

THE PET GUIDE

The background features a stylized illustration of four people in a professional setting. A woman with glasses is on the left, looking at a laptop. In the center, a man in a white shirt and tie is talking on a mobile phone. On the right, a man in a dark suit is holding a document. Another woman is partially visible behind him. The scene is overlaid with a network of blue lines and dots, suggesting data flow and connectivity. The overall color palette is dominated by various shades of blue, with accents of orange, red, and white.

THE UNITED NATIONS GUIDE ON
PRIVACY-ENHANCING TECHNOLOGIES
FOR OFFICIAL STATISTICS.

2023





**THE UNITED NATIONS
GUIDE ON
PRIVACY-ENHANCING
TECHNOLOGIES
FOR OFFICIAL STATISTICS**





MISSION STATEMENT OF UNITED NATIONS COMMITTEE OF EXPERTS ON BIG DATA AND DATA SCIENCE FOR OFFICIAL STATISTICS

The Committee will provide a strategic vision, direction and coordination for a global programme on Big Data and Data Science for official statistics, including for indicators of the 2030 Agenda for Sustainable Development; the Committee will promote practical use of Big Data sources, including cross-border data, while building on existing precedents and finding solutions for the many existing challenges, including:

- Methodological issues, covering quality concerns and data analytics,
- Legal and other issues in respect of access to data sources,
- Privacy issues, in particular those relevant to the use and reuse of data, data linking and re-identification,
- Security, information technology issues and management of data, including advanced means of data dissemination, assessment of cloud computing and storage, and cost-benefit analysis

The Committee will also promote capacity-building, training and sharing of experience and will foster communication and advocacy of the use of Big Data and Data Science for policy applications, especially for the monitoring of the 2030 Agenda for Sustainable Development; and, finally, the Committee will strive to build public trust in the use of Big Data and Data Science for official statistics.

DISCLAIMER

The designations employed and the presentation of the material in the present publication do not imply the expression of any opinion whatsoever on the part of the United Nations concerning the legal status of any country or its authorities or the delimitations of its frontiers. The term “country” as used in this publication also refers, as appropriate, to territories or areas. The designations of country groups in the publication are intended solely for statistical or analytical convenience and do not necessarily express a judgment about the stage reached by a particular country, territory or area in the development process. Mention of the names of firms and commercial products does not imply endorsement by the United Nations.

The views expressed in this publication are those of the authors and do not necessarily reflect those of the United Nations or its senior management, or of the experts whose contributions are acknowledged.

Copyright © United Nations, 2023

All rights reserved.

Suggested citation: United Nations, 2023, *United Nations Guide on Privacy-Enhancing Technologies for Official Statistics*, United Nations Committee of Experts on Big Data and Data Science for Official Statistics, New York.

Website: <https://unstats.un.org/bigdata>

CONTENTS

FOREWORD	09
ACKNOWLEDGEMENTS	11
ACRONYMS	14
EXECUTIVE SUMMARY	16
CHAPTER 1 INTRODUCTION TO PRIVACY-ENHANCING TECHNOLOGIES	18
1.1 Motivations for the use of Privacy-Enhancing Technologies	18
1.2 Challenges in using Privacy-Enhancing Technologies for Official Statistics	20
1.3 A New Approach to International Collaboration on PETs	24
CHAPTER 2 METHODOLOGIES AND APPROACHES	28
2.1 Secure Multi-party Computation	28
2.2 Homomorphic Encryption	32
2.3 Differential Privacy	36
2.4 Synthetic Data	40
2.5 Distributed Learning	43
2.6 Zero Knowledge Proofs	48
2.7 Trusted Execution Environments and Secure Enclaves	51
2.8 Practical Considerations of PETs	53
CHAPTER 3 CASE STUDIES	61
CASE STUDY 1 Boston Women’s Workforce Council: Measuring salary disparity using secure multi-party computation	66
CASE STUDY 2 European Statistical System: Developing Trusted Smart Surveys	69
CASE STUDY 3 Eurostat: Processing of longitudinal mobile network operator data	71
CASE STUDY 4 Indonesia Ministry of Tourism: Confidentially sharing datasets between two mobile network operators via a trusted execution environment	73
CASE STUDY 5 Italian National Institute of Statistics and Bank of Italy: Enriching data analysis using privacy-preserving record linkage	76
CASE STUDY 6 Office for National Statistics: Trialling the use of synthetic data at the United Kingdom’s national statistics institute	78
CASE STUDY 7 Samsung SDS (Korea): Data aggregation system	80
CASE STUDY 8 Statistics Canada: Measuring the coverage of a data source using a private set intersection	82

CASE STUDY 9 Statistics Canada: Training a machine learning model for private text classification using leveled homomorphic encryption	84
CASE STUDY 10 Statistics Canada: Trialling the use of synthetic data	86
CASE STUDY 11 Statistics Korea: Developing a privacy-preserving Statistical Data Hub Platform	88
CASE STUDY 12 Statistics Netherlands: Developing privacy-preserving cardiovascular risk prediction models from distributed clinical and socioeconomic data	90
CASE STUDY 13 Statistics Netherlands: Measuring effectiveness of an eHealth solution using private set intersection	92
CASE STUDY 14 Twitter and OpenMined: Enabling Third-party Audits and Research Reproducibility over Unreleased Digital Assets	95
CASE STUDY 15 United Nations Economic Commission for Europe: Trialling approaches to privacy-preserving federated machine learning	98
CASE STUDY 16 United Nations PET Lab: International Trade	100
CASE STUDY 17 United States Census Bureau: Deploying a differentially private Disclosure Avoidance System for the 2020 US Census	103
CASE STUDY 18 United States Department of Education: Analysing student financial aid data using privacy-preserving record linkage	105
CHAPTER 4 STANDARDS	109
4.1 Introduction	109
4.2 Key Standards	110
4.3 Related Standards	114
4.4 Standards Under Development	117
CHAPTER 5 LEGAL AND REGULATORY ISSUES	122
5.1 Introduction	122
5.2 The Legal and Regulatory Outlook	123
5.3 Challenges and Risks when using PETs	124
5.4 Opportunities and Affordances of PETs	125
5.5 Challenges of Crossing Jurisdictions	129
5.6 Advice to Regulators	129

“

GOVERNMENTS UNDERSTAND THE GREAT SOCIETAL AND ECONOMIC VALUE THAT CAN BE UNLEASHED BY A MORE WIDE-SPREAD USE OF DATA ON TOPICS LIKE HEALTH, TAXES OR SOCIAL SECURITY. THIS PET GUIDE CAN PAVE THE WAY FOR A BETTER UNDERSTANDING OF AND GREATER CONFIDENCE IN USING PRIVACY-ENHANCING TECHNOLOGIES TO SAFELY UTILIZE SENSITIVE DATA.



FOREWORD

RELEVANCE OF OFFICIAL STATISTICS / ACCESS TO SENSITIVE DATA

In recent years almost every government has been faced with very serious challenges, such as the global health pandemic, increasing occurrences of severe weather causing flooding, drought or fires, environmental degradation, supply chain disruption, increasing numbers of refugees and migrants, rising energy and food prices, and economic stagnation. To handle these crises in the right way, our leaders need the right data at the right time.

National statistical offices (NSOs), and other institutes of the national statistical system, are called upon by their governments to provide these trusted, relevant, timely and high-quality data, which support evidence-based decision making. In many cases, NSOs themselves collect sensitive data on persons and businesses through surveys and censuses, such as data from a population census or from household or business surveys. However, to act swiftly on emerging issues, NSOs are almost always obliged to supplement those data with additional secondary data sources such as administrative data (for example, tax records, social security data, health records or customs administration records) or private sector data (for example, mobile phone records or transactional credit card information).

On the basis of a national Statistics Act, NSOs are often entrusted by society to have access to these kinds of sensitive data. In practice, however, the administrative authorities or private sector companies are very reluctant to “hand over” their raw data. Institutional arrangements are complicated, and require additional legal approvals and guidance which may take a long time to finalize. The difficulties mount further when more partners are involved in the processing and analysis of the data. In addition, the national data protection authority may want to provide input in cases of data sharing, since they want to make sure that the privacy of persons and businesses is protected.

THE COVID EXCEPTION

When COVID-19 hit as a global health pandemic in the early part of 2020, many governments wanted to limit the spread of the very contagious virus and therefore invoked measures to limit the mobility of people. To monitor if these measures were successful, mobile phone data proved to be very useful. With these data, it could be shown almost in real-time, if and where the population stayed mostly at home and in which part of the country movement was still happening.

Getting access to the mobile phone data was possible, but by no means easy. In countries such as Ghana and the Gambia, negotiations with the mobile phone companies regarding data access had already been ongoing for a few years, so in March 2020 agreements to access the mobile phone records could be signed very quickly. Access was still restricted and data would still remain on the premises of the company, but analyses could be done, which were fit-for-purpose. In other countries, telecom companies were much more reluctant to come to agreements, and often access was only given to highly aggregated data, which would not allow for fine-grain analyses.

THE ROLE OF PETS IN OPENING A PATHWAY TO BETTER ACCESS TO DATA

Governments, companies and the public in general are worried that sensitive personal or business information could possibly be leaked and misused, if data were accessed by external partners, including NSOs. However, if data could be accessed without revealing any sensitive information and without possibilities of de-identification, would that take away the privacy concerns? Encryption has already been widely used in banking and internet data transfer, and has proven to be highly reliable. Could privacy-enhancing technologies (PETs) also be highly reliable and be used in a similar way for accessing, for example, health records, tax records or credit card data by NSOs?

This guide will exactly deal with this issue: can PETs guarantee the safe sharing of data?

THE CONTEXT OF THE UNCEBD, THE PET TASK TEAM AND PUBLICATIONS

As the digital society emerged over the last 20 years or so, the global community of official statistics saw the increasing need to explore benefits and challenges of new data sources, new methods and new technologies. The United Nations Statistical Commission, which is the highest global governing body for official statistics, therefore established in 2014 the UN Committee of Experts on Big Data and Data Science for Official Statistics (UNCEBD). This committee explored benefits and challenges of the use of a variety of Big Data sources and their application to various statistical domains. It became clear very soon that getting access to these data was one of the main challenges.

At the beginning of 2018, UNCEBD created a task team to look into the possibilities of using privacy-enhancing technologies. The objectives of this task team were to develop principles, policies and open standards for data sharing, taking full account of data privacy, confidentiality and security issues when designing methods and procedures for the collection, processing, storage and presentation of data. A first document on those issues was released in 2019.

Since 2019, many data sharing projects using PETs have been carried out, showing a diversity in data sets, project objectives and the kind of PETs used. The PET task team prepared a second document, which is this guide on PETs. It contains several new techniques (synthetic data and distributed learning), new international collaboration initiatives (like the UN PET Lab), a review of standards and legal and regulatory aspects of the use of PETs, and especially the descriptions of 18 use cases.

FORWARD LOOKING

The current global crises need a coordinated international response, which demands timely access to often sensitive data shared with multiple partners, of which some are in other countries. For understandable privacy concerns those partners cannot be given full access to all data.

Going forward, we should develop smart ways in which to elicit the essential information from the original data to arrive at the appropriate responses to recover from existing global crises. Application of PETs will help us in designing those smart methods. There are specific characteristics of persons, businesses or locations, which help us in formulating, driving and monitoring policies. We can make sure through the use of PETs that we extract those characteristics without identifying individual persons, businesses or locations.

ACKNOWLEDGEMENTS

The United Nations Guide on Privacy-Enhancing Technologies for Official Statistics was prepared by the Task Team on Privacy-Enhancing Technologies of the United Nations Committee of Experts on Big Data and Data Science for Official Statistics. We would like to acknowledge the valuable contributions of many experts, who voluntarily dedicated time and effort in the preparation of this document. The overall guidance was given by the editorial board under leadership of Matjaž Jug (Statistics Netherlands). The editorial board further consisted of Jess Stahl (OpenMined), Jack Fitzsimons (Oblivious AI, Ireland), Robert Pisarczyk (Oblivious AI, Ireland), Julian Padget (University of Bath, United Kingdom), Ronald Jansen (United Nations Statistics Division), David Buckley (Centre for Data Ethics and Innovation (CDEI), United Kingdom), Editorial board was responsible for organizing work on chapters, drafting of the Foreword and the Executive Summary, and for reviewing all chapters. The Editorial Board wish to thank external experts Hema Krishna Murty (OpenMined) and Fabio Ricciato (Eurostat) for useful and insightful suggestions during the editorial process, Augusto Cesar Fadel (Brazilian Institute of Geography and Statistics) for compiling the list of acronyms and Adrian McLoughlin (Swerve, Ireland), who was responsible for style and formatting.

CHAPTER 1 INTRODUCTION TO PRIVACY-ENHANCING TECHNOLOGIES

Contributors were Maxime Agostini (Sarus Technologies, France), Jack Fitzsimons (Oblivious AI, Ireland), Ronald Jansen (United Nations Statistics Division), Matjaž Jug (Statistics Netherlands), Saeid Molladavoudi (Statistics Canada), Monica Scannapieco (Italian National Institute of Statistics (ISTAT))

CHAPTER 2 METHODOLOGIES AND APPROACHES

Contributors were Will Abramson (Edinburgh Napier University, United Kingdom), Maxime Agostini (Sarus Technologies, France), David Archer (Galois, Inc, United States), Jack Fitzsimons (Oblivious AI, Ireland), Nicolas Grislain (Sarus Technologies, France), Hema Krishna Murty (OpenMined), Saeid Molladavoudi (Statistics Canada), Julian Padget (University of Bath, United Kingdom), Robert Pisarczyk (Oblivious AI, Ireland), Wade Shen (Actuate, United States), Andrew Trask (OpenMined, United States).

CHAPTER 3 CASE STUDIES

The overall guidance to and coordination of chapter 3 was given by David Buckley (CDEI, United Kingdom) and Matjaž Jug (Statistics Netherlands)

CASE STUDY 1 BOSTON WOMEN'S WORKFORCE COUNCIL: MEASURING SALARY DISPARITY USING SECURE MULTI-PARTY COMPUTATION

Specific contributors to case study 1 were Kinan Dak Albab (Brown University, United States), Azer Bestavros (Boston University, United States), Ran Canetti (Boston University, United States), Rawane Issa (Boston University, United States), Frederick Jansen (Nth party, United States), Andrei Lapets (Nth party, United States), Lucy Qin (Brown University, United States), Shannon Roberts (UMass Amherst, United States), Mayank Varia (Boston University, United States), and Nikolaj Volgushev (Elastic, Germany).

CASE STUDY 2 EUROPEAN STATISTICAL SYSTEM: DEVELOPING TRUSTED SMART SURVEYS

Specific contributors to case study 2 were Joeri van Etten (Statistics Netherlands), Sulaika Duijsings-Mahangi (Statistics Netherlands), Rob Warmerdam (Statistics Netherlands), Matjaž Jug (Statistics Netherlands), and Fabrizio De Fausti (ISTAT, Italy).

CASE STUDY 3 EUROSTAT: PROCESSING OF LONGITUDINAL MOBILE NETWORK OPERATOR DATA

Specific contributors to case study 3 were Fabio Ricciato (Eurostat), and Baldur Kubo (Cybernetica, Estonia).

CASE STUDY 4 INDONESIA MINISTRY OF TOURISM: CONFIDENTIALLY SHARING DATASETS BETWEEN TWO MOBILE NETWORK OPERATORS VIA A TRUSTED EXECUTION ENVIRONMENT

Specific contributors to case study 4 were Siim Esko (Positium, Estonia), Erki Saluveer (Positium, Estonia), Jaak Randmets (Cybernetica, Estonia), Angela Sakh (Cybernetica, Estonia) and Baldur Kubo (Cybernetica, Estonia), Addin Maulana, (Ministry of Tourism, Indonesia), Norman Sasono, (Ministry of Tourism, Indonesia).

CASE STUDY 5 ITALIAN NATIONAL INSTITUTE OF STATISTICS AND BANK OF ITALY: ENRICHING DATA ANALYSIS USING PRIVACY-PRESERVING RECORD LINKAGE

Specific contributors to case study 5 were Mauro Bruno (ISTAT, Italy), Massimo De Cubellis (ISTAT, Italy), Fabrizio De Fausti (ISTAT, Italy) and Monica Scannapieco (ISTAT, Italy).

CASE STUDY 6 OFFICE FOR NATIONAL STATISTICS: TRIALLING THE USE OF SYNTHETIC DATA AT THE UNITED KINGDOM'S NATIONAL STATISTICS INSTITUTE

Specific contributors to case study 6 were Owen Daniel (Office for National Statistics, United Kingdom) and Ioannis Kaloskampis (Office for National Statistics, United Kingdom).

CASE STUDY 7 SAMSUNG SDS (KOREA): DATA AGGREGATION SYSTEM

Specific contributors to case study 7 were Jihoon Cho (Samsung SDS, Republic of Korea), Hyojin Yoon (Samsung SDS, Republic of Korea) and Kyoohyung Han (Samsung SDS, Republic of Korea).

CASE STUDY 8 STATISTICS CANADA: MEASURING THE COVERAGE OF A DATA SOURCE USING A PRIVATE SET INTERSECTION

Specific contributors to case study 8 were Abel Dasyuva (Statistics Canada) and Jean-François Beaumont (Statistics Canada).

CASE STUDY 9 STATISTICS CANADA: TRAINING A MACHINE LEARNING MODEL FOR PRIVATE TEXT CLASSIFICATION USING LEVELED HOMOMORPHIC ENCRYPTION

Specific contributors to case study 9 were Saeid Molladavoudi (Statistics Canada), Benjamin Santos (Statistics Canada) and Zachary Zanussi (Statistics Canada).

CASE STUDY 10 STATISTICS CANADA: TRIALLING THE USE OF SYNTHETIC DATA

Specific contributors to case study 10 were Héloïse Gauvin (Statistics Canada), Claude Girard (Statistics Canada), Isabelle Michaud (Statistics Canada), Kenza Sallier (Statistics Canada) and Steven Thomas (Statistics Canada).

CASE STUDY 11 STATISTICS KOREA: DEVELOPING A PRIVACY-PRESERVING STATISTICAL DATA HUB PLATFORM

Specific contributors to case study 11 were Kyeongwon Choo (Statistics Korea), Keunkwan Ryu (Seoul National University, Republic of Korea), Jung Hee Cheon (Seoul National University & Cryptolab, Republic of Korea), and Jaebeom An (Seoul National University, Republic of Korea).

CASE STUDY 12 STATISTICS NETHERLANDS: DEVELOPING PRIVACY-PRESERVING CARDIOVASCULAR RISK PREDICTION MODELS FROM DISTRIBUTED CLINICAL AND SOCIOECONOMIC DATA

Specific contributors to case study 12 were Andre Dekker (Maastricht University, Netherlands), Inigo Bermejo (Maastricht University, Netherlands), Florian van Daalen (Maastricht University, Netherlands), Anke Bruninx (Maastricht University, Netherlands), Paul Grooten (Statistics Netherlands), Johan van der Valk (Statistics Netherlands), Bart Scheenstra (MUMC+, Netherlands) and Arnoud van't Hof (MUMC+, Netherlands).

CASE STUDY 13 STATISTICS NETHERLANDS: MEASURING EFFECTIVENESS OF AN EHEALTH SOLUTION USING PRIVATE SET INTERSECTION

Specific contributors to case study 13 were Tjerk Heijmens Visser (CZ, Netherlands), Martijn Antes (Zuyderland, Netherlands), Martine van de Gaar (Linksight, Netherlands), Ralph Schreijen (Statistics Netherlands), and Sulaika Duijsings-Mahangi (Statistics Netherlands).

CASE STUDY 14 TWITTER AND OPENMINED: ENABLING THIRD-PARTY AUDITS AND RESEARCH REPRODUCIBILITY OVER UNRELEASED DIGITAL ASSETS

Specific contributors to case study 14: the following members of OpenMined Laura Ayre, Jack Bandy, Irina Bejan, Tudor Ceberu, Phil Culliton, Kien Dang, Kyoko Eng, Ronnie Falcon, Bennett Farkas, Stephen Gabriel, Baye Gaspard, Shubham Gupta, Madhava Jay, Ionesio Junior, Yemissi Kifouly, Osam Kyemenu-Sarsah, Teo Milea, Ishan Mishra, Curtis Mitchell, George Muraru, Ivoline Ngong, Thiago Porto, Mark Rode, Rasswanth S., Jess Stahl, Kellye Trask, Andrew Trask and Gatha Varma, as well as the following staff of Twitter: Rumman Chowdhury, Vijaya Gadde, Aaron Gonzales, Kristian Lum, Nick Matheson, Nick Pickles, Tylea Richard, Jutta Williams and Patrick Woody.

CASE STUDY 15 UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE: TRIALLING APPROACHES TO PRIVACY-PRESERVING FEDERATED MACHINE LEARNING

Specific contributors to case study 15 were Saeid Molladavoudi (Statistics Canada), Benjamin Santos (Statistics Canada), Zachary Zanussi (Statistics Canada), Massimo De Cubellis (ISTAT, Italy), Fabrizio De Fausti (ISTAT, Italy), Matjaž Jug (Statistics Netherlands), Joeri van Etten (Statistics Netherlands) and Alex Noyvirt (Office for National Statistics, United Kingdom).

CASE STUDY 16 UNITED NATIONS PET LAB: INTERNATIONAL TRADE

Specific contributors to case study 16 were the following members of OpenMined: Laura Ayre, Jack Bandy, Irina Bejan, Tudor Cebere, Phil Culliton, Kien Dang, Kyoko Eng, Ronnie Falcon, Bennett Farkas, Stephen Gabriel, Baye Gaspard, Shubham Gupta, Madhava Jay, Ionesio Junior, Yemissi Kifouly, Osam Kyemenu-Sarsah, Teo Milea, Ishan Mishra, Curtis Mitchell, George Muraru, Ivoline Ngong, Thiago Porto, Mark Rode, Rasswanth S., Jess Stahl, Kellye Trask, Andrew Trask and Gatha Varma, Francesco Amato (ISTAT, Italy), Mauro Bruno (ISTAT, Italy), Massimo De Cubellis (ISTAT, Italy), J.A. van Etten (Statistics Netherlands), Jack Fitzsimons (Oblivious AI, Ireland), Ronald Jansen (United Nations Statistics Division), Matjaž Jug (Statistics Netherlands), Luke Keller (Census Bureau, United States), Karoly Kovacs (United Nations Statistics Division), Clarence Lio (United Nations Statistics Division), Sean Lovell (United Nations Statistics Division), Katelyn McCall Kiley (Census Bureau, United States), Saeid Molladavoudi (Statistics Canada), Alex Noyvirt (Office for National Statistics, United Kingdom), Benjamin Santos (Statistics Canada), Rob Warmerdam (Statistics Netherlands), Peter Zandbergen (Statistics Netherlands) and Zachary Zanussi (Statistics Canada).

CASE STUDY 17 UNITED STATES CENSUS BUREAU: DEPLOYING A DIFFERENTIALLY PRIVATE DISCLOSURE AVOIDANCE SYSTEM FOR THE 2020 US CENSUS

Specific contributors to case study 17 were Amy O'Hara (Georgetown University, United States) and Wade Shen (Actuate, United States).

CASE STUDY 18 UNITED STATES DEPARTMENT OF EDUCATION: ANALYSING STUDENT FINANCIAL AID DATA USING PRIVACY-PRESERVING RECORD LINKAGE

Specific contributors to case study 18 were David Archer (Galois, Inc., United States), Amy O'Hara (Georgetown University, Massive Data Institute, United States), Rawane Issa (Galois, Inc., United States) and Stephanie Straus (Georgetown University, Massive Data Institute, United States).

CHAPTER 4 STANDARDS

Contributors: Julian Padget (University of Bath, United Kingdom), Wo Chang (National Institute of Standards and Technology, United States). We thank the various British Standards Institute committees that reviewed and commented on a draft of the material presented in this document. We acknowledge the IEEE Standards Association (IEEE SA) for permission to reproduce extracts of Project Authorization Request (PAR) documents. We also acknowledge that permission to reproduce extracts from ISO standards is granted by BSI Standards Limited (BSI). No other use of this material is permitted. We are grateful to all the experts participating in the various standards bodies that have contributed to the standards and standards in development cited in this document.

CHAPTER 5 LEGAL AND REGULATORY ISSUES

Contributors: Sulaika Duijsings-Mahangi (Statistics Netherlands), Kuan Hon (United Kingdom), Yoichiro Itakuri (Higari Sogoh Law Offices, Japan), Julian Padget (University of Bath, United Kingdom), Robert Pisarczyk (Oblivious AI, Ireland), Loretta Pugh (CMS, United Kingdom), Andrew Sellars (Boston University, United States), Mayank Varia (Boston University, United States), Alexandra Wood (Harvard University, United States).

On behalf of the United Nations Committee of Experts on Big Data and Data Science for Official Statistics, we would like to thank all those who have contributed in smaller and larger ways to this guide on privacy-enhancing technologies for official statistics.

ACRONYMS

2PC	Secure Two-party Computation
ABUEA	Attribute-Based Unlinkable Entity Authentication
AI	Artificial Intelligence
API	Application Programming Interface
AWS	Amazon Web Services
BFV	Brakerski-Fan-Vercauteren (HE scheme)
BGV	Brakerski-Gentry-Vaikuntanathan (HE scheme)
BU	Boston University
BWWC	Boston Women's Workforce Council
CA	Central Authority
CARRIER	Coronary ARtery disease: Risk estimations and Interventions for prevention and EaRly detection
CART	Classification and Regression Tree
CBS	Centraal Bureau voor de Statistiek / Statistics Netherlands
CCPA	California Consumer Privacy Act
CDEI	Centre for Data Ethics and Innovation
CDR	Call Data Record
CEN	European Committee for Standardization
CENELEC	European Committee for Electrotechnical Standardization
CKKS	Cheon, Kim, Kim and Song (HE algorithm)
CPU	Central Processing Unit
CRS	Common Reference String
CSV	Comma Separated Values
DAS	Disclosure Avoidance System
DP	Differential Privacy
DP-SGD	Differentially Private Stochastic Gradient Descent
EDPB	European Data Protection Board
EDPS	European Data Protection Supervisor
EEA	European Economic Area

EHR	Electronic Health Record
ENISA	European Union Agency for Cybersecurity
FHE	Fully Homomorphic Encryption
GAN	Generative-Adversarial Network
GDPR	General Data Protection Regulation
GPS	Global Positioning System
GPU	Graphics Processing Unit
GSBPM	Generic Statistical Business Process Model
GUI	Graphical User Interface
HE	Homomorphic Encryption
HI	Hardware Isolation
HLG-MOS	High-Level Group for Modernization of Official Statistics
HTTPS	Hypertext Transfer Protocol Secure
IEC	International Electrotechnical Commission
IMDB	Internet Movie Database
IMSI	International Mobile Subscriber Identity
IND-CPA	indistinguishability Chosen Plaintext Attack
IND-CCA	indistinguishability Chosen Ciphertext Attack
IPP	UNECE Input Privacy Preservation Techniques project
ISI	International Statistical Institute
ISMS	Information Security Management System
ISO	International Organization for Standardization
ISTAT	Italian National Institute of Statistics
IT	Information Technology
JTC	Joint Technical Committee
LAN	Local Area Network
LSS	Linear Secret Sharing
LWE	Learning With Errors
META	Twitter's ML Ethics, Transparency, and Accountability team
ML	Machine Learning

MLP	Multi-Layer Perceptron
MNO	Mobile Network Operator
MOOC	Massively Online Open Courseware
MRI	Magnetic Resonance Imaging
MSS	Management System Standards
NIST	National Institute of Standards and Technology
NPSAS	National Postsecondary Student Aid Study group
NSLDS	National Student Loan Data System
NSO	National Statistical Office
NTTS	New Techniques and Technologies for Statistics
OECD	Organisation for Economic Co-operation and Development
ONS	Office for National Statistics
PAR	Project Authorization Request
PDSI	Private Data Sharing Interface
PET	Privacy Enhancing Technology
PIA	Privacy Impact Assessment
PII	Personally Identifiable Information
PIMS	Privacy Information Management System
PoC	Proof of Concept
PPML	Privacy Preserving Machine Learning
PROM	Patient-Reported Outcome Measures
PSI	Private Set Intersection
PUMF	Public-Use Microdata File
PWI	Preliminary Work Item
RLWE	Ring Learning With Errors
SA	Supervisory Authority
SDC	Statistical Disclosure Control
SFE	Secure Function Evaluation
SGX	Software Guard Extensions

sMPC	Secure Multi-Party Computation
SNARG	Succinct Non-Interactive Argument
SQL	Structured Query Language
SSN	Social Security Act
TDA	TopDown Algorithm
TEE	Trusted Execution Environment
TFHE	Fast Fully Homomorphic Encryption over the Torus
TLS	Transport Layer Security
TSS	Trusted Smart Survey
TTP	Trusted Third Party
UN	United Nations
UNCEBD	UN Committee of Experts on Big Data and Data Science for Official Statistics
UNECE	United Nations Economic Commission For Europe
VAE	Variational Auto-Encoders
VPN	Virtual Private Network
W3C	World Wide Web Consortium
WYSIWYG	What-You-See-Is-What-You-Get (model)
ZK	Zero Knowledge
ZKP	Zero Knowledge Proof

EXECUTIVE SUMMARY

This document presents methodologies and approaches to mitigate privacy risks when using sensitive or confidential data, which are collectively referred to as *privacy-enhancing technologies* (PETs). National Statistical Offices (NSOs) are entrusted with data that has the potential to drive innovation and improve national services, research, and social benefit. Yet, there has been a rise in sustained cyber threats, complex networks of intermediaries motivated to procure sensitive data, and advances in methods to re-identify and link data to individuals and across multiple data sources. Data breaches erode public trust and can have serious negative consequences for individuals, groups, and communities. This document focuses on PETs that protect data during analysis and dissemination of sensitive information so that the benefits of using data for official statistics can be realized while minimizing privacy risks to those entrusting sensitive data to NSOs.

This document explores current approaches to data protection (e.g., data de-identification, input party computation, contractual controls and agreements) and their associated limitations. In order to facilitate experimentation on pilot projects and effective collaboration on “real world” use cases, the UN Privacy-enhancing Technologies Task Team founded the UN PET Lab.^a

The team identified three core components to accelerate the adoption of PETs within the NSO community:

1. **Experimentation (PET Lab):** a series of active proofs-of-concept and pilot projects focused on the evaluation of PETs for real-world use cases in the official statistics community.
2. **Outreach & Training:** focus on sharing learnings and insights from the use of PETs with the wider statistical community through training, public events, and educational material
3. **Support Services:** a mechanism to enable those utilizing PETs to engage with the committee and its collaborators for support and advice

The goals and progress of these three pillars are discussed along with their respective plans for the future.

