Privacy Enhancing Technology to Address Requirements of Anonymization - Synthetic Data and Pseudonymization

Vancouver International Privacy & Security Summit

February 22, 2023

BACKGROUND

Requirements on the collection, use, and disclosure of data in privacy laws across the globe are triggered when data can identify a specific person

The requirements around the protection of personal information and the technologies available to anonymize such personal information are constantly evolving.

Canada's federal privacy legislation, the *Personal Information Protection and Electronic Documents Act* (*PIPEDA*), Canada's proposed Consumer Privacy Protection Act Bill C-27), and provincial legislation, such as British Columbia's *Personal Information Protection Act* (*PIPA*) may soon have anonymization requirements.

AGENDA

Introductions

"Rules of engagement" (we're not technical experts, questions at the end of the session)

Definitions

What challenges do we hope to solve

DESCRIPTIONS OF TERMS

ANONYMIZING DATA

Anonymizing data is a process with the aim of irreversibly preventing the identification of the individual to whom it relates.

PSEUDONYMIZATION (or Pseudonymized?)

Similar to anonymization.
It differs from anonymization as there is a way to restore the previous data and re-identify users.

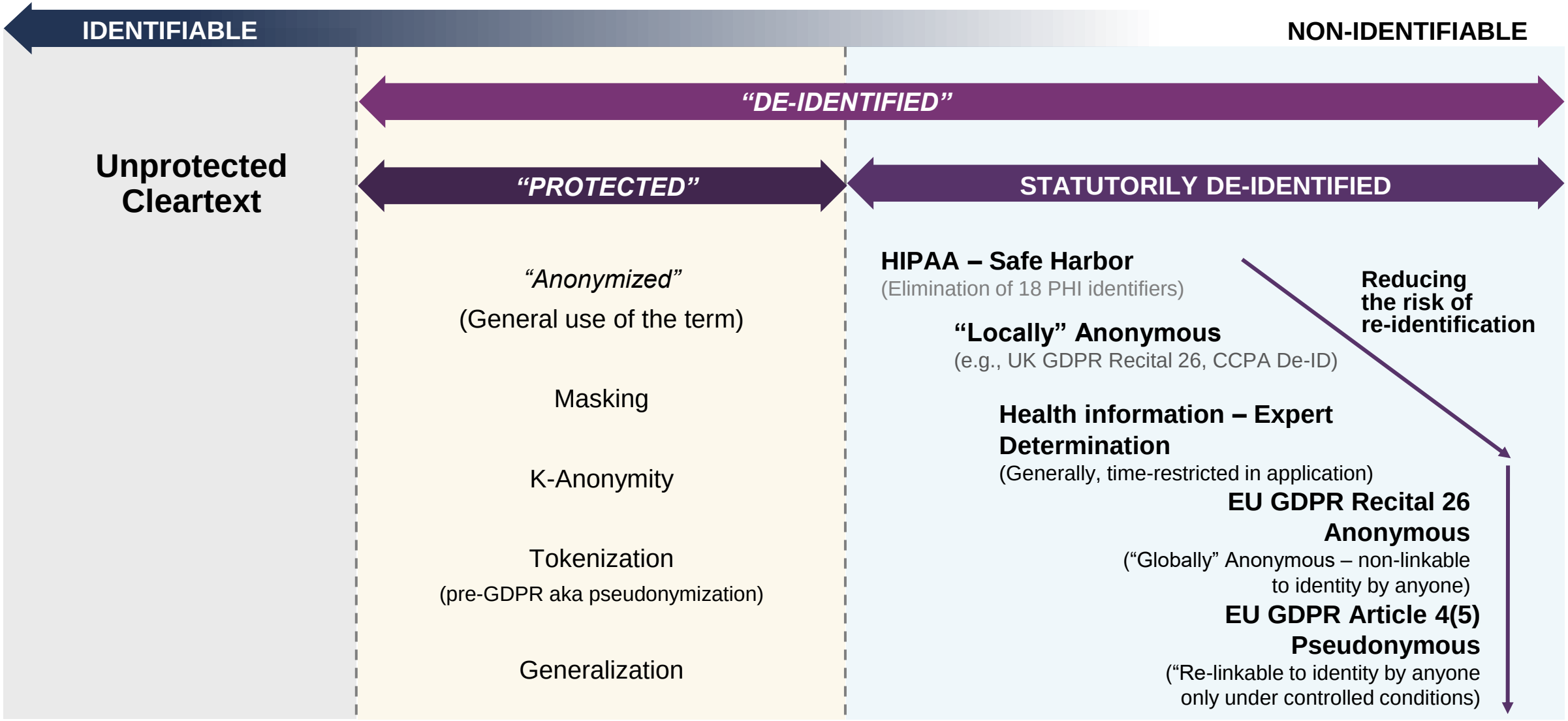
Pseudonymization is defined in the GDPR as:

Pseudonymized data remains “personal data” and is therefore remains subject to the requirements of the GDPR.

