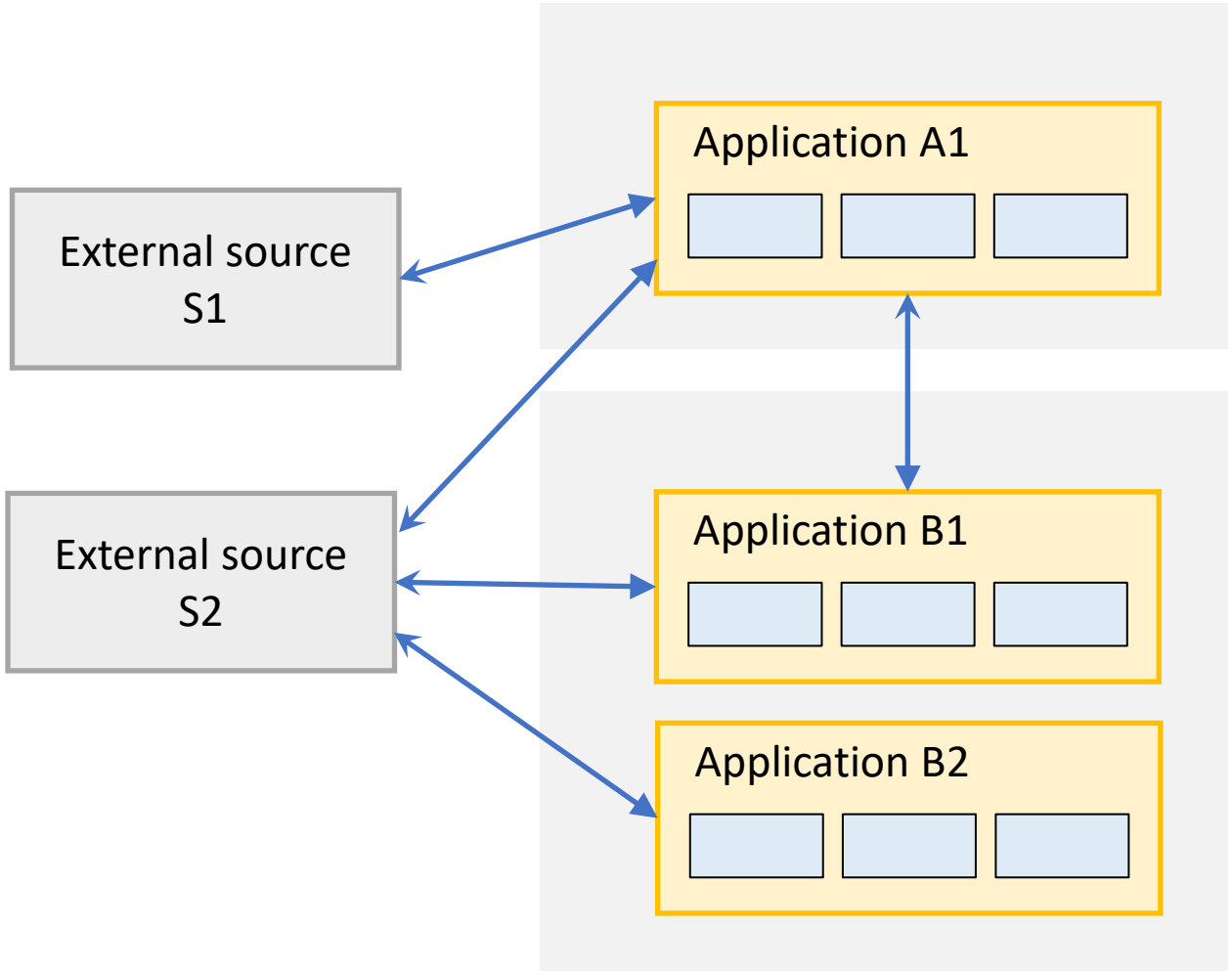
80 ***In the testing phase, sampling of real personal data should be avoided****, as such data cannot be used for purposes for which it was not collected and using it in testing environments may result in making personal data available to unauthorised individuals.*

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.*

Smals ICT for society

https://edps.europa.eu/data-protection/our-work/publications/guidelines/it-governance-and-it-management_en



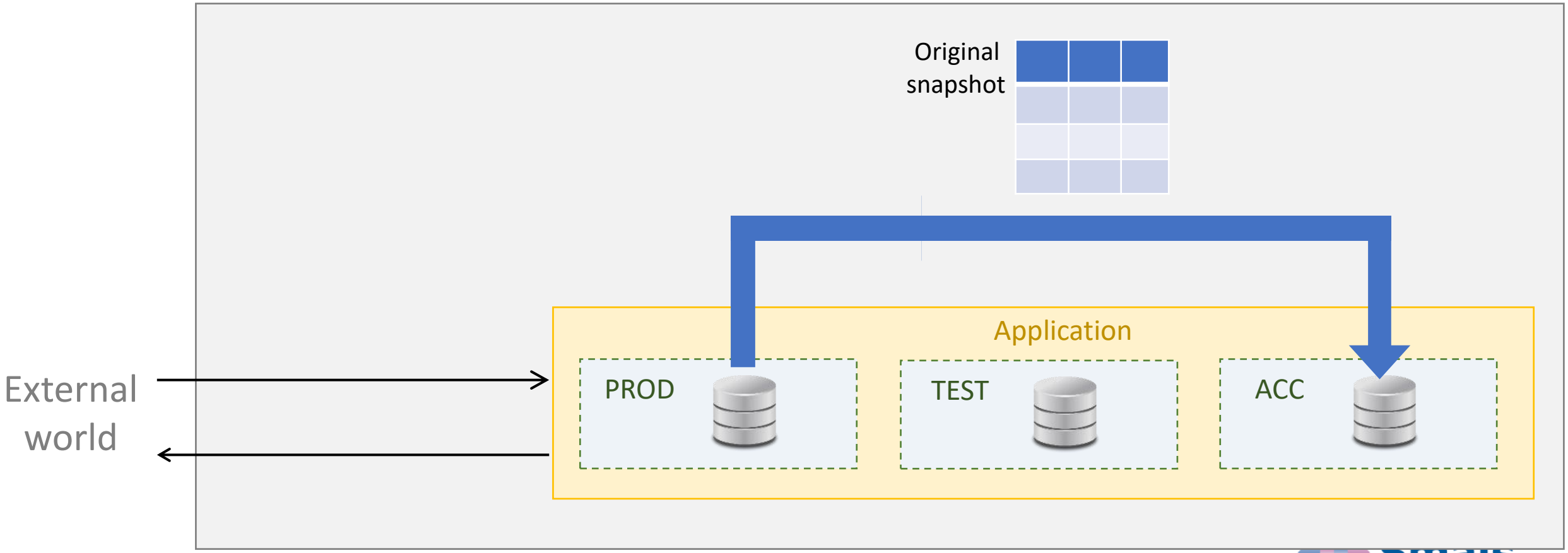
Completely fictional data — an option

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

Smals ICT for society

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

Smals ICT for society



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

Replace structured identifier by format-preserving pseudonym

-
-

Column-wise permutation

-
-

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

Pseudonymise

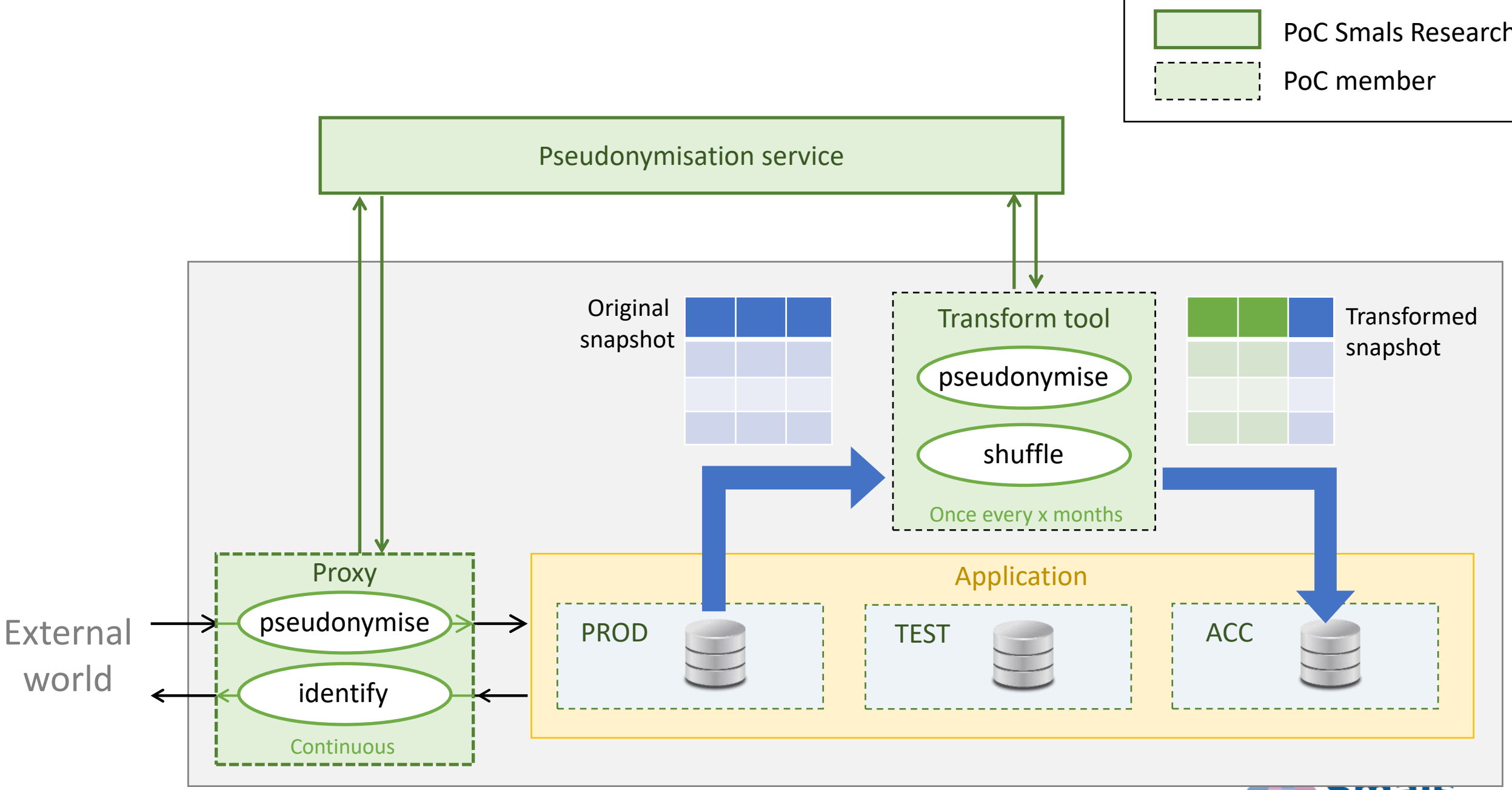
Shuffle

Shuffle

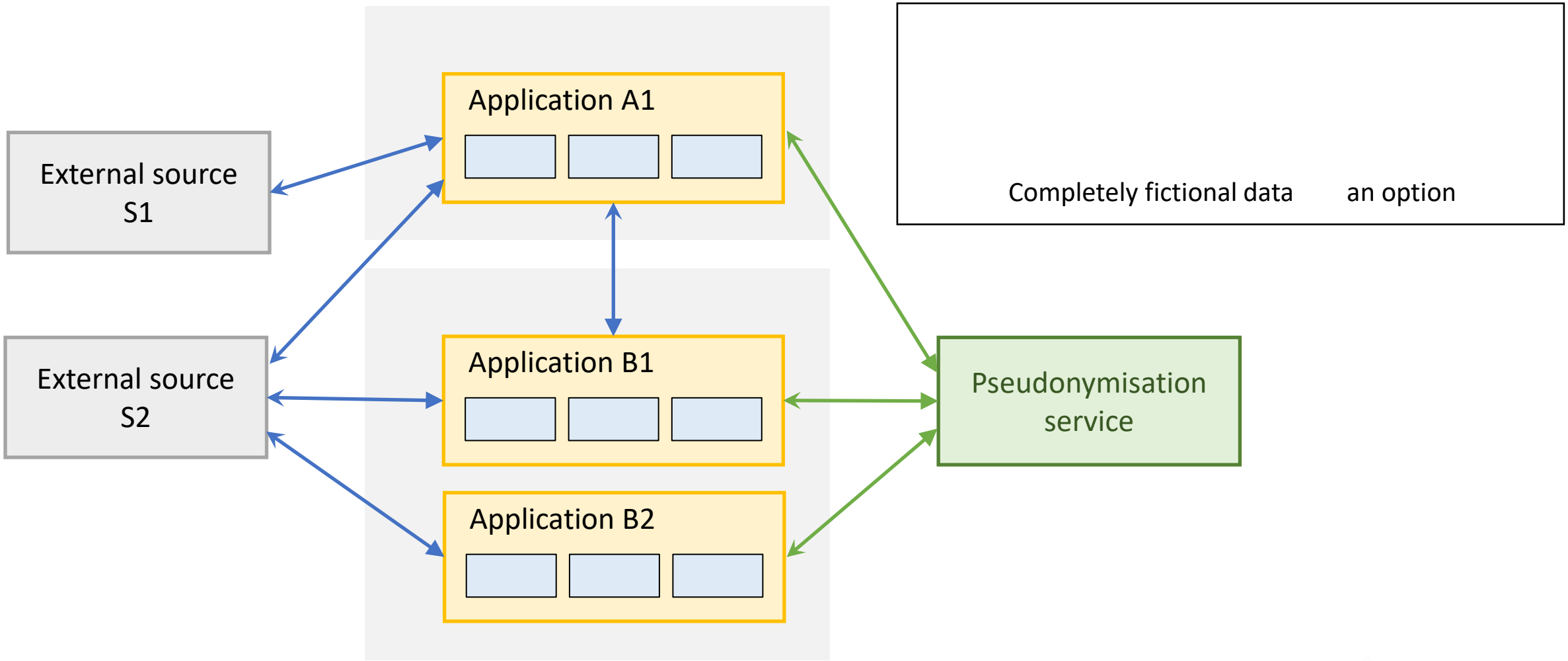
30.03.30-213.23	Melchior	Beernaert
66.08.15-286.27	Baltazar	de Brouckère
22.51.14-602.20	Kasper	Rogier

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

Smals ICT for society



Smals ICT for society



Smals ICT for society

Pseudo service maintains per environment

Pseudon. service – Instance 1

84.05.20-104.55	18.32.08-902.42
76.01.13-206.75	30.02.06-981.94
37.09.11-002.47	72.43.27-109.21
50.11.12-213.85	58.28.16-291.62

18.32.08-902.42	30.43.30-213.41
30.02.06-981.94	66.08.15-286.27
72.43.27-109.21	22.51.14-602.20

79.27.28-621.96	01.28.06-013.53
93.26.17-802.47	50.49.16-167.67

Pseudon. service – Instance 2

84.05.20-104.55	18.32.08-902.42
76.01.13-206.75	30.02.06-981.94
37.09.11-002.47	72.43.27-109.21
50.11.12-213.85	58.28.16-291.62

18.32.08-902.42	30.43.30-2
30.02.06-981.94	66.08.15-2
72.43.27-109.21	22.51.14-6

79.27.28-621.96	01.28.06-0
93.26.17-802.47	50.49.16-1

Pseudon. service – Instance 3

84.05.20-104.55	18.32.08-902.42
76.01.13-206.75	30.02.06-981.94
37.09.11-002.47	72.43.27-109.21
50.11.12-213.85	58.28.16-291.62

- Synchronization between instances
- Storage (backup, expensive)