Then, two broad categories of PETs (e.g., input privacy, output privacy) are introduced, including secure multiparty computation, homomorphic encryption, differential privacy, synthetic data, distributed learning, zero-knowledge proof, and trusted execution environments.

Each section defines a problem that the respective technique can solve and offers its overview and history. With NSO professionals in mind, the primary security considerations and cost of using each technology are presented along with an example use case taken from an NSO domain along with a discussion of practical considerations for choosing appropriate PETs.


Detailed case studies are presented that comprise a diverse range of use cases across sectors, leverage combinations of PETs, and involve collaboration among parties (such as multiple NSOs working together, NSOs working with other government agencies, and NSOs working with private sector organizations). Fifteen of the case studies describe implementations that are in the concept or pilot stage and three that have been deployed in production environments.

This document provides an overview of standards-making activities and identifies several new standards relevant to the processing of datasets, including standards under development and some that are a product of the precautionary principle applied to standards-making for artificial intelligence (AI). There has been a significant increase in standards-related activity relevant to PETs and data in AI, and more specifically, machine learning (ML), since the Privacy Preserving Techniques Handbook.¹ In the case of AI/ML, earlier approaches to standardization sought to draw upon practice and experience collected over a period of time to benefit from *hindsight* whereas the current driver is *foresight* with the goal to prevent potential harms (“known-knowns” and “known-unknowns”).

Given the expansion of activity dealing with PETs and the context in which they may be applied, standards are presented in two parts. The first identifies essential standards with sections on encryption and security techniques. The second considers indirectly related

^a <https://officialstatistics.org/petlab/>

¹ Archer, David W., Borja de Balle Pigem, Dan Bogdanov, Mark Craddock, Adria Gascon, Ronald Jansen, Matjaž Jug, Kim Laine, Robert McLellan, Olga Ohrimenko, Mariana Raykova, Andrew Trask and Simon Wardley (2023). UN Handbook on Privacy-Preserving Computation Techniques. doi: [10.48550/arXiv.2301.06167](https://doi.org/10.48550/arXiv.2301.06167).



standards that could affect the environment – technical and organizational – in which PETs may be deployed, with subtopics on cloud computing, big data, governance, AI, and data quality. For those interested in the “bigger picture”, there is an additional section on Related Standards.

There is increasing awareness of PETs across governmental, commercial and private organizations. The security and privacy properties they offer clearly connect with the values that are increasingly being embedded in legislative and regulatory frameworks. However, because PETs are new and do not map cleanly onto existing laws and regulations, it can be problematic to determine whether they are acceptable to use in any specific scenario. Indeed, that very issue imposes a substantial barrier to the adoption of PETs. Therefore, the final chapter offers an introduction to some of the key issues and underscores the importance of timely integration of legal advice into NSO projects.

1. INTRODUCTION TO PRIVACY-ENHANCING TECHNOLOGIES

1.1 MOTIVATIONS FOR THE USE OF PRIVACY-ENHANCING TECHNOLOGIES

Official statistics are a trusted source of information for governments around the world to make informed and data-driven decisions. As such, the breadth of information is collected from a range of data sources such as household and business surveys, population, economic or agricultural censuses, a variety of administrative records or even private sector data. Those data sources are the inputs for the compilation of statistics and indicators on the economy, the environment and the society. In many ways, official statistics offer a snapshot of a country's development and rate of progress.

Naturally, the more fine-grained the level of input data, the more nuanced the official statistics can be. However, the collection, processing, and dissemination of often sensitive data need to protect the privacy of persons and businesses. Additionally, looking at National Statistical Offices (NSOs) as part of national and international data ecosystems, NSOs could potentially share much more data if able to protect their privacy.

This inevitable tradeoff is the focus of this document, or more concisely: how can we use technology to mitigate privacy risks and give provable privacy guarantees throughout the collection, processing, analysis and distribution life-cycle of potentially sensitive information.

DATA PRIVACY

Protecting data from unauthorized access, processing or distribution is the simple goal of privacy-enhancing technologies. With such broad applicability to the day-to-day processes of official statistics, it is in the interest of every NSO to have an adequate level of understanding of privacy-enhancing technologies.

To highlight this point, let us sketch a simple example from census and survey information. Often governments will try to understand the income levels, job positions, education, race, and religion of their citizens, and the places where they live. This allows the government to monitor the growth or decline of social inequality and injustice in the country and take action accordingly. However, as a citizen

of the country, you may have valid questions and doubt about the security and privacy of the use and dissemination of the data. At first glance, one might assume that this hesitation may be from tax evaders or criminals, but in fact, quite the opposite could be true. Honest citizens may also have a range of fears, from revealing their personal financial positions to neighbors and acquaintances to the fear of persecution due to their ethnicity or religious beliefs. They may even fear that the information collected may be used by private corporations to target marketing campaigns at them without their consent.

NSOs that actively promote and utilize privacy-enhancing technologies have the opportunity to build greater trust with the public and hence unlock new opportunities associated with more accurate and complete data collection.

KEEPING DATA PRIVATE & SECURE

There are many points at which the privacy and security of data used for official statistics may be compromised, from the point at which data is collected, transmitted between parties, stored, processed, and ultimately shared with decision-makers and the public. To mitigate potential risks at each of these stages of the data life cycle, different tools are available to the NSOs.

Some tools required may be very familiar, such as ensuring TLS channels (HTTPS) when transmitting data between entities or ensuring that data is encrypted when it is stored in databases or as flat files on a server. An experienced IT security officer will be able to give many such examples of when encryption, authentication, authorization, and validation can be used in order to make sure data is not inadvertently exposed to inappropriate parties. These are mature domains and are not the focus of this guide.

Despite encryption during transit and at rest being mature, there are still many areas in which data is left insecure and without guarantees of how it is used.

PRIVACY-ENHANCING TECHNOLOGIES

Privacy-enhancing technologies (PETs) are technologies designed to safely process and share sensitive data. As discussed in the next section, there are two broad categories of PETs, namely PETS for input and for output privacy. Input privacy focuses on how one or multiple parties can process data in a manner that guarantees the data is not used outside of that strict context. Output privacy focuses on modifying the results of a computation such that the output data cannot be used to reverse engineer the original inputs. By using these technologies intelligently, safe data life cycles can be constructed, enabling collaboration, trust and providing confidence to data subjects.

THE FOCUS OF THE REPORT

There are, of course, many aspects of the data life cycle that pertain to the management and protection of personal or private information. However, this report focuses on the analysis and dissemination of sensitive information:

- How can we perform analysis and extract insights from data which should not be disseminated?
- How can we aggregate data between parties who may have conflicts of interest in sharing plaintext data with one another?
- How can we guarantee how data has been used?

Equally important is what this report is not about. In some communities, privacy technology implies the means to track and map the usage of data to consent forms, cookie policies, and other legal restrictions. While these topics remain important, the scope of this report excludes them as most of these problems can be addressed with traditional software development and do not require the advanced cryptographic and statistical constructs outlined in the following chapters.

1.2 CHALLENGES IN USING PRIVACY-ENHANCING TECHNOLOGIES FOR OFFICIAL STATISTICS

Privacy-Enhancing Technologies, also referred to as Privacy-Preserving Techniques or even privacy technology, encompass a broad range of technologies that endeavor to achieve the privacy goals set out in the previous section.

In practice, this collection of technologies forms the intersection of two prominent fields: statistics and cryptography. Statisticians typically present the statistical methodologies and constructs that they would like to investigate within a set of privacy constraints, while cryptographers endeavor to craft a set of protocols and mechanisms, which protect each party within the provided formal set of constraints.

As can be imagined, this leads to an ongoing tug-of-war between the flexibility of statistical analysis which can be performed and the enforceability of privacy constraints. Indeed, this is one of the age-old trade-offs seen under the guise of data governance, how to balance data usability with security and compliance.

While in practice there is a never-ending list of possible scenarios in which privacy and statistics interface, for the sake of simplicity and conciseness, we classify techniques into two broad categories: input and output privacy.

INPUT PRIVACY

Input privacy endeavors to allow two or more parties to submit data into a calculation without the other respective parties seeing data in clear. This is actually trickier to achieve in practice than it may first appear.

An example of input privacy is the case where two or more NSOs wish to reconcile their cross-border trade statistics. For each pair of countries, import data compiled by one country can be compared with the export data of the partner country. Whereas neither country is allowed to share transaction-level trade information, it may be possible to exchange useful information regarding, for example, the number of transactions per traded product, the number of transactions per border crossing or the number of transactions per mode of transport. If specific traded products show a large discrepancy, more targeted information could be shared, for example, on the number

of transactions of that product per month, the aggregated trade value per month maybe broken down by border crossing. Some information could be shared on the number of companies involved in the trade of a specific product, on the condition that a minimum number of companies (3 or more) trade in that product. Similarly, information could be made available on the average value per kilogram or the average value per unit of the product. What should not be revealed is the identity of a trading company, or the unit price of a product traded by a specific company.

Broadly speaking, there are three popular approaches to input privacy:

1. Finding a trusted third party or using a trusted legal entity, such as a national court system, to enforce pre-agreed contractual terms of use.
2. Using pure cryptographic-based approaches.
3. Leveraging trusted execution environments.

The trusted third-party approach looks straightforward. In essence, the two parties, which want to share data, would find a trustworthy third party which would receive the sensitive data and perform the calculations as desired. However, this approach does not work for most NSOs, since most NSOs are by law not allowed to share sensitive data with any third party. So, this approach will not be further discussed in this guide.

The use of pure cryptographic protocols is growing in popularity. In chapter 2, we describe Secure Multi-Party Computation (sMPC) and Homomorphic Encryption (HE) in detail. These approaches both use cryptographic primitives to perform calculations on sensitive input data through rounds of communication between the parties. Overall, these approaches offer theoretical guarantees at the protocol level, which may be desirable in some settings. These approaches can be computationally expensive; hence specific attention should be paid to this aspect when applying these PETs, especially when large datasets are involved in the protocols.

Finally, Trusted Execution Environments (TEEs), namely secure enclaves, endeavor to mimic the behavior of a trusted third party by attesting the functionality performed by hardware or by a cloud provider. These approaches

require the trust of the hardware/cloud provided but offer a more flexible functionality over pure cryptography-based approaches. TEEs are also discussed in detail in chapter 2.

OUTPUT PRIVACY

Output privacy is a concept, which is familiar to most official statisticians and is generally known as statistical disclosure control.¹ Output privacy aims to conceal sensitive individual data from being identified or re-identified from the disseminated output.

There are many approaches to output privacy, as can be seen by the rich literature on statistical disclosure controls. However, where these approaches meet the strict formalism of cryptographic research is with the use of differential privacy which offers a concise definition of output privacy that can be calculated for and combined over multiple diverse operations and disclosures. In chapter 2, section 2.3, we will discuss differential privacy in more detail.

CHALLENGES FACED BY PRIVACY TECHNOLOGY

PETs are not yet widely used. We will discuss three major categories of challenges, which currently limit the use of PETs, namely collaboration, the pace of cryptography development and cost.

COLLABORATION

The first big challenge is that privacy technology is often required in domains where there are many stakeholders from multiple organizations. Despite the best of intentions, technology that requires cross- and inter-organizational collaboration can often take a long time to be scoped, built, and ultimately used in a production setting. Friction does not just arise through different norms and processes between organizations, but very often between the language used between stakeholders from different communities, such as the technology, compliance, or legal communities.

THE PACE OF CRYPTOGRAPHY DEVELOPMENT

Privacy technology, at its core, is a form of data security. As such, it requires the scrutiny warranted by other

areas of computer security and cryptography. This is a slow-moving field, similar to other critical science domains such as aviation and pharmaceuticals. Small mistakes or overlooked circumstances can have large consequences. In the case of privacy technology, guarantees are made to the data subjects, that their data are being handled securely and privately. This also tends to lead to only few standards being developed and used commonly, as is seen with Transport Layer Security (TLS) and some other security standards.

The notorious slow pace of R&D in computer security is determined by the time needed for the core research being performed by academics, all the way through to the engineers who carefully implement the software packages for use in production.

COST

Cost can be a major factor when it comes to newly popularized technologies. Like most things of economic value, as something becomes more widely used its costs typically reduce. Some privacy technologies are not widely adopted yet and as a result, a lot of security design and analysis must be performed before they can be used in production. This can make the overhead of one-time technologies very expensive. The hope is that as these tools continue to grow in popularity, some of them will become more widely available, cheaper to use, and easier to support.

CLASSICAL DATA PROTECTION APPROACHES

Although policy or statute often restricts sensitive data sharing among organizations, some sharing does take place, and attempts are made to assure the privacy of the shared data in various ways. Below, we explore current approaches and their shortcomings.

DATA DE-IDENTIFICATION

Input Privacy and Output Privacy are often supposedly protected by de-identifying or anonymizing data—removing portions of the data that might be used to link the remainder to specific individuals—prior to sharing it. Unfortunately, de-identification can often be ineffective and insecure due to potential re-identification attacks.

¹ Hundepool et al., Statistical Disclosure Control [\[2012\]](#).

A variety of techniques have been shown to be able to expose seemingly anonymized data from which personally identifiable information or key attributes have been removed. They include linkage attacks that leverage joint information from external data sources or homogeneity attacks that exploit scarcity of data.

In addition, de-identification will impair the usability of data, lowering the value of statistical results to decision-makers. De-identification can further be expensive because it is often human reviewers who must survey the data and make decisions about which attributes to remove. Finally, de-identification is often specific to the intended computations to be performed, and so must be re-done prior to each distinct use of data.

CONFIDENTIALITY

Confidentiality refers to the legal, ethical and practical obligations that bind the NSOs not to disclose any sensitive information. Statistical agencies currently use various statistical disclosure mitigation techniques to protect the confidentiality and output privacy of their data subjects, while disseminating information of analytical values to their users. There is a direct correlation between the analytical utility of statistical products and the disclosure risk pertaining to the data subjects. For instance, whereas disclosure risk and analytical value are low for global summary statistics, they are both fairly high for multi-dimensional tables with micro-data. The main goal of statistical disclosure methods is to balance the trade-off between data utility and disclosure risks.

Various types of statistical disclosure risks can be identified, including identity, attribute, and inferential disclosures.² Additionally, multiple factors can influence these disclosure risks ranging from the data sources, such as census or survey data, to the analytical outputs, such as micro-data, analytical tables or graphs. Depending on the re-identification risk measures, different disclosure control techniques may be applied, which can be classified into non-perturbative and perturbative approaches. For example, among methods that are appropriate for micro-data dissemination include non-perturbative techniques, such as recoding and sub-sampling, as well as perturbative methods, such as data shuffling and injecting random noise to the data. Please refer to

section 2.3 for more details on noise injection methods, in particular differential privacy. Other existing approaches include coarsening, post-randomization and suppression methods depending on the data type and characteristics.³

The common pitfall of all disclosure control methods is that they have a negative impact on the quality of the products. More explicitly, data suppression methods reduce the information provided to the external users and data perturbation methods modify the data before dissemination, while retaining the information content as much as possible. Even when less information is accessible to the user, there is still some disclosure risk present. In addition to statistical disclosure techniques, NSOs use non-statistical or physical disclosure methods to evaluate and reduce the risks. These approaches include imposing and regulating access control to the data and using secure settings, license agreements (see below) and safe practices to reduce disclosure risks.

INPUT PARTY COMPUTATION

In some cases, input parties may perform computation on behalf of result parties directly, and then pass the results to those parties without the need for distinct compute parties. For example, a telecom company (input party) could perform on-demand computations for a NSO or for a research institute and pass only the results of the computations to them. While this approach provides strong Input Privacy guarantees, it also requires substantial effort on behalf of input parties which need to be willing to invest significant computational resources and may lack the expertise to perform complex computations. Input parties may also not have the scalability of resources to support analysis on behalf of multiple result parties. In addition, this approach requires that result parties provide the methodology and details of the analyses to be performed to the input parties, which result parties may be unwilling to do.

CONTRACTUAL CONTROLS AND AGREEMENTS

The most popular current approach to achieving privacy goals is to rely on legal terms and accountabilities. Input parties may require that compute and result parties contractually agree to keep input data private

² Hundepool et al., Statistical Disclosure Control (2012).

³ Gartner, The State of Privacy and Personal Data Protection, 2020-2022 (2020).

and to strictly control access to computation outputs. Such agreements are ineffective and unsafe, if not also supported by implementation solutions relying on PETs, although they do allow for attribution of blame and potentially for assignment of financial responsibility. In many cases, the financial remedy is often of no use to individuals whose data are compromised and could be assigned to organizations who collected the data rather than to those individuals. In addition, contractual control is ineffective against insider threats or compromise of systems by a cyber-attack.

A PROMISING HORIZON

Despite slow adoption, the future of privacy technology has never looked brighter. Today, there is an influx of investment by both large technology giants and venture capitalists alike, funding new approaches and endeavors in this space. Chip manufacturers continue to increase the effectiveness of secure enclaves, all major cloud providers now support trusted execution providers, and open-source frameworks are becoming extremely popular. A recent whitepaper by Lunar Venture amplified these points.⁴

In parallel to this, commercial and non-commercial entities alike are endeavoring to comply with the ever-increasing global regulation pertaining to privacy technology. By 2023, it is estimated that 65% of the world's population will have their personally identifiable information protected by modern privacy regulations and laws.⁵ Further by 2024, this is said to affect 80% of organizations worldwide. While these regulations to date do not strictly specify or recommend a specific technology to leverage, their enforcement leads to wider investigation and adoption of privacy technology by those looking to be best in class.

⁴ Lawrence Lundy-Bryan, Privacy Enhancing Technologies. Part 2 ([2020](#)).

⁵ Scannapieco et al., "Input Privacy" ([2021](#)).

1.3 A NEW APPROACH TO INTERNATIONAL COLLABORATION ON PETS

The 2019 UN Handbook on Privacy-Preserving Computation Techniques⁶ mostly gave insights on the concepts of privacy technology. The current document also provides information on the methodologies and approaches of these techniques, but in addition elaborates on a large number of use cases, which are described in the chapter 3. The practical application of the PETs shows their value. Will we be able to share sensitive data while protecting privacy? If we manage to do so, we could potentially create value for our societies out of sensitive data, such as health records, population census data, mobile phone records or tax records.

In addition to describing use cases, which have been designed and conducted by others, the members of the task team on PETs also wanted to collaborate on actual use cases themselves. For this purpose it created the UN PET Lab which is described in this section. The objectives of the UN PET Lab are experimentation on pilot projects, learning by doing, and offering support services to those who want to be early adopters of PETs. The UN PET Lab was officially launched on 26 January 2022 at the EXPO 2020 event in Dubai.

UN PET LAB

Over the past months, the task team on PETs has increased its efforts across three core pillars to accelerate the adoption of PETs within the community of official statistics, namely through:

- 1. Experimentation:** Experimentation is advanced through a series of active proofs-of-concept and pilot projects, focused on the evaluation of PETs for real-world use cases in the official statistics community.
- 2. Outreach & Training:** Outreach and training are promoted by spreading shared learnings and insights from the use of PETs to the wider statistical community through training, public talks, and educational collateral.

- 3. Support Services:** Finally, support services are offered through a mechanism to enable those using or intending to use PETs to engage with the UN PET Lab and its collaborators for support and advice.

The combination of people, processes, and systems in place to drive these three pillars are referred to as the United Nations Privacy-Enhancing Technologies Lab, or the UN PET Lab for short. In this section, we outline the goals and progress of these three pillars in more detail, along with their respective plans for the future.

EXPERIMENTATION

The mission of the first pillar of this international collaborative effort is to enable practitioners from the national and international official statistics communities to get hands-on with using PETs. There is a wide range of benefits that come from practically trialing a technology within the context of a known problem space, including:

- 1. Proving Value:** While it is easy to describe hypothetical value creation from privacy technology at a high level, it is important to dig into the nuances of potential projects to understand and demonstrate the full value of these technologies in the context of real-world problems of current interest. Understanding all of the benefits involved helps the community to better evaluate the risk-to-reward calculations involved in kicking off fully-fledged projects.
- 2. Understanding practical challenges:** Unfortunately, utilizing privacy technology in many scenarios brings unforeseen challenges, such as those presented in Chapter 2.8 on practical considerations of PETs. By facilitating the usage of privacy technology within safe experimental environments, participants and collaborators can better understand potential issues, risks, and considerations before committing to using such technologies in production.
- 3. Engaging with stakeholders:** There are many stakeholders involved in any domain of data

⁶ Archer et al., *UN Handbook on PPTs* (2023)

governance, each weighing in with a different perspective from technical feasibility to data security and legal considerations. By running experimental trials and projects within a safe environment, these stakeholders can express their concerns and views ahead of production usage. The learnings from these help us to mitigate such frictions where possible and ultimately reduce the number of unknown barriers to entry for production-level usage.

- 4. Building a privacy technology literacy:** Finally, and certainly not least, active usage of privacy builds a level of literacy within the community. Those involved both, directly and indirectly, develop holistic knowledge about the technology space and associated issues. This development of skills and knowledge will help to grow the community and act as an asset to the wider international statistics community.

In the above benefits, one caveat that is emphasized is that trials are performed in a safe and flexible environment. Two ways in which the PET Lab creates such an environment are by leveraging non-sensitive data initially and by bootstrapping with general-purpose infrastructure. The first point is important in order to reduce the red tape in kicking off work in the first place, as well as eradicating the risks associated with data leakage. One such example has been the use of COMTRADE⁷ data (see chapter 3) and other such publicly available datasets. These datasets are desirable as they often represent data that is known at a more nuanced, and correspondingly more sensitive, level but are not currently used as such.

The second point is important as it allows the group to spin-up ad hoc servers and infrastructure to accommodate various privacy-enhancing technology, especially those approaches which benefit from a third semi-trusted party involved. Fortunately, the UN Global Platform⁸ for Official Statistics infrastructure is available for exactly such scenarios.

These experiments are documented and reviewed by the PET Lab experts, building towards a shared repository of experience and use cases.

OUTREACH & TRAINING

The second focus of the collaboration is aimed at sharing knowledge more broadly within the global community of official statistics. This has been a long-time focus of the PETs task team and a motivation for this very document. However, the PET Lab formalizes these efforts in a structured fashion. There are four types of educational resources disseminated:

- 1. Official Guides & Overviews:** These are documents that give formal collateral to those interested in learning about PETs. This document is an example of such material.
- 2. Talks & Presentations:** This involves active efforts to speak to subgroups and communities within the international statistics community about PETs. The goal of this is to put the discipline on the radar of practitioners who may not be familiar with the space and who would benefit from insights and awareness. Over the past couple of years, the group has been involved in presenting at Eurostat Conference on New Techniques and Technologies for Statistics (NTTS), 63rd ISI World Statistics Congress, and the Road to EXPO Dubai workshop series.
- 3. Use Case Repository:** This is an online resource that gives details of use cases in PETs within the context of official statistics from around the world. This repository, or wiki, is regularly updated and welcomes updates from any person or team who would like to contribute.
- 4. Collaborating with Massively Online Open Courseware (MOOCs):** Finally, in order to spread the knowledge to a broad audience, and in order to certify practitioners based on their learnings, the PETs task team and PET Lab collaborate on MOOCs to disseminate widely the accumulated knowledge.

These ongoing efforts have already brought great tangible results to the community. For example, the collaboration with OpenMind's MOOC on Foundations of Private Computation⁹ has led to over 9,000 learners registering and participating in formalized training on privacy technology. This is one of the largest public disseminations of training on privacy technology at a global scale today.

^{a.} <https://comtrade.un.org/data>

^{b.} <https://unstats.un.org/bigdata/un-global-platform.cshhtml>

^{c.} <https://courses.openmined.org>

SUPPORT SERVICES

Lastly, the PET Lab has begun to open its doors for a free consultation to institutes seeking to utilize PETs. The idea here is to enable those who are actively engaged in the planning, development or deployment of privacy technology to have access to the team to ask questions, pose topics for debate and discussion, request collaboration, and other related activities.

The first example is collaboration with the UNECE Input Privacy Preservation (IPP) Techniques project which has collected and investigated a number of statistical use cases that require protection on the input side. The IPP team initially developed methodology and a template¹⁰ to document use cases and is working on practical experimentation on techniques such as Private Set Intersection and Private Machine Learning. During the experimentation phase teams organized presentations and joint sessions and some IPP project tracks used the UN PET Lab for practical testing ([see use cases in chapter 3](#)).

To expand this to the wider statistical community practitioners can fill in a web form to request for such collaboration. Once the form has been submitted, it is automatically filed and will be reviewed by the PET Lab at the next meeting, typically within one month of submission. The reviewers will confirm the appropriateness of the request for support and assign 1-2 members to take an initial call with the applicant team. From there, the appropriate personnel will be asked to be involved ad hoc, depending on how the team can best support the applicant and the availability of the members.

In order to provide such support, there must be clear limitations to the scope of the request. Given that experts of the PET Lab all contribute on a voluntary basis, the requests should not be highly time-consuming. Equally importantly, the group is unable to take on any of the project liabilities for the applicant party due to the nature of support that can be provided. Nevertheless, the expectation is that this support will be helpful to the wider statistical community.

FUTURE DIRECTIONS

The ultimate goal of these endeavors is to work toward the creation of a community of practitioners, in which members of the community who are actively using PETs can support one another, organize conferences and share knowledge and support at an international level. This model has been successful in the domain of data science, and it is believed that as the usage of PETs continues to increase, a self-supporting community becomes viable.

⁷ Scannapieco et al. ([2021](#))

CHAPTER 1. INTRODUCTION TO PRIVACY-ENHANCING TECHNOLOGIES

BIBLIOGRAPHY

Archer, David W., Borja de Balle Pigem, Dan Bogdanov, Mark Craddock, Adria Gascon, Ronald Jansen, Matjaž Jug, Kim Laine, Robert McLellan, Olga Ohrimenko, Mariana Raykova, Andrew Trask and Simon Wardley (2023). *UN Handbook on Privacy-Preserving Computation Techniques*. arXiv publication: 2023, originally published 2019. doi: [10.48550/ARXIV.2301.06167](https://doi.org/10.48550/ARXIV.2301.06167).

Gartner (2020). *The State of Privacy and Personal Data Protection, 2020-2022*. url: <https://www.gartner.com/en/documents/3989495>. Accessed 2022-06-13.

Hundepool, Anco, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Eric Schulte Nordholt, Keith Spicer and Peter-Paul de Wolf (July 2012). *Statistical Disclosure Control*. ISBN: 978-1-118-34821-5. Wiley.

Lawrence Lundy-Bryan (2020). *Privacy Enhancing Technologies. Part 2 - The Coming Age of Collaborative Computing*. Lunar Ventures. url: <https://docsend.com/view/db577xmkswv9ujap?submissionGuid=650e684f-93eb-4cee-99e8-12a92d5d88a0>. Accessed 2022-06-13.

Scannapieco, Monica, Fabrizio De Fausti, Massimo De Cubellis, Matjaž Jug, Saeid Molladavoudi and Dennis Ramondt (2021). "Input Privacy: Towards a Logical Framework for Defining Official Statistics Scenarios". In: 63rd ISI World Statistics Congress. url: <https://www.isi-web.org/files/docs/papers-and-abstracts/56-day2-ips047-input-privacy-towards-a-logica.pdf>. Accessed 2022-07-01.

2. METHODOLOGIES AND APPROACHES

2.1 SECURE MULTI-PARTY COMPUTATION

PROBLEM DEFINITION

Secure multi-party computation (also called *sMPC*) is a cryptographic technique that mitigates the problem of *input privacy* when two or more (mutually distrusting) parties wish to compute an agreed-on function on data that they (or possibly other parties) provide to that computation, but are unwilling to disclose to others. In other words, *sMPC* is a technology that allows computation over data while preventing any participant from learning anything about the data except what can be learned from the output of the computation. *sMPC* also mitigates the problem of *code assurance* when parties need to know what function is computed on their shared data. That is, *sMPC* assures (depending on the specific choice of protocol) that the function computed on the data is the same as that agreed on by the parties

EXAMPLE USE CASE

sMPC has been applied to many use cases.¹ An illustrative use case is that of sharing individually identifiable data among a group of several government agencies to compute statistics and make policy decisions based on those statistics. For example, a recent use case² allowed five distinct agencies in County Government in the USA to share their unique data and compute the answers to queries such as, “How many persons that were incarcerated during a certain period had previously taken advantage of publicly provided mental health services or public housing?” The data provided by each agency included personal identifiers (for example, Social Security numbers), along with personal data such as mental health visit records and criminal records. *sMPC* was used to allow queries to be answered while keeping the input data strictly confidential to each party that provided it.

In another use case, the Italian National Institute of Statistics (ISTAT) and the Bank of Italy have run a private set intersection protocol with analytics to enrich their

statistics using information from both organisations such as age, number of children from ISTAT and mortgage information from the Bank of Italy.³ It enabled them to perform analytics on the joint subset of individuals, identified by a unique tax code without sharing directly any of this sensitive data.

OVERVIEW

sMPC computation is based on one of several technologies. The most common technology choices are *circuit garbling* and *linear secret sharing*. The former is typically used in the case of two parties, while the latter may be used for groups of two to many parties. In both technologies, parties first agree on a function to be computed, and express that function as a *logic circuit*. While many functions can be described as circuits, some cannot, so not all functions are practically computable in *sMPC*. While a given set of *sMPC* primitives can technically be Turing complete, typical *sMPC* protocols are non-branching, fixed-length programs in order to be reasonably efficient. This behavioural property may be likened to the difference between a sequencer, which does not support data determined branch conditions and uses a fixed number of gates for processing, and a general purpose computer.

Circuit garbling protocols typically involve two parties. After agreeing on the function to be computed, one party assumes the role of *garbler*, while the other assumes the role of *evaluator*. The garbler takes the agreed-on circuit and creates one or more *encryptions* of the circuit. A circuit encryption defines a randomly chosen value to represent the nominal logic values on each *wire* in the circuit. In addition, circuit encryption encrypts the functions of the logic gates in the circuit. The garbler can then communicate the encrypted circuit to the evaluator, but does not communicate the encryption keys to the evaluator. Thus the evaluator can evaluate the circuit without knowing the actual values on the circuit's wire signals. The garbler also sends encryptions of her input

¹ Archer et al., “Applications of *sMPC*” (2018).

² Hart et al., *Privacy-Preserved Data Sharing* (2019).

³ See [Case Study 5](#) in Chapter 3.

data, using the same keys, to the evaluator. Through an additional cryptographic protocol, the evaluator can work with the garbler to encrypt the evaluator's inputs to the circuit, in such a way that the garbler learns nothing about those inputs. The evaluator then evaluates the encrypted circuit on the encrypted inputs, achieving an encrypted output. That output is returned to the garbler to be decrypted.