SYNTHETIC DATA

Synthetic data is information that's artificially generated rather than produced by real-world events. Typically created using algorithms, synthetic data can be deployed to validate mathematical models and to train machine learning models.

Spectrum of Identifiable/Protected/Statutorily Non-Identifiable Data



1. De-Linking

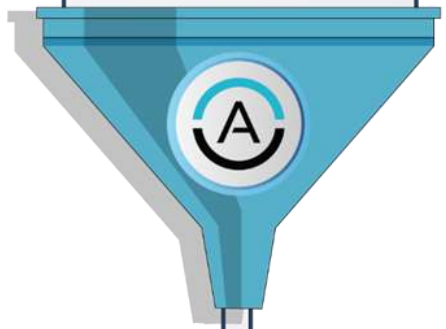
Separating Information Value from Identity

Master Index: Identifying PII

Unique Pseudonym	Name	Email	Telephone	SSN	Credit Card #	Gender	Date of Birth	Zip Code	Latitude	Longitude	Reliability Score

Record-Level Dynamic Pseudonymization

Assign a unique de-identifier for each record (within and between datasets) to defeat linkage and inference attacks. Each pseudonym maps to the Master Index to relink back to identity when ready.



Usable Data: Indirect Identifiers

Unique Pseudonym	Gender	Date of Birth	Zip Code	Latitude	Longitude	Reliability Score

Enables Dynamic De-Risking and Controlled Relinking

2. De-Risking

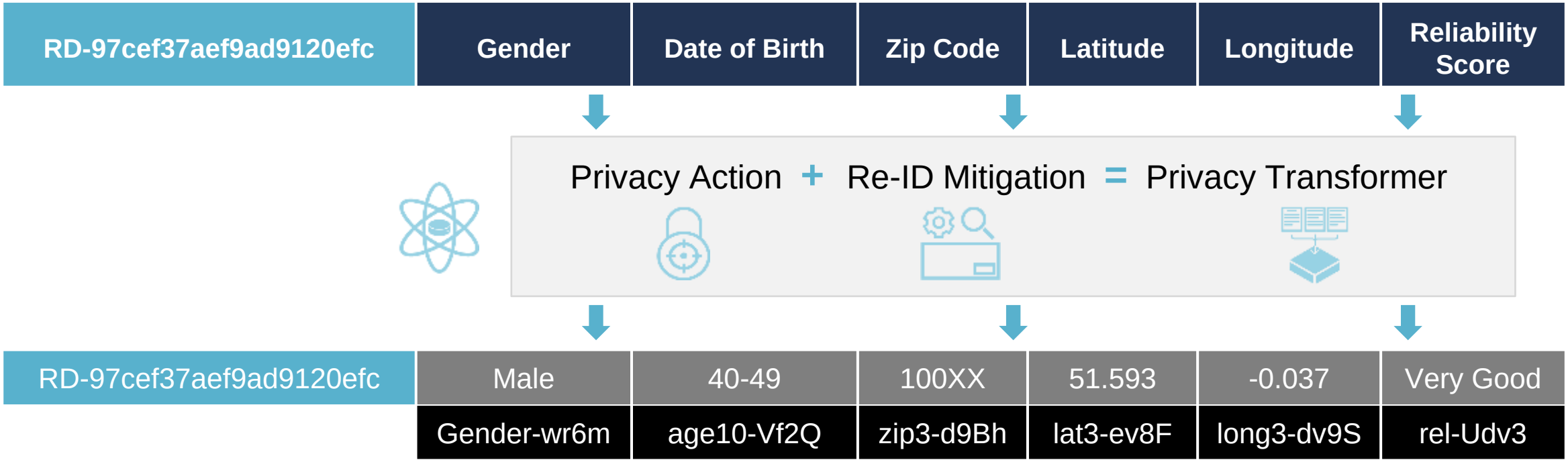
Technical Controls that Travel with the Data

Embedding protection into the data:

1. Anonymization Techniques
2. Field-Level Pseudonymisation
3. Re-Identification Risk Management

Protection against:

- ✓ Inference attacks
- ✓ Linkage attacks
- ✓ Singling-out attacks



3. Dynamic Embedded Controls

Privacy Transformer Scales Variant Twin Creation

Privacy Transformer Enforces Policy

RD-97cef37aef9ad9120efc	gender-wr6m	age10-Vf2Q	zip3-d9Bh	lat3-ev8F	long3-dv9S	rel-Udv3