More data is harder to secure

Pseudonymisation service maintains per environment

Key size is 32 bytes

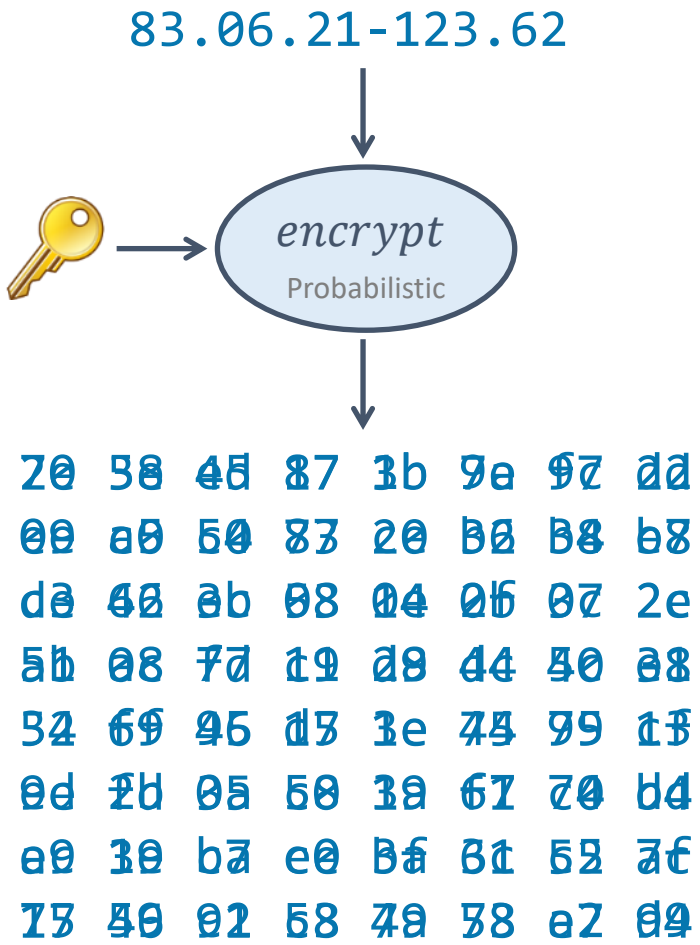
```
a4 71 c3 e0 9f 79 b3 64
3f 89 42 24 16 a1 9d 1e
6f f0 f6 4e 87 ea 34 03
68 a4 4e ee c0 14 dd 2d
```

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

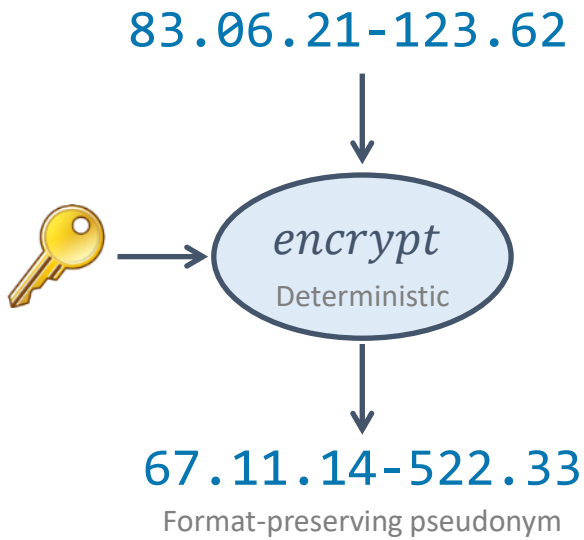
✓
- Minimal storage required
- Synchronization hugely simplified
- All keys derivable from single master key

Smals ICT for society

Traditional encryption



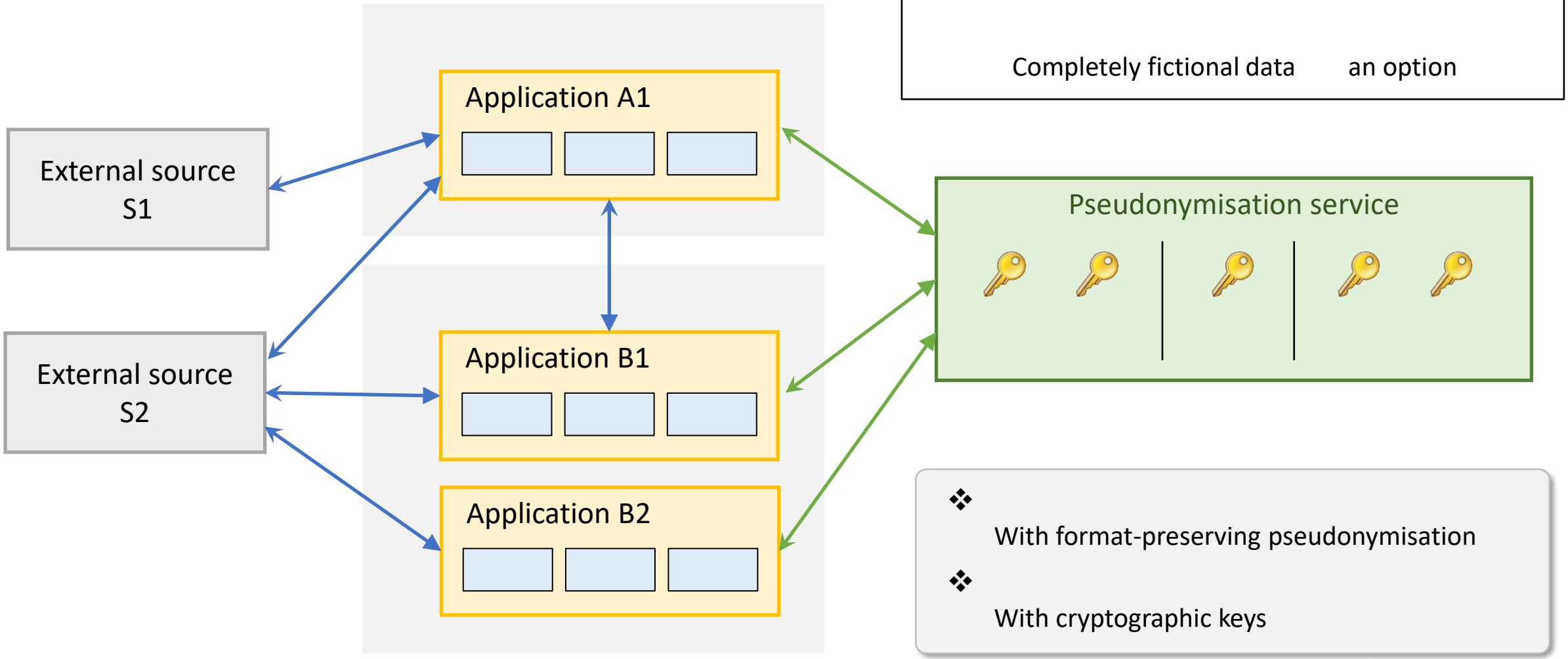
Format-preserving encryption



- 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)

Smals ICT for society

Completely fictional data an option



- With format-preserving pseudonymisation
- With cryptographic keys

Smals ICT for society

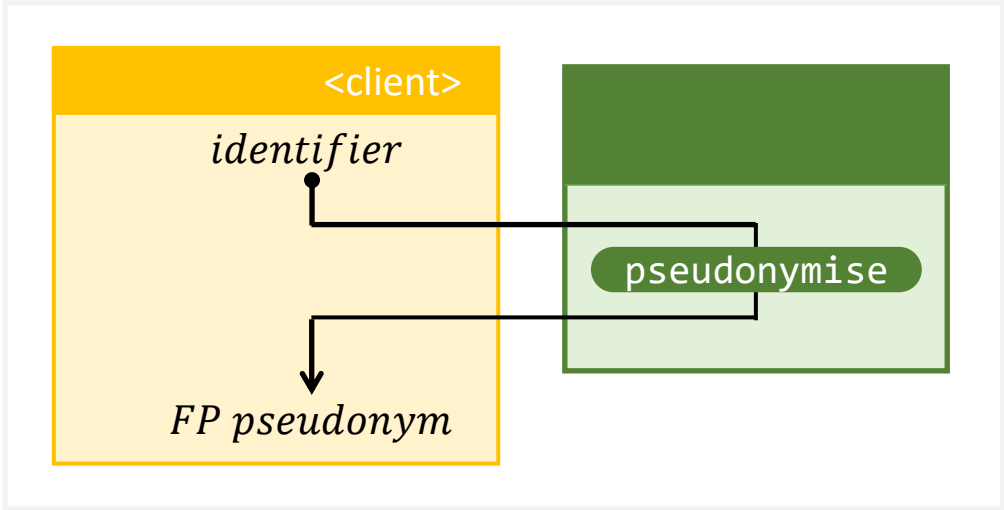
- Problem statement
- Concept & PoC
- **Experimental service**
- Conclusion

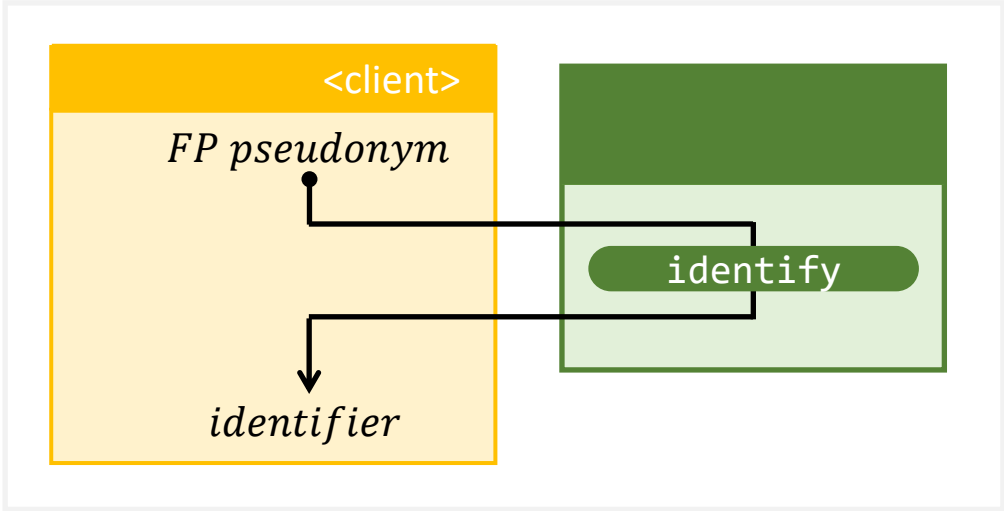
Smals ICT for society

Built by Smals Research

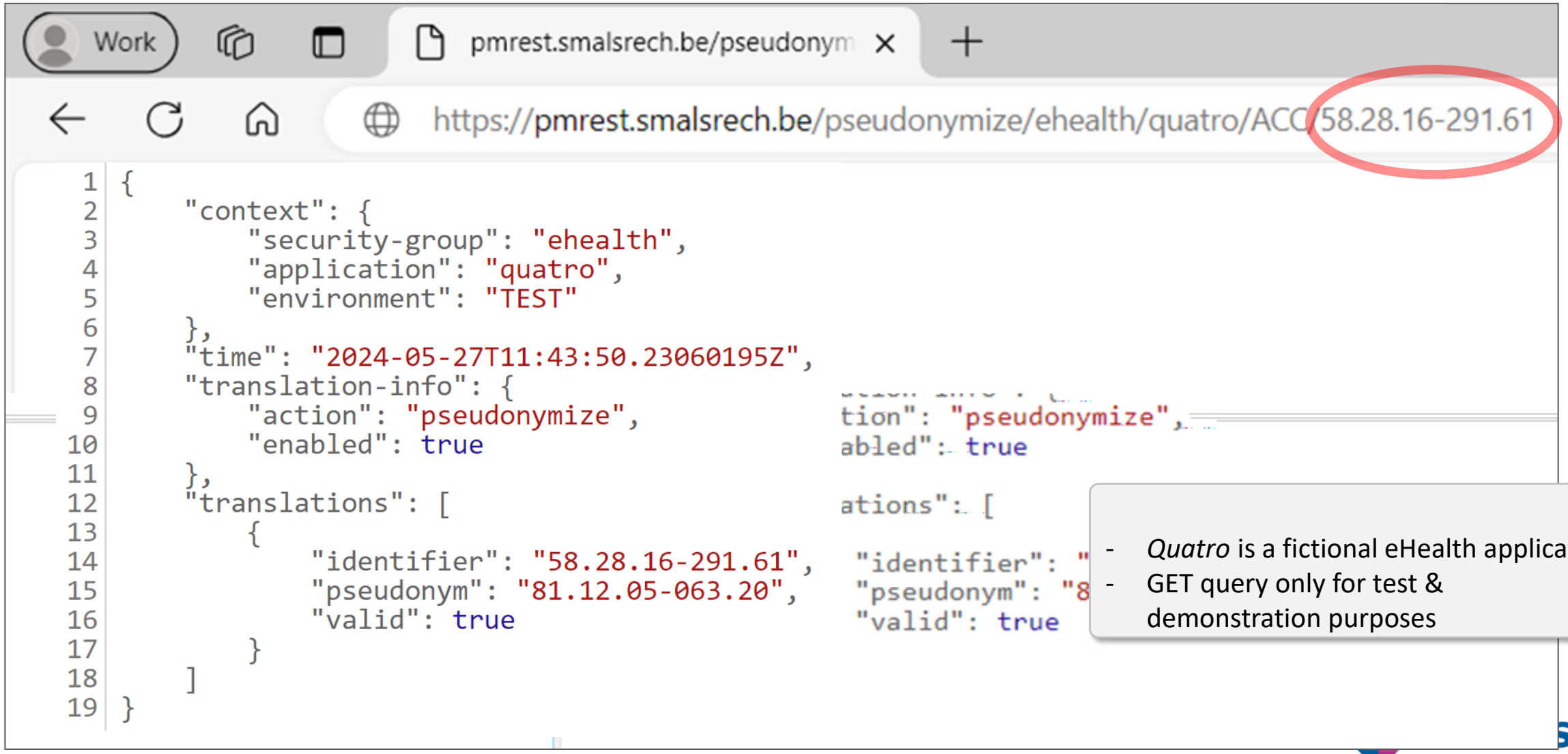
- ✓ *Pseudonymise & Identify*
- ✓ *GET and POST*
- ✓ *Also batch (POST only)*

- ✓
  - Currently, only RRN, BIS, INSZ
  - KBO number, Bank account numbers, …
- ✓