There are several open source software libraries that implement garbled circuit technology. Some operate only on *Boolean* gates - gates that have only logical 0 or 1 as input and output - while some operate on *arithmetic* gates that may have many possible input and output values.

Linear secret sharing (LSS) protocols proceed by dividing each input from a party into *secret shares* that are themselves random, but when combined (for example, by addition) recover the original data. sMPC relies on dividing each data input item into two or more such shares, and distributing these to compute parties. The homomorphic properties of addition and multiplication allow for those parties to compute on the shares to attain *shared results*, which when combined produce the correct output of the computed function. To perform the shared computation required for sMPC, all participating compute parties follow

a *protocol*: a set of instructions and intercommunications that when followed by those parties implements a distributed computer program.

There are several open source software libraries that implement LSS technology. As with circuit garbling, these libraries may operate on Boolean values, arithmetic values, or both, including floating point values.

It should be noted that all sMPC protocols use communication among the compute parties frequently. In fact, estimations of run-time for sMPC protocols can be quite accurate using communication cost as the only estimating factor (that is, ignoring estimates of computation delay at compute parties entirely). Thus the complexity of computation is most easily seen in sMPC by its impact on *network communication cost*.

Many modern applications of multiparty computation endeavour to leverage the benefits of more than one sMPC approach such as efficiency or functionality, thus switching between linear secret sharing and garbled circuits. An example of a popular open source library which performs this is the ABY framework by the Cryptography and Privacy Engineering Group at TU Darmstadt and the corresponding compiler for it, EZPC, by Microsoft Research.

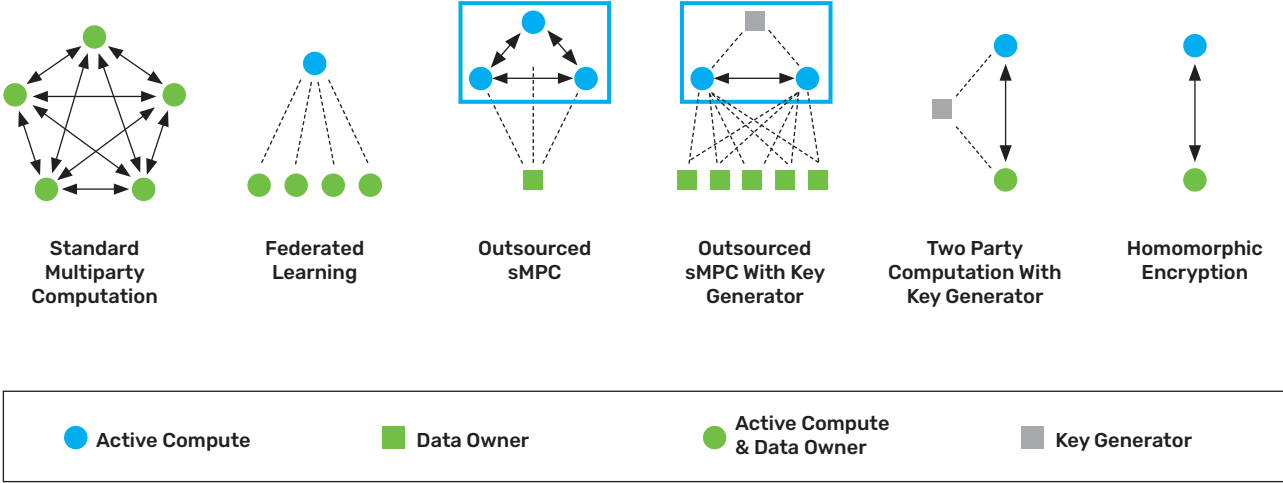


Figure 2.1.1: An overview of some of the ways sMPC and related technologies can be leveraged to preserve privacy under different settings.

HISTORY

sMPC was first formally introduced as secure two-party computation (2PC) in 1982 (for the so-called Millionaires' Problem), and in more general form in 1986 by Andrew Yao.^{3,4} The area is also referred to as Secure Function Evaluation (SFE). The two-party case was followed by a generalization to the multi-party case by Goldreich, Micali and Wigderson.⁵

The high reliance on both available network bandwidth and network latency between parties kept sMPC mainly a theoretical curiosity until the mid 2000's when major protocol improvements led to the realisation that sMPC was not only possible, but could be performed for useful computations on an internet latency scale. sMPC can be now considered a practical solution to carefully selected real-life problems (especially ones that require mostly local operations on the shares with not much interactions among the parties). Distributed voting, private bidding and auctions, sharing of signature or decryption functions and private information retrieval are all applications that exhibit these properties. The first large-scale and practical application of multiparty computation (demonstrated on an actual auction problem) took place in Denmark in January 2008.⁶

A characterisation of available commercial and Government sMPC solutions would be almost immediately out of date, as would cataloguing the plethora of academic sMPC research tools. Instead, for the purposes of providing some practical illustrations of the technology, we point towards some well known open-source sMPC frameworks and use cases documented by private companies.

As sMPC continues to grow in popularity, so too do the number of academic-developed open source frameworks, which are typically used for experimental implementations and testing. One of the more popular of those is the

ABY framework from TU Darmstadt,^b a framework that supports mixed primitive sMPC. Microsoft Research has also built a compiler for ABY call EZPC.^c It is worth noting that while the compiler on top of ABY does not explicitly emphasize its experiential nature, any framework built on ABY will inherit its caveats for production environments. There is also a growing number of public domain complete sMPC systems. These are either general libraries, general purpose systems or systems that solve a specific application problem. In each of these three categories, we note

- SCAPI (from Bar-Ilan University)^d - an API over various sMPC primitives
- the SCALE-MAMBA (from KU Leuven)^e - a complete sMPC system
- swanky (from Galois Inc.)^f - a set of Rust libraries for secure sMPC with garbled circuit, oblivious transfer, private set intersection protocol
- Motion (from TU Darmstadt, Aarhus University and the University of Hamburg)^g - a mixed protocol sMPC framework
- JIFF (from Boston University)^h - a library allowing users to build applications JavaScript on top of sMPC protocols
- CrypTen (from Facebook)ⁱ - secure training and inference of machine learning models using sMPC

Examples of such systems in commercial settings include the Sharemind statistical analysis system by Cybernetica, and cryptographic key management systems from Sepior and Unbound Tech. Other companies offer design consultancies in specific areas based on sMPC technology. For example, Partisia helps design market mechanisms based on sMPC on a bespoke basis and Oblivious deployed sMPC as part of the contact-tracing effort for COVID-19 in India.

³. Yao, "Protocols for secure computations" (1982).

⁴. Yao, "How to generate and exchange secrets" (1986).

⁵. Goldreich et al., "A Completeness Theorem for Protocols" (2019).

⁶. Bogetoft et al., "Secure Multiparty Computation Goes Live" (2009).

^b <https://github.com/encryptogroup/ABY>

^c <https://github.com/mpc-msri/EzPC>

^d <https://cyber.biu.ac.il/scapi/>

^e <https://homes.esat.kuleuven.be/~nsmart/SCALE/>

^f <https://github.com/GaloisInc/swanky>

^g <https://github.com/encryptogroup/MOTION>

^h <https://github.com/multiparty/jiff>

ⁱ <https://crypten.ai/>

SECURITY MODEL

Because sMPC assumes the possibility of mutually distrusting parties, it also assumes a new class of adversary: one that controls one or more participants in the computation. Such an adversary might be an insider threat, or might be a Trojan or other penetrative, long-lived attack from outside an organization. This new class of adversary is typically described in terms of several traits in the literature: degree of honesty, degree of mobility, and proportion of compromised compute parties.

Honesty. In the *semi-honest* or *honest-but-curious* adversary model, such control is limited to inspection of all data seen by the corrupted participants, as well as an unlimited knowledge about the computational program they jointly run. In the *covert* model, an adversary may extend that control to modifying or breaking the agreed-upon protocol, usually with the intent of learning more than can be learned from observation alone. However, in this model the adversary is motivated to keep its presence unobserved, limiting the actions it might take. In the *malicious* model, an adversary may also modify or break the agreed-upon protocol, but is not motivated to keep its presence hidden. As a result, a malicious adversary may take a broader range of actions than a covert adversary. When non-technical stakeholders consider encryption as a risk mitigator, they typically assume a covert or malicious security model. Thus the honesty model should ideally be clearly communicated to all stakeholders to confirm its suitability for purpose.

Mobility. A stationary adversary model assumes that the adversary chooses a priori which participants to affect. Such a model might represent for example that one compute participant is compromised, but others are not. Stronger versions of this adversary mobility trait allow for an adversary to move from participant to participant during the computation. At present, a real-world analog of such an adversary is hard to imagine.

Proportion of compromised parties. sMPC adversary assumptions fall into one of two classes: honest majority, and dishonest majority. Just as there are a variety of participant adversary models for sMPC, there are also diverse sMPC protocols that provide security arguments that protect against those adversaries. Security is typically argued by showing that a real execution of an sMPC protocol is indistinguishable from an idealized simulacrum where all compute parties send their private inputs to a trusted broker who computes the agreed-upon function and returns the output. The number of parties that can

be compromised is highly protocol dependent, but as a general rule-of-thumb, the greater the proportion one wishes to defend, the higher the protocol overheads.

COSTS OF USING THE TECHNOLOGY

sMPC technology performance depends heavily on the functions to be securely computed. A typical metric for sMPC performance is *computational slowdown* – the ratio of the latency of computation in sMPC to the latency of the same computation done without sMPC security. For general computation such as the calculations needed to process typical relational database query operators, recent results show a slowdown up to 10,000 times.

Within the research community, the performance of sMPC is often benchmarked via a number of metrics such as the numbers of rounds of communication, the volume of data communicated, and the complexity or latency of the computation involved. sMPC such as linear additive protocols or garbled circuits, are considered efficient when compared to homomorphic encryption. However, this is achieved due to a greater number of parties being involved and, typically, a larger number of rounds of communication being required.

While it remains tricky to give guidance on where sMPC might be performant and where it might not, we can offer some general guidelines. Computations that rely heavily on addition, such as summations, are typically faster than general computation, while computations that rely on division or other more complex functions are typically much slower. sMPC is typically designed to operate on integers via Galois fields. This can be easily extended to fixed-point arithmetic, but floating point operations are much less easily represented and can require orders of magnitude more resources. As such, they are typically avoided. Computations that rely on generative functions such as random number generation are also typically slow. In contrast to homomorphic encryption, a specific 2-PC technique discussed in the next chapter, which currently only supports polynomial functions, general sMPC offers a broader set of possible operations.

2.2 HOMOMORPHIC ENCRYPTION

PROBLEM DEFINITION

Homomorphic encryption (HE) is a cryptographic technology that allows for direct computation (addition and multiplication) on encrypted data. It enables a party that provides data to outsource a computation. That is, no party aside from the party providing the data learns anything about the data in homomorphic encryption computations. Furthermore, only the party providing the data has the key with which to decrypt the output. Code assurance is not yet practical in homomorphic encryption, and code privacy is not possible in these technologies. Typically, the HE security model offers privacy to the input provider, but not from the algorithm provider. Under such a scenario, there may be an ahead of time agreement to apply a particular circuit function, but there is no mechanism to confirm that the agreed circuit was indeed applied.

EXAMPLE USE CASE

A commonly cited class of applications for homomorphic encryption is in the medical domain, where regulations enforce strict patient data privacy measures, but hospitals and medical clinics may nevertheless want to enable third-party service providers to analyze, evaluate, or compute on their data without directly sharing such data. For example, a service provider may offer an image analysis service for detecting tumors in MRI scans. A predictive model can be evaluated directly on homomorphically encrypted data, avoiding the issue of medical data leaking to the service provider.

For data storage providers a potential application is in performing analytics on encrypted customer data. For example, a customer may want to store a large encrypted database using a cloud storage service and not have to download the entire database for simple computational queries, as this creates unnecessary logistical challenges and potentially exposes the full dataset to a potentially

low security computation environment. Instead, all possible aggregation of the data should be performed in encrypted form directly by the cloud storage provider avoiding unnecessary exposure of the data to the client's machine.

In a similar context, Statistics Canada has used homomorphic encryption to train a neural network to classify product descriptions from scanner data.^j The data came from retailers whose brands name and prices of various products were sensitive. Using HE has increased the security and privacy levels while allowing the cloud provider to be the compute party.

OVERVIEW

Homomorphic encryption refers to a family of encryption schemes with a special algebraic structure that allows computations to be performed directly on encrypted data. Homomorphic encryption offers post-quantum security, but can result in a high computational overhead and large expansion of data representation. Thus, ideal applications have a relatively small but critical encrypted computation component, include a persistent storage aspect, and are hard or impossible to implement using other methods.

The most commonly used (fully) homomorphic encryption schemes at this time are the Brakerski-Gentry-Vaikuntanathan (BGV)⁷ and the Brakerski-Fan-Vercauteren (BFV)^{8,9} schemes. Both allow encrypted computation on vectors of finite field elements. The trade-offs between the different schemes are complicated and can be difficult to understand even for experts in the field. For very large and very small computations the BGV scheme has a performance advantage over the BFV scheme, but in many other cases the difference is negligible with modern optimisation techniques. On the other hand, the BGV scheme is more complicated and has a steeper learning curve than the BFV scheme. Other schemes have been proposed, but some have been shown to be insecure.

^j See [Case Study 9](#) in Chapter 3.

⁷ Brakerski et al., "Leveled Fully Homomorphic Encryption" (2014).

⁸ Fan et al., *Somewhat Practical Fully Homomorphic Encryption* (2012).

⁹ Brakerski, "Fully Homomorphic Encryption" (2012).

Another recently popular approach is the CKKS algorithm,¹⁰ implemented in many open-source frameworks. It provides approximate arithmetic on real or complex numbers. This is a promising direction of research, but with the caveat that recent attacks have been shown to be devastating to the regime unless mitigating nuances are correctly employed as part of the protocol.¹¹ As always when it comes to cryptography, we should be extremely cautious to deploy relatively new mechanisms without thorough investigation and assessment.

While in principle fully homomorphic encryption schemes allow arbitrary computation on encrypted data, in practice almost all efficient implementations use a so-called *levelled mode* where the encryption scheme is configured to support computations of only a specific or bounded size, typically resulting in significant performance improvements. For simplicity, in this handbook we freely use the term Homomorphic Encryption (HE) to refer to either Fully Homomorphic Encryption (FHE) or Levelled Fully Homomorphic Encryption.

HISTORY

Encryption schemes that support one single type of arithmetic operation (addition or multiplication) have been known since the 1970's¹² and are often said to be *singly* or *partially* homomorphic. The practical value of such a "homomorphic property" was first recognised and explored by Rivest, Adleman, and Dertouzos.¹³ In 2009 Craig Gentry described the first so-called *fully* homomorphic encryption scheme¹⁴ that allows both additions and multiplications to be performed on encrypted data. This was a significant invention, because in principle such an encryption scheme can allow arbitrary Boolean and arithmetic circuits to be computed on encrypted data without revealing the input data or the result to the party that performs the computation. Instead, the result would

be decryptable only by a specific party that has access to the secret key – typically the owner of the input data. This functionality makes homomorphic encryption a powerful tool for cryptographically secure cloud storage and computation services and also a building block for higher-level cryptographic primitives and protocols that rely on such functionality.

While theoretically powerful and academically interesting, the first homomorphic encryption schemes quickly turned out to be unusable in terms of performance and key size. A significant amount of work was done over the next few years in inventing and implementing both simpler and faster homomorphic encryption schemes. This work culminated in the release of the homomorphic encryption library HElib¹⁵ by IBM Research, which improved the performance over prior homomorphic encryption implementations by several orders of magnitude. Today there are multiple open source homomorphic encryption libraries available implementing a variety of homomorphic encryption schemes suitable for different applications. These include

- Microsoft SEAL^k – implementing both BFV and CKKS schemes. For the latter, Microsoft has also released a Python compiler that takes charge of choosing appropriate encryption parameters, rescaling and relinearization operations.
- PALISADE^l – supporting a range of different schemes and variants thereof including not BFV, BGV, CKKS, Levelled Somewhat HE and others
- TFHE (from Inpher)^m – a implementation of TFHE – Fast Fully Homomorphic Encryption over the Torus
- Concrete (from Zama.ai)ⁿ – implementing a variant of TFHE

¹⁰ Cheon et al., "HE for Arithmetic of Approximate Numbers" (2017).

¹¹ B. Li et al., "Security of HE on Approximate Numbers" (2021).

¹² Rivest et al., "Digital Signatures and Public-Key Cryptosystems" (1978).

¹³ Rivest et al., "On data banks and privacy homomorphisms" (1978).

¹⁴ Gentry, "A fully homomorphic encryption scheme" (2009).

¹⁵ Halevi et al., *Design and implementation of HElib* (2020).

^k <https://github.com/microsoft/SEAL>

^l <https://palisade-crypto.org/>

^m <https://inpher.io/tfhe-library/>

ⁿ <https://github.com/zama-ai/concrete>

SECURITY MODEL

Today, all homomorphic encryption schemes with close to practical performance are based on the *Learning With Errors*¹⁶ (LWE), or *Ring Learning With Errors*¹⁷ (RLWE) problems. In other words, one can show that if these homomorphic encryption primitives can be efficiently broken, then either LWE or RLWE can be efficiently solved for specific parameterisations. As LWE and RLWE have been studied extensively and are believed to be very hard to solve, there is strong reason to believe that the corresponding homomorphic encryption schemes are secure.

As homomorphic encryption refers only to a type of encryption primitive and not a protocol, its security definition states merely that, given a ciphertext, an adversary without the secret key cannot obtain any information about the underlying plaintext. However, for secure uses of homomorphic encryption it is critical that no information about decrypted data is ever communicated back to the source of the corresponding encrypted data, unless that source is trusted not to misbehave; this includes seemingly innocuous information, such as a request to repeat a protocol execution, refusing to pay for a service, or revealing any change in behavior that can be expected to depend on the outcome of the encrypted computation. As a result, outsourced storage and computation involving a single data owner should be considered as the primary use-case of homomorphic encryption. After receiving the result, the secret key owner must not perform any action that is observable to the service provider based on the decrypted result to avoid the attacks described above.

In technical terms it means that HE is typically proven to be secure under the indistinguishability Chosen Plaintext Attack (IND-CPA) model and not indistinguishability Chosen Ciphertext Attack (IND-CCA1) or IND-CCA2. What this means in non-technical terms is that HE does not give security guarantees if the adversary gets hold of decryptions of selected cipher texts. The aforementioned attack on CKKS scheme has exploited this and shown that IND-CPA security in this approximate scheme is not sufficient in practical scenarios. An adversary with access

to a decryption of a cipher text can recover the client's private key.

Another subtlety is that most homomorphic encryption schemes do not provide *input privacy* for more than a single party, since there is only one secret (decryption) key: if a computation depends on the private encrypted input of two or more parties, the encryption scheme is not guaranteed to protect these inputs *from the owner of the secret key*. Homomorphic encryption is also *malleable* by nature, so anyone intercepting a ciphertext can modify the underlying plaintext unless, for example, the ciphertext is cryptographically signed by the sender.

It is important to understand that homomorphic encryption is a low level cryptographic primitive and building secure protocols from it is not possible without the help of a cryptography expert. Even in the simplest cases such protocols can result in unexpected or unintended security gaps. Most homomorphic encryption based protocols can be proved to be secure only in a semi-honest security model, although there are exceptions where a stronger security model is achieved by combining homomorphic encryption with other primitives.¹⁸

COSTS OF USING THE TECHNOLOGY


The use of homomorphic encryption comes with at least three types of costs: message expansion, computational cost, and engineering cost.

In HE systems, encrypted data is typically significantly larger than unencrypted data due to encoding inefficiency (converting real data into plaintext elements that can be encrypted) and inherent expansion from the encryption scheme (ratio of ciphertext size to plaintext size). Depending on the use-case, encoding inefficiency can range from the ideal case (no expansion at all) to an expansion rate measured in the tens or hundreds of thousands when the encoding method is poorly chosen. Thus in most cases, one should not think of encrypting large amounts of data with homomorphic encryption, but instead carefully consider what data exactly will be

¹⁶ Regev, "On Lattices and Cryptography" (2009).

¹⁷ Lyubashevsky et al., "On Ideal Lattices over Rings" (2013).

¹⁸ H. Chen et al., "Labeled Private Set Intersection" (2018).



needed for the desired encrypted computations and encrypt only that.

The computational cost of homomorphic encryption is significant *compared to unencrypted computation*. The exact cost depends strongly on the parameterisation of the encryption scheme and whether throughput or latency is measured. Namely, most homomorphic encryption schemes support natively high-dimensional vectorized computations on encrypted data, and if this vectorisation can be fully utilised it can increase the throughput by a factor of up to 1,000 or so.

Developing complex systems with homomorphic encryption can be challenging and should always be done with the help of an expert, making the initial cost for such solutions potentially high. There are two reasons for this: the security model – as discussed earlier – can be hard to comprehend and evaluate without special expertise, and the available homomorphic encryption libraries can be hard to use to their full potential without a deep understanding of how they work. It should also be noted that homomorphic encryption can be hard or impossible to integrate with existing systems. Instead, sophisticated applications of this technology can require substantial changes in existing data pipelines, data manipulation procedures and algorithms, and data access policies.

2.3 DIFFERENTIAL PRIVACY

PROBLEM DEFINITION

Differential privacy (DP) provides an information-theoretic notion of Output Privacy. Its goal is to quantify the maximum amount of information about individual records in a database that could be leaked by releasing the result of any computation on that database. Keeping this amount small ensures that the individuals are protected irrespective of any side knowledge or post processing by an attacker.

DP provides a more general notion of privacy as it covers any type of information derived from a database, contrary to other specialized definitions such as k -anonymity¹⁹ or l -diversity²⁰ which only apply to the release of aggregates. Furthermore, DP was designed to avoid pitfalls that previous attempts to define privacy loss incurred, especially in the context of either multiple releases or when adversaries have access to side knowledge. We note that such pitfalls also affect less sophisticated attempts at privacy preservation, such as aggregation alone or *ad hoc* noise addition to aggregate results.

EXAMPLE USE CASES

Differential privacy is just over 15 years old as of this report and is being implemented in more and more industrial applications in database analysis, statistics, and machine learning. In recent years, some generic DP systems have been open sourced or made commercially available providing the first production-ready implementations. The interest generated by the solid principles behind DP and the growing concerns about online privacy have led to a number of real-world deployments, typically using *ad-hoc* algorithms for specific applications.

Two well-known applications of DP are its use in Google Chrome and Apple's iOS/OSX to collect usage statistics in a privacy-preserving way. These applications follow the local model of DP, where each individual user privatizes their own data before sending it to a centralized server

for analysis. For example, Chrome used this approach to discover frequently visited pages in order to improve its caching and pre-fetch features, while iOS uses it to discover words and emojis frequently used in a texting application to improve the language models used in typing assistance. Additionally, Microsoft also announced that they employ DP in the local model to collect telemetry data from devices running their operating systems.

The most well-known usage of the curator model is by the U.S. Census Bureau, who has released the results of the 2020 Census with differential privacy controls.⁰ This was motivated by research showing that without the kind of protection provided by differential privacy it is sometimes possible to recover accurate information about individuals from Census data through aggregate statistics at different levels of granularity alone.

OVERVIEW

Differential privacy specifies a property that a data analysis algorithm must satisfy in order to protect the privacy of its inputs. In this sense, DP is a privacy standard, rather than a single tool or algorithm. The DP property is stated in terms of an alternate world where the input of a particular individual has been removed from or added to a database. DP requires that the outputs produced by the algorithm in the real and alternate world are statistically indistinguishable. Being a property of the algorithm means that such indistinguishability must hold regardless of what the database is and which individual we choose to remove or add. DP is therefore not a property of the output, and cannot be measured directly by looking at the output of the algorithm on a given input database. Another crucial remark about the definition of DP is that the indistinguishability requirement is too strong to be satisfied by any deterministic algorithm. Randomness is therefore an indispensable ingredient in the design of any differentially private algorithm.

The need for a robust definition of privacy becomes more

¹⁹ Sweeney, "k-Anonymity: A Model for Protecting Privacy" (2002).

²⁰ Machanavajjhala et al., "L-diversity: Privacy beyond k-anonymity" (2007).

⁰ <https://www.census.gov/library/fact-sheets/2021/comparing-differential-privacy-with-older-disclosure-avoidance-methods.html>

compelling as access to side-knowledge has become more widespread. Most anonymization techniques previously assumed that attackers would not be able to re-identify a user from data points that are not obviously identifying, such as names, dates of birth, or addresses. But the proliferation of public information makes it possible to leverage data sources that were once unconceivable. Researchers famously re-identified anonymized Netflix history by using public data from IMDB.²¹ Since then, the risk has only compounded with the proliferation of information made public on social networks or stored in public or private data warehouses.

Differential privacy has become a natural standard that researchers use when evaluating the privacy risks of releasing the output of a computation. It is versatile enough to be applied to any data processing flow. For instance, a common scenario is when data is produced on a terminal node (a user device), then sent to a trusted party (curator or aggregator) that does some computation before releasing some output. One could study the output privacy of the transfer of information from the terminal node to the aggregator just as they could study the output privacy of what the aggregator releases. The former is sometimes referred to as local privacy, the latter as global privacy. Both address different threat models but can be studied in the formalism of DP, that is to say the level of access of information and trust associated with each party differs. If the curator is trusted, individuals may send their information directly to them for the purpose of running a differentially private data analysis algorithm whose output is released. We note that the curator model can be combined with input privacy-preserving techniques such as multi-party computation, a technique that protects the input data between the terminal nodes and the curator.

The interested reader should consult the recent paper by Nissim et al.²² for a more extensive non-technical introduction to DP. Additionally, monographs by Dwork and Roth²³ and Vadhan²⁴ provide a comprehensive account of the basics of differential privacy from a technical perspective.

HISTORY

Historically, DP is related to the privacy models classically studied in the literature on statistical disclosure control and statistical databases. These methods were to release statistics based on aggregates to retain privacy of individuals. However, in the 1990s a range of attacks on datasets such as linkage attacks have been performed. Famously, Latanya Sweeney joined voter registration list with the supposedly anonymised hospital data to de-anonymise health records of individuals in Massachusetts. She also introduced k-anonymity, which makes sure that any record is indistinguishable from at least the other k-1 records in the dataset, to deal with such attacks. However, k-anonymity has also been shown to be insufficient for a range of applications. In 2003, Irit Dinnur and Kobbi Nissim presented reconstruction attacks that reconstruct database records from multiple queries with noisy answers.²⁵ This paved the way for the formal introduction of differential privacy in 2006 by Dwork et al.²⁶

Since then, differential privacy has been heavily researched, with a range of mechanisms being introduced and implemented. Currently, several libraries are offering open source implementations of the main differentially-private primitives. Those primitives are the toolbox upon which more complex solutions are built. The libraries are mostly used in academic research or for experimentation. Some libraries include higher level mechanisms such as a SQL engine, which intercept SQL queries and add appropriate noise to return differentially private outputs. Others, focus on differentially private ML model training. The mechanisms there are based on appropriate modification of stochastic gradient descent, wherein the gradient is clipped, to limit its dependence on individual points and it is also perturbed by noise addition. The higher-level libraries combine several differentially-private primitives and typically come with a privacy accountant to manage the composition of privacy parameters across several queries.

²¹ Narayanan et al., "Robust De-anonymization of Large Sparse Datasets" (2008).

²² Wood et al., "Differential Privacy: A Primer" (2018).

²³ Dwork et al., "The Algorithmic Foundations of Differential Privacy" (2014).

²⁴ Vadhan, "The Complexity of Differential Privacy" (2017).

²⁵ Dinur et al., "Revealing information while preserving privacy" (2003).

²⁶ Dwork et al., "Calibrating Noise to Sensitivity in Private Data Analysis" (2006).

The main libraries available as of this writing are:

- OpenDP/SmartNoise Core (DP primitives, accountant) and SmartNoise SDK (SQL engine over Smartnoise Core)
- Google DP (DP primitives, accountant, SQL engine)
- TensorFlow Privacy (DP-SGD)
- Pytorch Opacus (DP-SGD)
- IBM Diffprivlib (DP primitives, some machine learning models)
- Diffpriv (DP primitives)

These open-source libraries are designed to experiment with differential privacy but do not constitute production ready solutions. They focus on being able to output the result of a differentially private mechanism. They leave it to the implementer to enforce that the application is actually differentially private. The Private Data Sharing Interface (PDSI) developed by the Privacy Tools project led by Harvard University implements a generic methodology for providing differentially private access to sensitive datasets. PDSI focuses on typical use cases in the social sciences, allowing researchers to upload sensitive datasets, release a set of selected statistics with differential privacy, and allow other researchers to create their own DP queries against the dataset. The tool comes with a graphical user interface that guides data owners through the release process, helping them create an appropriate privacy budget and also select from a number of readily available statistics.

SECURITY MODEL

Differential privacy offers a mathematical guarantee to individuals contributing sensitive data to a database on which certain queries will be performed. The guarantee takes the form of a bound on the risk incurred by individuals contributing their data, and builds upon the intuition that queries that are invariant to removal of any single record of a database are immune to attacks regardless of the side-knowledge of the adversary including reconstruction attacks, which reconstruct data from aggregates or membership-inference attacks which can determine if particular records were in the training datasets. In other words, differential privacy provides a convincing argument for a user to contribute data to a database, as it guarantees that query results will be very similar regardless of whether the user joins the database or not. This characteristic of DP protects individuals against attacks where an adversary is allowed to query the database and has access to unlimited side knowledge.

More formally, the release of a computational result on a

database is differentially private if an adversary observing this release will not be able to determine if any particular record was present in the database. This guarantee takes a statistical flavor: since DP requires that the data analysis algorithm must be randomized, the adversary's inability of determining the presence of a record in the database is measured in terms of the similarity between the probability distributions over outputs when the record is either present or missing in the database. This similarity measure is parametrized numerically (typically represented by greek letters epsilon and delta), with smaller values of these parameters representing a stronger privacy protection. Although these values have a very precise statistical interpretation, there is no general application-agnostic recipe for choosing appropriate values of these parameters – one of the current limitations in usability of DP.

DP provides privacy even in the context of adversaries with access to arbitrary side-knowledge. Side-knowledge may even include data sources that could be made available in the future and any computational capabilities of an adversary. It can be demonstrated that protecting against side-knowledge requires some randomization in the algorithm. A user with access to the output of a deterministic algorithm on two datasets differing by one individual may be able to learn something about this individual with absolute certainty, which would be a blatant privacy breach. All deterministic algorithms are therefore subject to re-identification attacks with the use of side-information. Traditional anonymization techniques like removing fields, reducing accuracy or aggregating values are often deterministic constructs. They therefore need to make strong assumptions about the side-information accessible to the recipient in their threat model. Differential privacy makes no such assumption and can therefore protect against much stronger attackers.