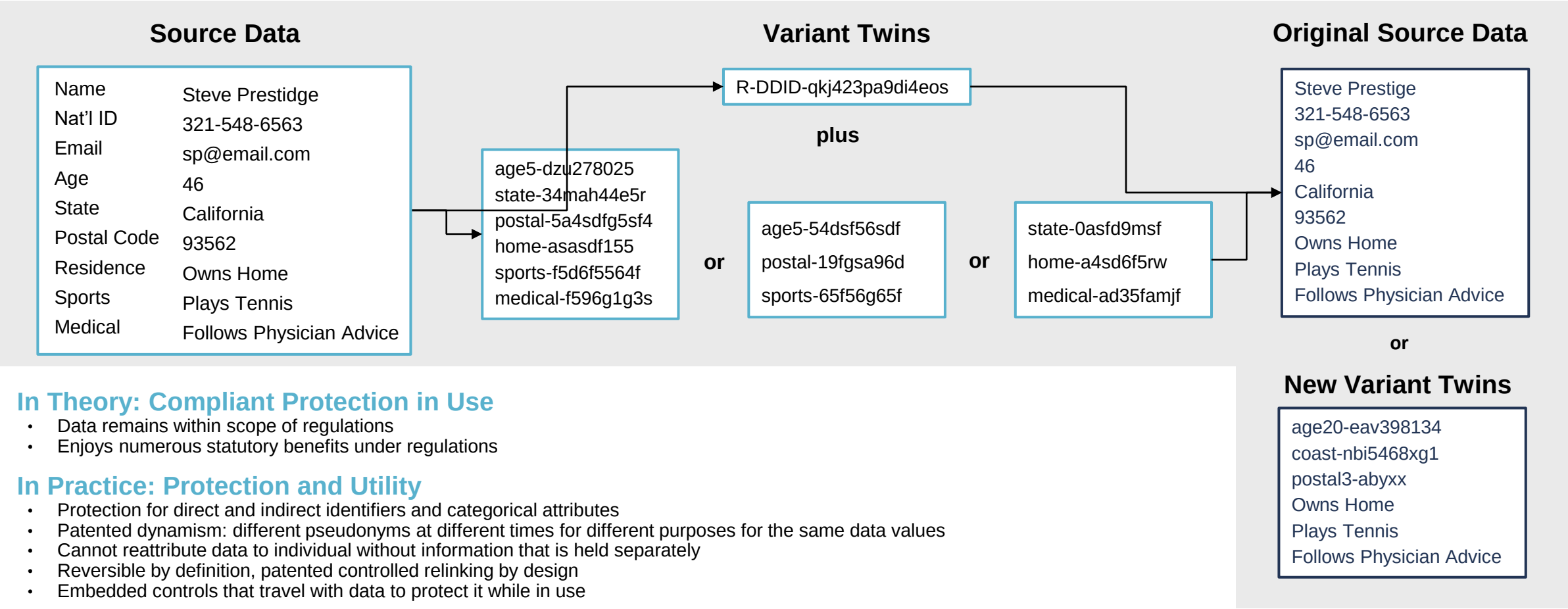
Variant Twin Embodies Policy

Unique Pseudonym	Gender	Age_10	Zip_3	Lat_3	Long_3	Reliability Score
RD-97cef37aef9ad9120efc	gender-wr6m	age10-Vf2Q	zip3-d9Bh	lat3-ev8F	long3-dv9S	rel-Udv3
RD-c75dd862e63ed8d259b0	gender-wr6m	age10-0z4S	zip3-1cgh	lat3-dv0J	long3-dv2X	rel-Udv3
RD-9c015cba189493b9cac8	gender-wr6m	age10-qPTL	zip3-d9Bh	lat3-ev8F	long3-dv9S	rel-sc6K
RD-80d74c7536e5bc706f8a	gender-OrWg	age10-1fcQ	zip3-uy4c	lat3-iob4	long3-iev5	rel-Udv3
RD-b6ff1a08bf59ecc70f15	gender-OrWg	age10-aMpl	zip3-d9Bh	lat3-5jAn	long3-7eeG	rel-j9dV

Variant Twins enable scalable sharing, combining and enriching of data for analytics, AI and ML.

4. Controlled Relinking

Relinking of Variant Twins to Source Data



In Theory: Compliant Protection in Use

- Data remains within scope of regulations
- Enjoys numerous statutory benefits under regulations

In Practice: Protection and Utility

- Protection for direct and indirect identifiers and categorical attributes
- Patented dynamism: different pseudonyms at different times for different purposes for the same data values
- Cannot reattribute data to individual without information that is held separately
- Reversible by definition, patented controlled relinking by design
- Embedded controls that travel with data to protect it while in use

In Reality:

- 100% precision and accuracy in analytics, ML and AI; compliant international data transfers

QUESTIONS?

Thank You

Jay Loder CIPP/C, CIPM, FIP
Privacy Officer
FortisBC
Jay.Loder@fortisbc.com

Tahir Latif, FIP CIPP/E/A/US
Global Practice Lead
Data Responsibility & Privacy
Data+ Group
Tahir.Latif@cognizant.com

Aaron Stevens CIPP, CIPT, FIP
Data Privacy Evangelist
Aaron.stevens@anonos.com