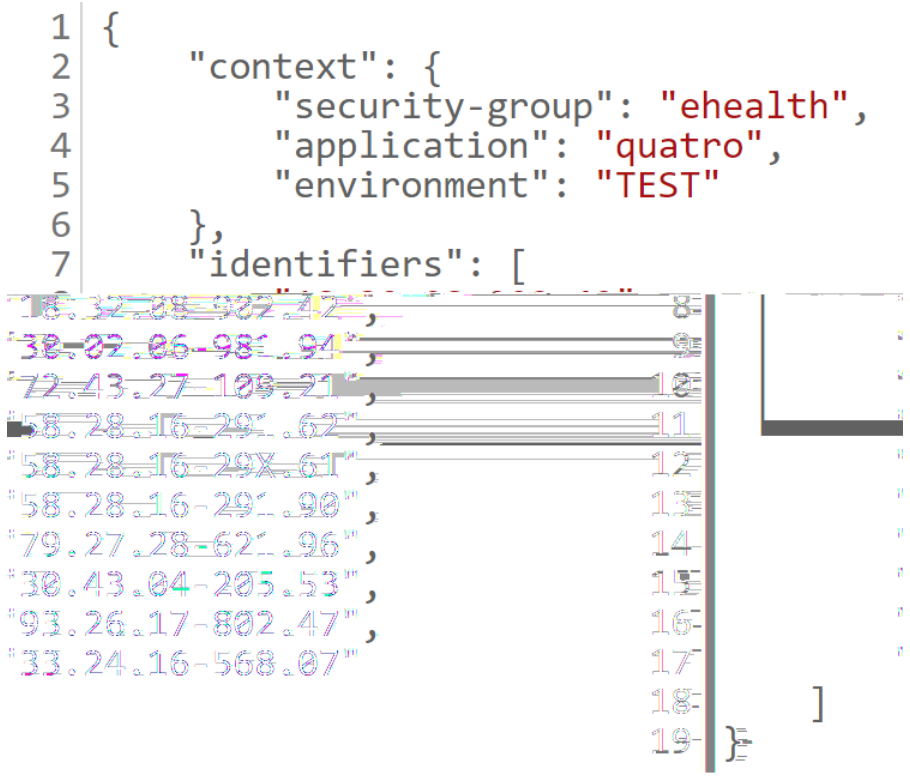




Smals ICT for society



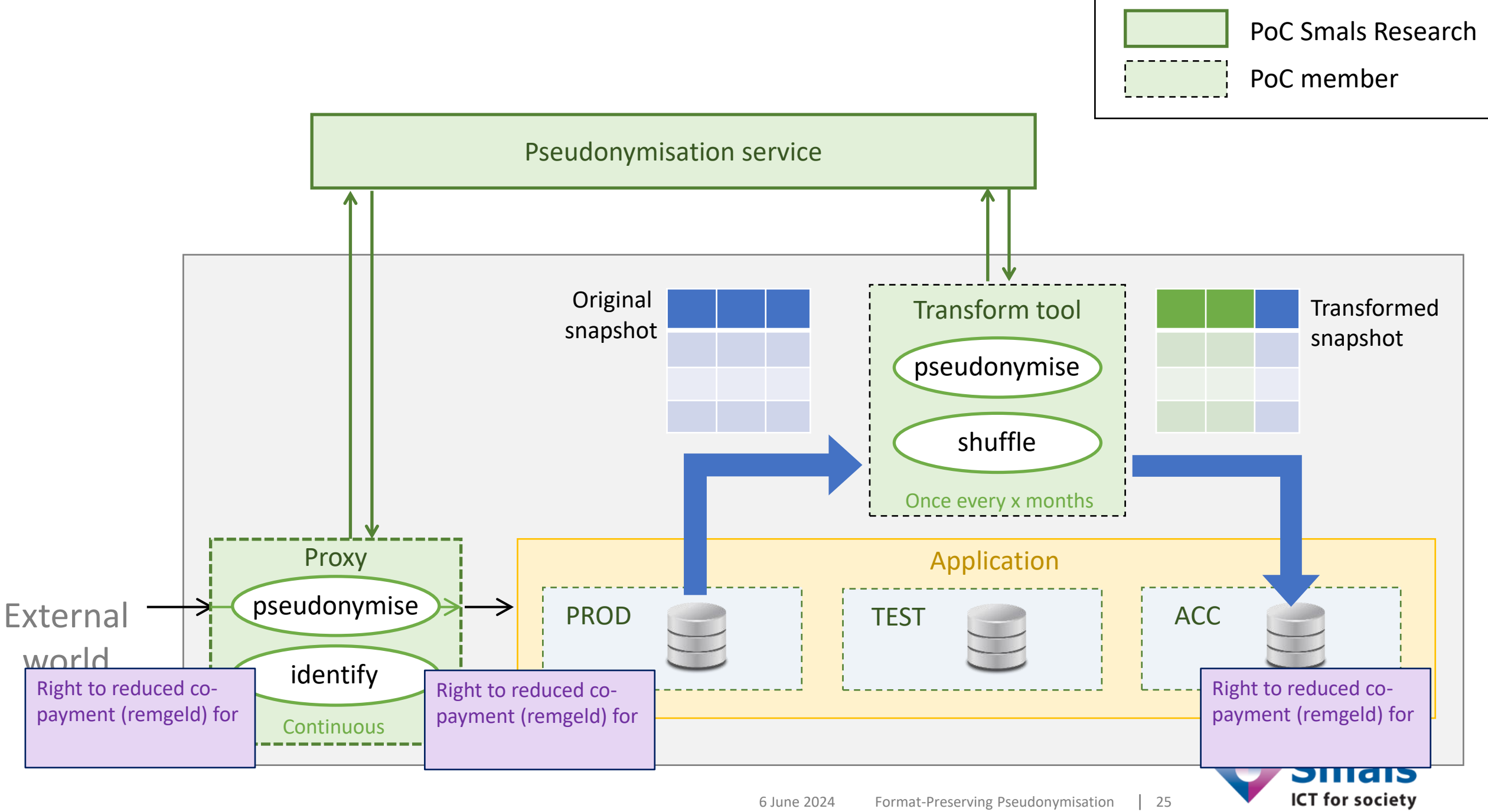
- *Quatro* is a fictional eHealth application
- GET query only for test & demonstration purposes



```
{
    "context": {
        "security-group": "ehealth",
        "application": "quatro",
        "environment": "TEST"
    },
    "identifiers": [
        "18.32.08-902.42",
        "30.02.06-981.94",
        "72.43.27-109.21",
        "58.28.16-29[illegible]",
        "58.28.16-29[illegible]",
        "58.28.16-29[illegible]",
        "58.28.16-29[illegible]",
        "79.27.28-62[illegible]",
        "30.43.04-205.53",
        "93.26.17-802.47",
        "33.24.16-568.07"
    ]
}
```

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

```
{
    "context": {
        "security-group": "ehealth",
        "application": "quatro",
        "environment": "TEST"
    },
    "time": "2024-01-08T08:20:39.128207895Z",
    "translation-info": {
        "action": "pseudonymize",
        "enabled": true
    },
    "translations": [
        {
            "identifier": "18.32.08-902.42",
            "pseudonym": "30.43.30-213.41",
            "valid": true
        },
        {
            "identifier": "30.02.06-981.94",
            "pseudonym": "66.08.15-286.27",
            "valid": true
        },
        {
            "identifier": "72.43.27-109.21",
            "pseudonym": "22.51.14-602.20",
            "valid": true
        },
        {
            "identifier": "58.28.16-291.62",
            "pseudonym": "null",
            "valid": false,
            "error": "checksum"
        }
```



Smals ICT for society

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

Smals ICT for society

-

- Citizen known under different pseudonym in each environment

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

- Relatively simple service
-

- Advanced PoC (Extensibility, unit tests, error handling, Smals standards, …)
- running
- No logging, access control, …

Gegevensbescherming m.b.v. structuurbehoudende pseudonimisatie van rijksregisternummers

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

Conversion from citizen identifiers to pseudonyms

Format-Preserving Pseudonymisation

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

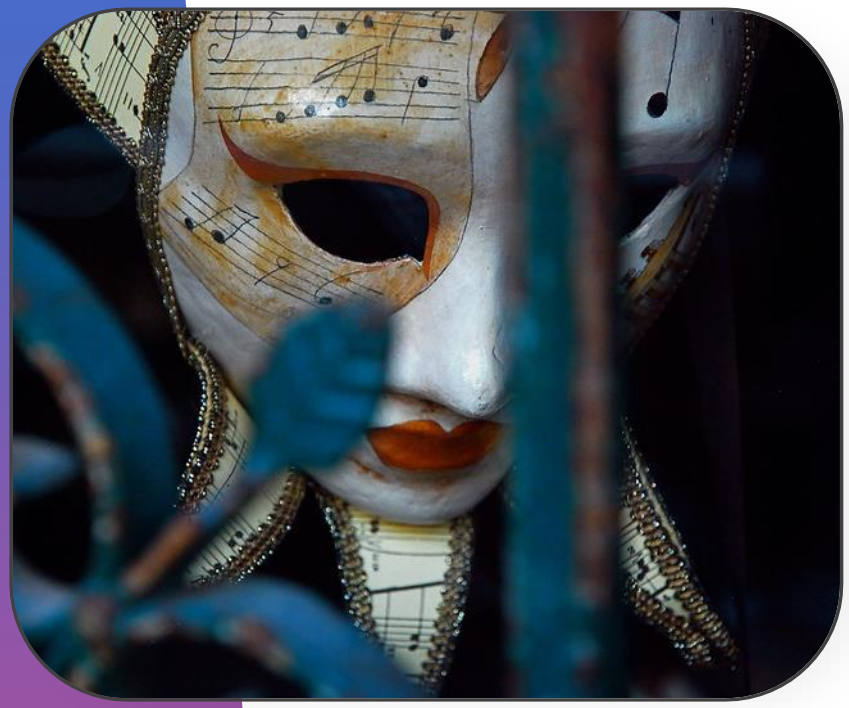
eHealth Blind Pseudonymisation

Proactive protection of personal data in applications
Privacy by Design

Oblivious Join

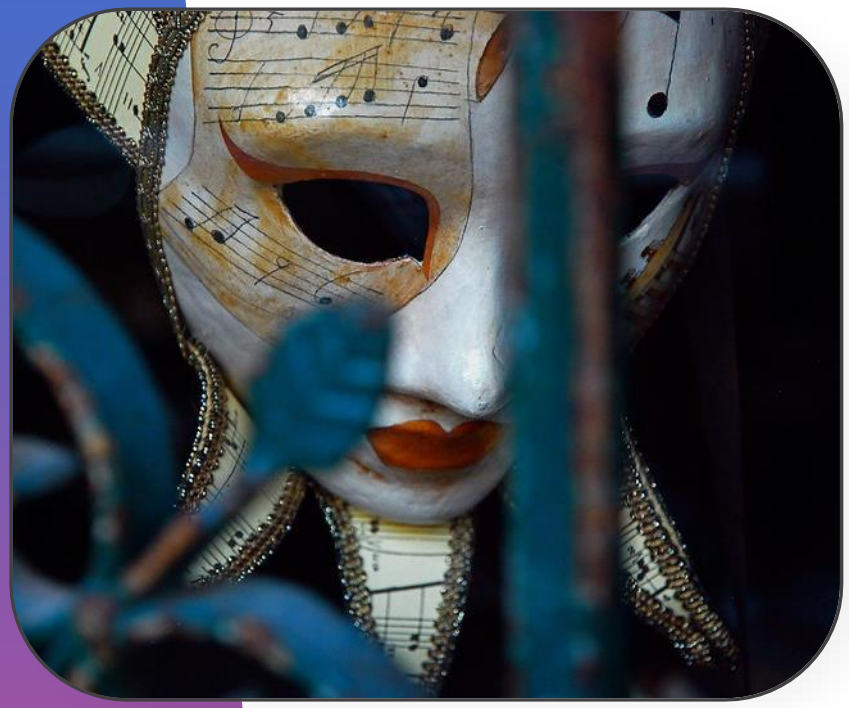
Non-trivial join & pseudonymise projects for research purposes
Distributed & no integration