The DP formalism can be applied to any computation from a single database query to all the iterative steps required to train a machine learning model. Composition theorems make it possible to analyze the differential privacy parameters resulting from complex computational processes. This makes it possible to apply DP to various threat models. For instance, DP can be applied to a scenario where the attacker may access the output of a single query or when the attacker may access the output of an arbitrary number of queries. Obtaining the same privacy guarantee in those two threat models will lead to different parametrization of DP at the query-level.

DP only addresses the privacy of an output of a flow of information (Output Privacy). It does not solve the privacy risks when managing input data between where it is collected, stored, and eventually processed (Input Privacy). Even when using DP, the full threat model should consider the trust assumptions on the overall system. For instance, in the local and curator models, the threat model will differ in that the latter assumes a trusted party will collect the data to be analyzed and release the results of such analysis using DP, while the former makes no trust assumptions on the entity collecting the data. From a statistical perspective, local privacy can be much harder to achieve in the strictest application of DP and require the addition of significant noise. This stems from the fact that the output from a terminal node to the centralized party should be indistinguishable to what would be sent hadn't the terminal node existed at all. To make it more tractable, implementations usually make assumptions on the observables accessible to the receiver and their range of possible values.

PRIVACY BUDGET

Differential Privacy is achieved by the introduction of random noise as the privacy mechanism. This works well when only one query is made to a database but it can break down when the querier dynamically poses questions to the database. This is simply because the effect of noise reduces with the number of samples observed. In other words, if someone poses only one question the level of noise required to adhere to Differential Privacy with fixed parameterization, is different than if they ask two or three queries.

Luckily, the composition of two DP mechanisms is still a DP mechanism and it is possible to derive the parametrization of the composed mechanism from the original mechanisms. From there, one can study the DP parameters of the mechanism consisting of the sequence of all DP releases of information on the dataset. This quite naturally leads to the definition of a privacy loss budget as the limit of the DP mechanism of all subsequent queries onto the data.

To enforce the privacy of the whole dataset across many queries, privacy budgets are typically maintained by a technical component called a privacy accountant. These budgets take into account the previous queries made and how information from these queries can compound with one another to leak a greater level of information than each individually in isolation.

COSTS OF USING THE TECHNOLOGY

The main cost of using differential privacy is a loss in terms of output accuracy with respect to solutions for the same problem that do not provide output privacy. Typically, this cost depends on the level of privacy required (more privacy incurs more loss in accuracy), the number of individuals in the dataset (increasing the amount of data available reduces the accuracy loss), the number of queries to be made on the data, and the range of possible values for each individual.

In addition, for a given privacy protection, the accuracy is also influenced by the amount of information being released. For example, releasing a single statistic about a dataset can typically be done with more accuracy than releasing a large number of detailed statistics, a more complex object, such as a machine learning model or a synthetic dataset. Moreover, an important observation that motivates the definition of differential privacy is that one can't hope to query a database indefinitely without ultimately revealing a large percentage of its contents. This makes deployments where the queries are not fixed *a priori* especially challenging. Algorithms that best utilize information released in the past to answer any number of queries with minimum privacy risk are a domain of active research.

On the computational side, differential privacy generally incurs only a moderate increase in complexity over non-private alternatives.

2.4 SYNTHETIC DATA

PROBLEM DEFINITION

Synthetic data is one of the output privacy techniques that aim at providing privacy guarantees when releasing information to a third party. The underlying principle is to transform a sensitive dataset into a new dataset with similar statistical properties without revealing information on individuals from the original dataset. It is often useful when your organisation wishes to share information externally with contractors or external stakeholders while having privacy guarantees about sensitive data.

It aims to meet two objectives:

1. Utility for statistical analysis: one should be able to study the statistics of the original data, potentially including complicated patterns, directly from the synthetic data.
2. Privacy: the synthetic data should not reveal information about individuals from the original dataset.

Synthetic data can also refer to data augmentation or the creation of data for validation and verification purposes, but these are not within the scope of PET. Data augmentation can be achieved through building an artificial model and generating data from it. For example, a 3D model can be used to generate many pictures of an object and then train neural networks. Likewise, an agent-based model can be used to simulate the salient features of possible worlds and generate data from the interactions that occur between agents. These forms of synthetic data are not discussed further, although the techniques that are described here can often be used for such purposes as well.

EXAMPLE USE CASES

Synthetic data have a range of valuable use cases wherever sharing sensitive data is necessary. Here are some illustrative use cases: an organization may want to disclose the list of all the datasets that they possess in order to initiate data collaboration opportunities. It can take months to grant access to the original data to an external party and sharing just the metadata is unlikely to provide enough information to assess whether this process is worth it. Synthetic datasets can be used to give fine-grain

understanding of the original data without the risks and compliance hurdles. Another use case is when validating a proof-of-concept or evaluating third-party solutions. This is the case when the data owners would want to assess the value of the vendor's technology on their own data. This process is compliance heavy and may block the project altogether. Validating the solution on synthetic data can be a way to work around such challenges. Next is the use case to expand training datasets for AI systems that typically benefit from large training sets. Particularly, when the training data is sensitive, synthetic data might be the only way to provide large data sets at scale.

Sometimes companies' production data may not be available for training or performing engineering tests. Using synthetic data, companies can achieve those objectives without the risk of exposing individual information. Synthetic data may also be a good enough alternative to monetize data assets in cases where organizations would want to monetize their data in data marketplaces, but they may not be able to share sensitive information with third parties. Lastly, data protection regulations often limit the retention of data to the minimum required for the performance of the service. Transforming old data into synthetic data is a way to keep the benefit of using the data for potential future studies.

To give an example of a range of uses of synthetic data, the Office for National Statistics in the UK, has used synthetic data for applications such as^p

- public releases, by synthesising its UK Annual Business Survey
- for testing of load balances in the preparation of the ONS Census data
- Training machine learning models
- Testing of Covid-19 transmissions through synthesized version of mobile phone data

Similarly, Statistics Canada^q produced synthetic data from Census information, Canadian Cancer Registry and Canadian Vital Statistics Death Database for the purpose of organising hackathons and being able to derive new analytical insights without compromising on privacy.

^p See [Case Study 6](#) in Chapter 3.

^q See [Case Study 8](#) in Chapter 3.

OVERVIEW

Historically, anonymization of a dataset is done by altering input data either by masking some fields or perturbing values up to a point where records can no longer be identified in the altered data (perturbation method). However, this approach fails to provide formally verifiable protection and the amount of perturbation needed becomes prohibitive as the number of columns grows due to the curse of dimensionality. Using a machine learning model to generate brand new records has emerged as the preferred alternative for generating synthetic data. This is therefore known as a “model-based approach”, which is the focus of this section.

It leverages advanced traditional inference and deep learning techniques that endeavour to learn the distribution of the original input data. This distribution is captured by a generative model that is used to sample from the learned distribution in order to create new data points that together exhibit the macro characteristics of the original data. This is an application of unsupervised machine learning and techniques include the use of copulas, generative-adversarial networks (GANs) and variational auto-encoders (VAEs), amongst others. The model or approach taken depends very much on the data domain. For example, generative adversarial models can be highly effective when synthesizing image or audio data. However, simpler and more transparent methods, such as copula based approaches, work well with tabular data.

The objective of providing quantifiable privacy guarantees in such a dataset is not so straightforward to satisfy, as discussed in the security model section. One may distinguish between synthetic data that has *provable* privacy guarantees (see Section 2.3 on Differential Privacy) and synthetic data that does not make such claims.

HISTORY

The scientific community has used the creation of random and synthetic data for a long time. Very popular statistical methods such as Monte Carlo simulations rely heavily on the sampling of random variables to perform inference and approximate distributions. However, the practice

of creating synthetic data for the purposes of privacy is very much rooted in the domain of official statistics, when Donald Rubin²⁷ and Roderick Little created a synthetic dataset for dissemination from US Census data, in the early 1990s. Over the years, synthetic data has grown in popularity and has influenced modern approaches to data imputation, data masking and other statistical data disclosure control approaches.

Currently, there are several open-source libraries that implement synthetic data generators with and without quantifiable privacy guarantees, depending on the type of data and applications. For example, the Synthetic Data Vault^r offers different generators for tabular and time series datasets albeit without provable privacy. On the other hand, Smart Noise^s provides synthesizers with mechanisms that are differentially private.

SECURITY MODEL

Synthetic data is often used as a blanket term to describe all fake or dummy data generation. In and of itself, synthetic data offers no privacy or security guarantees. The generator may remember some personal information, especially when the original data is sparse, which is likely in high dimensional datasets such as images, text, or series of events, and the model has a large learning capacity, which is the case of most neural-network-based generative models. Very flexible models can “overfit”, leading to potentially sensitive information influencing the synthetic data generation and hence to reidentification of certain samples. To illustrate the point, consider a census dataset where just one individual has a given profession. When the generator outputs a row with this profession, one may want to look at the other fields: are they close to their values in the original data? If no particular care has been taken during training, it is likely that this fake line will reveal quite precise information on that actual individual.

The absence of privacy guarantees used to be neglected because models had a limited learning capacity and the risk of overfitting one individual was deemed small. Also, before differential privacy emerged as a standard for measuring privacy risk, assessing such risk seemed intractable. As this pitfall became more obvious, practitioners started to include differential privacy in the process of building

²⁷ Rubin, “Statistical disclosure limitation” (1993).

^r <https://sdv.dev/>

^s <https://github.com/opensdp/smartnoise-sdk/tree/main/synth>

their synthetic dataset. The most straightforward way of achieving differentially private synthetic data is to train the generator with a differentially private learning algorithm, as described in Section 2.3. The synthetic data from such a generator inherits the privacy properties from the generator thanks to composition theorems.

Differential privacy is becoming the natural way of measuring the privacy risk of a synthetic dataset. Its parametrization (i.e. the amount of randomness that needs to be added to the training phase) dictates the guarantees that the synthetic dataset will inherit. It is a common misconception that synthetic data is safe because the data is fake. It is also a common misconception that differentially private synthetic data is safe because it uses differential privacy. It should be deemed only as safe as the privacy bounds, in the differential privacy sense, that have been used in the parametrization. If the upper bound is high, differential privacy provides little benefit.

Furthermore, differentially private synthetic data has the same requirements in terms of maintaining differential privacy budgets as differentially private statistics, that is to say, if synthetic data is produced ad hoc, then the organisation should manage the quantity and frequency of its generation and disclosure to ensure privacy guarantees.

COSTS OF USING THE TECHNOLOGY

The main cost of using synthetic data is loss of utility. The achievable quality depends on the dimension of the data, and the number of rows in the dataset but as a general rule, unless the synthetic data *is* the original data, some queries on the synthetic data will differ from queries on the original dataset. However, if sufficient utility has been retained, the original data may be considered superfluous and deleted, as long as a suitable metric for utility can be defined. In general, the model that generates data has a finite learning capacity and the learning phase will focus on optimizing against a learning objective. The choice of the learning objectives is always a tradeoff and high utility for one objective typically comes at the cost of lower utility for another. Synthetic data can be faithful for a limited number of predefined objectives but cannot be universally faithful. Even for a given objective, for instance the cumulative distribution function of a field, synthetic data should not be better than the differentially private version of the objective on the original data.

Perfect synthetic data that preserves privacy has been proven impossible, as there is an obvious trade off between the similarity of the fake data with the real data while also preventing reidentification of the real data. Synthetic datasets that are both differentially private and can be used broadly for training machine learning models or statistical analyses are considered beyond the frontier of modern research. It is more reasonable to consider the use of synthetic data for a finite subset of learning objectives that have been used to train the generator.

A corollary of privacy-preserving synthetic data is that no synthetic record can be linked at the individual level, since that would contradict the claim that an individual cannot be re-identified. For this reason, synthetic data is not an option when one wants to ask questions in the future which are beyond the scope of the current requirements, as the synthetic data algorithm cannot guarantee that the specific characteristics required to answer such future questions will be preserved by the generating model.

From a computing perspective, training a synthetic data generative model is a one-off exercise. Depending on the dataset size and the type of learning procedure it can incur a significant cost, but once the data is generated, using synthetic data for analysis or model training is identical to using the real data from the user's perspective.

2.5 DISTRIBUTED LEARNING

PROBLEM DEFINITION

Distributed learning is a class of protocols that aim to train a Machine Learning model, in particular a neural network, on input data that is owned by multiple parties who want to keep their data private. Two protocols that implement this process in slightly different ways are known as *Federated Learning* and *Split Learning*, each with their own pros and cons. Both of these protocols begin with the same setup; multiple parties have access to data that they consider sensitive, and there is a central not-fully-trusted authority server who will assist them. The parties agree on a neural network architecture they would like to train, as well as other particulars such as hyperparameters. In the following, we will see how the two ideas diverge.

Recall that in a vanilla and centralized deep learning model, the goal is to build a high-dimensional neural network by composing smaller parametric functions. The simplest neural network architecture or the so-called Multi-Layer Perceptron (MLP) is made up of a number of layers that the unstructured input data is passed through. At every layer, an affine transformation followed by an element-wise non-linear function (also known as activation function) is applied to the input data and this continues until the last layer. This step is called forward propagation. An optimization algorithm, such as stochastic gradient descent, is then applied to minimize a selected cost function that depends on the independent affine transformations, as learnable parameters. In practice, at the end of each forward pass, the optimization algorithm uses a reverse-mode automatic differentiation method to update the parameters based on the evaluation of the cost function on a limited number of data points. This is where the stochasticity comes from and this step is known as backward propagation. The combination of forward and backward propagations continues iteratively for a number of rounds (or epochs) on the entire training data. In the remainder of this section we will assume a basic knowledge of how to train a neural network.²⁸

Distributed learning is an Input Privacy technique in the sense that it improves the security of the input data between the input parties and the compute parties. However, it can be used in conjunction with other Input Privacy techniques, such as Secure Multi-party Computation and Homomorphic Encryption to enhance the privacy and security of the data. In itself, it does not provide output privacy guarantees since the design does not prevent personal information from being shared through the output. In some cases, the process may have some output privacy benefits provided that the parties carefully craft the learning parameters (e.g. input data allowed to be used, types of models and number of parameters for the model, number of iterations, injection of noise into the data or the parameters that are sent ...). These output privacy considerations are not core to the distributed learning value proposition but are supplemental features that need to be addressed as part of the output privacy design.

EXAMPLE USE CASE

Distributed learning and optimization can be a privacy preserving model-training solution for data that is stored across a heterogeneous network of distributed edge devices, e.g. mobile phones. In addition to edge devices, distributed learning can also leverage sensitive smaller individual datasets that are stored locally on a network of entities or organizations with limited resources to collaboratively train a machine learning solution in a variety of domains, such as health-care, finance, logistics, etc.

An example application is that of Trusted Smart Surveys.[†] Data collected from smartphone sensors could be used to supplement the National Statistical Office in their current surveys and be applied to regular production of statistics as well as public and private data collaboration.

²⁸ Zanussi, A Brief Survey of Privacy Preserving Technologies [2021].

[†] <https://unstats.un.org/wiki/pages/viewpage.action?spaceKey=UGTTOPPT&title=Trusted+Smart+Surveys>

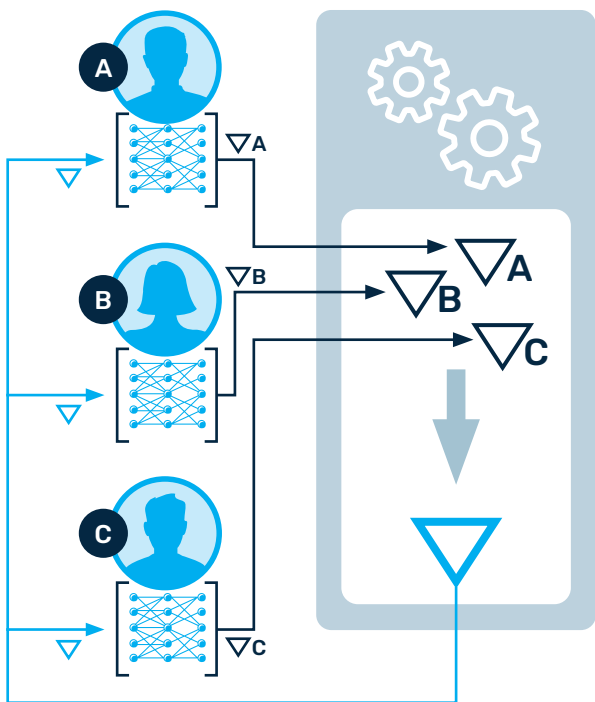


Figure 2.5.1: In federated learning, each data holder computes and updates weights on their data and sends it up to a central authority who computes and distributes it down to each party. In this way, each party can obtain a neural network that has been trained on the union of their datasets, without sharing their data.

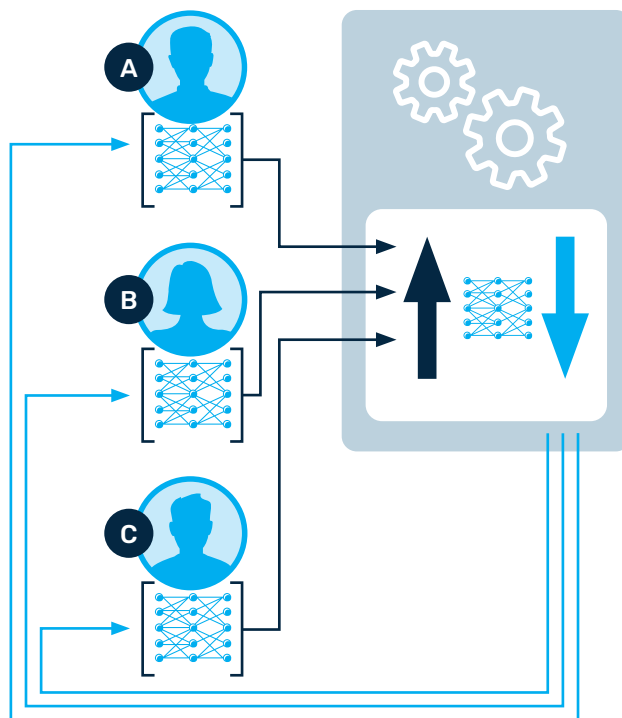


Figure 2.5.2: In split learning, the desired network is “split” between the parties and the server. Forward propagation is shown going up in dark blue, and backward propagation goes down in magenta. Each party performs forward propagation up to the split and sends the result to the server, who propagates forward and back again, sending the updated weights back to their respective parties who can then update their networks.

OVERVIEW

In federated learning, each party holds an identical local copy of the neural network they are training. They each perform one or more epochs of training on their network and send their model updates (i.e. parameters or weights) to the authority. The authority coordinates these weights, which could be as simple as an aggregation, and instructs each of the parties on how to update their local models by combining the insights gained by every party’s data. The process then repeats for the desired number of epochs, until finally the authority and every party holds a trained version of the network that they can use for inference on new data. Each copy of the network is identical, and the process reveals no more about the data than the sequence of weights computed by each party. This could potentially facilitate an avenue for attack that needs to be considered when implementing a federated learning scheme.

In split learning, the neural net is split by the authority at a certain layer, and layers after the split are shared with the parties. Each party propagates their data forward from the first layer up to the cut, then sends the activations at the cut layer to the server. The server finishes the forward propagation on the rest of the network, performs backward propagation up to the cut, then sends the weights to the parties who can then each finish back propagation and update their copy of the network. After the desired number of epochs, the authority distributes its half of the network to each of the parties, and then each party has their own copy of the total network. The only data leaked are those which can be inferred from the activations and weights exchanged at each epoch. The layers before the split serve to alter the data enough that they are protected (sometimes called “smashing” the data), while still allowing the server to gather insights from it.

HISTORY

Google first introduced Federated Learning in 2016²⁹ to address some of the privacy challenges of the decentralised model training. Since then it has been incorporated in some of Google's products, for instance in Gboard on Android devices.³⁰ Apple has also integrated Federated Learning in some of its products, such as QuickType keyboard and the speech recognition applications, such as Siri.³¹ Moreover, other applications of Federated Learning solutions have been explored in various areas, such as medical research,³² COVID-19,³³ financial risk management,³⁴ and manufacturing.³⁵ For recent surveys on Federated Learning and its applications, interested readers can refer to.^{36,37} On the other hand, split learning was first introduced for distributed deep learning in 2018³⁸ and since then it has been refined and proposed for use in the health-care domain.³⁹ It is worth noting that distributed learning models, as a field, are younger than other privacy-enhancing technologies discussed in the handbook, in particular the encryption-based approaches, such as secure multi-party computation and homomorphic encryption.

There are a range open source enabling distributed learning:

- Syft + Grid (from OpenMined)^u - Syft combined federated learning, differential homomorphic encryption and multi-party computation to enable private distributed learning. Grid provides an API to deploy Syft.
- Flower^v - a flexible framework for federated learning compatible with many ML frameworks (PyTorch, TensorFlow, MXNet and others).

- TensorFlow Federated^w - Python library supported and used by Google
- IBM Federated Learning^x - Python framework supporting a range of models including neural networks (in Keras, TensorFlow and PyTorch), linear regressions, decision trees.
- OpenFL^y - another Python library for federated learning from Intel.

SECURITY MODEL

In the field of privacy preserving machine learning and in particular deep learning, there are two main competing objectives. On one hand, successful deep learning models are the ones that find faithful representations of unstructured input data in high dimensional spaces. It is shown that in some extreme cases for neural networks and deep learning models to reach near-optimal accuracy in natural prediction problems, they need to memorize their input training data⁴⁰. On the other hand, the same property of deep learning models could pose privacy risks and legal challenges related to sensitive input data.

Two types of privacy-related risks can be considered in federated learning: first, local or gradient-level risk, which is related to the fact that sharing model updates, i.e. weights instead of the inputs can nonetheless leak sensitive information about the underlying data to the authority as well as to the clients indirectly via their querying functionalities on particular datasets.⁴¹ More precisely, an adversary (either the server or clients) can passively observe the weights to learn more about the

²⁹. McMahan et al., "Communication-Efficient Learning of Deep Networks" (2017).

³⁰. Hard et al., "Federated Learning for Mobile Keyboard Prediction" (2018).

³¹. Apple Differential Privacy Team, *Learning with Privacy at Scale* (2017).

³². Rieke et al., "The Future of Digital Health with Federated Learning" (2020).

³³. Dayan et al., "Federated learning for predicting COVID-19 outcomes" (2021).

³⁴. Federated AI Ecosystem (FEDAI), *WeBank and Swiss Re MOU* (2019).

³⁵. Musketeer Project, *Data Economy meets Industry 4.0* (2020).

³⁶. Yang et al., "Federated Machine Learning: Concept and Applications" (2019).

³⁷. Kairouz et al., "Advances and Open Problems in Federated Learning" (2021).

³⁸. Gupta et al., "Distributed learning of deep neural networks" (2018).

³⁹. Vepakomma et al., "Split learning for health" (2018).

^u <https://github.com/OpenMined/PySyft>

^v <https://flower.dev/>

^w <https://www.tensorflow.org/federated>

^x <https://github.com/IBM/federated-learning-lib>

^y <https://github.com/intel/openfl>

⁴⁰. Brown et al., "Memorization of irrelevant training data" (2021).

⁴¹. Phong et al., "Privacy-Preserving Deep Learning" (2018).

features, distribution of the training data and class representations. Disclosure control methods, such as differential privacy can mitigate the gradient-level risk significantly. However, due to the high-dimensionality of the data in deep learning models, there are challenges to achieve a proper implementation of these techniques.

Second, global or model-level risk, which is related to the learned model to protect it against membership inference and reconstruction attacks. To protect the global model against such attacks and mitigate the risk, the server can apply similar disclosure control techniques, such as differential privacy. However, federated learning approaches are still vulnerable to multiple avenues of attacks, such as model poisoning and model inversion attacks at both local and global levels, due to non-privacy related risks. An active attack for instance would consist of malicious or Byzantine clients injecting adversarial training examples (e.g. backdoor and non-independent and identically distributed data points) to poison the global model.⁴² There are mitigation strategies, such as alternative model aggregation methods that are robust to contributions of tail users or outliers.⁴³ However, they come at the price of suppressing or discarding their contributions with a negative impact on global model accuracy and fairness, especially for tail users.⁴⁴

Similarly, the smashed vectors in split learning should represent and reveal important information about the input training data to serve its purpose. However, there has been some research in split learning that shows information-theoretic guarantees (in the sense of distance correlations) can prevent information leakage about the input data from the smashed feature vector, while preserving the utility of it.⁴⁵ Moreover, deployment of split learning on a large number of edge devices with potentially small individual datasets can reveal information about their sizes, e.g. in extreme cases of single data points. Similar to federated learning, split learning scenarios are also vulnerable to malicious clients to collude and corrupt the global model's performance.

Depending on the privacy requirements, a common approach to strengthen the security and privacy of

a distributed learning model is to combine it with other privacy-enhancing technologies. For instance, a combination of Paillier (or additive) homomorphic encryption ([see section 2.2](#) of the handbook) and federated learning will provide a guarantee that the incoming weights from participating parties are protected against a malicious central coordinator, who would perform the aggregation of weights homomorphically. Similarly, multiple servers can coordinate a secure multi-party computation protocol ([see section 2.1](#)) to aggregate the updated weights in a secure way. Finally, it is worth noting that distributed learning and their security vulnerabilities are relatively new and active areas of research.

COST OF USING THE TECHNOLOGY

The main costs in distributed learning are associated with communication and computation on both clients and server sides. For instance, federated learning requires transferring model updates (weights) from the clients (e.g. edge devices) to the central authority back and forth multiple rounds that would add up to the communication and computation costs on the clients' side. The client-side computational cost is lower in split learning because the operations that can be performed are limited to a portion of the neural network that belongs to the client (before the cut layer). Instead, the sequential nature of split learning combined with queuing delays add to the latency of training and make it slower overall compared to federated learning, which allows parallel model training across a large number of nodes. Other important factors in determining costs are related to data, namely type and sizes, neural network architecture and the decentralized network properties, for instance number of clients and their relationships, i.e. their hierarchy and topology of the network.

Some research activities have proposed the use of quantized and compressed ML models and neural networks to reduce the volume of the model's parameters and hence the communication costs, in particular during training on edge devices with energy constraints.⁴⁶ Additionally, alternative approaches focus on network


⁴² Bagdasaryan et al., "How To Backdoor Federated Learning" (2020).

⁴³ Data et al., *Byzantine-Resilient High-Dimensional Federated Learning* (2020).

⁴⁴ H. Wang et al., *Attack of the Tails* (2020).

⁴⁵ Vepakomma et al., "NoPeek: Information leakage reduction" (2020).

⁴⁶ Konečný et al., *Strategies for Improving Communication Efficiency* (2016).



resource management and probabilistic device selection for an adaptive and efficient allocation of network resources to reduce the bandwidth and communication overhead, while improving the convergence time and model performance.^{47,48,49} As mentioned, more research is currently being undertaken to overcome and mitigate some of the outlined challenges of distributed learning models, such as device and data heterogeneity, privacy-related issues and communication costs.

⁴⁷. S. Wang et al., “Adaptive Federated Learning” (2019).

⁴⁸. M. Chen et al., “Communication-efficient federated learning” (2021).

⁴⁹. T. Li et al., “Challenges, Methods, and Future Directions” (2020).

2.6 ZERO KNOWLEDGE PROOFS

INTRODUCTION

Zero knowledge (ZK) refers to a class of cryptographic technologies that allows one party (called the *prover*) to convince another party (called the *verifier*) of the veracity of a claim that depend on secret information known to the prover without revealing those secrets to the verifier. Unlike typical proofs, which are often constructed as a sequence of statements where each statement can be fully verified by the prover to decide whether the claimed fact is true, ZK typically convinces the verifier that the claim is true with very high probability, but not mathematical certainty. A simple example of such a claim is “I am in possession of the private key, K, that was used to produce this signature, S, on this message M”. A more complex statement might be “we executed a machine learning prediction model on the whole input portfolio and past transaction history of a company to prove its solvency, and obtained the following result.” Note that in the former case, there is a (very) small probability that the same signature could be produced using a different key – yet the claim likely still convinces a reasonable verifier, so long as they trust the digital signature scheme in use.

ZK proofs are not so much aimed at general computation that preserves privacy, but instead focus on offering proofs of specific statements, that may or may not involve significant computation, whose verification only reveals minimal information – only the veracity of the claim itself. Specifically, a zero-knowledge proof has three salient properties:

- **Completeness:** If the statement is true and both the prover and the verifier follow the protocol; the verifier will accept the proof provided by the prover
- **Soundness:** If the statement is false, and the verifier follows the protocol; the verifier will not be convinced by the proof.
- **Zero-knowledge:** If the statement is true and the prover follows the protocol; the verifier will not learn any confidential information from the interaction with the prover except that the statement is true.

EXAMPLE USE CASES

In recent years there has been an increasing number of practical applications that leverage zero knowledge proofs. Many of these applications have been motivated in the context of cryptocurrencies. In the cryptocurrency setting, zero knowledge can allow for adding encrypted transactions to the ledger and then proving that those transactions are consistent with ledger policy, for example preclusion of double spending. The crypto currency ZeroCash^z was one of the first adopters of this functionality, and remains a primary illustrative use case.

Another nascent use case for ZK is in authentication systems that protect the authentication credentials, such as Direct Anonymous Attestation.^{aa} A credential can be thought of as a signature on some set of messages. Rather than disclosing this signature to a verifier, a holder of a credential can use the signature to create a fresh unlinkable zero knowledge proof of knowledge each time they wish to present the credential. Verification of this proof gives confidence in the authenticity and integrity of the attributes disclosed. Zero knowledge protocols can also be applied to the attributes themselves, the canonical example “I am over the age of 21” is dependent on there existing a signed attribute attesting to the persons age that they can use as the input to produce this proof. In this way zero knowledge is used to minimise the disclosure of information, protecting input privacy, whilst retaining the ability for input and output verification found in traditional signature schemes. Signature schemes supporting efficient zero knowledge proof protocols exist in both theory and practice, for example the BBS+ signature⁵⁰ scheme which has at least two independent implementations^{ab,ac} and is beginning to be adopted by organisations deploying with W3C Verifiable Credentials.

Another promising but preliminary use case for zero knowledge proofs is the proof of properties about software programs. Current research shows that it is possible to construct zero knowledge proofs of the existence of

^z <http://zerocash-project.org/>

^{aa} <https://tokenzoo.github.io/> is a good source of information for ZK authentication systems.

