



## **Data Anonymisation and Privacy-Enhancing Technologies (PETs) Course for Data Sharing**



## **Data Anonymisation and Privacy-Enhancing Technologies (PETs) Course for Data Sharing**

### **Introduction:**

Data anonymisation and privacy-enhancing technologies PETs for data sharing are the subject of this 5-day course for data science, data sharing, statistical release and privacy engineering staff, who produce a Data Release and PETs Selection Plan for a case organisation. Units that hold valuable microdata often refuse partner requests or release extracts with only names removed, because nobody can measure how easily records could be re-identified. Nominees already prepare or approve datasets at work and practise quasi-identifier treatment, attack testing and noise calibration in hands-on lab exercises. CoreConcept Training Center delivers this data anonymisation and PETs course.

### **Course Objectives:**

- Classify dataset attributes into direct identifiers, quasi-identifiers and sensitive values using the ISO/IEC 20889 technique taxonomy
- Apply generalisation hierarchies, suppression, top-coding and record swapping to reach target k-anonymity, l-diversity and t-closeness levels
- Measure re-identification risk through prosecutor, journalist and marketer attack models and linkage tests against auxiliary sources
- Calibrate differential privacy mechanisms, sensitivity and epsilon allocation for counts, sums and published query interfaces
- Compare trusted execution environments, secure multi-party computation, homomorphic encryption and federated learning for joint analysis without raw data exchange
- Assemble a Data Release and PETs Selection Plan that matches each sharing scenario to a release route and residual risk rating

### **Target Audience:**

- Data science staff responsible for preparing extracts and analytical files for partners or researchers
  - Data sharing and data access staff responsible for assessing incoming requests and approving release conditions
  - Statistical production staff responsible for publishing frequency tables, magnitude tables and public-use microdata
  - Privacy engineering staff responsible for selecting and configuring disclosure protection techniques
  - Research data service staff responsible for operating controlled access environments and output checking
-

## Course Outline:

### Day 1: Disclosure Risk Foundations and Dataset Profiling

- Identity, Attribute and Membership Disclosure Types Compared
- ISO/IEC 20889 De-identification Technique Classification Overview
- Direct Identifier, Quasi-Identifier and Sensitive Attribute Tagging
- Pseudonymisation Versus Anonymisation Through Keyed Hashing and Token Vaults
- Release Inventory Profiling of Uniqueness and Sample Frequencies

### Day 2: Privacy Models and Release Route Architecture

- ISO/IEC 27559 De-identification Framework Context and Risk Thresholds
- k-Anonymity Equivalence Classes and Minimum Group Size Selection
- l-Diversity and t-Closeness Against Homogeneity Attacks
- Release Route Spectrum From Open Files to Enclave Access
- ISO/IEC 29100 Privacy Principles Mapped to Release Decisions

### Day 3: Microdata Treatment and Table Protection Lab

- Generalisation Hierarchies for Age, Postcode and Occupation Codes
- Local Suppression, Top-Coding and Record Swapping on Microdata
- tau-Argus Primary and Secondary Cell Suppression on Magnitude Tables
- Controlled Rounding and Small Count Treatment in sdcTable
- Information Loss Metrics and Analytical Validity Checks

### Day 4: Attack Testing, Differential Privacy and Cryptographic PETs

- Linkage Attack Simulation Against Auxiliary Public Registers
- Membership Inference and Reconstruction Attack Testing on Aggregates
- Laplace and Gaussian Mechanisms With Sensitivity and Epsilon Allocation
- Trusted Execution Environments and Secure Multi-Party Computation Trade-offs
- Homomorphic Encryption and Federated Learning for Cross-Organisation Analysis

### Day 5: Lab Case Work and the Data Release and PETs Selection Plan

- Hospital Admissions Case Microdata Treatment and Risk Score
  - Retail Transaction Case Differential Privacy Query Release Design
  - SACRO Output Checking of Researcher Results From Secure Enclave
  - Release Route and PETs Decision Matrix for Five Sharing Requests
  - Data Release and PETs Selection Plan Completion and Peer Challenge
-

## Skills You Will Gain:

- Quasi-Identifier Analysis
- Generalisation Hierarchy Design
- Table Cell Suppression
- Linkage Attack Testing
- Privacy Budget Calibration
- Secure Computation Selection
- Information Loss Measurement
- Release Route Design

## Why Attend This Course:

- Deliver a Data Release and PETs Selection Plan to the data owner and the data sharing approval board, covering treatment, risk scores and release routes
- Decide whether a requested dataset can go out as public-use microdata, through a controlled enclave or only as noisy aggregates
- Prevent re-identification incidents, refused partnerships and unusable over-suppressed files that follow untested releases
- Coach analysts and statisticians on quasi-identifier tagging, suppression settings and epsilon allocation for routine releases

## Conclusion:

Once back at work, the participant hands the Data Release and PETs Selection Plan to the data owner and the data sharing approval board. They use its risk scores, treatment settings and release routes to approve or decline partner requests, to choose between open files, enclave access and computation on encrypted data, and to set privacy budgets for published query services. After the first release under the plan, the unit should compare measured re-identification risk and information loss against the targets and adjust group sizes, suppression rules and epsilon values.

## Frequently Asked Questions FAQ:

### What should participants know before a data anonymisation and privacy-enhancing technologies course?

Participants should already prepare, query or approve datasets at work and read basic descriptive statistics. Lab exercises use spreadsheets and open statistical tools, so light scripting experience in R or Python helps, although step-by-step files are supplied for every exercise.

## **How does a data anonymisation and PETs course differ from a privacy law or database security course?**

It teaches the technical treatment and testing of datasets before release. Privacy law courses cover legal duties, governance and rights handling, while database security courses harden engines, roles and encryption, and neither measures re-identification risk or tunes noise for published statistics.

## **Is removing names and identity numbers enough for data anonymisation before data sharing?**

No. Combinations of quasi-identifiers such as birth date, area and occupation often single out individuals, and linkage with public sources can restore identities. Anonymisation needs treatment of those combinations, measured risk and, for repeated queries, mechanisms such as differential privacy.

## **What do participants take back from the data anonymisation and privacy-enhancing technologies course?**

Participants take back a Data Release and PETs Selection Plan with attribute tagging, treatment settings, risk scores, a release route for each request and a technology choice for joint analysis, together with lab files they can reuse on their own datasets.

---