ICT for society



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

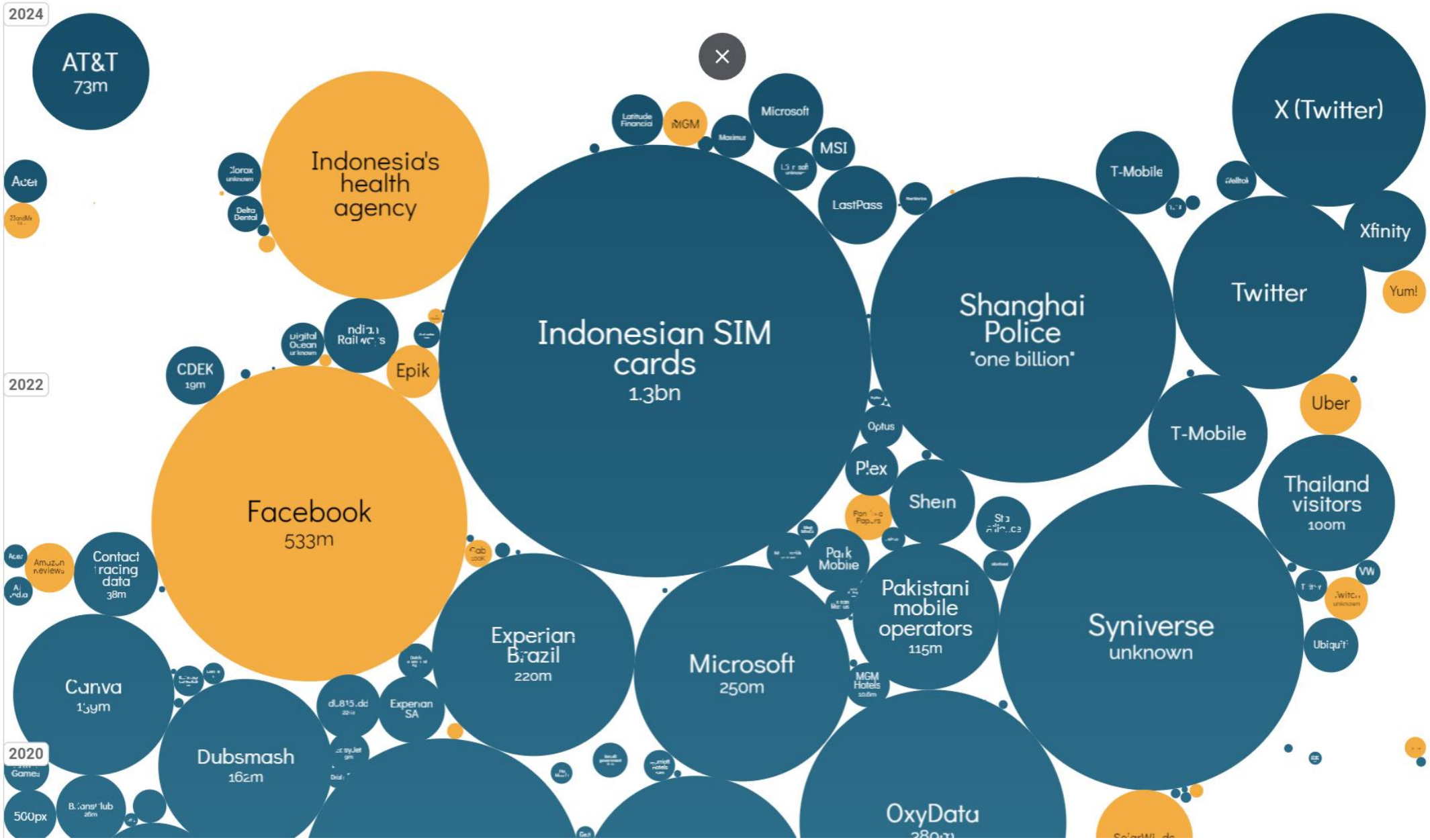
Smals ICT for society



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

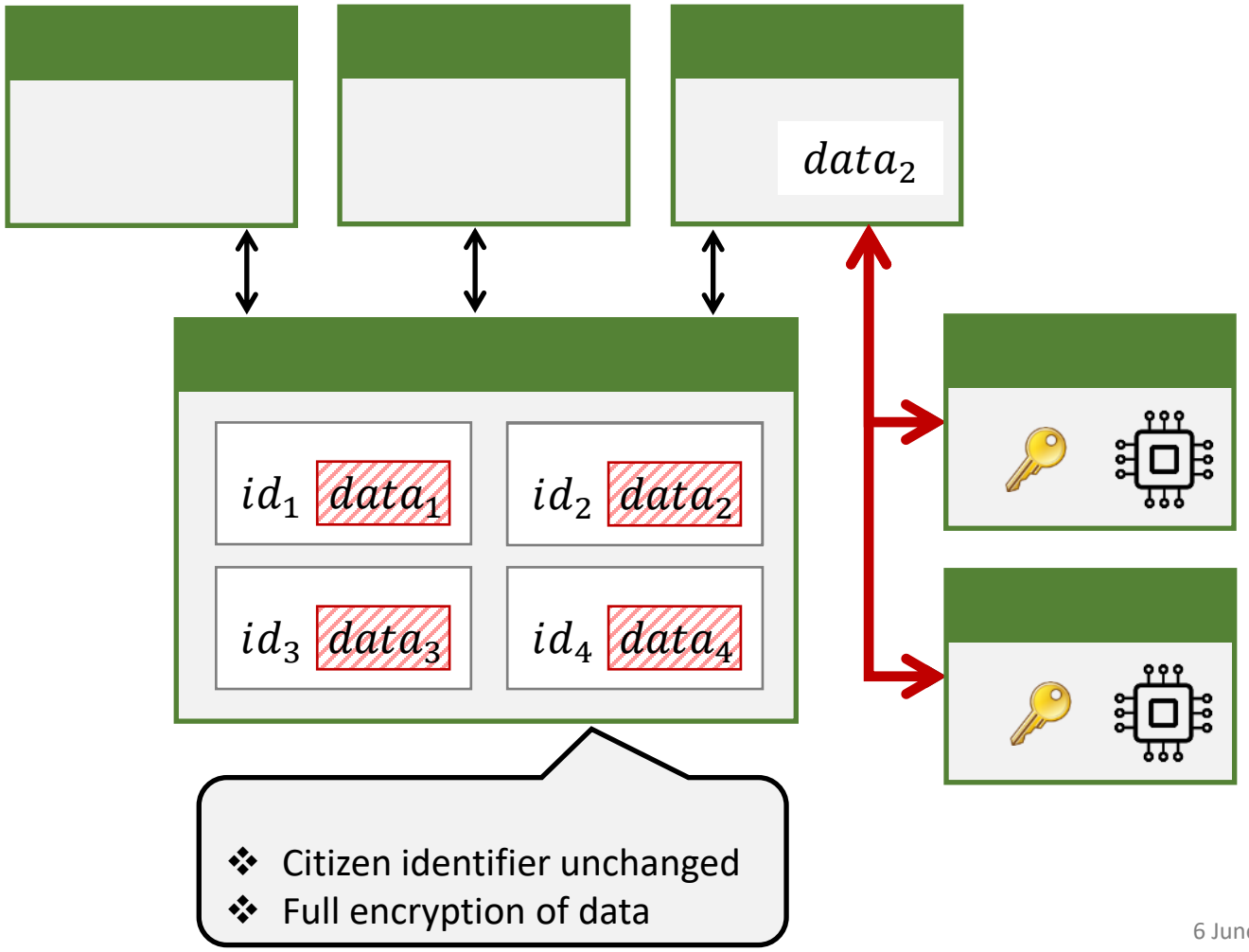
Smals ICT for society

World's Biggest Data Breaches & Hacks



mals for society

https://informationisbeautiful.net/visualizations/worlds-biggest-data-breaches-hacks/



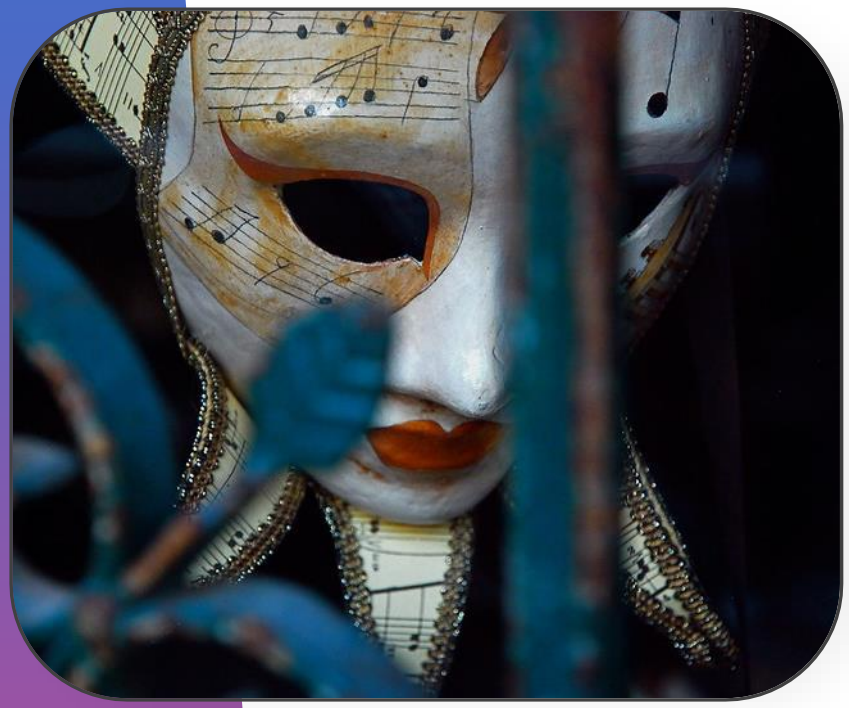
❖ Privacy should be taken into account when designing and building products and services

❖ Entity managing encryption keys should not have access to protected data (and vice versa)

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

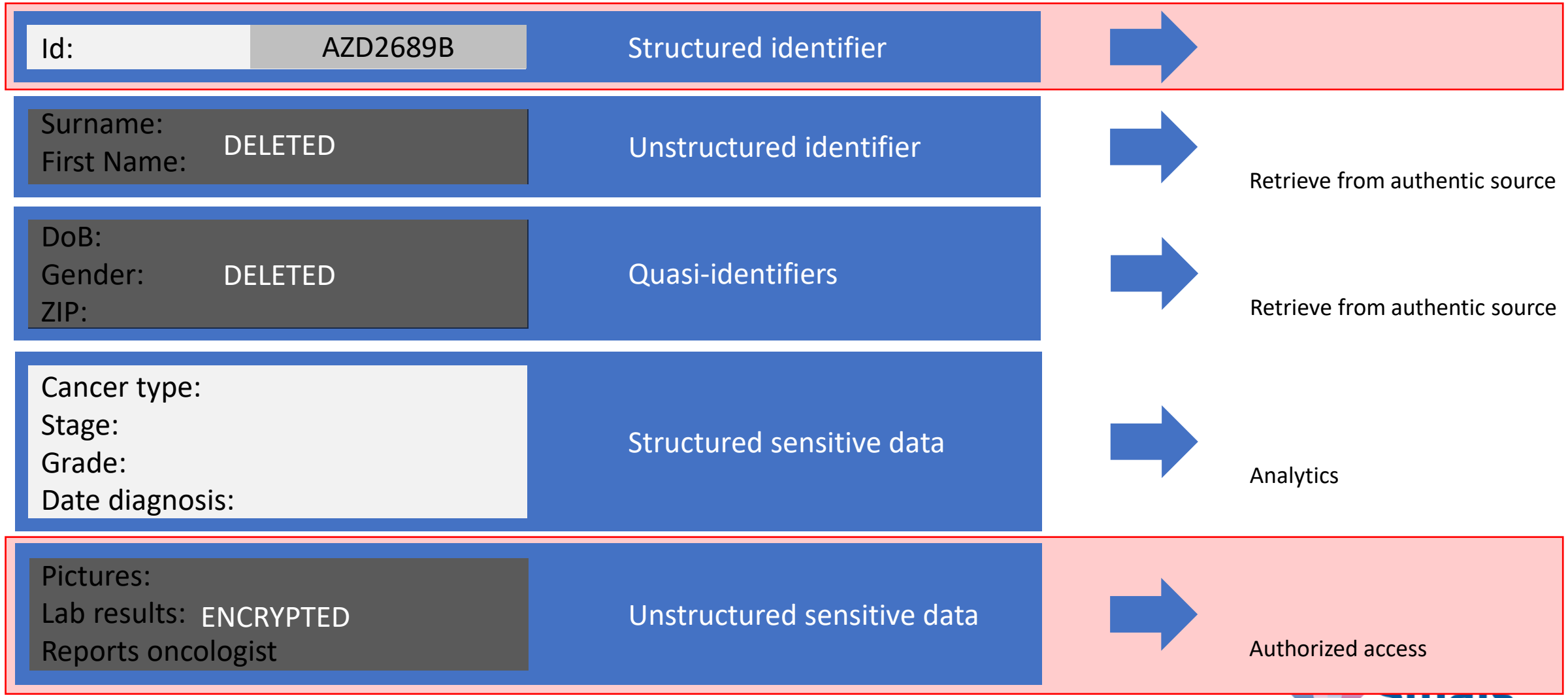
✓ High security

✗ Full encryption limits functionality
Input verification, statistics, analytics



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

Smals ICT for society



ICT for society

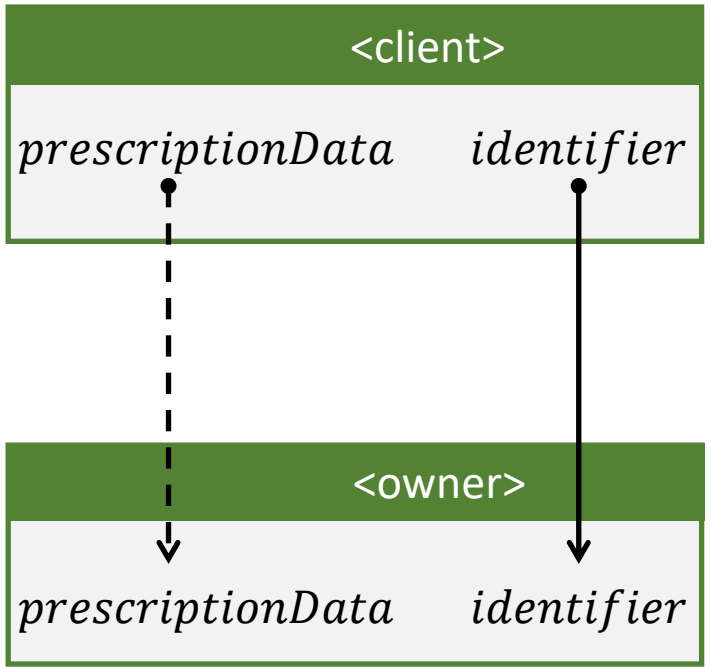
Use case 1 - Live

Unaddressed Health Message Exchange Platform

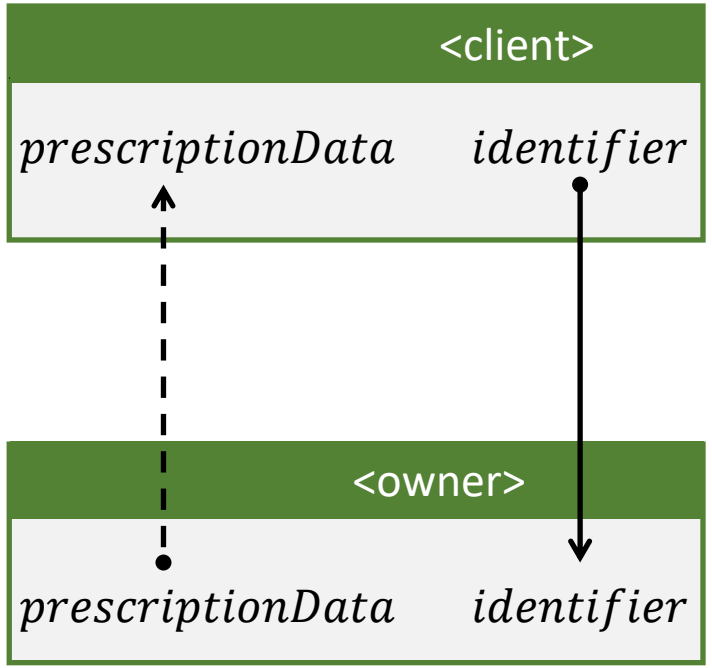
A certificate to start a certain treatment (e.g. physiotherapist, dieticians, speech therapists). Without a referral prescription issues by a doctor, the treatment may not be started.

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

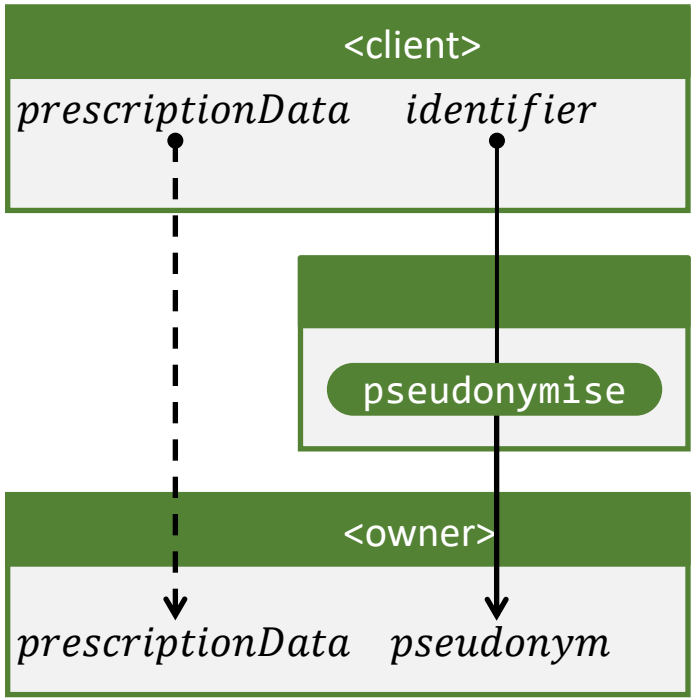
Doctor (client) requests UHMEP (owner) to register prescription