⁵⁰ Camenisch et al., “Anonymous Attestation” (2016).

^{ab} <https://github.com/hyperledger/ursa>

^{ac} <https://github.com/docknetwork/crypto>

vulnerabilities in software. For example, it is possible to construct a proof of the following kind of statement: “I know an input that triggers a vulnerability in this program, that results in undefined program behavior.” At present this kind of proof is still in the realm of research, and only practical for relatively small programs or components in software libraries. However, current research is focused on scaling up the complexity of proofs that can be made practical, offering the promise that proofs of vulnerability for larger programs will be practical in the near future.

OVERVIEW

Some applications of ZK proofs are at the time of this writing well understood and adopted in practical use, while others are still emerging from the research phase of development. Proofs for different classes of statements require diverse proof structures, so there is little notion of generalization in ZK proof methods at this time. A key challenge when applying ZK proofs is the translation from a human problem into a set of statements that can be mathematically proven under a given scheme. Considerable effort is being invested in developing ZK proof efficiency as well, because at present ZK proofs are only practical for simple proof statements such as, “the funds required for this currency transaction were sufficient to cover the amount transferred”. More complex statements, such as “This software program has the property of memory safety” are as yet not possible at any practical scale.

Despite these limitations, the pace of innovation has increased dramatically over recent years, primarily stimulated by the interest from blockchain based projects. Eli Ben Sasson refers to this as a Cambrian explosion in zero knowledge proof systems in a post that is well worth a read for those seeking to understand the nuances and limitations of the existing proof systems.^{ad} Furthermore, these blockchain projects have accelerated the synthesis of cryptographic protocols into software artifacts and are applying them to a diverse set of use cases. Examples include emerging developer toolkits,

libraries and languages such as Aleo, Arkworks, Cairo and Zokrates. In addition to this considerable work is underway in developing a taxonomy and standardization for ZK proofs. We refer the reader to <https://docs.zkproof.org/reference.pdf> for more information, but we caution that the mathematics underlying the ZK proof paradigm are very complex.

HISTORY

Zero knowledge proofs were introduced in the 1980s in a publication by Goldwasser, Micali and Rackoff who defined a computational complexity theory of “knowledge” contained within a statement and introduced the concept of interactive proofs, proofs that require communication between the prover and verifier.⁵¹ Following this it was demonstrated that interactive proof systems could be transformed into non-interactive versions if both the prover and verifier had knowledge of a common reference string (CRS).⁵² The Fiat-Shamir transformation provided an example of how this common string could be produced using a hash function.⁵³ This approach was then used by Schnorr to define a zero knowledge authentication scheme.⁵⁴ This knowledge-of-exponent approach formed an important part of the privacy preserving cryptographic credential libraries UProve (from Microsoft) and Idemix (from IBM) developed in the early 2000s. However, producing efficient ZK proof systems for more generic computations remained infeasible until more recently.

FORMAL LIMITATIONS TO THE ZK PROOF SECURITY MODEL

As we noted in the Introduction to this section, ZK proofs offer 3 security guarantees: completeness, soundness and zero-knowledge.

These security properties are achieved by relying on diverse mathematical hardness assumptions, as is true with most cryptographic constructions. Sometimes, the constructions used in ZK proof systems go beyond

^{ab} <https://github.com/hyperledger/ursa>

^{ac} <https://github.com/docknetwork/crypto>

^{ad} <https://nakamoto.com/cambrian-explosion-of-crypto-proofs/>

⁵¹ Goldwasser et al., “Knowledge Complexity of Interactive Proof” (1989).

⁵² Blum et al., “Non-interactive zero-knowledge and its applications” (1989).

⁵³ Fiat et al., “Practical Solutions to Identification and Signature” (1986).

⁵⁴ Schnorr, “Efficient Signature Generation by Smart Cards” (1991).



those typically used in other cryptographic solutions. In particular, we point out that ZK proof systems often rely on *non-falsifiable assumptions* and the *Fiat-Shamir heuristic*. Use of these techniques should be taken into account when developing a full understanding of the security stance of zero knowledge systems.

In addition, most zero knowledge systems are proven in the setting of a single execution where at any time the prover is executing a single proof with a single verifier, and similarly the verifier is interacting with a single prover. Such security proof does not guarantee that the system preserves its security properties in concurrent executions, where there are many proofs being executed in parallel. Such concurrency issues are mostly relevant for interactive ZK proofs.

COSTS OF USING THE TECHNOLOGY

There are several costs to consider when using a zero knowledge system. These include efficiency of the proof generation by the prover, efficiency of proof verification by the verifier, size of the proof, and whether the verification requires interaction between the prover and the verifier. For example, *Succinct Non-Interactive Argument* (SNARG) proof systems provide proofs of small constant size (usually a few hundred bytes), which requires very little communication between prover and verifier. Verification is very efficient, usually taking a few milliseconds (dependent on the length of the input from the verifier). However, the SNARG prover incurs overhead for the computation of the proof. Usually the runtime for this computation is several orders of magnitude slower than the computation of the statement “in the clear”.

There are also many other types of zero knowledge systems apart from SNARGs. They offer different trade-offs: they may require interaction between the prover and the verifier, may have longer proofs (logarithmic, square root or linear in the statement size), or be more expensive to verify. The main advantage of such systems is that they impose substantially less overhead on the prover, which is useful in cases when this is the bottleneck for the application.

2.7 TRUSTED EXECUTION ENVIRONMENTS AND SECURE ENCLAVES

PROBLEM DEFINITION

A Trusted Execution Environment (TEE) is a feature of a modern CPU that mitigates the problems of *input privacy*, *code privacy*, and *code assurance*. Input privacy and code assurance have been introduced previously. Code privacy is the problem of assuring that the code being used to operate on data, along with confidential non-data information such as cryptographic keys, is not visible to potential adversaries. A TEE is typically implemented partly in the hardware of a CPU and partly in associated software libraries. Intel Software Guard Extensions (SGX) and AMD TrustZone are examples of popular TEEs. More recently, AWS Nitro Enclaves offer a hypervisor-based approach to TEEs which offers an alternative security model but allows for very flexible deployment.

EXAMPLE USE CASE

An illustrative use case for TEEs is that of borrowing data from a repository of sensitive information in order to use it for research purposes. For example, a recent use case allowed researchers to select network traffic data files from a catalog and perform research analysis on the selected files. As the files contained information such as complete network packets or packet headers, the files were deemed sensitive, since they might reveal for example the web browsing habits of network users. In this use case:

1. A researcher created a TEE on their research workstation and downloaded into it a set of research queries approved by the data provider.
2. The TEE executed an *attestation* protocol with a server owned by the data provider, proving to that server that the code contained in the TEE was exactly the approved code.
3. The data provider then transmitted the encrypted data to the TEE..
4. Next, the data provider's server distributed the decryption key for that data to the TEE over a secure (TLS) channel.

Thus the data was not decryptable by any party except the TEE. A characteristic of TEE solutions is that they prevent

anyone – even users with control privileges on the host where the TEE runs – from learning anything about the code, data, or execution of that code inside the TEE. With the decryption key and the encrypted data file available, the TEE decrypted the data inside its own memory and ran the relevant, approved queries, transmitting *only the query results* out of the enclave to be viewed by the researcher. Thus the researcher was able to achieve answers to only those queries approved by the data provider but could learn nothing about the data except those query results.

In another real-life application, Eurostat in partnership with Cybernetica has leveraged TEE and in particular, Intel SGX chips to process sensitive data coming from mobile network operators (MNOs).⁹⁸ Such data typically contain extremely sensitive information including locations and call detail records. The project provided a proof-of-concept of how national statistical offices (NSOs) can partner with MNOs, who provide longitudinal data about individual mobile phone users, to derive new official statistics. In that scenario, NSO can run its analytics with its chosen parameters and obtain results without having direct access to the data. Furthermore, the attestation process guarantees that only the pre-approved computation with output privacy guarantees in the form of k-anonymity is run.

OVERVIEW

TEE technology typically requires hardware support *inside* a CPU. That support includes the use of dedicated on-chip memory in which to store data used frequently during the computation; specific features in virtual memory management that prevent other processes on the CPU from accessing the memory space used by the TEE; hardware encryption support to encrypt and decrypt any data that must be moved out of the CPU and into system main memory; precautionary limits on advanced CPU features such as *speculative execution or branch prediction*, and so on. In addition, TEEs do not permit software to easily call operating system services, so a key component of a TEE is a software library that provides commonly used system calls *inside* the TEE. TEEs may suffer from reduced performance due to the limited amount of memory that they can access

⁹⁸ https://ec.europa.eu/eurostat/cros/sites/default/files/unece2021_estat_cybernetica_v6.pdf

without resorting to encryption and decryption as data migrates on and off the CPU. While TEEs are traditionally focused on the aforementioned hardware-based partitioning, there have been efforts by cloud providers such as AWS to provide software-based alternatives to TEEs. AWS Nitro Secure Enclaves are hypervisor-managed secure enclaves. These can be seen as a sort of third-party hosted (AWS) sandbox with limited input-output and physically partitioned memory and dedicated CPUs.

HISTORY

The Intel SGX TEE was first introduced in the Intel *Skylake* microarchitecture, which debuted in 2015. Note that TEE architectures are proprietary to CPU vendors, and so code developed for one hardware-based TEE is not necessarily likely to work on another vendor's TEE, or possibly even on the TEE of a different generation within the same CPU family. The Intel SGX TEE architecture has been subject to many security threats since its introduction and continues on to the present day. In some cases, hardware mitigations have been practical. In other cases, CPU performance has been intentionally sacrificed in order to mitigate these threats. Recently, Nvidia has announced its plans to release GPU-based enclaves that would make some of the computations more efficient than the current CPU-based enclaves.

This is not the case for non-hardware-based TEEs such as AWS Nitro, which run regular code in a virtualized environment. In these cases, development is highly portable and is developed as Docker containers which are later deployed into independent virtual machines. Attestation and other services are still consistent.

SECURITY MODEL

As with some other technologies described in this document, TEEs assume distrusting parties. In particular, a party who may send data to a remote (not locally controlled) TEE may well distrust the remote owner of the CPU which hosts that TEE. In addition, the owner of a TEE may distrust the provider of code to run in the TEE, fearing that the code provider may use the enclave as a kind of secure "trojan horse" with which to monitor or attack other processes running on that processor.

Ultimately, the security model of TEEs is not cryptographic in nature (despite leveraging hashes as an attestation). Rather, the security model is ultimately tied to trust in the physical hardware and/or hypervisor design. Hardware will always have some vulnerabilities if an untrusted

party has physical access and can take arbitrary physical measurements. In the case of hypervisor-based TEEs, trust is in both the software security (ie that there are no bugs by the provider of the TEE) and that the hypervisor owner, typically a cloud provider, will not maliciously attack the system. These respective risks are typically accepted by users of TEEs.

TEEs are also a what-you-see-is-what-you-get (WYSIWYG) model. As arbitrary code can run in the TEE, both parties have to understand and agree on exactly what is acceptable. This may include the specific versioning of libraries and frameworks, the sources they are from, and other security considerations. Importantly, TEEs have no guarantees against timing-based attacks, and as such users should be exceptionally careful not to run code that signals specific proprietary sensitive input based on the number of branches created (how many times a loop runs or similar).

COSTS OF USING THE TECHNOLOGY

Hardware-based TEE technology has a small but noticeable performance penalty for applications that fit within the on-chip memory limits of the TEE. For example, in the first-generation Intel SGX, about 96 MBytes of application memory were available. Applications that stayed within that memory footprint encountered performance penalties on the order of 20%. However, applications that substantially exceeded that limit encountered penalties of up to 16X slowdown or more, increasing as a linear function of memory sizes in typical cases. Unlike Multi-Party Computation or homomorphic encryption technologies, SGX performance is not unevenly affected by the mix of operations performed. Instead, the performance impact is generally a function of overall memory size, and the degree and frequency of input-output that must cross the TEE's memory boundary. There are now a number of open source and proprietary frameworks to develop for hardware-based TEEs, easing the development life-cycle. Further many of the early hardware restrictions, such as memory, have since been relaxed.

Software-based TEE does not have such limitations as memory can be dynamically allocated at the point that the TEE is created. The software that runs inside the TEE is built from a standard Docker container, thus the development life cycle is expedited. AWS, one such cloud platform widely offering this technology, does not charge any additional cost for using these services over regular cloud instances.

2.8 PRACTICAL CONSIDERATIONS OF PETS

In this section, we describe how to choose a PET for a privacy-preserving application. While there is no standardized process, there are fairly standard steps that will help in understanding the privacy preservation needs of the use case, the constraints that a selected PET must meet, and the practical utility needed in the resulting application.

FIRST STEPS - CONSTRUCTING USER STORIES

A good place to begin any application development is with a set of user stories that capture what it is you want to get done and why. A good user story identifies.

1. the diverse roles and numbers of users involved in using the application
2. the goals of each type of user
3. the kind and quantity of data to be processed
4. where that data comes from
5. what computation resources must or may be available
6. what outcome is expected in each use of the application
7. and which party or parties receive the output of the computation

In addition, the need for a PET implies that some (or even all) of the data to be used is sensitive. So, a user story should also capture

8. what data is sensitive
9. who is allowed to see it
10. who *must* not see it, and
11. what the owner or provider of the data needs in terms of assurance that it stays private
12. the stakeholder-agreed acceptance criteria that indicate the conditions under which the user story can be considered “done”.

The remainder of this section offers an example of this kind of user story and further questions that one may want to consider when designing a system built with PETs.

JOINING PRIVATE DATABASES

As a first example, imagine a scenario where two parties have sensitive data and a third party would like to query

their combined data to generate a report. Below is a brief schematic and description of an exemplar high-level ideal functionality.

In steps 1 and 2 in Figure 2.8.1, two database owners have access to the data in their respective databases. In steps 3 and 5, they provide only selected views over their database content to a PET. In steps 4 and 6, the owners receive a notification whenever the provided data is queried by the Querier, but do not receive the content of the queries nor the data selected by the queries. Step 7 shows the Querier asking a query of the combined data by providing that query to the PET. Step 8 shows the PET responding to the query with an answer that is compliant with the privacy restrictions specified by the database owners. Note that for some queries, the PET may return only partial data, or no data at all, if those restrictions prevent full responses.

Using our outline from above, and expanding somewhat on the description of the use case, we might arrive at the following use case description:

- **Roles and goals:**
 - Two database owners (in general, there might be more) with the goal of sharing certain portions of their data, subject to rules that they define, with a querier and an approved audience. It will be important to carefully understand the constraints on which data can be revealed and to whom, including which statistics over that data. The choice of a PET will be heavily influenced by those constraints

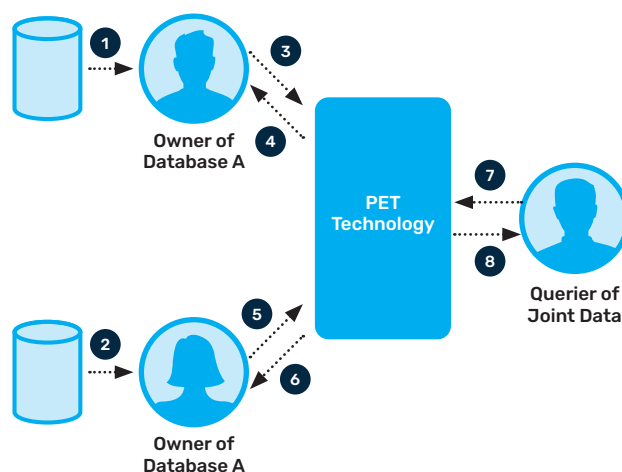


Figure 2.8.1: The use of PETs in joining two databases

- One querier of the resulting combined data, with the goal of reporting the answers to queries of interest to the audience.

- **Kind and quantity of data to be processed:**

- Relational database contents (it will be important to know the rough volume of data because diverse PETs will perform differently depending on how much data must be processed)

- **Where that data will come from:**

- Data comes from the two data owners over secure channels

- **What computation resources may be available**

- **What outcome is expected in each use of the application:**

- Queries from the querier are answered, providing that they comply with restrictions specified by the data owners. Owners are notified after each query that their data has been queried but do not learn the queries

- **Which party or parties receive the output of the computation:**

- The querier, and the audience who will receive the querier's report

- **What data is sensitive:**

- You'll need to discover and define which portions of the data must be kept private. In this example, all input data from the data owners are sensitive and must be kept private, except for what can be revealed in the query answers according to the rules defined. In addition, the queries are to be kept private from the data owners.

- **Who is allowed to see data, and who is not:**

- In this case, the data from the database owners must be seen only by the PET

- Queries and allowed query answers may be seen by the Querier (and by whatever audience the data owners agree may see the resulting report)

Note that the above schematic and description are strictly limited to only two data providers, rather than some arbitrary number of data providers. These small nuances may change the appropriateness of different PETs accordingly.

The benefit of breaking down the entities, data, and information flow in this way is that one begins to diagnose the privacy challenges and the feasibility of a proposition, creating a tangible technical specification

to assess various PETs against. This concept steps from the field of cryptography whereby algorithm designers formalize such relationships precisely in mathematics to remove any ambiguity of what an acceptable outcome looks like. This is known as the "ideal functionality" of the system. Of course, such a qualitative approach will not match a cryptographer's rigor, it acts as an excellent communication tool to engage the broader stakeholders of the project, from privacy experts to legal and regulatory.

CHOOSING A PET

Now that we have defined the use case, we can choose an appropriate PET or combination of PETs to achieve the desired functionality.

What needs to be protected, and from whom? If there is a lot of sensitive data coming to the PET from separate parties who do not wish to share data with one another, then *input privacy* techniques, such as sMPC, homomorphic encryption, and secure enclaves may be the right approach. If the output from the PET may contain sensitive information (either because that information is contained in the input data, or because the algorithm used for processing is sensitive), then *output privacy* approaches such as differential privacy may be required.

How many parties are providing input to the computation, and are the parties computing on the data trusted or not? Some PETs can accommodate significant numbers of data providers, while other PET cannot. This is particularly true for input privacy PETs, where the party performing the computation must not be able to view the data (unless it is first encrypted in an appropriate way). For example, trusted execution enclaves can generally support multiple data providers. Some forms of secure multi-party computation such as linear secret sharing can support a limited number of data providers. Homomorphic encryption can generally support only one data provider. For output privacy techniques, when the parties performing the computation are trusted but those receiving the output are not, accommodating a significant number of data providers is generally workable.

What level of privacy assurance is required? Differential privacy, homomorphic encryption, and sMPC are all mathematically derived protocols that offer provable guarantees to the end user provided that implementation assumptions are correct. Secure enclaves on the other hand mix hardware and software to provide similar outcomes. Which is "better" is still an open question. Recently a JASON report to the US Census suggested

avoiding secure enclaves in favor of sMPC,⁵⁵ while in Europe the Gaia-X consortium appears to intend on heavily leveraging secure enclaves. Ultimately, this will be for you to decide in conjunction with security and legal colleagues.

What is the environment in which the application will be performed? Similar to the last question, differential privacy, homomorphic encryption, and sMPC are all pure software-based approaches and thus are highly portable in different environments. Secure enclaves, on the other hand, require both software and specific hardware (such as Intel processors equipped with SGX enclaves). Today, secure enclaves are offered on all major cloud providers such as Google Cloud Platform, Azure, AWS, and IBM to name but a few. However, if you are not planning to leverage a cloud environment, a purchase of enclave-supported servers may be required.

How flexible should the solution be to scope creep? Scope creep is unfortunately a reality of many projects and this may happen for a multitude of reasons. Some technologies differ in terms of implementation a lot depending on the protocol required, others do not. For example, the implementation of homomorphic encryption and sMPC may require a different proof or a change in basis which may have a serious knock-on effect in terms of the time and cost of the project. Secure enclaves are far more flexible in this respect and are much more similar to writing conventional software.

How fast should the result be? Output privacy techniques are typically fast, especially differentially private statistics which require only a constant time overhead to add privacy-protective noise. Synthetic data may require training a machine learning algorithm for the data generation, which may be costly. While this may add some time during a preprocessing phase, it will likely be fast when synthetic data is actually needed.

Input privacy techniques are a different story. sMPC and homomorphic encryption can be many orders of magnitude slower than plaintext calculations. This means a query that would typically take seconds to run in an unsafe environment may take days to be performed. Secure enclaves, however, impose only limited slowdown over non-privacy-preserving approaches, provided the volume of data is reasonable

How will the solution integrate with other authentication, identity management, and key management systems? Systems leveraging privacy-enhancing technologies rarely live in isolation and often need to be integrated into other security and privacy management systems within an organization. This may affect the sub-network they are hosted on, whether a VPN is used, and how authentication, access control, key management services, and other systems are integrated.

How many resources are you able to dedicate to the solution? This may be an important question for you to consider as it relates to the maturity of different technologies. Homomorphic encryption and sMPC are still not widely used in practice and typically require expert cryptographers to implement them well. This limits the number of providers who can support you on the PET development and will likely increase the cost of development, deployment, and maintenance. Differential privacy, while still relatively new, has received a lot of attention from industry and startups. There are many service providers and typically off-the-shelf solutions that may be fit for purpose without problematic re-development time. Secure enclaves are somewhere in the middle of the cost spectrum. They are growing in popularity and their widespread cloud support is helping the spread of knowledge in development. While at the time of writing, these are still relatively few solutions, it is likely that this will change in the coming years.

How easy is it to mix and match these technologies? Some PETs play nicely with one another, while others do not. Whenever there is a shared notion of privacy and security between two or more technologies, their integration is generally possible without too much difficulty. For example, synthetic data generation followed by some differentially private statistical reporting may complement one another well as the privacy budgets would hopefully be able to be combined. This, of course, will depend on them using the same definition of differential privacy as there are now many variants and flavors.

Similarly, homomorphic encryption and different flavors of sMPC have been used in some scenarios together and are often referred to as mixed protocols. These can provide a balance between the speed of some techniques and the flexibility of operations of others.

⁵⁵ JASON Science Advisory Panel, *Secure Computation for Business Data* (2020).

Secure enclaves can also easily deploy output privacy techniques within the enclave. However, sMPC and homomorphic encryption may be a little trickier and require detailed research in order to combine them with output privacy.

FURTHER CONSIDERATIONS

Eyes-off data handling, as this guide endeavors to describe, has a multitude of privacy and security benefits. However, users should be aware that these benefits come with some corresponding challenges.

Fairness and Societal Impacts: One such challenge associated with not directly seeing input data as it is processed, is one's ability to ensure algorithmic fairness and other disparities in performance that rely on frequency statistics at run-time. Over the past number of years, the impact of algorithms on society has become pressing and is an active research topic. Fairness is context-dependent: an algorithm is neither fair nor unfair, but its outputs may be considered one or the other, and the same output may be fair in one context and not in another. Thus, what is fair will be determined by the requirements for the system of which the algorithm is a part, so maintaining human accountability for system outcomes. A model may however be biased as a consequence of any, some or all of data, human-cognitive or architecture bias. The issue of concern here is that input privacy and, to a lesser extent, output privacy to limit or prevent a model owner's capability to monitor the behavior of a deployed model. This problem may be mitigated through the presentation of partial macro statistics to a designated party, but this inevitably changes the security regime and trust model accordingly.

Bias and Adversarial Robustness: Similar to the topic of fairness, are issues pertaining to biases in training models and their ability to withstand adversarial perturbations to the data that can create misclassifications or outcomes. Often when machine learning models are developed, the owners of the models would like some level of assurance that the future performance of the system will be similar to that of the training. This is especially true if data drift, that is the changing of the behavior of data over time, occurs. Adversarial attacks, as they are referred to, are when data designed to trick the system into poor performance are presented at run time. Models can be resilient against such attacks if considerations are made during training and a level of stress testing is performed. However, as is the case with challenges associated with algorithmic

fairness, when less visibility of the underlying data is enforced by design, it may be more challenging to ensure highly robust models through training and deployment.

Mismatches in Data Cleaning: One of the most frustrating challenges comes when different parties, who do not have direct access to one another's data, have pre-processed data differently. While many privacy-enhancing technologies may work perfectly theoretically for their specific requirements, if the data does not conform strictly to a pre-agreed schema or set of pre-processing steps, then there can be a disparity between expected performance and the resulting outcomes. This can also be an issue in the context of input privacy when the linkage is desired, such as performing an SQL-like Join.

The above considerations are not absolute in nature, and in many cases, there may be mechanisms or design choices that can be made to alleviate these challenges. However, they should be thought through prior to undertaking a major privacy-enhancing technology project

CHAPTER 2. METHODOLOGIES AND APPROACHES

BIBLIOGRAPHY

Apple Differential Privacy Team (2017). *Learning with Privacy at Scale*. Apple. url: <https://machinelearning.apple.com/research/learning-with-privacy-at-scale>. Accessed 2022-07-01.

Archer, David, Dan Bogdanov, Yehuda Lindell, Liina Kamm, Kurt Nielsen, Jakob Illeborg Pagter, Nigel P Smart and Rebecca N Wright (Sept. 2018). "From Keys to Databases—Real-World Applications of Secure Multi-Party Computation". In: *The Computer Journal* 61.12, pp. 1749–1771. issn: 0010-4620. doi: [10.1093/comjnl/bxy090](https://doi.org/10.1093/comjnl/bxy090).

Bagdasaryan, Eugene, Andreas Veit, Yiqing Hua, Deborah Estrin and Vitaly Shmatikov (2020). "How To Backdoor Federated Learning". In: *The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020, 26–28 August 2020, Online [Palermo, Sicily, Italy]*. Ed. by Silvia Chiappa and Roberto Calandra. Vol. 108. Proceedings of Machine Learning Research. PMLR, pp. 2938–2948. url: <http://proceedings.mlr.press/v108/bagdasaryan20a.html>.

Blum, Manuel, Paul Feldman and Silvio Micali (2019). "Non-interactive zero-knowledge and its applications". In: *Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali*. Ed. by Oded Goldreich. ACM, pp. 329–349. doi: [10.1145/3335741.3335757](https://doi.org/10.1145/3335741.3335757).

Bogetoft, Peter, Dan Lund Christensen, Ivan Damgård, Martin Geisler, Thomas Jakobsen, Mikkel Krejgaard, Janus Dam Nielsen, Jesper Buus Nielsen, Kurt Nielsen, Jakob Pagter, Michael Schwartzbach and Tomas Toft (2009). "Secure Multiparty Computation Goes Live". In: *Financial Cryptography and Data Security*. Ed. by Roger Dingledine and Philippe Golle. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 325–343. isbn: 978-3-642-03549-4. doi: [10.1007/978-3-642-03549-4_20](https://doi.org/10.1007/978-3-642-03549-4_20).

Brakerski, Zvika (2012). "Fully Homomorphic Encryption without Modulus Switching from Classical GapSVP". In: *Advances in Cryptology – CRYPTO 2012*. Ed. by Reihaneh Safavi-Naini and Ran Canetti. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 868–886. isbn: 978-3-642-32009-5. doi: [10.1007/978-3-642-32009-5_50](https://doi.org/10.1007/978-3-642-32009-5_50).

Brakerski, Zvika, Craig Gentry and Vinod Vaikuntanathan (July 2014). "(Leveled) Fully Homomorphic Encryption without Bootstrapping". In: *ACM Transactions on Computation Theory* 6.3. issn: 1942-3454. doi: [10.1145/2633600](https://doi.org/10.1145/2633600).

Brown, Gavin, Mark Bun, Vitaly Feldman, Adam D. Smith and Kunal Talwar (2021). "When is memorization of irrelevant training data necessary for high-accuracy learning?" In: *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21–25, 2021*. Ed. by Samir Khuller and Virginia Vassilevska Williams. ACM, pp. 123–132. doi: [10.1145/3406325.3451131](https://doi.org/10.1145/3406325.3451131).

Camenisch, Jan, Manu Drijvers and Anja Lehmann (2016). "Anonymous Attestation Using the Strong Diffie Hellman Assumption Revisited". In: *Trust and Trustworthy Computing – 9th International Conference, TRUST 2016, Vienna, Austria, August 29–30, 2016, Proceedings*. Ed. by Michael Franz and Panos Papadimitratos. Vol. 9824. Lecture Notes in Computer Science. Springer, pp. 1–20. doi: [10.1007/978-3-319-45572-3_1](https://doi.org/10.1007/978-3-319-45572-3_1).

Chen, Hao, Zhicong Huang, Kim Laine and Peter Rindal (2018). "Labeled PSI from Fully Homomorphic Encryption with Malicious Security". In: *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. CCS '18. Toronto, Canada: Association for Computing Machinery, pp. 1223–1237. isbn: 9781450356930. doi: [10.1145/3243734.3243836](https://doi.org/10.1145/3243734.3243836).

Chen, Mingzhe, Nir Shlezinger, H. Vincent Poor, Yonina C. Eldar and Shuguang Cui (2021). "Communication-efficient federated learning". In: *Proceedings of the National Academy of Sciences* 118.17, e2024789118. doi: [10.1073/pnas.2024789118](https://doi.org/10.1073/pnas.2024789118).

Cheon, Jung Hee, Andrey Kim, Miran Kim and Yongsoo Song (2017). "Homomorphic Encryption for Arithmetic of Approximate Numbers". In: *Advances in Cryptology – ASIACRYPT 2017*. Ed. by Tsuyoshi Takagi and Thomas Peyrin. Cham: Springer International Publishing, pp. 409–437. isbn: 978-3-319-70694-8. doi: [10.1007/978-3-319-70694-8_15](https://doi.org/10.1007/978-3-319-70694-8_15).

Data, Deepesh and Suhas Diggavi (2020). *Byzantine-Resilient High-Dimensional Federated Learning*. doi: [10.48550/ARXIV.2006.13041](https://doi.org/10.48550/ARXIV.2006.13041). Accessed 2021-03-01.

