

CLAIMS

1. A computer implemented method for securely sharing datasets that include information and an original unique ID associated with each individual, where the datasets
5 originate from several different data contributors and are to be shared with a central party or recipient, while concealing or masking the original unique ID;

in which each contributor encrypts and sends the original unique ID to an intermediary party, together with the information associated with each individual;

the intermediary party performs a homomorphic crypto operation that
10 deterministically but irreversibly maps the encrypted original unique IDs to an encrypted and obscured form, and sends that encrypted and obscured data to a central or recipient party;

the central party (i) decrypts the encrypted and obscured data received from the intermediate party to generate unique IDs that are one-to-one mapped to but cannot be
15 used to recover each original unique ID and then (ii) joins together the information for each individual from each contributor.

2. The method of Claim 1 in which the original unique ID, once encrypted, is never sent in the clear.

20

3. The method of any preceding Claim in which the intermediary party generates and raises the ciphertext to the power of a secret, random number exponentiation key, $\mathbf{k}$, to generate the encrypted and obscured form and the central party decrypts this to generate $\mathbf{m}^{\mathbf{k}}$, where $\mathbf{m}$ is the original unique ID and the decisional Diffie-Hellman
25 assumption implies that it is cryptographically hard for the central party to generate $\mathbf{m}$ using $\mathbf{m}^{\mathbf{k}}$.

4. The method of any preceding Claim in which key generation, encryption and decryption operations are based on an ElGamal encryption system and the intermediary
30 party conducts the operation on the ciphertexts exploiting the fact that ElGamal encryption is homomorphic under exponentiation.

5. The method of any preceding Claim –in which the central party generates an elliptical curve parameter set and shares that with all contributors and the intermediary

party; private and public key pairs are then generated at the central party and the intermediary party and these public keys are shared with each data contributor; each data contributor then converts each original unique ID **m** to a point on an elliptical curve and encrypts that with both public keys to generate pairs of encrypted points; these encrypted
 5 points are then sent to the intermediary party for partial decryption with its private key, and obscuring with the secret exponentiation key, **k**.

6. The method of Claim 5 in which the central and intermediary party sign their public key with their private key and share these signatures with the contributors,
 10 proving that they have obeyed the key generation protocol.

7. The method of any preceding Claim in which the method eliminates the requirement for the sharing of a unique ID for an individual that is also shared between some or all data contributors, and that is also used by the central party.

15

8. The method of any preceding Claim in which an individual has more than one original unique ID such as address, email, and postcode.

9. The method of Claim 8 in which an individual's original unique IDs are joined
 20 together.

10. The method of any preceding Claim in which biometric signatures are used as original unique ID such as voice signatures, digitised fingerprints, retina scans or iris scans.

25

11. The method of any preceding Claim in which an original unique ID is a combination of multiple identifiers, and it is the combination of identifiers that is unique.

12. The method of any preceding Claim in which each contributor includes with the
 30 information sent to the intermediary party a schema, or structural description of the information, and the intermediary party is able to use the schema from each contributor to identify columns in the data that correspond to the original unique IDs.

13. The method of any preceding Claim in which the information associated with each individual, that is sent from each contributor to the intermediary party, is encrypted.

14. The method of any preceding Claim in which original unique IDs are stored in a column in the datasets held by a contributor.

15. The method of any preceding Claim in which each contributor uses the same unique ID for an individual, or more than one original unique ID with a common identifier.

16. The method of any preceding Claim in which the information that is joined together for each individual from each contributor is anonymised or generalised to less specific values to preserve privacy and is then published.

17. The method of Claim 16 in which publication is by a computer-based system that processes a sensitive dataset and publishes a derivative dataset such that privacy is preserved in the derivative dataset by generalising data values to less specific values by transforming columns in a table of data.

18. The method of Claim 17 in which the system enables a user to (a) define the required level of k-anonymity and/or l-diversity, and the system then automatically generalises some or all of the data in the sensitive dataset so that the derivative dataset achieves the required level of k-anonymity and/or l-diversity, and also enables the user to (b) manually configure how to generalise each column in the table.

19. The method of Claim 18 in which the anonymised or generalised form of data is then shared back with one or more of the contributors.

20. The method of Claim 19 in which the sharing back is done by a privacy preserving query system for querying a database in which a single access point for configuration of the query system is provided, and in which the system's privacy controls is configured by a data holder from this single access point, and in which the database is queried by a data analyst from this single access point.

21. The method of any preceding Claim in which each contributor is independent of all other contributors and cannot share personal information, with other contributors.

22. The method of any preceding Claim in which each contributor is:

5 a different bank or other financial institution and the datasets include personal financial information, or

a different holder of medical data and the datasets include medical information, or

10 a different telecommunications service provider and the datasets include call-related information, or

a different internet service provider and the datasets include internet or web browsing-related information, or a different social network and the datasets include social network-related information.

15 23. The method of any preceding Claim in which each contributor holds geo-location information for multiple individuals and/or holds sensitive personal information for different individuals.

20 24. The method of any preceding Claim in which the central party is a governmental organisation.

25 25. The method of any preceding Claim in which the central party data mines the information for multiple individuals, without being able to identify any specific individual.

26. The method of any preceding Claim in which the central party assumes the role of the intermediary party.

30 27. The method of any preceding Claim which provides a representation of the dataset(s) in an aggregate form wherein the representation prevents any original unique ID for any individual to be re-generated.

28. The method of any preceding Claim in which the representation further provides a comprehensive view of information held by the dataset(s) that specifically relates to

health, pension, financial or education data.

29. A software system that implements the computer implemented methods as defined in any one of Claims 1 – 28.

5

30. A dataset that has been generated using the computer implemented methods as defined in any one of Claims 1 – 28.

31. A cloud computing infrastructure that implements the computer implemented
10 methods as defined in any one of Claims 1 – 28.

32. A method of accessing cloud-hosted resources, including the step of implementing the methods as defined in any one of Claims 1 - 28.