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



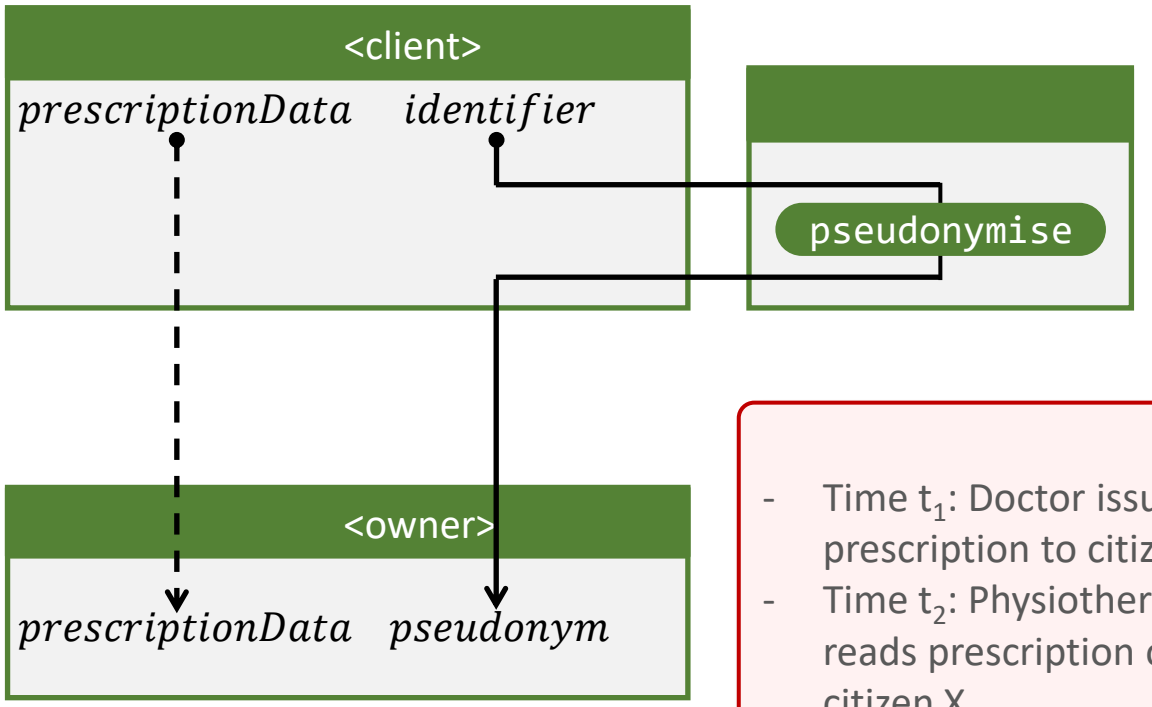
Smals ICT for society



Data leaks to one entity

To be combined somehow

Previously known as *'dienst codage'*



Data leaks to two entities

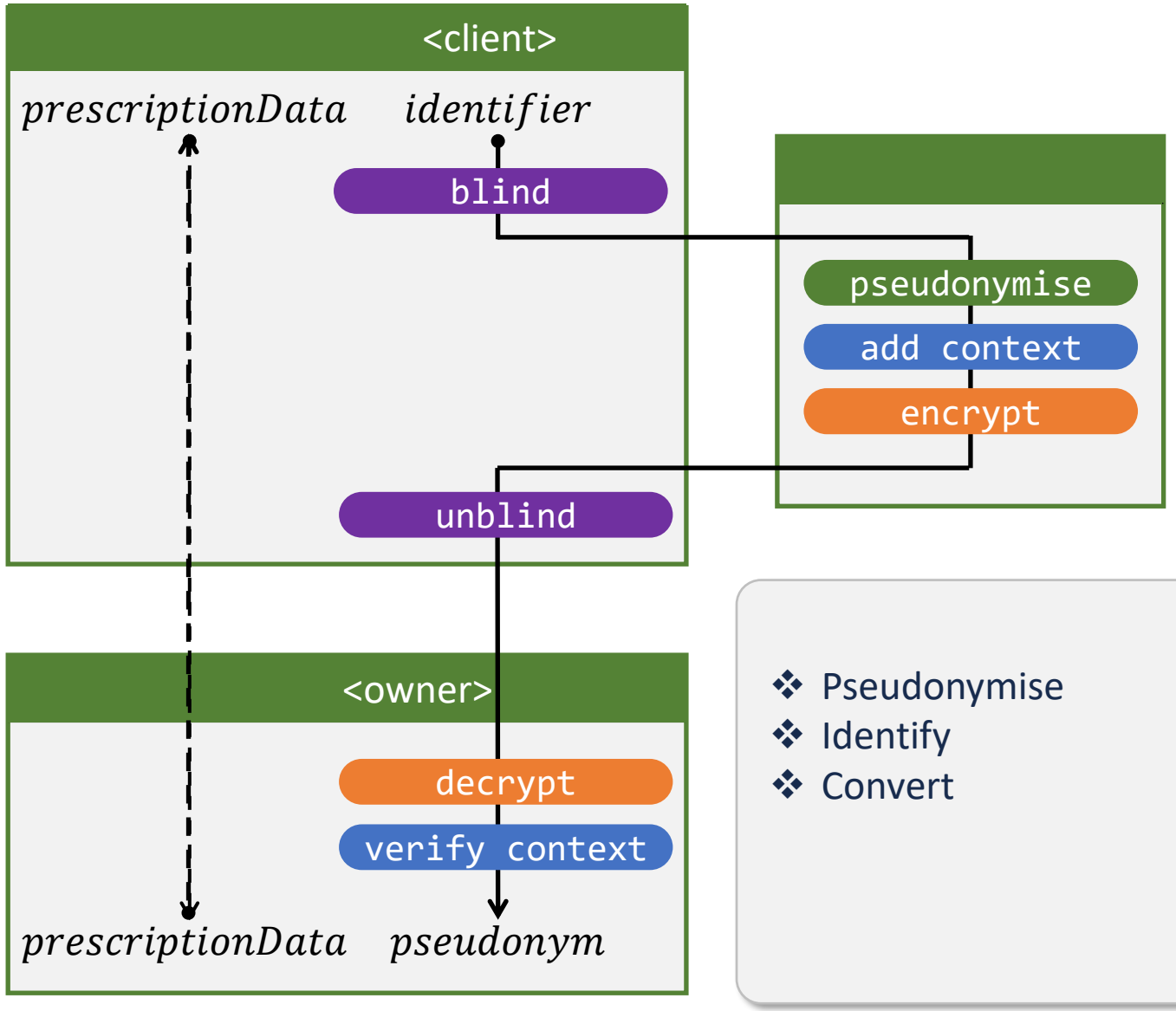
- Time t_1: Doctor issues prescription to citizen X
- Time t_2: Physiotherapist reads prescription of citizen X
- Time t_3: Doctor issues prescription to citizen X
- Time t_4: Data of citizen X involved in research project on diabetes
- ...

(AV+VXF9H5LdTe4b1 SSC7bHjp6b2enJmf plC6a3/jCR5fUHxX RSaRniYR8h7ugNqa lGvP49cZnv6lf9B7 2RUG0rA/, eSmII52CEtsZzSseU DY3YKLtSgqhq1wLPm 9ncHBzGiv1wMIxmc1 jSmpW36GhTt/s1P5s hZGhG8ncoWKSGkJDy fw=)

- Pseudo service on the sidelines
- Only direct communication between healthcare professional and UHMEP backend

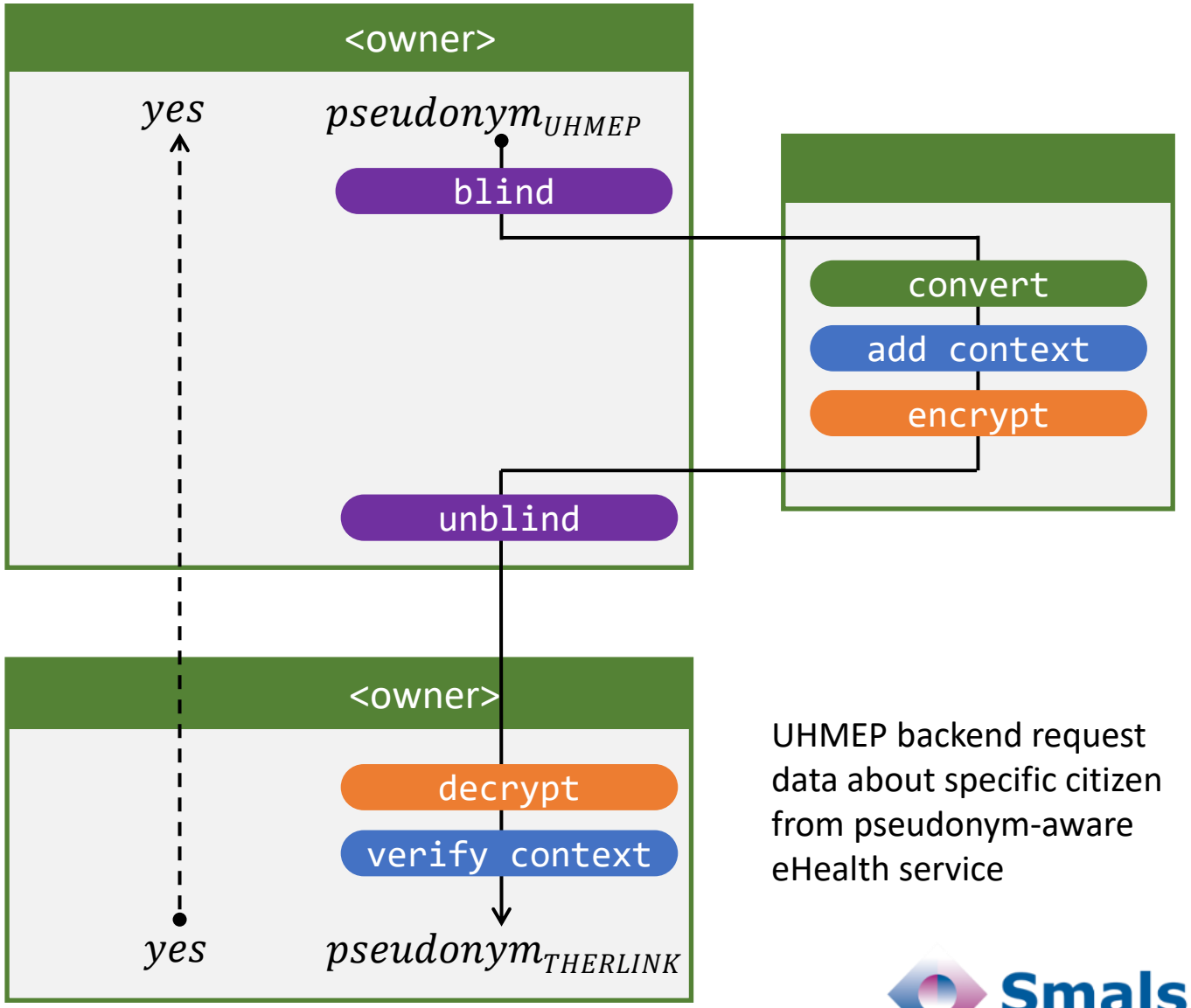
- No extra keys requires
- Relatively simple implementation

	Sender		Translator		Receiver	
	identifier	pseudonym	identifier	pseudonym	identifier	pseudonym
Seals	●	●	●	●	○	●
TTP	●	○	●	●	○	●
Blind	●	○	○	○	○	●