Dayan, Ittai, Holger R. Roth, Aoxiao Zhong, Ahmed Harouni, Amilcare Gentili, Anas Z. Abidin, Andrew Liu, Anthony Beardsworth Costa, Bradford J. Wood, Chien-Sung Tsai, Chih-Hung Wang, Chun-Nan Hsu, C. K. Lee, Peiyang Ruan, Daguang Xu, Dufan Wu, Eddie Huang, Felipe Campos Kitamura, Griffin Lacey, Gustavo César de Antônio Corradi, Gustavo Nino, Hao-Hsin Shin, Hirofumi Obinata, Hui Ren, Jason C. Crane, Jesse Tetreault, Jiahui Guan, John W. Garrett, Joshua D. Kaggie, Jung Gil Park, Keith Dreyer, Krishna Juluru, Kristopher Kersten, Marcio Aloisio Bezerra Cavalcanti Rockenbach, Marius George Linguraru, Masoom A. Haider, Meena AbdelMaseeh, Nicola Rieke, Pablo F. Damasceno, Pedro Mario Cruz e Silva, Pochuan Wang, Sheng Xu, Shuichi Kawano, Sira Sriswasdi, Soo Young Park, Thomas M. Grist, Varun Buch, Watsamon Jantarabenjakul, Weichung Wang, Won Young Tak, Xiang Li, Xihong Lin, Young Joon Kwon, Abood Quraini,

Andrew Feng, Andrew N. Priest, Baris Turkbey, Benjamin Glicksberg, Bernardo Bizzo, Byung Seok Kim, Carlos Tor-Díez, Chia-Cheng Lee, Chia-Jung Hsu, Chin Lin, Chiu-Ling Lai, Christopher P. Hess, Colin Compas, Deepeksha Bhatia, Eric K. Oermann, Evan Leibovitz, Hisashi Sasaki, Hitoshi Mori, Isaac Yang, Jae Ho Sohn, Krishna Nand Keshava Murthy, Li-Chen Fu, Matheus Ribeiro Furtado de Mendonça, Mike Fralick, Min Kyu Kang, Mohammad Adil, Natalie Gangai, Peerapon Vateekul, Pierre Elnajjar, Sarah Hickman, Sharmila Majumdar, Shelley L. McLeod, Sheridan Reed, Stefan Gräf, Stephanie Harmon, Tatsuya Kodama, Thanyaweeurl Puthanakit, Tony Mazzulli, Vitor Lima de Lavor, Yothin Rakvongthai, Yu Rim Lee, Yuhong Wen, Fiona J. Gilbert, Mona G. Flores and Quanzheng Li (Oct. 2021). "Federated learning for predicting clinical outcomes in patients with COVID-19". In: *Nature Medicine* 27:10, pp. 1735–1743. issn: 1546-170X. doi: [10.1038/s41591-021-01506-3](https://doi.org/10.1038/s41591-021-01506-3).

Dinur, Irit and Kobbi Nissim (2003). "Revealing information while preserving privacy". In: *Proceedings of the Twenty-Second ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 9-12, 2003, San Diego, CA, USA*. Ed. by Frank Neven, Catriel Beeri and Tova Milo. ACM, pp. 202–210. doi: [10.1145/773153.773173](https://doi.org/10.1145/773153.773173).

Dwork, Cynthia, Frank McSherry, Kobbi Nissim and Adam D. Smith (2006). "Calibrating Noise to Sensitivity in Private Data Analysis". In: *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings*. Ed. by Shai Halevi and Tal Rabin. Vol. 3876. Lecture Notes in Computer Science. Springer, pp. 265–284. doi: [10.1007/11681878_14](https://doi.org/10.1007/11681878_14).

Dwork, Cynthia and Aaron Roth (2014). "The Algorithmic Foundations of Differential Privacy". In: *Found. Trends Theor. Comput. Sci.* 9:3-4, pp. 211–407. doi: [10.1561/04000000042](https://doi.org/10.1561/04000000042).

Fan, Junfeng and Frederik Vercauteren (2012). *Somewhat Practical Fully Homomorphic Encryption*. Cryptology ePrint Archive, Paper 2012/144. url: <https://eprint.iacr.org/2012/144>. Accessed 2022-06-07.

Federated AI Ecosystem (FEDAI) (2019). *WeBank and Swiss Re signed cooperation MOU*. FedAI. url: <https://www.fedai.org/news/webank-and-swiss-re-signed-cooperation-mou/>. Accessed 2022-06-07.

Fiat, Amos and Adi Shamir (1986). "How to Prove Yourself: Practical Solutions to Identification and Signature Problems". In: *Advances in Cryptology - CRYPTO '86, Santa Barbara, California, USA, 1986, Proceedings*. Ed. by Andrew M. Odlyzko. Vol. 263. Lecture Notes in Computer Science. Springer, pp. 186–194. doi: [10.1007/3-540-47721-7_12](https://doi.org/10.1007/3-540-47721-7_12).

Gentry, Craig (2009). "A fully homomorphic encryption scheme". PhD. Stanford University.

Goldreich, Oded, Silvio Micali and Avi Wigderson (2019). "How to Play Any Mental Game, or a Completeness Theorem for Protocols with Honest Majority". In: *Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali*. New York, NY, USA: Association for Computing Machinery, pp. 307–328. isbn: 9781450372664. doi: [10.1145/3335741.3335755](https://doi.org/10.1145/3335741.3335755).

Goldwasser, Shafi, Silvio Micali and Charles Rackoff (1989). "The Knowledge Complexity of Interactive Proof Systems". In: *SIAM Journal on Computing* 18:1, pp. 186–208. doi: [10.1137/0218012](https://doi.org/10.1137/0218012).

Gupta, Otkrist and Ramesh Raskar (2018). "Distributed learning of deep neural network over multiple agents". In: *J. Netw. Comput. Appl.* 116, pp. 1–8. doi: [10.1016/j.jnca.2018.05.003](https://doi.org/10.1016/j.jnca.2018.05.003).

Halevi, Shai and Victor Shoup (2020). *Design and implementation of HElib: a homomorphic encryption library*. Cryptology ePrint Archive, Paper 2020/1481. url: <https://eprint.iacr.org/2020/1481>. Accessed 2022-06-07.

Hard, Andrew, Kanishka Rao, Rajiv Mathews, Françoise Beaufays, Sean Augenstein, Hubert Eichner, Chloé Kiddon and Daniel Ramage (2018). "Federated Learning for Mobile Keyboard Prediction". In: CoRR abs/1811.03604. arXiv: [1811.03604](https://arxiv.org/abs/1811.03604). url: <http://arxiv.org/abs/1811.03604>.

Hart, Nicholas, David Archer and Erin Dalton (Mar. 2019). *Privacy-Preserved Data Sharing for Evidence-Based Policy Decisions: A Demonstration Project Using Human Services Administrative Records for Evidence-Building Activities*. Technical Paper. Available at SSRN 3808054. Bipartisan Policy Center. doi: [10.2139/ssrn.3808054](https://doi.org/10.2139/ssrn.3808054).

JASON Science Advisory Panel (2020). *Secure Computation for Business Data*. url: <https://irp.fas.org/agency/dod/jason/secure-comp.pdf>. Accessed 2022-06-13.

Kairouz, Peter, H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista A. Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, Rafael G. L. D'Oliveira, Hubert Eichner, Salim El Rouayheb, David Evans, Josh Gardner, Zachary Garrett, Adrià Gascón, Badih Ghazi, Phillip B. Gibbons, Marco Gruteser, Zaïd Harchaoui, Chaoyang He, Lie He, Zhouyuan Huo, Ben Hutchinson, Justin Hsu, Martin Jaggi, Tara Javidi, Gauri Joshi, Mikhail Khodak, Jakub Konečný, Aleksandra Korolova, Farinaz Koushanfar, Sanmi Koyejo, Tancrede Lepoint, Yang Liu, Prateek Mittal, Mehryar Mohri, Richard Nock, Ayfer Özgür, Rasmus Pagh, Hang Qi, Daniel Ramage, Ramesh Raskar, Mariana Raykova, Dawn Song, Weikang Song, Sebastian U. Stich, Ziteng Sun, Ananda Theertha Suresh, Florian Tramèr, Praneeth Vepakomma, Jianyu Wang, Li Xiong, Zheng Xu, Qiang Yang, Felix X. Yu, Han Yu and Sen Zhao (2021). "Advances and Open Problems in Federated Learning". In: *Found. Trends Mach. Learn.* 14:1-2, pp. 1–210. doi: [10.1561/22000000083](https://doi.org/10.1561/22000000083).

Konečný, Jakub, H. Brendan McMahan, Felix X. Yu, Peter Richtárik, Ananda Theertha Suresh and Dave Bacon (2016). *Federated Learning: Strategies for Improving Communication Efficiency*. doi: [10.48550/ARXIV.1610.05492](https://doi.org/10.48550/ARXIV.1610.05492). Accessed 2021-11-01.

Li, Baiyu and Daniele Micciancio (2021). "On the Security of Homomorphic Encryption on Approximate Numbers". In: *Advances in Cryptology – EUROCRYPT 2021*. Ed. by Anne Canteaut and François-Xavier Standaert. Cham: Springer International Publishing, pp. 648–677. isbn: 978-3-030-77870-5. doi: [10.1007/978-3-030-77870-5_23](https://doi.org/10.1007/978-3-030-77870-5_23).

Li, Tian, Anit Kumar Sahu, Ameet Talwalkar and Virginia Smith (2020). "Federated Learning: Challenges, Methods, and Future Directions". In: *IEEE Signal Process. Mag.* 37.3, pp. 50–60. doi: [10.1109/MSP.2020.2975749](https://doi.org/10.1109/MSP.2020.2975749).

Lyubashevsky, Vadim, Chris Peikert and Oded Regev (Nov. 2013). "On Ideal Lattices and Learning with Errors over Rings". In: *J. ACM* 60.6. issn: 0004-5411. doi: [10.1145/2535925](https://doi.org/10.1145/2535925).

Machanavajjhala, Ashwin, Daniel Kifer, Johannes Gehrke and Muthuramakrishnan Venkatasubramanian (2007). "L-diversity: Privacy beyond k-anonymity". In: *ACM Trans. Knowl. Discov. Data* 1.1, p. 3. doi: [10.1145/1217299.1217302](https://doi.org/10.1145/1217299.1217302).

McMahan, Brendan, Eider Moore, Daniel Ramage, Seth Hampson and Blaise Agüera y Arcas (2017). "Communication-Efficient Learning of Deep Networks from Decentralized Data". In: *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20–22 April 2017, Fort Lauderdale, FL, USA*. Ed. by Aarti Singh and Xiaojin (Jerry) Zhu. Vol. 54. Proceedings of Machine Learning Research. PMLR, pp. 1273–1282. url: <http://proceedings.mlr.press/v54/mcmahan17a.html>.

Musketeer Project (2020). *Best Success Story Contest Data Economy meets Industry 4.0 to create the next generation of Smart Manufacturing thanks to Federated Learning – Full story*. url: <https://musketeer.eu/wp-content/uploads/2020/04/Best-Success-Story-Contest-Data-Economy-meets-Industry-4.0-to-create-the-next-generation-of-Smart-Manufacturing-thanks-to-Federated-Learning.pdf>. Accessed 2022-06-12.

Narayanan, Arvind and Vitaly Shmatikov (2008). "Robust De-anonymization of Large Sparse Datasets". In: *2008 IEEE Symposium on Security and Privacy (S&P 2008), 18–21 May 2008, Oakland, California, USA*. IEEE Computer Society, pp. 111–125. doi: [10.1109/SP.2008.33](https://doi.org/10.1109/SP.2008.33).

Phong, Le Trieu, Yoshinori Aono, Takuya Hayashi, Lihua Wang and Shiho Moriai (2018). "Privacy-Preserving Deep Learning via Additively Homomorphic Encryption". In: *IEEE Trans. Inf. Forensics Secur.* 13.5, pp. 1333–1345. doi: [10.1109/TIFS.2017.2787987](https://doi.org/10.1109/TIFS.2017.2787987).

Regev, Oded (Sept. 2009). "On Lattices, Learning with Errors, Random Linear Codes, and Cryptography". In: *J. ACM* 56.6. issn: 0004-5411. doi: [10.1145/1568318.1568324](https://doi.org/10.1145/1568318.1568324).

Rieke, Nicola, Jonny Hancox, Wenqi Li, Fausto Milletari, Holger Roth, Shadi Albarqouni, Spyridon Bakas, Mathieu N. Galtier, Bennett A. Landman, Klaus H. Maier-Hein, Sébastien Ourselin, Micah J. Sheller, Ronald M. Summers, Andrew Trask, Daguang Xu, Maximilian Baust and M. Jorge Cardoso (2020). "The Future of Digital Health with Federated Learning". In: *CoRR* abs/2003.08119. arXiv: [2003.08119](https://arxiv.org/abs/2003.08119). url: <https://arxiv.org/abs/2003.08119>.

Rivest, Ronald, Leonard Adleman and Michael L. Dertouzos (1978). "On data banks and privacy homomorphisms". In: *Foundations of Secure Computation*. Ed. by Richard A. DeMillo, Richard J. Lipton, David P. Dobkin and Anita K. Jones. Academic Press, Inc., pp. 169–180. isbn: 0122103505.

Rivest, Ronald, Adi Shamir and Leonard Adleman (Feb. 1978). "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems". In: *Commun. ACM* 21.2, pp. 120–126. issn: 0001-0782. doi: [10.1145/359340.359342](https://doi.org/10.1145/359340.359342).

Rubin, Donald B. (1993). "Statistical disclosure limitation". In: *Journal of Official Statistics* 9.2, pp. 461–468.

Schnorr, Claus-Peter (1991). "Efficient Signature Generation by Smart Cards". In: *J. Cryptol.* 4.3, pp. 161–174. doi: [10.1007/BF00196725](https://doi.org/10.1007/BF00196725).

Sweeney, Latanya (2002). "k-Anonymity: A Model for Protecting Privacy". In: *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* 10.5, pp. 557–570. doi: [10.1142/S0218488502001648](https://doi.org/10.1142/S0218488502001648).

Vadhan, Salil P. (2017). "The Complexity of Differential Privacy". In: *Tutorials on the Foundations of Cryptography*. Ed. by Yehuda Lindell. Springer International Publishing, pp. 347–450. doi: [10.1007/978-3-319-57048-8_7](https://doi.org/10.1007/978-3-319-57048-8_7).

Vepakomma, Praneeth, Otkrist Gupta, Tristan Swedish and Ramesh Raskar (2018). "Split learning for health: Distributed deep learning without sharing raw patient data". In: *CoRR* abs/1812.00564. arXiv: [1812.00564](https://arxiv.org/abs/1812.00564). url: <http://arxiv.org/abs/1812.00564>. Accessed 2021-02-01.

Vepakomma, Praneeth, Abhishek Singh, Otkrist Gupta and Ramesh Raskar (2020). "NoPeek: Information leakage reduction to share activations in distributed deep learning". In: *2020 International Conference on Data Mining Workshops (ICDMW)*. IEEE, pp. 933–942.

Wang, Hongyi, Kartik Sreenivasan, Shashank Rajput, Harit Vishwakarma, Saurabh Agarwal, Jy-yong Sohn, Kangwook Lee and Dimitris Papailiopoulos (2020). *Attack of the Tails: Yes, You Really Can Backdoor Federated Learning*. doi: [10.48550/ARXIV.2007.05084](https://doi.org/10.48550/ARXIV.2007.05084). Accessed 2021-03-01.

Wang, Shiqiang, Tiffany Tuor, Theodoros Salonidis, Kin K. Leung, Christian Makaya, Ting He and Kevin Chan (2019). "Adaptive Federated Learning in Resource Constrained Edge Computing Systems". In: *IEEE J. Sel. Areas Commun.* 37.6, pp. 1205–1221. doi: [10.1109/JSAC.2019.2904348](https://doi.org/10.1109/JSAC.2019.2904348).

Wood, Alexandra, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, James Honaker, Kobbi Nissim, David O'Brien, Thomas Steinke and Salil Vadhan (2018). "Differential Privacy: A Primer for a Non-Technical Audience". In: *Vanderbilt Journal of Entertainment & Technology Law* 21.17. Also published as Berkman Klein Center Research Publication No. 2019-2. doi: [10.2139/ssrn.3338027](https://doi.org/10.2139/ssrn.3338027).

Yang, Qiang, Yang Liu, Tianjian Chen and Yongxin Tong (Jan. 2019). "Federated Machine Learning: Concept and Applications". In: *ACM Trans. Intell. Syst. Technol.* 10.2. issn: 2157-6904. doi: [10.1145/3298981](https://doi.org/10.1145/3298981).

Yao, Andrew Chi-Chih (1982). "Protocols for secure computations". In: *23rd Annual Symposium on Foundations of Computer Science*, pp. 160–164. doi: [10.1109/SFCS.1982.38](https://doi.org/10.1109/SFCS.1982.38).

Yao, Andrew Chi-Chih (1986). "How to generate and exchange secrets". In: *27th Annual Symposium on Foundations of Computer Science*, pp. 162–167. doi: [10.1109/SFCS.1986.25](https://doi.org/10.1109/SFCS.1986.25).

Zanussi, Zachary (2021). *A Brief Survey of Privacy Preserving Technologies*. Data Science Network for the Federal Public Service newsletter. url: <https://www.statcan.gc.ca/eng/data-science/network/privacy-preserving>. Accessed 2021-02-01.

3. CASE STUDIES

3.1 CASE STUDIES IN OFFICIAL STATISTICS

INTRODUCTION

A number of national statistical offices (NSOs) and government agencies are leveraging PETs to enable rich and innovative statistical analysis, whilst protecting the privacy and confidentiality of sensitive information contained within their datasets.¹ This section provides 18 detailed case studies of PETs being leveraged in this way.

The case studies cover a diverse range of use cases that cut across different sectors, leverage different combinations of PETs, and involve collaboration with different types of parties (such as multiple NSOs working together, NSOs working with other government agencies, and NSOs working with private sector organisations). 15 of the case studies describe implementations that are at

concept or pilot stage, and 3 that have been deployed in production environments.

Table 3.1 provides summary information of each of the 18 case studies. Full, detailed case studies follow providing background information, details of the implementation, and a description of project outcomes and lessons learned.

Case studies will also be made available in an online repository^a. In time, we hope that the repository will become a 'live' resource, containing up-to-date case study information, thereby enabling knowledge sharing amongst practitioners of the opportunities and challenges of using PETs in the real world.

^a <https://unstats.un.org/bigdata/task-teams/privacy/case-studies/>

¹ For information about further use cases beyond national statistics, see the UK Centre for Data Ethics and Innovation's use case repository: <https://cdeiuk.github.io/pets-adoption-guide/repository>

TABE 3.1 SUMMARY TABLE OF CASE STUDIES

	PURPOSE	DATASETS	PETS USED	APPLICATION	IMPLEMENTATION STATUS
CASE STUDY 1. Boston Women's Workforce Council: Measuring salary disparity using secure multi-party computation	To measure the gender and racial wage gaps throughout the greater Boston area every 1-2 years.	Real demographic and payroll data from companies and non-profit organisations, large and small, throughout the greater Boston area.	Secure Multi Party Computation	Secure vector addition	Production
CASE STUDY 2. European Statistical System: Developing Trusted Smart Surveys	To develop modern "trusted smart survey" (TSS) techniques that use sensors in smart devices to supplement existing data collection methods.	Sensor data collected on devices of survey participants.	Federated Learning, Secure Multi Party Computation, Homomorphic Encryption.	Privacy-preserving statistical analysis	Proof of Concept
CASE STUDY 3. Eurostat: Processing of longitudinal mobile network operator data	To enable a NSO to safely and confidently conduct analysis on longitudinal Mobile Network Operator (MNO) mobility data.	Summary of daily visited locations by individual (pseudonymised) subscribers extracted from CDR (call data record) or signalling data, for 100 million mobile subscribers.	Trusted Execution Environment	Privacy-preserving statistical analysis	Proof of Concept
CASE STUDY 4. Indonesia Ministry of Tourism: Confidentially sharing datasets between two mobile network operators via a trusted execution environment	To generate tourism statistics from the combined data of two mobile network operators (MNOs).	A list of IMSIs from the two MNOs in border areas for the same time period. The IMSIs were uniformly hashed from the 7th digit onwards.	Trusted Execution Environment	Privacy-preserving statistical analysis	Production
CASE STUDY 5. Italian National Institute of Statistics and Bank of Italy: Enriching data analysis using privacy-preserving record linkage	To enable enriched socio-economic analysis by augmenting data held by Bank of Italy with data held by ISTAT (and vice versa).	Socio-demographic and financial datasets (linkable via tax code common key)	Secure Multi Party Computation	Private Set Intersection with Analytics	Pilot

	PURPOSE	DATASETS	PETS USED	APPLICATION	IMPLEMENTATION STATUS
CASE STUDY 6. Office for National Statistics: Trialling the use of synthetic data at the United Kingdom's national statistics institute	To test engineering and analytical pipelines used by ONS staff and independent researchers.	UK Census data, linked census-mortality data, Covid Infection Survey.	Synthetic data, Differential Privacy	Generating high quality data to test engineering and analytical pipelines	Proof of Concept (Initial examples delivered, building future proof of concepts)
CASE STUDY 7. Samsung SDS (Korea): Data aggregation system	A data aggregation system without trusted 3rd party, to securely compute aggregation key and ratio of common data.	Randomly generated two datasets with various data sizes (from 1 m to 20 m).	Secure Multi Party Computation	Private Set Intersection	Proof of Concept
CASE STUDY 8. Statistics Canada: Measuring the coverage of a data source using a private set intersection	To measure the coverage of a third party data source relative to data held by an NSO.	A third party file consisting of a list of units in a domain of a target population. And a corresponding reference file held by the NSO, containing a similar list of the same units.	Secure Multi Party Computation	Exact privacy-preserving data matching with a keyed-hash function	Proof of Concept
CASE STUDY 9. Statistics Canada: Training a machine learning model for private text classification using leveled homomorphic encryption	To migrate machine learning workloads to a cloud environment whilst ensuring input privacy.	Synthetic product description data, similar to retailer scanner data.	Homomorphic Encryption	Supervised text classification	Proof of Concept
CASE STUDY 10. Statistics Canada: Trialling the use of synthetic data	To create synthetic data sets for training and testing purposes.	33 variables selected from a dataset made of the 2006 long-form Census linked to the 2015 Canadian Mortality Registry 47 variables selected from a dataset made of the 2006 long-form Census linked to the 2015 Canadian Cancer Registry and the 2014 Canadian Vital Statistics Death Database.	Synthetic data	Generating high quality data for training and hackathons	Pilot (synthetic datasets were successfully used as a training aid and to support a hackathon)

	PURPOSE	DATASETS	PETS USED	APPLICATION	IMPLEMENTATION STATUS
CASE STUDY 11. Statistics Korea: Developing a privacy-preserving Statistical Data Hub Platform	A cloud-based data system where Statistics Korea uses state-of-art cryptography to securely link data dispersed across government departments and public institutions.	Various kinds of data. Examples include statistical registers held by Statistics Korea.	Homomorphic Encryption, Secure Multi Party Computation, Differential Privacy	Descriptive statistics and logistic regression in the proof of concept	Pilot
CASE STUDY 12. Statistics Netherlands: Developing privacy-preserving cardiovascular risk prediction models from distributed clinical and socio-economic data	To develop cardiovascular risk prediction models from sensitive healthcare data.	Vertically partitioned datasets with primary care data, secondary care (hospital) data, and socioeconomic data.	Secure Multi Party Computation, Homomorphic Encryption, Secret Sharing, Federated Learning	Record linkage and development of machine learning models	Concept
CASE STUDY 13. Statistics Netherlands: Measuring effectiveness of an eHealth solution using private set intersection	To measure the effectiveness of a specific eHealth solution, without sharing patient information.	Medical data (Hospital), treatment cost data (Medical Insurance Company) and socio-economic data (NSO). Synthetic data was used in the initial Proof of Concept, and real data used for the Pilot phase	Homomorphic Encryption, Secret Sharing, Secure Multi Party Computation	Private set intersection with analytics	Pilot
CASE STUDY 14. Twitter and OpenMined: Enabling Third-party Audits and Research Reproducibility over Unreleased Digital Assets	To evaluate the efficacy of PETS for algorithmic transparency. If successful, the goal is to enable researchers outside of Twitter to perform research on data and models within the firm using privacy-enhancing technologies (without having direct access to the underlying information being studied).	The central datasets in the first project come from the paper Algorithmic amplification of politics on Twitter, in addition to synthetic reproductions of the private datasets therein for the purpose of development and testing. The largest synthetic dataset contains approximately 1 billion rows of data.	Remote execution (sometimes called federated learning/analytics), differential privacy, and secure multi-party computation.	Remote Data Science	Ongoing Proof of Concept

	PURPOSE	DATASETS	PETS USED	APPLICATION	IMPLEMENTATION STATUS
CASE STUDY 15. United Nations Economic Commission for Europe: Trialling approaches to privacy-preserving federated machine learning	To privately train a neural network model on isolated lifestyle data collected by smart devices.	Publicly available dataset on human activity recognition with smart devices' accelerometer and gyroscope data. The data was split into four subsets, one for each participating statistical office, for the purpose of the experiments.	Federated Learning, Homomorphic Encryption, Differential Privacy	Development of a machine learning model.	Proof of Concept
CASE STUDY 16. United Nations PET Lab: International Trade	Enable multiple national statistical offices (NSOs) to perform reconciliation and joint analysis on independently collected trade datasets.	The datasets involved were originally from the UN Comtrade Datasets and are now being extended to integrate third-party data sources.	Differential Privacy, Secure Enclaves, Secure Multi Party Computation	Reconciliation and joint trade analysis	Proof of Concept (ongoing)
CASE STUDY 17. United States Census Bureau: Deploying a differentially private Disclosure Avoidance System for the 2020 US Census	To protect against the disclosure of sensitive information collected by the census.	Data from the 2020 US census, and associated data from annual surveys (e.g. the American Communities Survey).	Differential Privacy	Statistical disclosure	Production (2020 result release subject to legal challenges)
CASE STUDY 18. United States Department of Education: Analysing student financial aid data using privacy-preserving record linkage	To compute statistics on average student loan and grant data across the US for 30 categories of undergraduate students	Student financial records for financial aid loans and grants	Secure Multi Party Computation	Private Set Intersection with Analytics	Pilot

CASE STUDY 1: BOSTON WOMEN'S WORKFORCE COUNCIL: MEASURING SALARY DISPARITY USING SECURE MULTI-PARTY COMPUTATION

Purpose	To measure the gender and racial wage gaps throughout the greater Boston area every 1-2 years
Datasets	Real demographic and payroll data from companies and non-profit organisations, large and small, throughout the greater Boston area
PETs used	Secure Multi Party Computation
Application	Secure vector addition
Details of computation	Organisations contribute a spreadsheet containing more than 600 cells of data. The Boston Women's Workforce Council receives the summation of each cell across all participating organisations.
Parties and trust relationship	More than 100 participating organisations act as input parties; the Boston Women's Workforce Council serves in a compute and output party role; Boston University serves as a compute party. Participants trust BU and the BWWC to behave semi-honestly, with the ability to audit and verify code.
Implementation status	Production
Resources	Boston Women's Workforce Council reports Data submission website Open-source code repository on GitHub Publications about the PET used in this project appear at SOUPS 2019 , COMPASS 2018 , SecDev 2016 , and the Communications of the ACM .

BACKGROUND

The Boston Women's Workforce Council (BWWC) is an organisation whose vision is to eliminate all gender and racial wage gaps in the metropolitan region of Boston, Massachusetts. It is a public-private partnership between the Boston Mayor's Office and a collection of employers in the greater Boston area. To date, over 250 Boston-area employers have signed a pledge called the "100% Talent Compact," in which they pledge to examine the root causes of the wage gap, discuss and use evidence-based methods to close the gap, and measure their progress over time.

CASE STUDY DESCRIPTION

The BWWC has used secure multiparty computation (sMPC) five times—in 2015, 2016, 2017, 2019, and 2021—to measure the city-wide wage gap using real salary information of approximately 1 in 6 employees in the greater Boston area. The BWWC receives the average wage grouped by gender, race, and ethnicity; however, it cannot see any individual employer or employee’s salary data.

In more detail, employer’s secret-share their payroll data between two entities: the BWWC and a team of researchers and software engineers at Boston University. The BWWC receives only the city-wide average statistics as output, and Boston University sees neither the inputs nor the outputs. All actors agreed on the civic value gained by learning city-wide wage gap statistics, and hence expressed interest in contributing toward its calculation. Still, sMPC was necessary to provide a safe and secure way to perform this measurement while protecting input salary data.

Usability and explainability were the core guiding principles

behind the design and deployment of sMPC for the wage gap analysis. For ease of use by input parties, the Boston University team designed a web application (viewable at 100talent.org) for organisations to contribute their data; the client browser then checks for common types of data entry error and (if none exist) splits all data into two secret shares. The BWWC can also connect over the web to perform its roles as a computing party and recipient of the output data; only the Boston University team must configure and maintain a web server. In order to spur adoption, it was also important that the data aggregation process was easy to describe to all participants such as IT staff, HR and diversity officers, general counsel, and executives at the participating companies. We used the figure below to explain the secure summation protocol to a wide range of participants, and also made all of our code open source (at <https://github.com/multiparty/>) so that any participant can audit and verify that the code adheres to the sMPC protocol described in the figure.

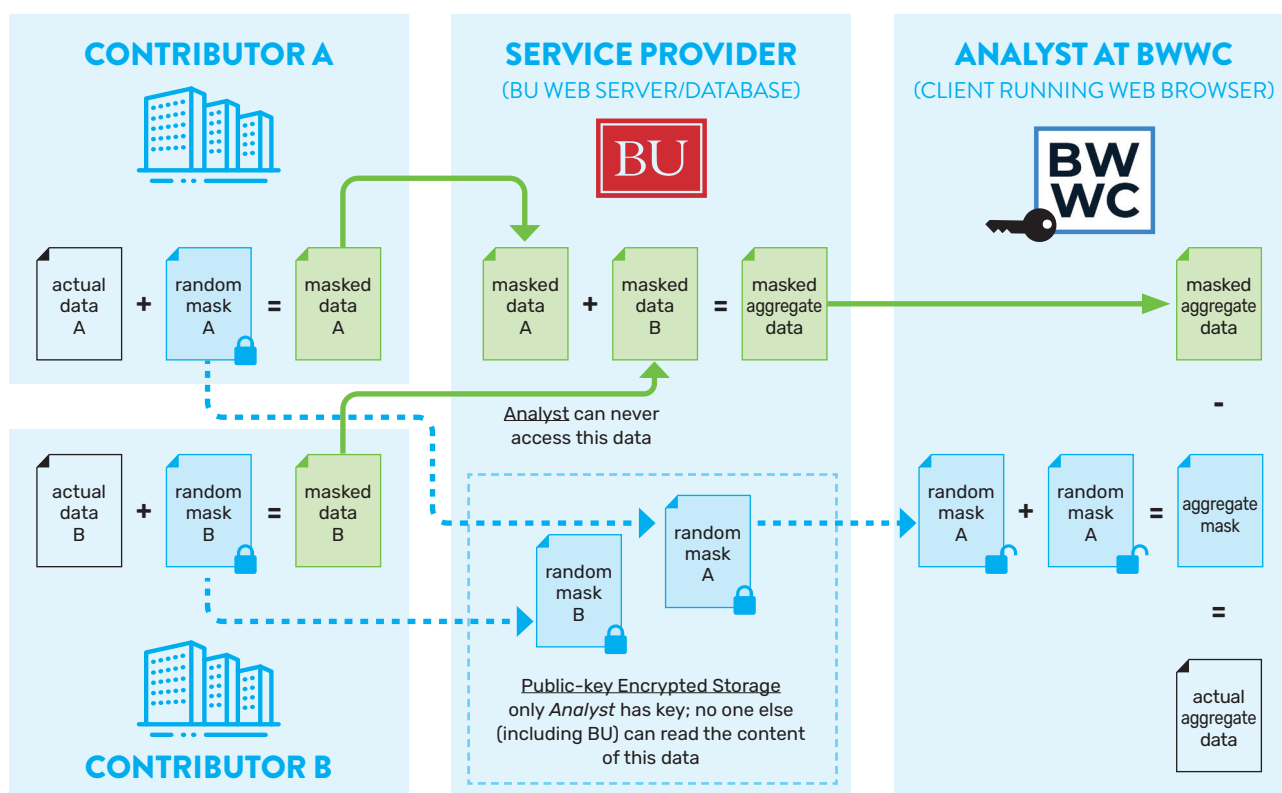


Figure 3.1: Diagram shown to data-contributing organisations to explain why only the sum of all contributors’ salary information is revealed to the Boston Women’s Workforce Council (BWWC), whereas neither Boston University nor the BWWC can see any inputs or intermediate information.

OUTCOMES AND LESSONS LEARNED

Our custom web-based sMPC platform has been deployed five times by the Boston Women's Workforce Council, and it has also been used by other non-profit entities such as the [Greater Boston Chamber of Commerce](#) who wanted a privacy-preserving yet accessible data science platform. We have learned several lessons throughout this process—working hand-in-hand with the BWWC and other members of the target community— and have iteratively improved the platform accordingly.

One major lesson learned was the importance of starting small: finding a single aggregate statistic of interest to a large community; choosing a semi-honest sMPC protocol that is easy to explain to the target audience; building an initial software implementation with a small number of lines of code for ease of auditing; and running small-scale pilots with fictitious data or data from a smaller cohort of community members to build awareness of how to use the front-facing website. After achieving initial adoption, we have slowly added features of interest to the target community, such as support for multiple-choice questions where the BWWC only learns the total number of people who selected each option, and improvements to the data entry process that we [we designed together with human factors experts](#).

The Boston University team has learned several other lessons as well. First, the bottleneck to the adoption of PETs rarely involves the (in)efficiency of the technology itself. The first instance of the wage gap calculation required nearly two years of discussions with the input parties, social scientists, and city officials and the data entry portal was open for weeks for the employers to contribute secret-shared data - the actual runtime of the sMPC algorithm was inconsequential in the grand scheme of things. Second, the BWWC turned to PETs after trying unsuccessfully to find someone to serve as a trusted third party. Using sMPC turned out to be quicker, cheaper, and safer than establishing new trust relationships involving sensitive data. Finally, it is strongly recommended to work with usability and human factors experts from the start when deploying a PET, both because these relatively new tools often lack the ease of use of existing systems and because the privacy features make it difficult (if not impossible) to recover from data entry errors after the fact.

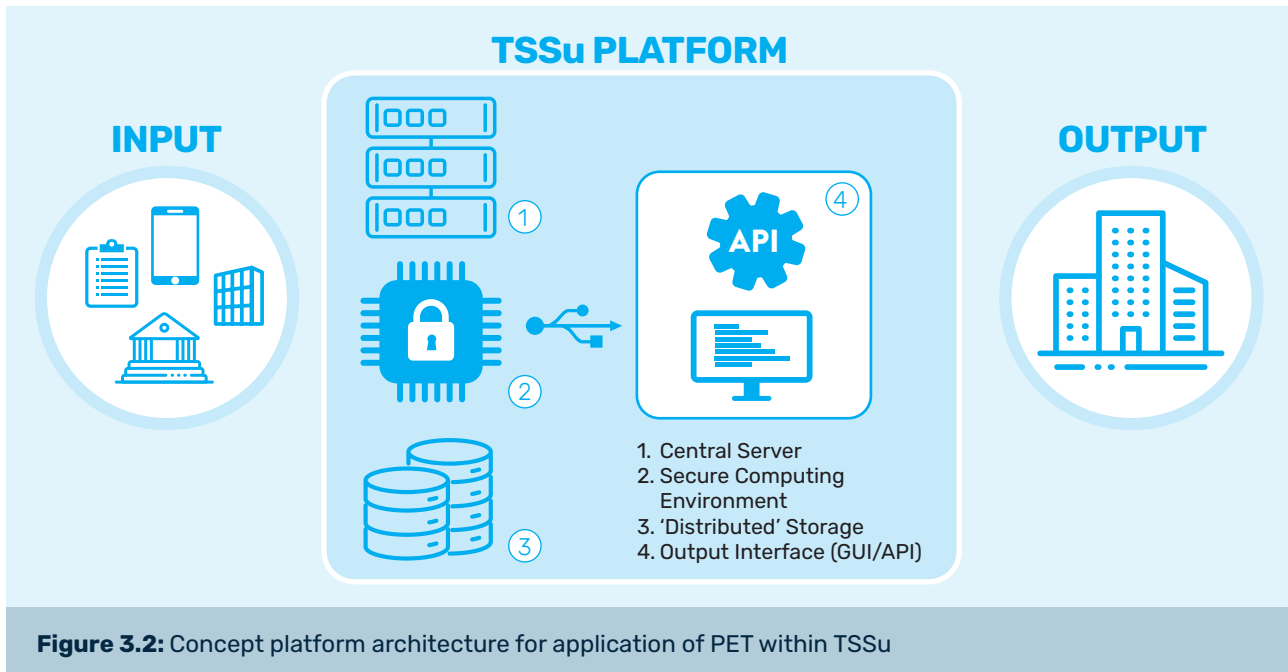
CASE STUDY 2: EUROPEAN STATISTICAL SYSTEM: DEVELOPING TRUSTED SMART SURVEYS

Purpose	To develop modern “trusted smart survey” (TSS) techniques that use sensors in smart devices to supplement existing data collection methods.
Datasets	Sensor data collected on devices of survey participants.
PETs used	Federated Learning, Secure Multi Party Computation, Homomorphic Encryption
Details of computation	After a local pre-processing stage (that reduces the complex set of data about a single individual down to a structured individual record) the main PET operation will be aggregation to produce statistics at municipal, regional or national level.
Parties and trust relationship	A large number of survey participants act as input parties; a single NSO acts as output party. Longer term, this could be extended so there are multiple NSOs acting as output parties using the same TSS backend.
Implementation status	Proof of Concept
Resources	https://ec.europa.eu/eurostat/cros/content/essnet-smart-surveys_en https://doi.org/10.1017/dap.2020.7 https://europa.eu/!DjbbHw https://europa.eu/!FU98Jt

BACKGROUND

The term “*trusted smart survey*” has been [proposed by Eurostat](#) as an augmentation of the “*smart survey*” concept by PET technologies. The term *smart surveys* has been used to refer to surveys based on smart personal devices, typically the smartphone. Smart surveys involve (continuous, low-intensity) interaction with the respondent and their personal device(s). They combine (inter)active data provided explicitly by the respondent (such as responses to queries, or shared images), together with passive data collected in the background by sensors (e.g. accelerometer, GPS) on their smart device(s) (e.g. smartphone, smartwatch, home-assistant). The term [trusted smart surveys](#) refers to an augmentation of the smart survey concept by technological solutions that increase trustworthiness, thereby promoting public confidence and participation.

CASE STUDY DESCRIPTION



Constituent elements of a trusted smart survey are the strong protection of personal data through privacy-preserving system design, and full transparency and auditability of the algorithms used to process the data that is collected. Appropriate platform architecture (e.g. using distributed computing) and the use of different PETs (e.g. sMPC, secure enclaves) can ensure the smart survey provides these elements of privacy, transparency, and auditability.

Privacy-preserving survey design/development is a significant engineering challenge. For example, implementing and testing the data pre-processing pipeline without granular access to the underlying raw data can be difficult. Hence, for new surveys, it might be necessary to perform small pilots with respondents who consent to the use of their data without PET to aid in proper design.

Varying legal requirements and survey specifications across countries complicate application of PET and require harmonisation (e.g. federated learning at European level will not work if classifications differ between countries)..

OUTCOMES AND LESSONS LEARNED

Privacy preserving analysis seems theoretically feasible given proper architecture and survey design. Concept architecture is shown in figure 1. Note that the architecture has not been put into practice at the time of writing.

Practical implementation will take considerable work by people with various kinds of expertise (cryptography, DevOps, software development, statistics, AI, legal and more).

For the federated learning approach, we learned the following lessons:

- For i.i.d data, federated learning performs as expected
- Secure aggregation of weights is quite readily adopted using partially homomorphic encryption (Paillier encryption scheme)

For the secure multi party computation approach, we learned the following lessons:

- Standard aggregate functions (e.g. sum, count) can readily be adapted to secret sharing based sMPC protocols
- Secret sharing based sMPC technically forbids use of data in branching operations (e.g. selecting values based on conditions). However, workarounds are possible without loss of privacy (though complexity and computational overhead quickly increases with the amount of 'branching' operations, see p. 48-51 of [ESSnet Smart Surveys document](#)).

CASE STUDY 3: EUROSTAT: PROCESSING OF LONGITUDINAL MOBILE NETWORK OPERATOR DATA

Purpose	To enable a NSO to safely and confidently conduct analysis on longitudinal Mobile Network Operator (MNO) mobility data.
Datasets	Summary of daily visited locations by individual (pseudonymised) subscribers extracted from (call data record) CDR or signalling data, for 100 million mobile subscribers.
PETs used	Trusted Execution Environment
Details of computation	Articulated workflow consisting of a chain of simple operations performed regularly. Longitudinal MNO analysis and integration of MNO and NSO data takes place within a secure enclave/trusted execution environment using a predefined set of algorithms that deliver aggregate (non-personal) data in output.
Parties and trust relationship	MNO and NSO act as both input and output parties, and their relationship is assumed honest-but-curious.
Implementation status	Proof of Concept
Resources	https://ec.europa.eu/eurostat/cros/content/eurostat-cybernetica-project_en

BACKGROUND

[Eurostat](#) has developed a proof-of-concept solution with a technology provider. The main goal of this project was to explore the feasibility of a Secure Private Computing solution for the privacy-preserving processing of Mobile Network Operator data. The technology of choice for this project was a Trusted Execution Environment (TEE) with hardware isolation (specifically, Intel SGX) in combination with the privacy-enhancing software [Sharemind HI](#) developed by [Cybernetica](#). The solution was tested on synthetic data emulating a population of up to 100 million mobile users. Detailed information about the project scenario and results are available from the [public project page](#). Eurostat is open to support NSOs and MNOs that are interested in testing the solution in the field.

CASE STUDY DESCRIPTION

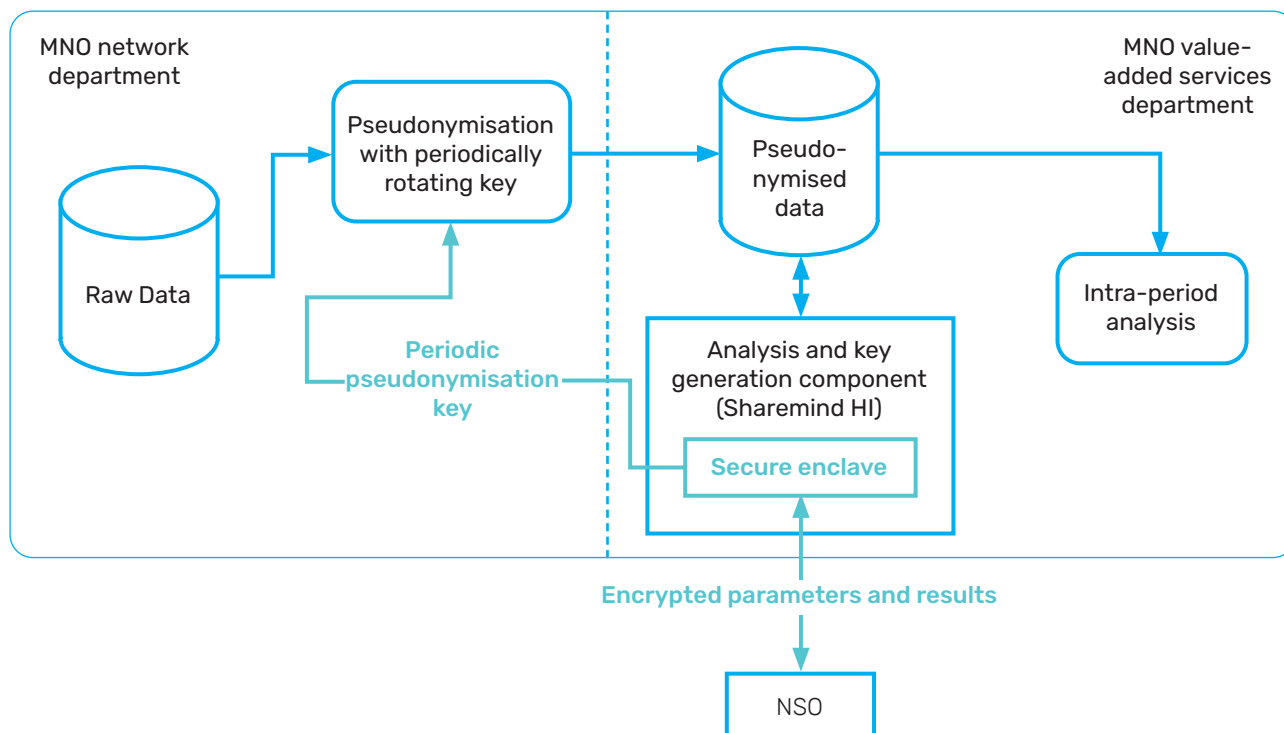


Figure 3.3: Block diagram of the developed solution.

The MNO collects records in the form $\langle user_pseudonym, time, location \rangle$ for the mobile users. In the reference scenario, the data protection policy is defined such that all pseudonyms are changed every period T , in order to reduce the risk and impact of re-identification attacks. However, the statistical methodology defined by the NSO requires observation of the mobile user longitudinally for a much larger window $W \gg T$ (e.g. $T=24$ hours and $W=3$ months) (e.g. $T=24$ hours and $W=3$ months) in order to reliably identify the few locations that constitute the “usual environment” of the mobile user.

It is also assumed that the NSO is set to receive only non-personal aggregate data (fulfilling some k -anonymity condition) but not raw data. Additionally, the NSO has other data that could be used to better calibrate the MNO statistics (e.g. census grid data at fine resolution) but cannot be passed to the MNO.

A Secure Private Computing solution is developed to ensure that longitudinal MNO analysis and integration of MNO and NSO data takes place exclusively on a predefined set of algorithms that deliver aggregate (non-personal) data in output.

OUTCOMES AND LESSONS LEARNED

The PoC showed that scalability is not a point of major concern in the considered scenario. Despite the limited amount of memory in the enclaves, the I/O bandwidth (with hardware accelerated encryption) proved to be sufficient in the test scenario. Alternative solutions based on other TEE technologies or Secure Multiparty Computation could be explored in the future.

The legal analysis conducted in the project revealed a complex interplay between statistical, telecoms and data protection regulations, with a marked heterogeneity across EU countries due to different national legislation.

Furthermore, during the project it became evident that bringing together a multi-disciplinary team of experts - including specialists in statistical methodologies, experienced security and privacy engineers, and legal experts - is a key success factor for inception projects in the field of Secure Private Computing.

CASE STUDY 4: INDONESIA MINISTRY OF TOURISM: CONFIDENTIALLY SHARING DATASETS BETWEEN TWO MOBILE NETWORK OPERATORS VIA A TRUSTED EXECUTION ENVIRONMENT

Purpose	To generate tourism statistics from the combined data of two Confidential sharing of datasets of two mobile network operators (MNOs).
Datasets	A list of IMSIs from the two MNOs in border areas for the same time period. The IMSIs were uniformly hashed from the 7th digit onwards.
PETs used	Trusted Execution Environment
Details of computation	Statistics are generated from the input data in a trusted execution environment (Intel SGX), through the Sharemind HI platform
Parties and trust relationship	Two MNOs act as input parties; Sharemind serve in a compute role; the Ministry of Tourism acts as output party.
Implementation status	Production
Resources	https://sharemind.cyber.ee/sharemind-hi/ https://netmob.org/assets/netmob19_withFCC.pdf

BACKGROUND

Timely and accurate statistics on cross-border tourism can be difficult to attain for various reasons, including privacy and confidentiality concerns if roaming information from telecom companies is used. Indonesia, led by its Ministry of Tourism, is one of the first countries in the world to use data from mobile network operators for measuring cross-border tourism activity. Positium - an analytics company specialising in mobile positioning data - has set up a system for the Ministry based on data from one of the mobile operators. The Ministry wanted to establish a true baseline for roaming market share, which is hard to estimate due to subscribers cross-roaming in the networks of different operators during a single visit. The challenge was how to compare data without explicitly sharing it, as location data contains sensitive information about customers, and confidential business information of the operators.

CASE STUDY DESCRIPTION

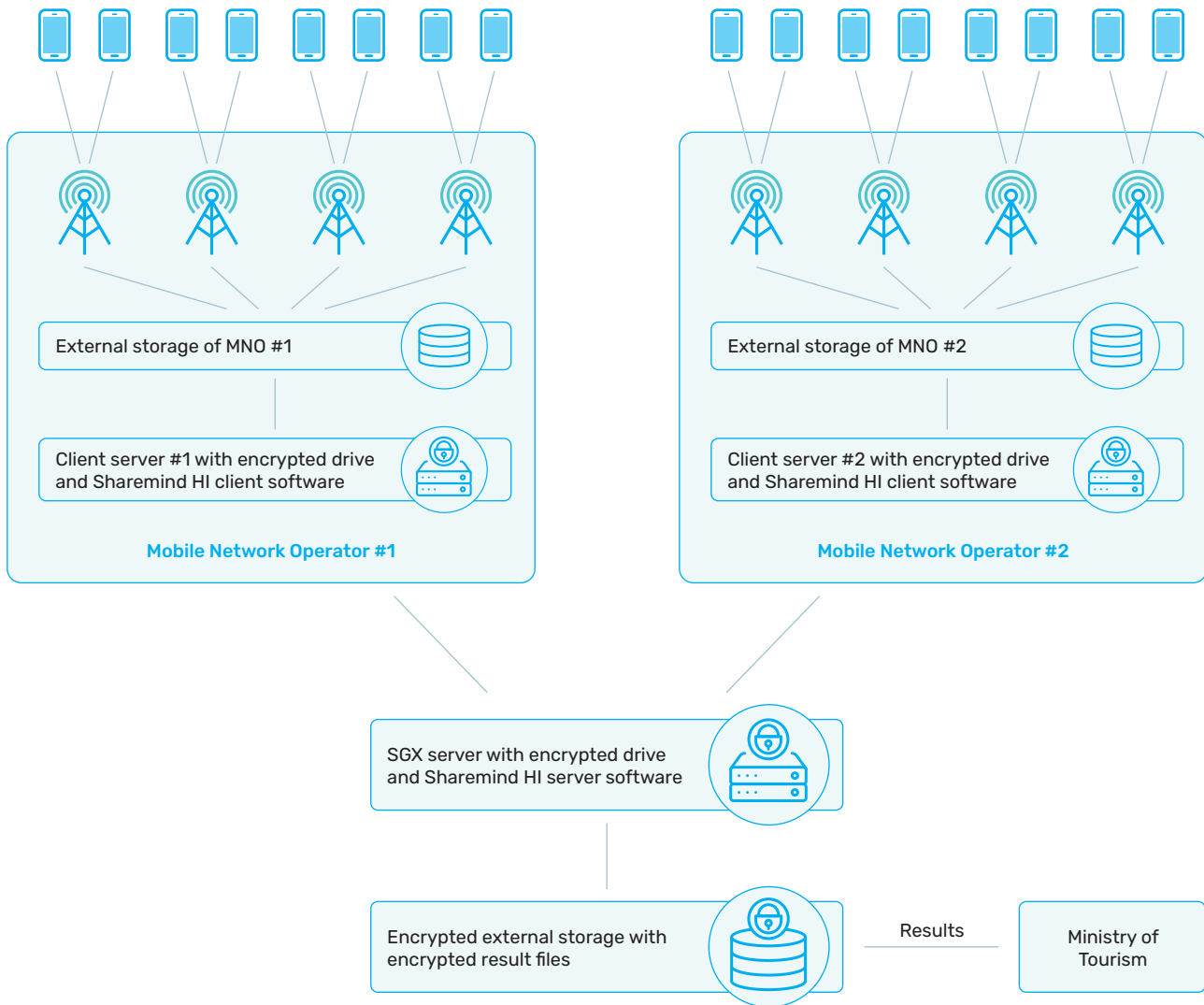


Figure 3.4: Simple local cloud setup for cross-roaming analysis

Mobile positioning data provides insights into the quantities and movements of tourists. As tourists move around, their mobile phones roam through multiple local mobile network operators (MNOs). Cross-roaming is a situation where a person might use two or more different MNOs in the country of reference. A complete understanding of the nature of cross-roaming can only be derived when unique subscriber information (IMSI) is compared across several operators. Because of privacy concerns, this is a complex task that requires uniform hashing of IMSIs over at least two operators.

Sharemind is a secure computing platform created to specifically reduce the risk of a privacy breach when processing confidential data. The data is encrypted at the source, by the data owner, and only then sent to the Sharemind service. The host of the service will not have access to the unencrypted data nor the encryption keys. The solution protects data at rest and in transit and surpasses state-of-the-art methods with protecting data in use. It does not remove data protections even while processing, so data remains protected by cryptographic means during the whole analysis. The Trusted Execution

Environment (TEE) technology used in Sharemind HI to implement privacy-preserving data processing is the Intel Software Guard Extensions (SGX) available in modern Intel processors. The three key concepts that SGX provides to protect data are enclaves, attestation and data sealing.

MNO data is transferred to the Sharemind HI environment, where analysis on the data is carried out, and the encrypted results are shared with the Ministry of Tourism. .

OUTCOMES AND LESSONS LEARNED

The result provided the Ministry of Tourism with information on roaming counts and roamer overlap between the two biggest telecom providers in Indonesia, enabling an accurate calculation of roaming market share. The result is still used in official statistics until today as Indonesia produces monthly tourism statistics indicators based on mobile phone data.

The rapid development and deployment of the solution depended heavily on the good working relationship between the ministry of tourism, statistical office, mobile operators, mobile positioning expert and privacy technology expert organizations. The solution made confidential sharing of private datasets possible within the existing legal and business environment. It is still the only known solution for understanding cross-roaming subscriber overlap of MNO datasets. The solution is extendable to multi-MNO settings and performance is good even on commercial off-the-shelf hardware.

CASE STUDY 5: ITALIAN NATIONAL INSTITUTE OF STATISTICS AND BANK OF ITALY: ENRICHING DATA ANALYSIS USING PRIVACY-PRESERVING RECORD LINKAGE

Purpose	To enable enriched socio-economic analysis by augmenting data held by Bank of Italy with data held by ISTAT (and vice versa).
Datasets	Socio-demographic and financial datasets (linkable via tax code common key)
PETs used	Secure Multi Party Computation
Application	Private Set Intersection with Analytics
Details of computation	ISTAT and Bank of Italy perform an Exact PSI using the shared tax code key. The intersection is encrypted and transferred to the third “linker” party. ISTAT and Bank of Italy submit queries to the linker, which can perform aggregation and counts against the data on-demand, with outputs transmitted to ISTAT and Bank of Italy.
Parties and trust relationship	Bank of Italy and ISTAT act as input, compute, and output parties; third “linker” party serves in a compute role. Organisations trust each other (“honest but curious” threat model).
Implementation status	Pilot
Resources	

BACKGROUND

Two parties, ISTAT and Bank of Italy, own databases D1 and D2 respectively, which contain socio-demographic and financial data. D1 and D2 have a common key (tax code), which can be exploited to perform an Exact PSI. The parties wish to enrich their information assets by learning the results of a statistical analysis applied to the intersection of their databases - a so-called Private Set Intersection with Analytics (PSI-A).

CASE STUDY DESCRIPTION

ISTAT and Bank of Italy are involved in a first phase of private set intersection; a third party named ‘Linker’ returns the results of the calculation (aggregation counting).

The process implemented requires the following four phases:

- Preliminary phase: agreement between the parties on base protocol parameters
- Exact PSI: private intersection of common database keys
- Loading: transmission of encrypted data to the Linker
- Query: submission of queries to the Linker and transmission of results to the two parties

We apply an additive-only scheme and a set intersection. Aggregation and counts can be performed on demand.

This use case can be generalised to two public organisations that want to generate insights from the join of their vertically split datasets. Such a solution could be provided by (transnational) infrastructure with strong security and protocol assurance, including assurance needed for data protection regulators.

The extensions of statistical analyses in the PSI-A framework could be explored with homomorphic cryptography.

PSI - PRIVATE SET INTERSECTION WITH ANALYTICS

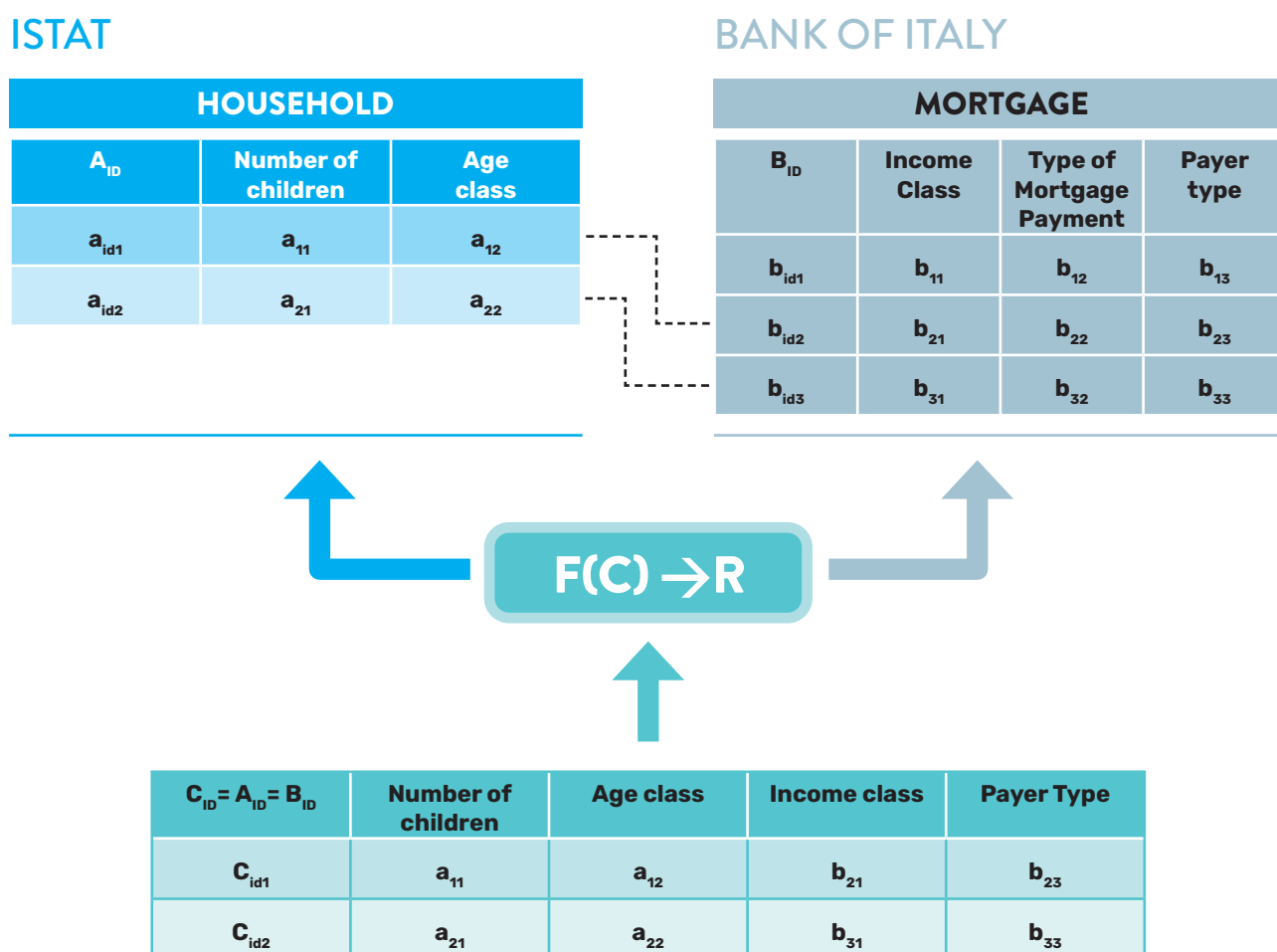


Figure 3.5: An example of Private Set Intersection with Analytics (PSI-A)

OUTCOMES AND LESSONS LEARNED

This use case has been developed in an experimental context with the use of synthetic data, so we haven't dealt with the legal considerations in relation to data access or data treatment, but only with the technical and functional aspects.

We have learned how important the preliminary investigation phase is for the application of the PSI protocol. Applying the techniques effectively requires contextual understanding of the specific use case, which can affect certain design and parameter choices. Specifically, specifying appropriate privacy requirements

is a far from trivial task, and has implications for all subsequent design choices.

In this multi-party scenario, we also learned how roles are distributed and the role of each individual party involved in the protocol. We have learned how important a third party is. In this case study, the data is encrypted with the same key (known by both data owners), and to perform the analytics function in a privacy-preserving manner, the support of a third party, who does not know the key, is required. However, in this use case the third party cannot always fully guarantee the desired results in terms of privacy preservation, as it does not check the number and type of queries that the two main parties perform.

CASE STUDY 6: OFFICE FOR NATIONAL STATISTICS: TRIALLING THE USE OF SYNTHETIC DATA AT THE UNITED KINGDOM'S NATIONAL STATISTICS INSTITUTE

Purpose	To test engineering and analytical pipelines used by ONS staff and independent researchers.
Datasets	UK Census data, linked census-mortality data, Covid Infection Survey.
PETs used	Synthetic data, Differential Privacy
Details of computation	Varies by application, but includes the use of Generative Adversarial Networks, and differential privacy.
Parties and trust relationship	Single input party (NSO) and many output parties (public or private entities) with no relationship
Implementation status	Proof of concept (Initial examples delivered, building future proof of concepts)
Resources	2021 Census Dress Rehearsal Using GANs to create synthetic data SynthGauge python package for evaluating synthetic data

BACKGROUND

The ONS Data Science Campus has been studying the problem of synthetic data generation since 2018. Our work is balanced between conducting research into methodologies for creating and evaluating synthetic data, and finding opportunities to apply synthetic data in practice..