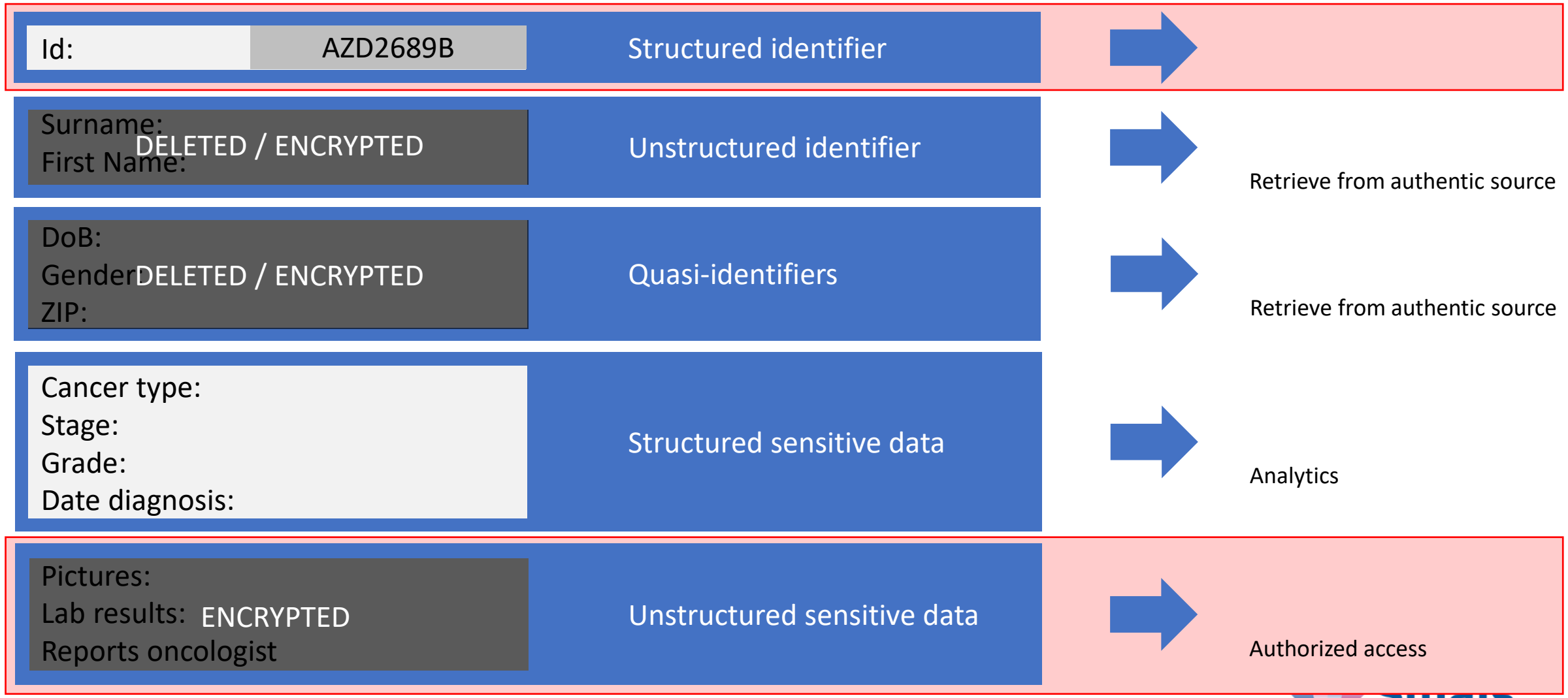
- Pseudonymise
- Identify
- Convert

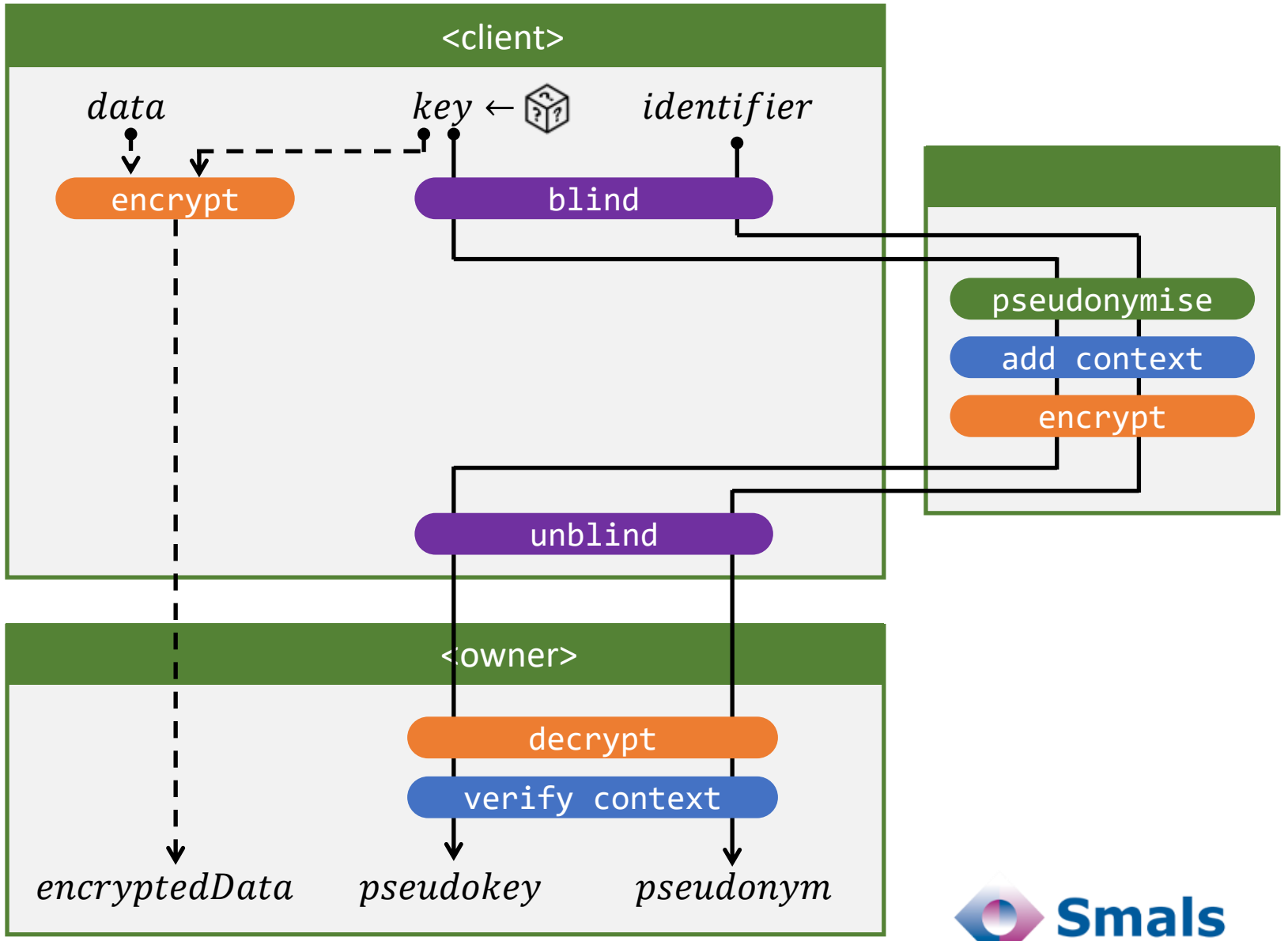
- Pseudonymise
- Identify
- Convert



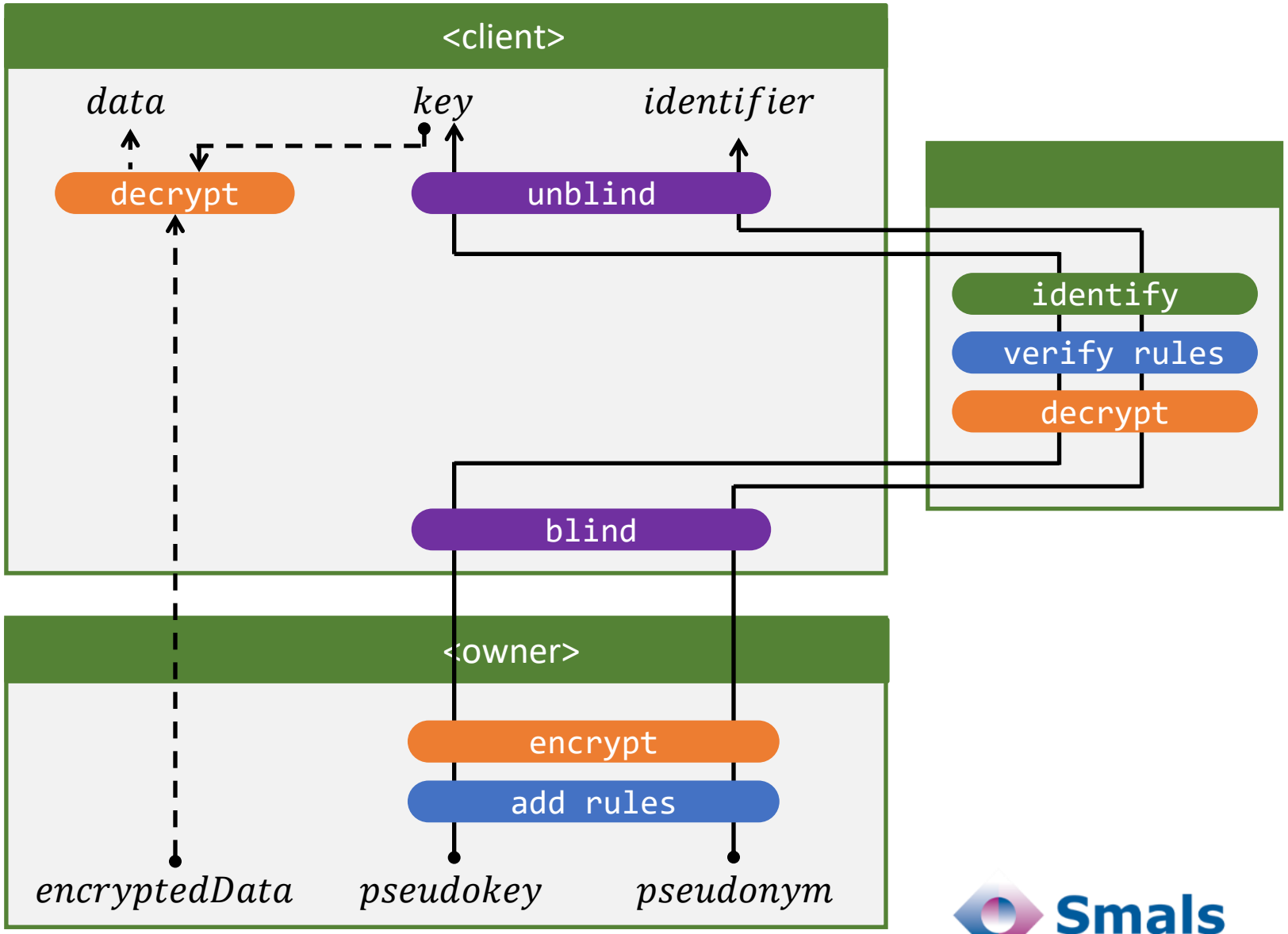
UHMEP backend request data about specific citizen from pseudonym-aware eHealth service

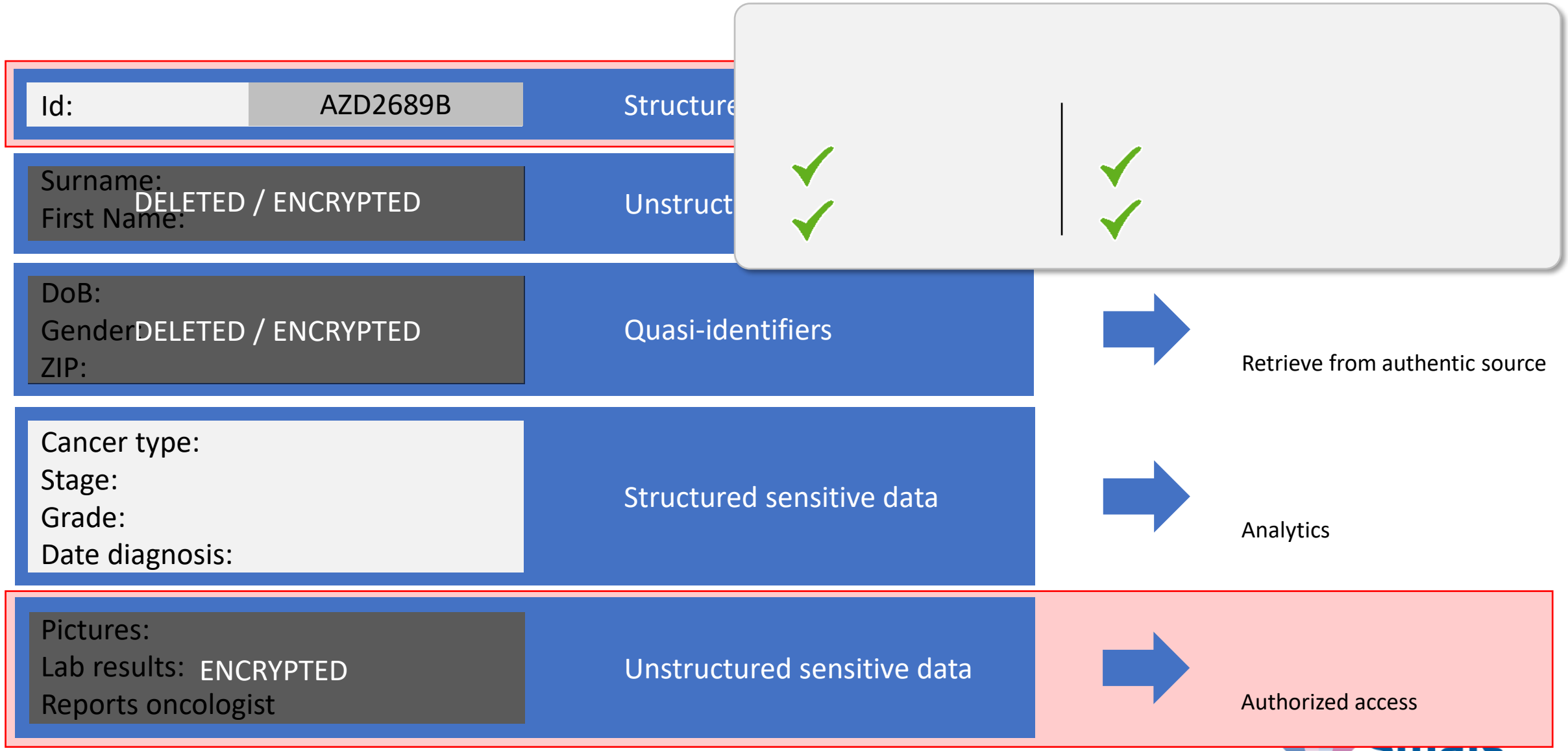
Smals ICT for society



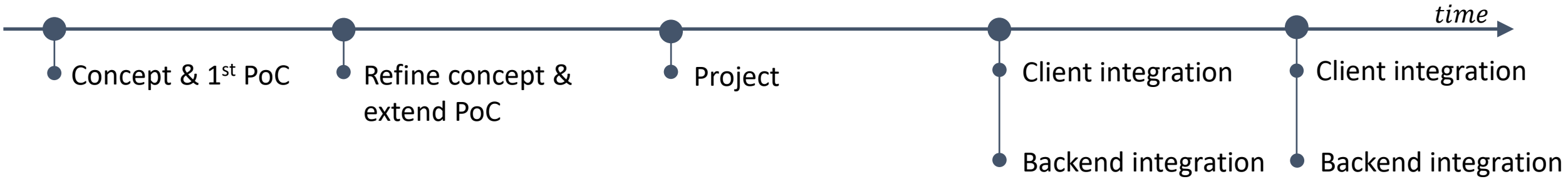


Smals ICT for society

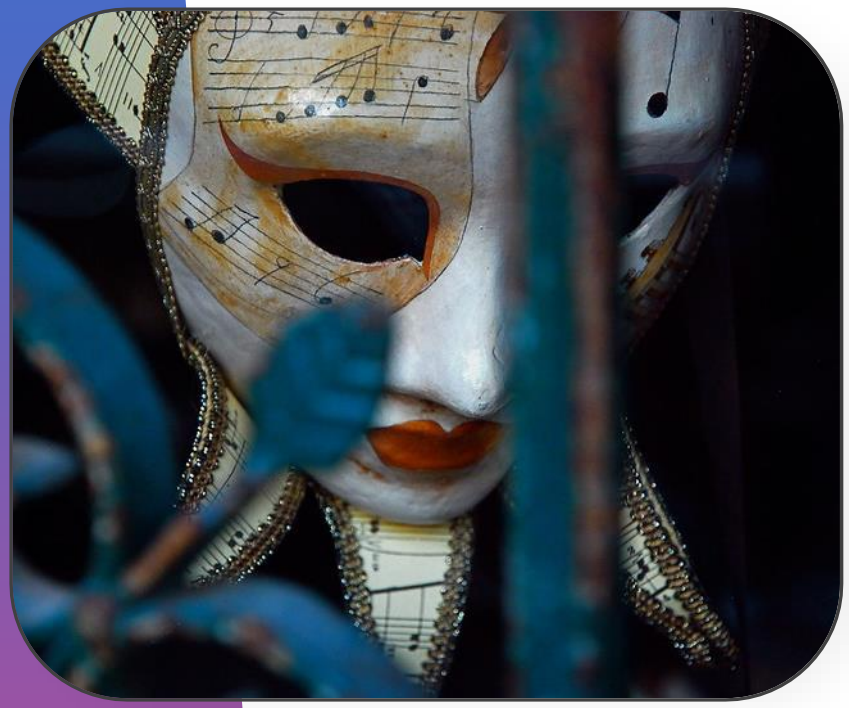

<client>
data
key
identifier
decrypt
unblind
identify
verify rules
decrypt
blind
<owner>
encrypt
add rules
encryptedData
pseudokey
pseudonym



ICT for society



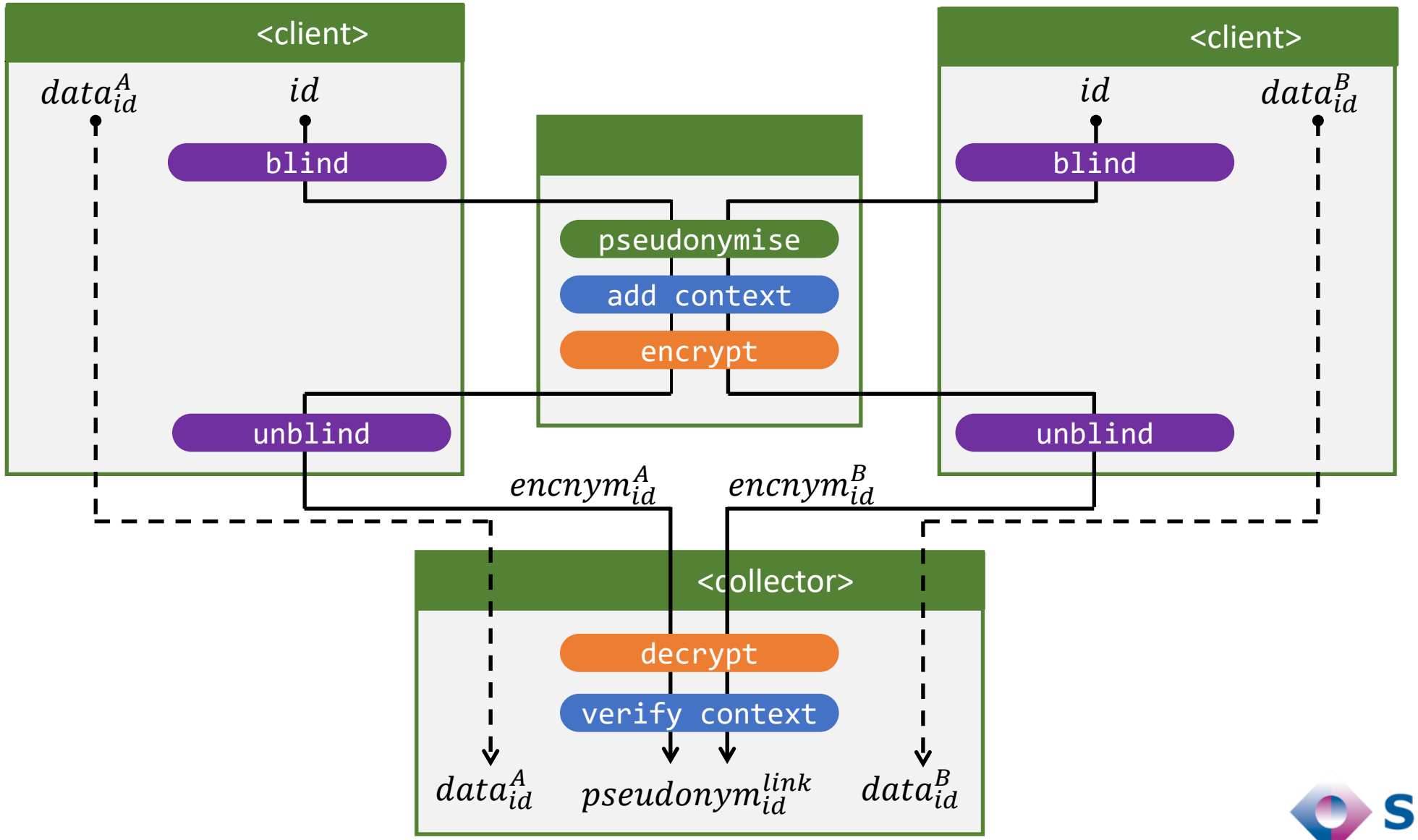
Smals ICT for society

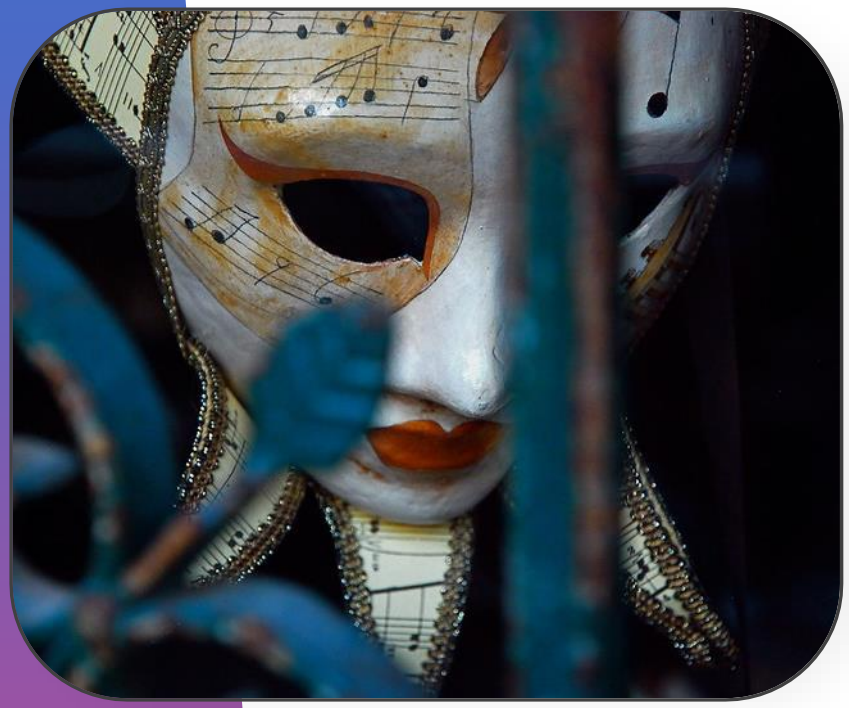


- Problem statement
- Secure records in live environments
- **Join & pseudonymise for research**
- Conclusion

Smals ICT for society

Use case 2 – Proposal





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

Smals ICT for society

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

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

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

Introductie tot de nieuwe eHealth pseudonimiseringsdienst

Introduction au nouveau service de pseudonymisation eHealth

Pseudonimisering & Anonimisering

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

ICT for society

- Problem statement
- Concept
- In practice
- Conclusion

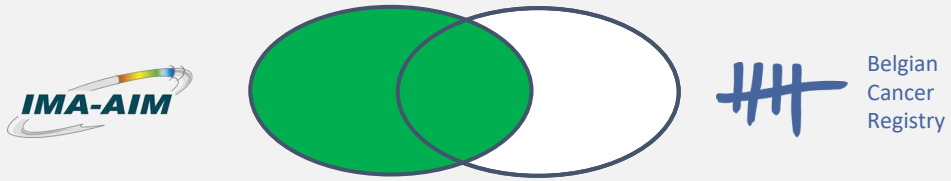
Smals ICT for society

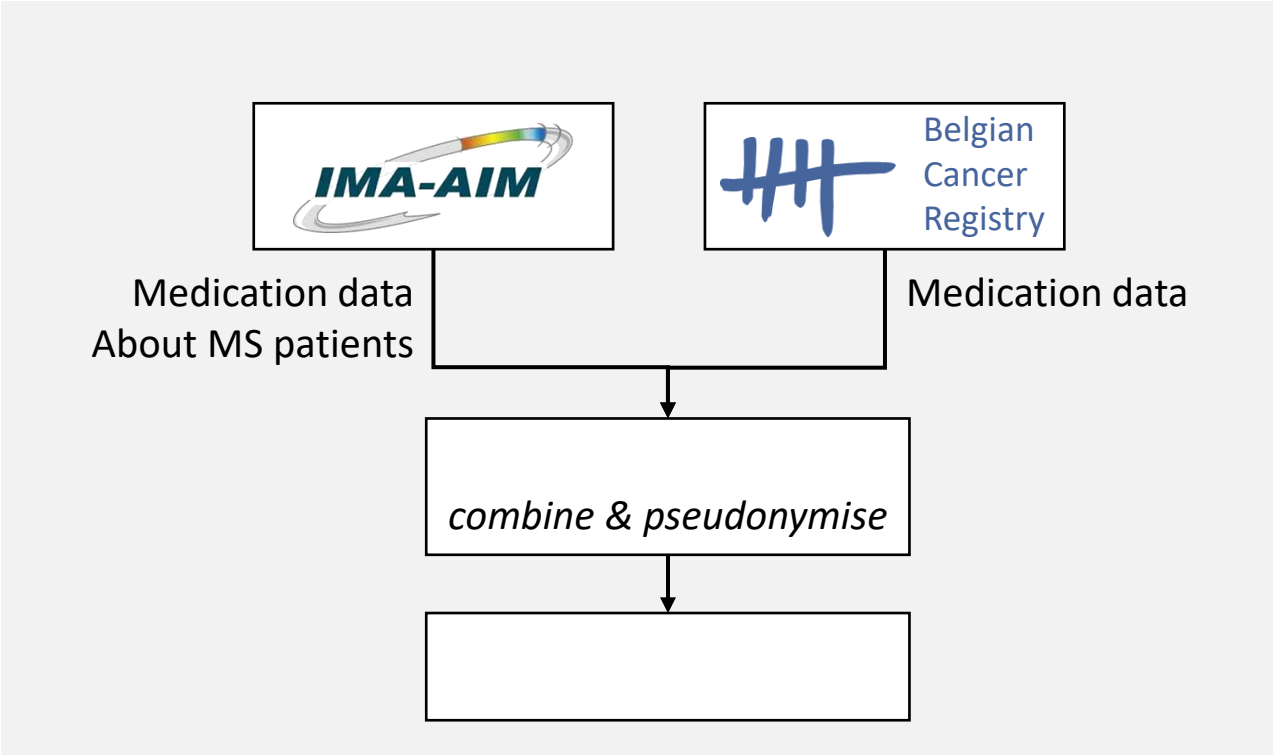
- **Problem statement**
- Concept
- In practice
- Conclusion

Smals ICT for society

Use case

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?



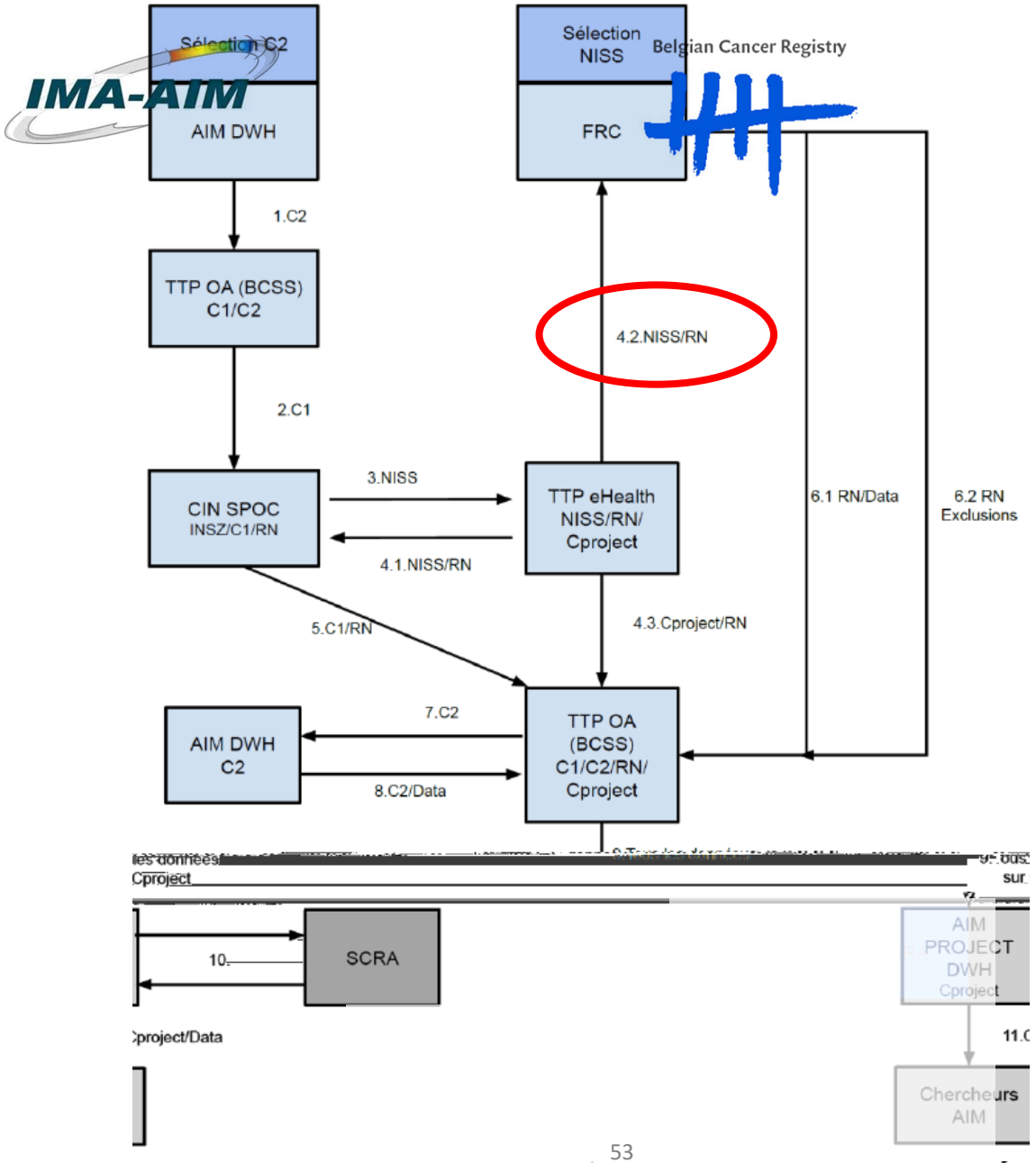


Smals
ICT for society

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

"Lasts weeks, months, even years"

"Requires an exorbitant amount of resources"

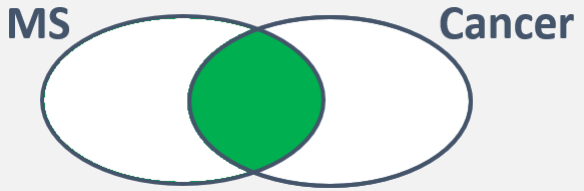


Can we for specific research projects
combine and pseudonymise personal data
originating from different sources

(because each research question is different)?

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

Extensible from there

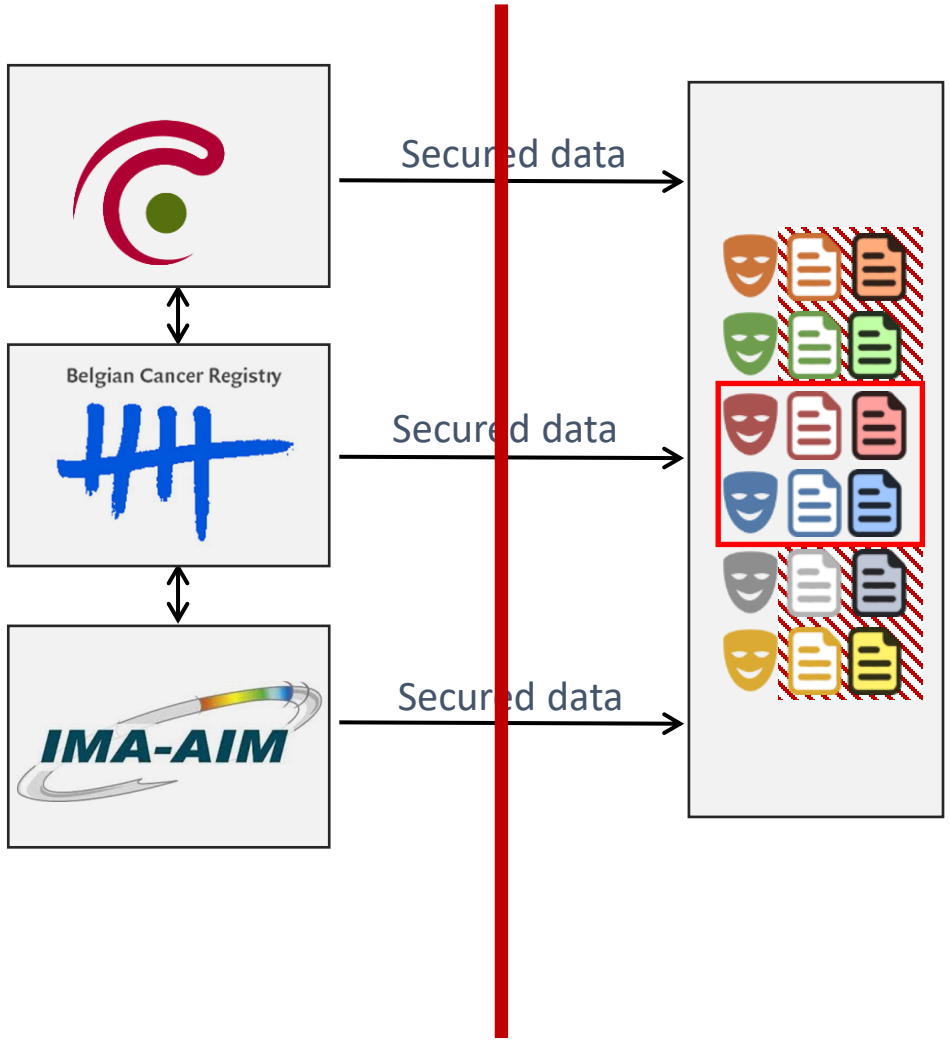


Smals
ICT for society

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

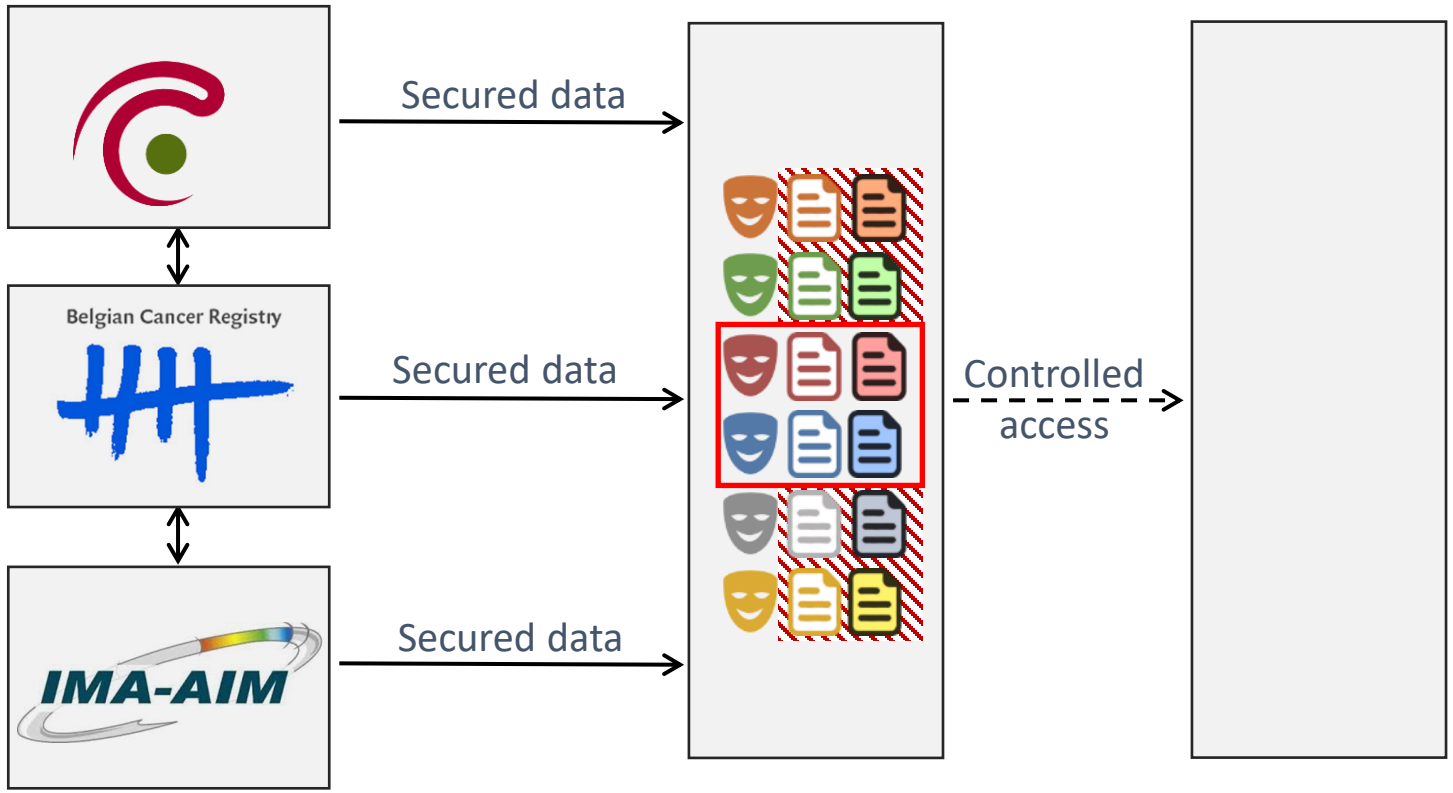
Smals ICT for society

Smals
ICT for society



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

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



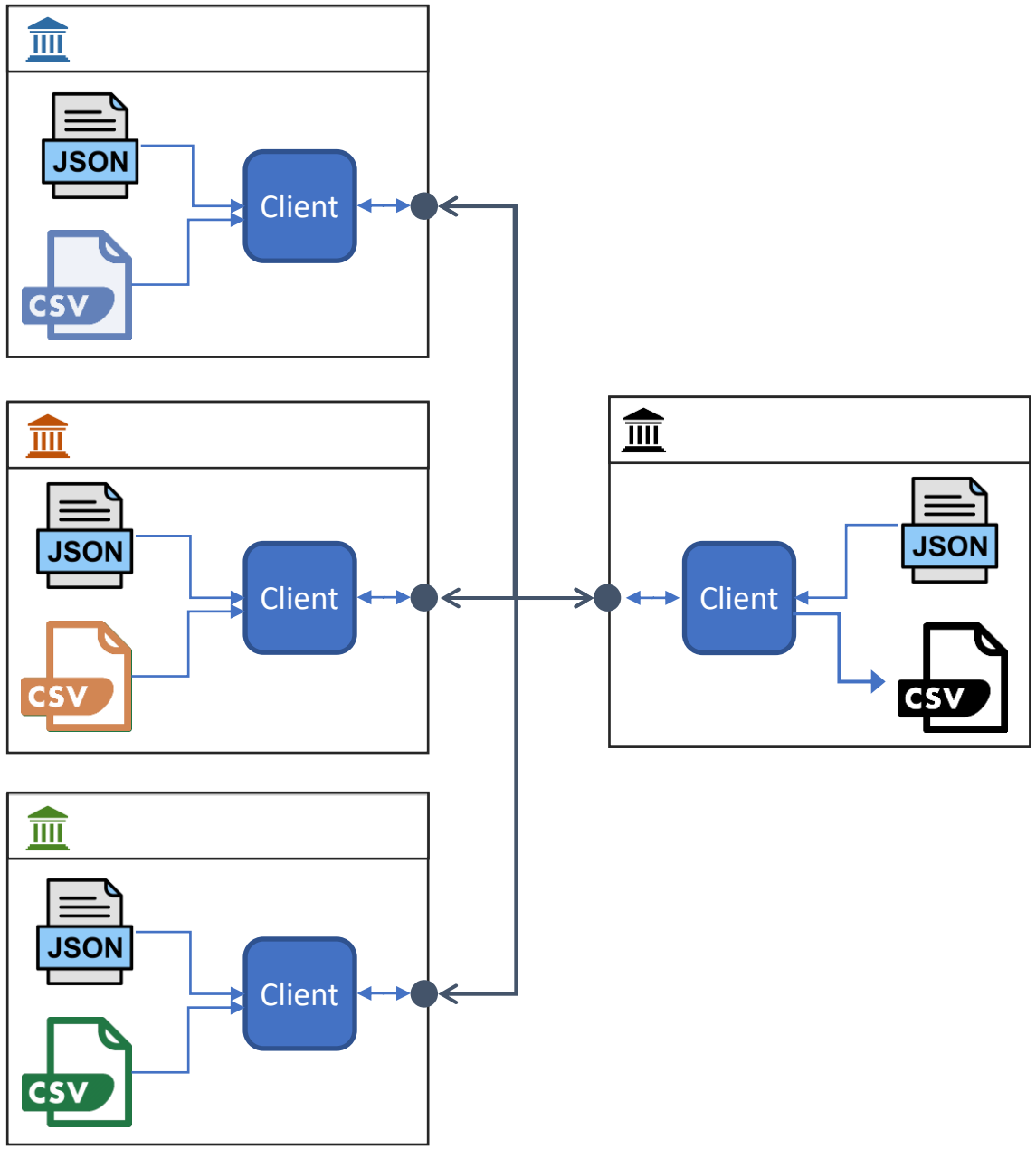
1. Deletes asap irrelevant ciphertexts
2. Can do additional checks on the data
3. Control access by researcher

▶

Smals ICT for society

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

Smals ICT for society



client

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

JSON

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

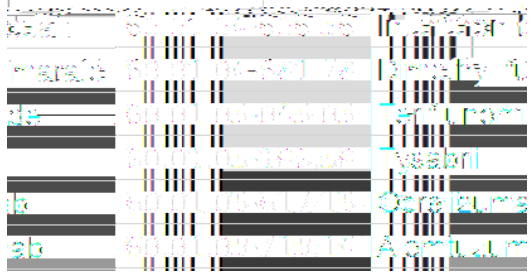
CSV

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

CSV

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

60.01.03-231.73	Teriflunomide
60.01.03-562.33	Alemtuzumab
60.01.03-697.92	Glatiramer acetate



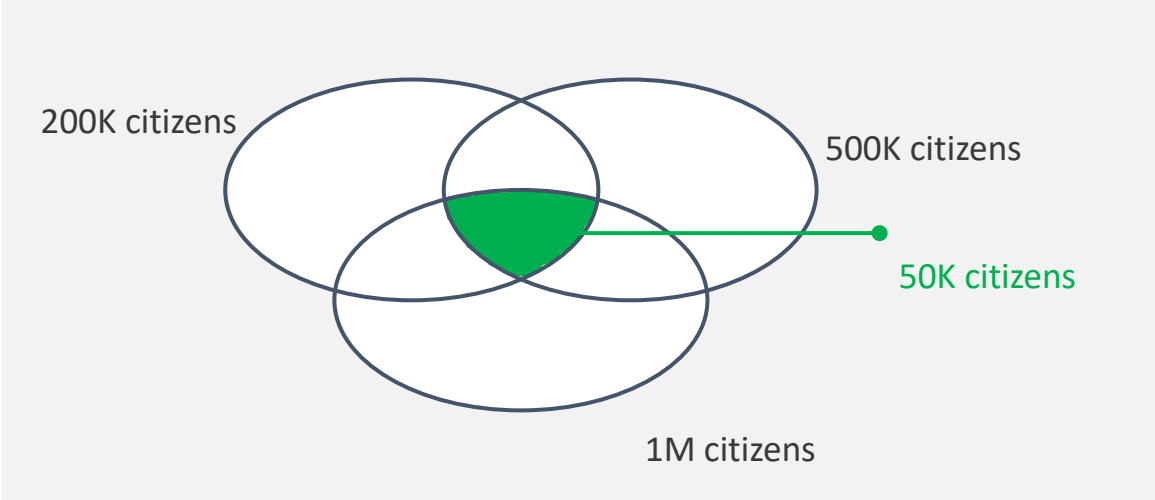
E.g. Citizens with MS

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

E.g. Citizens with cancer

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

E.g. Citizens with high-risk profile



- MinNbRecords: 10
- 128 bit security

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

-
- Excl. a few hundred MBs data transfer

60.01.03-231.73	Teriflunomide
60.01.03-562.33	Alemtuzumab
60.01.03-697.92	Glatiramer acetate

E.g. Citizens with MS

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

E.g. Citizens with cancer

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

E.g. Citizens with high-risk profile

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

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

Smals ICT for society

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

Smals ICT for society

Smals ICT for society

DistriNet KU LEUVEN

COSIC KU LEUVEN

UNIVERSITY OF WATERLOO

Goal is an A-tier confererence

Smals ICT for society

KU LEUVEN

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

Springer

CHAIN Connecting Humankind through Algorithms and Information Networks

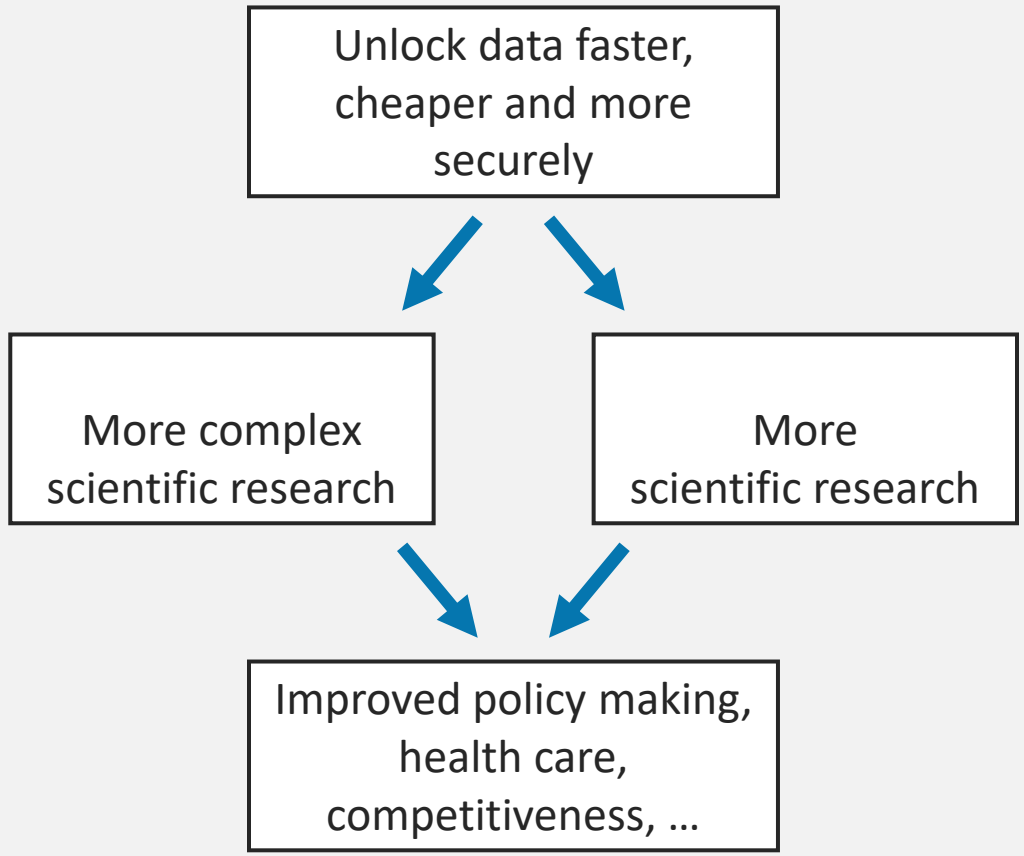
Utrecht University

https://chainresearch.eu/

https://www.uu.nl/en/events/conference-public-governance-and-emerging-technologies-values-trust-and-compliance-by-design

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

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



Smals ICT for society

Smals
ICT for society

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

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

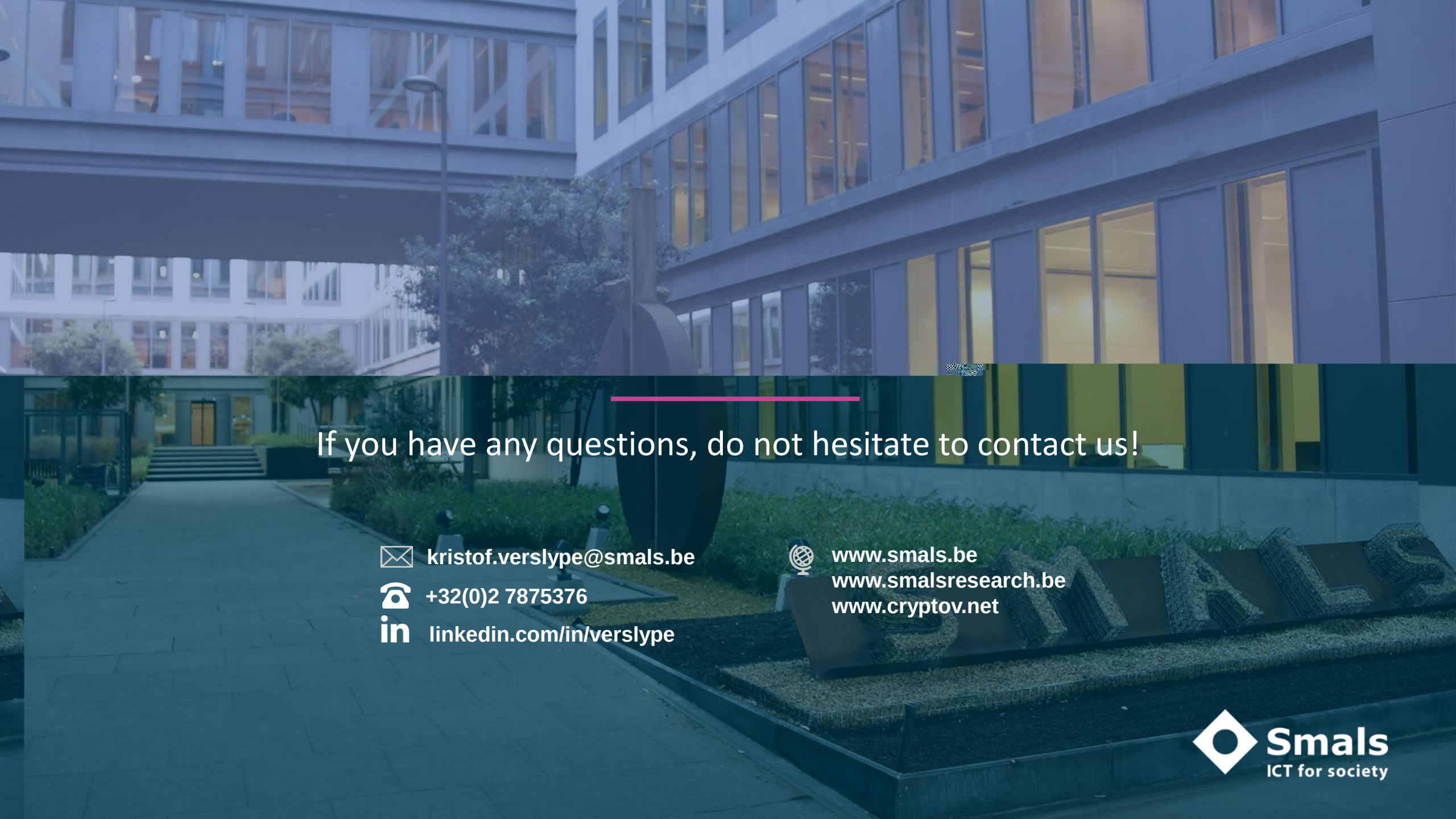
Format-Preserving Ps

eHealth Blind Ps

Oblivious

Unknown Privacy technology

Perfect fit your use case



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

kristof.verslype@smals.be

+32(0)2 7875376

linkedin.com/in/verslype

www.smals.be
www.smalsresearch.be
www.cryptov.net

Smals
ICT for society

Creative Commons
https://flickr.com/photos/peterscherub/53152339550/

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

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

Creative Commons
Flickr

Creative Commons
https://iconscout.com/free-icon/mask-126

Smals
ICT for society