At present, the ONS is not considering the use of synthetic data for decision making: our applications to date have focused on the use of synthetic data to facilitate data and machine learning pipelines. Our ongoing proof-of-concept projects are looking to extend this to producing synthetic datasets to enable researchers to understand our data whilst awaiting accreditation to see sensitive information.

CASE STUDY DESCRIPTION

The ONS Data Science Campus has recently delivered two successful projects demonstrating the benefit of synthetic data.

In preparation for the UK's 2021 Census, a rehearsal was conducted: we generated multiple synthetic datasets for testing the load balancing and functions used in the processing pipeline of the ONS Census. Some of these tests required that the synthetic data had distributions that are representative of the population in the true census.

Broadly, the work involved modifying the variables of the 2011 Census dataset to match the format of the 2021 Census and producing plausible synthetic data for use in testing.

In a second application, we generated a synthetic version of the UK's Covid Infection Survey to help debug a machine learning pipeline that was estimating national Coronavirus infection rates. The data needed synthesising so that it could be shared outside of teams cleared to access the data, and the method of synthesis needed to be easily explainable to facilitate the quick release of the data.

Going forward we are exploring the use of differential privacy for synthesising data. We are approaching this by constructing noisy marginal distributions of our data, and then building synthetic datasets with these distributions, testing methods developed as a part of the NIST Differential Privacy Challenge.



Figure 3.6: Depiction of our current approach using Differential Privacy to generate privatised synthetic data.

OUTCOMES AND LESSONS LEARNED

Our successful development of synthetic data for the Census rehearsal enabled pipelines to be built in advance of receiving real Census data, allowing greater time to prepare and enhance data processing.

The synthetic Covid Infection Survey data was tested in the modelling pipeline, and ran without encountering errors. This result was initially surprising as the data had been designed to trigger the same issues. Through identifying the statistical properties that were only present in the true data, analysts were quickly able to identify model misspecifications in edge cases, and debug their model.

One of the challenges that this project highlighted is the need to balance not only utility with privacy when generating synthetic data, but also the interpretability of the process used to synthesise data. Differential privacy offers the opportunity to allow stakeholders to simply express their risk tolerance by setting a few parameter values, but we require further research into how to make this process transparent and interpretable to non-subject matter experts.

CASE STUDY 7: SAMSUNG SDS (KOREA): PRIVACY-PRESERVING DATA AGGREGATION SYSTEM

Purpose	A data aggregation system without a trusted third party, to securely compute aggregation key and ratio of common data.
Datasets	Randomly generated two datasets of various sizes (from 1 m to 20 m)
PETs used	Secure Multi Party Computation
Application	Private Set Intersection
Details of computation	Two synthetic datasets are generated using the Faker python module, and are saved to two different computers. The PSI protocol is applied, and the results - including aggregation key and ratio of common data - are saved in CSV format.
Parties and trust relationship	Two organizations, not necessarily trusting one another, would like to perform vertical aggregation without relying on a trusted third party.
Implementation status	Proof of Concept
Resources	

BACKGROUND

Data aggregation can be classified into two types: vertical and horizontal aggregation. In the latter case, small sets of data with the same features or attributes can be aggregated to improve the statistical power of analyses. In vertical aggregation, different attributes related to a specific object (e.g. an individual) may reside in different organizations' datasets, and such data may need to be joined together for analysis. For example, EU citizens often move across borders within the EU, and there may exist a need to aggregate features or attributes on a group of people, where such data resides in two or more countries.

PET technologies are maturing, but are not yet fully integrated into statistical and analytics tools or services. For this reason, conventional de-identification technologies are heavily used. Without PETs, vertical integration requires a trusted third party (TTP) to securely aggregate the datasets. Establishing this trusted third party can be a slow, manual process. We thus suggest the use of privacy-enhancing technologies - specifically, private set intersection (PSI) - which removes the need for a TTP and allows two or more parties to securely and efficiently aggregate datasets

CASE STUDY DESCRIPTION

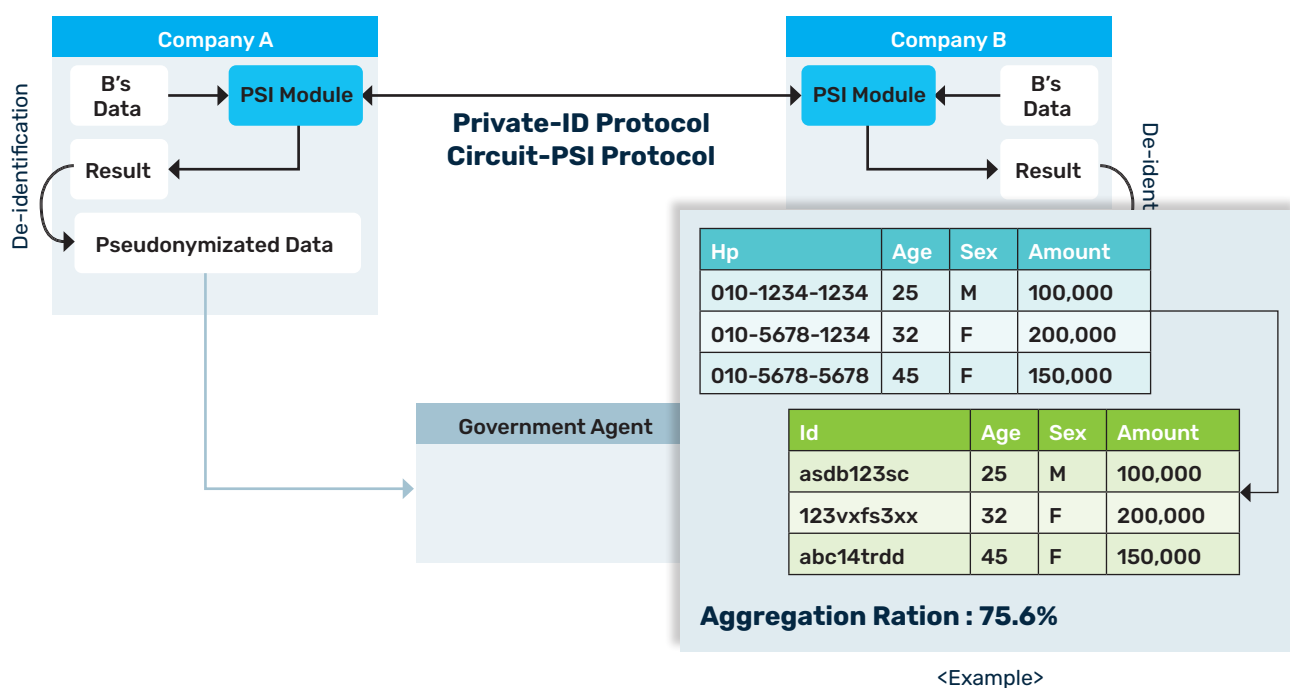


Figure 3.7: The overall process of the data aggregation system.

Two organizations have datasets which contain between one and twenty million identifiers. A “Private-ID” protocol is used to make a common ID, which cannot be linked to the original identifiers. A “Circuit-PSI” protocol is then used to compute the aggregation ratio - 75.6% in our example in the figure above. This means 75.6% of identifiers are shared between the two organizations. Throughout the process, there is no risk of private information being leaked. When tested in a LAN environment, the process takes between 2 and 30 mins to complete, depending on the size of the dataset. The accuracy of aggregation remains the same as in the case of conventional manual processes.

OUTCOMES AND LESSONS LEARNED

Samsung SDS’s project describes a secure and efficient data aggregation process using PETs to perform PSIs. The use case also eliminates the need for a trusted third party, which is often difficult to establish, particularly when using data held across borders.

There are existing data aggregation services operating today that use TTPs and conventional cryptographic and de-identification techniques, which often take a long time to set up, and can be prone to human error. This use case shows that using PSI can remove the need for a TTP, thereby removing the overhead associated with establishing one, whilst ensuring that results remain accurate.

CASE STUDY 8: STATISTICS CANADA: MEASURING THE COVERAGE OF A DATA SOURCE USING A PRIVATE SET INTERSECTION

Purpose	To measure the coverage of a third party data source relative to data held by an NSO.
Datasets	Proof of concept is using synthetic data generated from public census data: <ol style="list-style-type: none"> 1. a third party file consisting of a list of units in a domain of a target population 2. corresponding reference file held by the NSO, containing a similar list of the same units.
PETs used	Secure Multi Party Computation
Application	Exact privacy-preserving data matching with a keyed-hash function based on Christen (2012, Chapter 8.3.1).
Details of computation	A “Linker” party receives hashed records from the NSO and the third party, and evaluates the size of their intersection by applying record matching and statistical modelling (including a capture-recapture approach) whilst accounting for linkage errors.
Parties and trust relationship	A data holder at the NSO and the third party act as input parties; a “Linker” at the NSO serves in a compute role. The data holder is also the output party, and learns the result of the computation. The data holder and Linker do not collude, and the third party must trust the NSO to implement access control mechanisms that ensure there is no collusion.
Implementation status	Proof of Concept
Resources	<p>[1] Christen, P. (2012). Data Matching, Berlin:Springer.</p> <p>[2] Dasylyva, A. and Goussanou, A. (2020). “Estimating linkage errors under regularity conditions”, Proceedings of the Survey Methods Section, American Statistical Association.</p>

BACKGROUND

An NSO may need to measure the coverage of a third party dataset with respect to a target dataset or population, to inform its NSO data acquisition strategy or provide a value-added service (e.g. a data quality evaluation service) to the third party. To this end, the NSO may use the most basic form of PSI without any transfer of data beyond the linkage variables, where the third party has limited trust in the NSO or membership in the third party dataset is sensitive information. This basic use case is seen as an important first step towards more complex PSI scenarios.

CASE STUDY DESCRIPTION

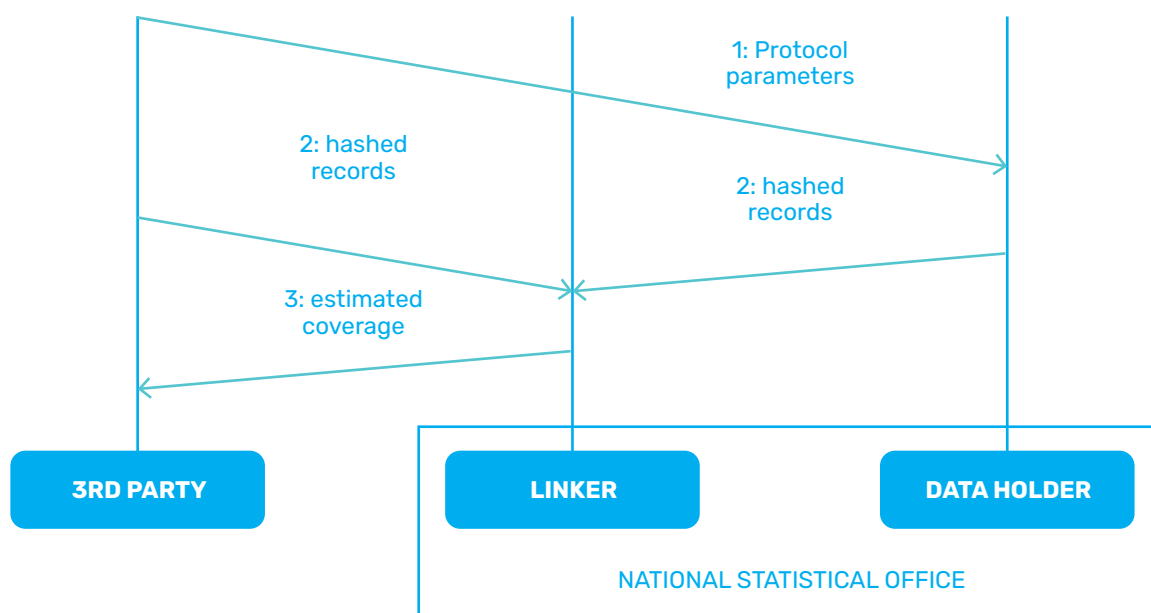


Figure 3.8: Message flow for private set intersection between two parties

The proposed solution adapts the simple three-party protocol by Christen (2012, Chapter. 8) [1], with exact comparisons, two non-colluding parties at the NSO and a statistical model to estimate the size of the intersection and the coverage of the third party source, while accounting for the linkage errors. The two NSO parties include a Linker and a data holder, where the Linker receives the hashed records from the data holder and the external third party, while the data holder has access to all the NSO data. The coverage of the third party source is evaluated without transferring any data beyond the hashed quasi-identifiers.

Python 3 implementation of the Blake2 hash function which is provided in the hashlib module.

It is also possible to implement simple PSI solutions that use quasi-identifiers and account for linkage errors, at least when the goal is to measure the coverage of the third party dataset without any transfer of data beyond the quasi-identifiers.

OUTCOMES AND LESSONS LEARNED

When designing a PSI solution, it is important to capitalize on the nature of the NSO as a public institution, to maximize the public good, i.e. to achieve the best balance of the security guarantees and the solution cost/complexity.

It is possible to implement PSI solutions that are software-based and use open source code. Here, we used the

CASE STUDY 9: STATISTICS CANADA: TRAINING A MACHINE LEARNING MODEL FOR PRIVATE TEXT CLASSIFICATION USING LEVELED HOMOMORPHIC ENCRYPTION

Purpose	To migrate machine learning workloads to a cloud environment whilst ensuring input privacy.
Datasets	Synthetic product description data, similar to retailer scanner data.
PETs used	Homomorphic encryption
Application	Supervised text classification
Details of computation	Input data is encrypted using a leveled homomorphic encryption scheme, and transferred to a cloud environment. A neural network for text classification is trained on the encrypted data, with encrypted model weights returned to the input party.
Parties and trust relationship	Multiple retailers act as input parties (a single input party was used for this proof of concept). The cloud provider acts as the compute party. Statistics Canada acts as the output party. Input parties trust the output party, but neither trusts the compute party.
Implementation status	Proof of Concept
Resources	Z. Zanussi, B. Santos and S. Molladavoudi, <i>Supervised text classification with leveled homomorphic encryption</i> , To appear in the proceedings of the 63rd ISI World Statistics Congress, 2021. Slides from UN ML Group presentation

BACKGROUND

In this use case, a leveled Homomorphic Encryption (HE) scheme was used to train an end-to-end supervised machine learning algorithm to classify synthetic product descriptions similar to retailer scanner data while preserving the privacy of the input data points. This was a proof of concept to explore the feasibility of migrating machine learning workloads to the cloud, whilst ensuring data remains protected.

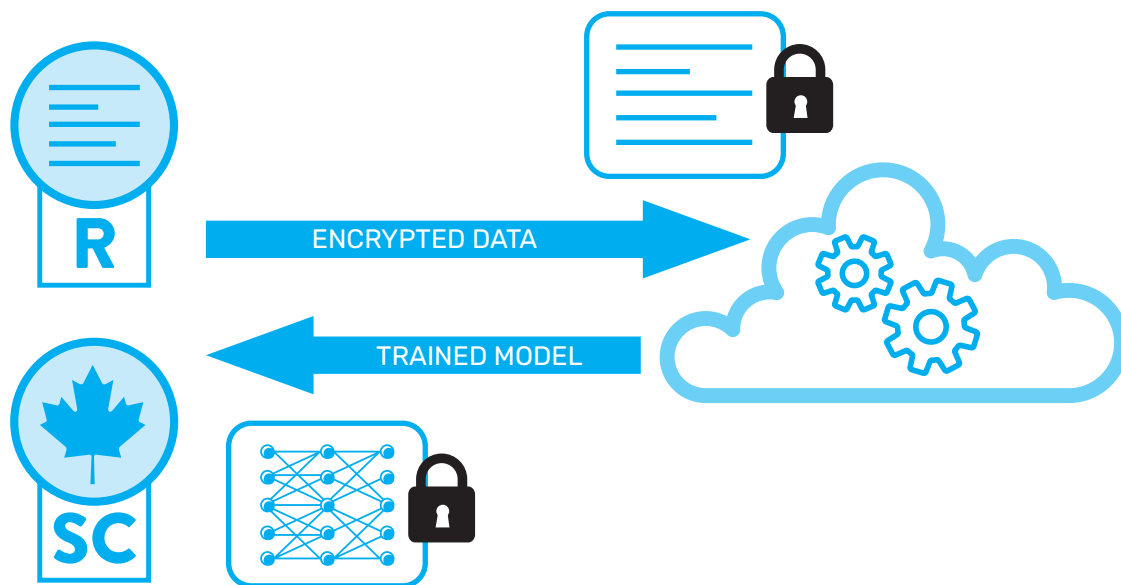


Figure 3.9: A schematic overview of the information flow for training a model in the cloud using encrypted training data.

The input data leaves the input party in an encrypted format, and remains encrypted throughout the computation by the compute party. Once the model (in this case, a neural network) has been trained, the resulting weights are communicated back to the input party, still in an encrypted format. The predictions throughout the training process and inference on the new data points by using the trained model all happen in the ciphertext space on the cloud. The input party is able to decrypt the results using the cryptographic key that was used for the original encryption.

This example use case can be generalised to secure (outsourced) processing of any sensitive data from different (private) data providers that do not trust each other nor the processing parties.

OUTCOMES AND LESSONS LEARNED

The main goal of the use case was to investigate the feasibility of using HE in computationally intensive ML tasks, such as training a neural network while preserving the privacy of the input dataset. Compared to the cleartext experiments, the results of experiments in the ciphertext domain prove that the performance degradation introduced by the inherent noise as well as the approximate computation of HE is manageable. To the best of our knowledge, our experiment is the largest encrypted text classification training problem with neural networks undertaken so far. It is worth noting that HE, as a technology, has advanced to a point where one can take an open-source library and solve a real problem in a reasonable amount of development and compute time.

CASE STUDY 10: STATISTICS CANADA: TRIALLING THE USE OF SYNTHETIC DATA

Purpose	To create synthetic datasets for training and testing purposes.
Datasets	33 variables selected from a dataset made of the 2006 long-form Census linked to the 2015 Canadian Mortality Registry 47 variables selected from a dataset made of the 2006 long-form Census linked to the 2015 Canadian Cancer Registry and the 2014 Canadian Vital Statistics Death Database
PETs used	Synthetic data
Application	Generating high quality data for training and hackathons
Details of computation	Fully Conditional Specification approach with CART and regression methods were used to create a synthetic dataset which allowed access to detailed information in a non-secure environment by students (non-trusted analysts).
Parties and trust relationship	Statistics Canada acts as input party; multiple students/researchers act as output parties. There is no assumption of trust between the parties. However, when the synthetic data was used during a hackathon, participants were asked to agree not to share any data outside the hackathon environment.
Implementation status	Pilot (synthetic datasets were successfully used as a training aid and to support a hackathon)
Resources	

BACKGROUND

Statistics Canada's recent experience with synthetic data is related to specific uses, such as providing datasets of high analytical value to hackathon participants. The participants were allowed to access the data in the hackathon setting under an agreement not to copy or share the data further. Analytical value was comparable to the original datasets. The disclosure risk was evaluated as if the produced synthetic datasets were Public-Use Microdata Files (PUMFs) with real respondents. The original microdata, e.g. census data, health variables and mortality indicators, contains sensitive information, and so could not be made available to outside researchers in an uncontrolled environment.

CASE STUDY DESCRIPTION

In the last two instances in 2018 and 2019, synthetic datasets were created using a mass imputation method – the fully conditional specification approach with the Classification and Regression Tree (CART) method – in order to preserve the analytical value of the original files. The files have been shared in the hackathons and can be offered to other trusted institutions looking for open data. They can also support access via remote desktop connection.

SYNTHETIC DATASETS FOR TRAINING PURPOSES

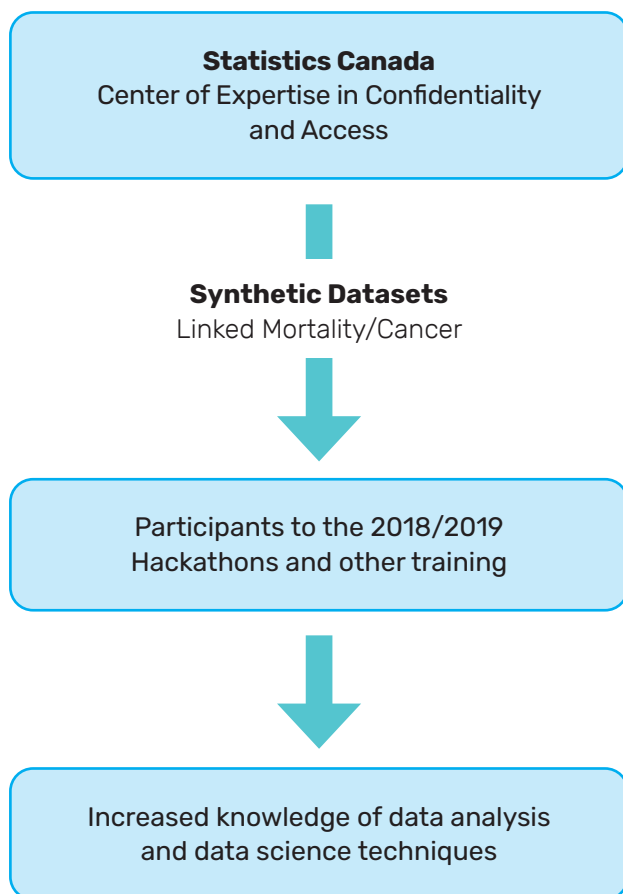


Figure 3.10: Synthetic datasets for training purposes

OUTCOMES AND LESSONS LEARNED

The two hackathon exercises have helped to develop Statistics Canada's experience with synthetic data with high analytical value. The projects have illustrated that synthetic data can be created that preserves the analytic utility of the original data while effectively reducing the risk of disclosure. Both hackathons met the goal of increasing the knowledge and experience of the attendees. Both instances have illustrated the challenges in understanding the risks (real or perceived) associated with creating these files. Finally, in terms of the data's utility, there are challenges associated with developing synthetic datasets that meet a generic analytic goal, i.e. without any prior assumptions on types of analyses to be performed by the users.

CASE STUDY 11: STATISTICS KOREA: DEVELOPING A PRIVACY-PRESERVING STATISTICAL DATA HUB PLATFORM

Purpose	A cloud-based data system where Statistics Korea, leveraging its unique database assets such as the integrated statistical register, uses state-of-the-art cryptography to link data dispersed across government departments and public institutions.
Datasets	Various kinds of data. Examples include statistical registers held by Statistics Korea.
PETs used	Homomorphic Encryption, Secure Multi Party Computation, Differential Privacy
Application	Multiple applications envisioned
Details of computation	In the pilot project, two datasets are linked in their encrypted state, and two computations performed: the generation of descriptive aggregate statistics related to shop space and turnover, and a logistic regression to model the impact of the Covid-19 pandemic on small businesses.
Parties and trust relationship	Central and local governments, as well as private and public companies and academic institutions (input); individuals, private companies, public organisations (output). In general, no trust relationship among the parties is assumed. However, the same entity can play multiple roles, e.g. as both input and output party.
Implementation status	Pilot
Resources	

BACKGROUND

Statistics Korea is promoting the establishment of a public big data system that leverages cutting-edge privacy-preserving techniques to enable the safe linkage and use of scattered governmental data. Enabling safe access to linked, high-quality, large scale datasets can drive innovation that enhances both economies of scope and scale. Accordingly, Statistics Korea aims to maximise the potential value of data by facilitating the linkage between governmental data through statistical registers on population, households, and establishments.

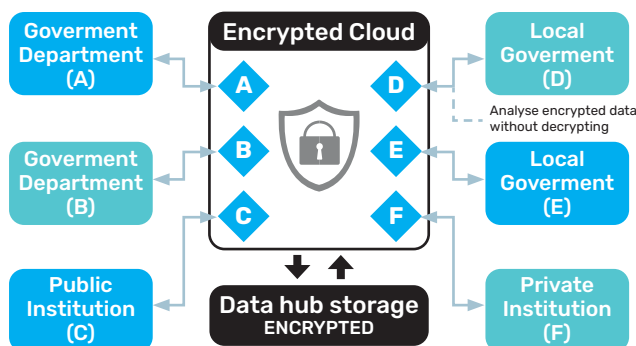
To achieve this, Statistics Korea is promoting the development of privacy-preserving techniques such as homomorphic encryption, differential privacy and synthetic data through national R&D projects, which will be leveraged to construct the Statistical Data Hub Platform between 2021 and 2024, in cooperation with the Ministry of Science and ICT. The development of the Statistical Data Hub Platform aims to incentivise academia and industry to advance and commercialise these technologies.

CASE STUDY DESCRIPTION

As a pilot project, Statistics Korea linked its statistical business register with small business information from Gyung-Gi Province, and performed analysis in their encrypted state to confirm the practicality of the Statistical Data Hub Platform. The two datasets have the following common fields: name of establishment, corporation

registration number, and administrative district code. By linking data in its encrypted state, Statistics Korea was able to confirm that it is possible to combine and use data without exposing sensitive information, thereby validating a critical premise of the Statistical Data Hub Platform.

Additionally, Statistics Korea tested the accuracy and efficiency of statistical analysis performed on



homomorphically encrypted data. Firstly, a descriptive statistical analysis was conducted on turnover and shop space information from encrypted linked data. No difference in accuracy was found between the results of the plaintext and ciphertext analyses, and the ciphertext analysis was efficient in terms of operation time and storage space.

Statistical operation	Time (Sec)	Storage space increase relative to plaintext (MB)
Count	10-11	52
Average	148-153	14
Standard deviation	395-406	28

Table 3.2: Results from statistical analysis on encrypted turnover and shop space data. Sample data: 8,576 franchisees (Korean foods restaurants) in Kyung-Gi province, established between 2015 and 2020

Secondly, a logistic regression was performed on the encrypted linked data. Independent variables were chosen related to the shop's turnover, shop space, type of business district, amongst others (see Table 3.3 for full list of variables). These were used to predict a dependent variable representing the survival of the establishment. An approximate function was derived for the logistic regression, which is necessary to perform logistic regression analysis on homomorphically encrypted data.

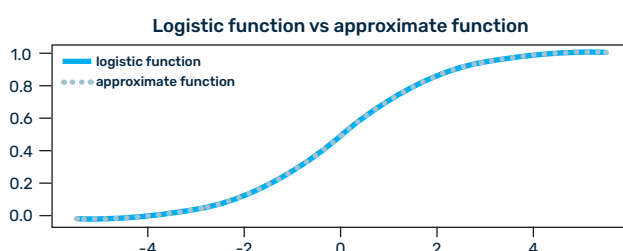


Figure 3.11: comparison of the outcome of performing a logistic regression on the plaintext data (solid line) and applying an approximate regression formula to the plaintext data.

	Logistic function on plaintext	Approximate function on plaintext	Approximate function on ciphertext
log(previous year's turnover) X observed turnover or not	-0.235** (0.021)	-0.247	-0.246
observed turnover or not	-0.557*** (0.049)	-0.540	-0.540
log(shop space) X observed shop space or not	-0.060*** (0.024)	-0.049	-0.049
observed shop space or not	-1.242*** (0.031)	-1.235	-1.235
Business district (1 for a developed zone or major retail area, 0 otherwise)	0.056* (0.032)	0.069	0.069
Local businesses (1 if shop is in a underdeveloped commercial area, 0 otherwise)	-0.006 (0.049)	-0.006	-0.006
COVID-19 period (1 if after March 2020, 0 otherwise)	-0.087** (0.042)	-0.031	-0.031
Constant	-2.499*** (0.079)	-2.478	-2.478

Table 3.3: Comparison of the results of performing logistic regression on plaintext and ciphertext data. The dependent variable is the monthly survival records of 8,576 franchisees (Korean food restaurants) in the Kyung-Gi province, which were established between 2015 and 2020.

Comparing the results of the logistic regression with plaintext results verified the accuracy and efficiency of the analysis run on encrypted data.

OUTCOMES AND LESSONS LEARNED

Statistics Korea's pilot project provided accurate results of descriptive statistics and logistic regression performed on homomorphically encrypted data. The project has also demonstrated the potential of homomorphic encryption to facilitate data cooperation between government agencies who do not necessarily have a trusted relationship.

We expect that if data linkage is encouraged through the Statistical Data Hub Platform, it will enable the production of high-quality statistics and analyses using pension, childcare, and employment data, amongst others. This could help enable timely, informed, and effective responses to important national issues.

CASE STUDY 12: STATISTICS NETHERLANDS: DEVELOPING PRIVACY-PRESERVING CARDIOVASCULAR RISK PREDICTION MODELS FROM DISTRIBUTED CLINICAL AND SOCIOECONOMIC DATA

Purpose	To develop cardiovascular risk prediction models from sensitive healthcare data.
Datasets	Vertically partitioned datasets with primary care data, secondary care (hospital) data, and socioeconomic data.
PETs used	Secure Multi Party Computation, Homomorphic Encryption, Secret Sharing, Federated Learning
Application	Record linkage and development of machine learning models.
Details of computation	<ul style="list-style-type: none"> - Record matching with pseudonyms provided by trusted third party - Scalar product - Additive operations - Machine learning algorithms: Bayesian network parameter learning, backpropagation of deep neural networks, federated ensembles, Cox hazard models - Options for global feature selection (e.g. information gain based selection) - On demand computing of the predefined workflows.
Parties and trust relationship	Four honest but curious parties (National Statistical Office, Hospital and University / TTP), in addition we assume that two or more parties will not collude against a third party. One output party (University).
Implementation status	Concept
Resources	https://commit2data.nl/projecten/carrier

BACKGROUND

The CARRIER (Coronary ARtery disease: Risk estimations and Interventions for prevention and EaRly detection) project concerns secondary processing of medical, lifestyle and other personal data that relates to citizens, and which is held by a number of organizations, namely: MUMC+, Zuyderland, Maastricht University/RNFM together with ZorgTTP, and Statistics Netherlands (CBS).

Considering the reuse of already gathered data, the project is heavily dependent on the legal basis on which the data was collected and other regulatory regimes that impact the processing. The final data governance framework needs to adhere to national laws (e.g. Wet op het Centraal bureau voor de statistiek) and European laws such as the GDPR (2016/679), whilst also being ethically sound.

One of the main challenges of the project is the linkage of data sets owned by the different parties. When linking datasets, there is risk of re-identification of subjects. This requires CARRIER to adhere to the highest standards of data security and privacy preserving measures.

CASE STUDY DESCRIPTION

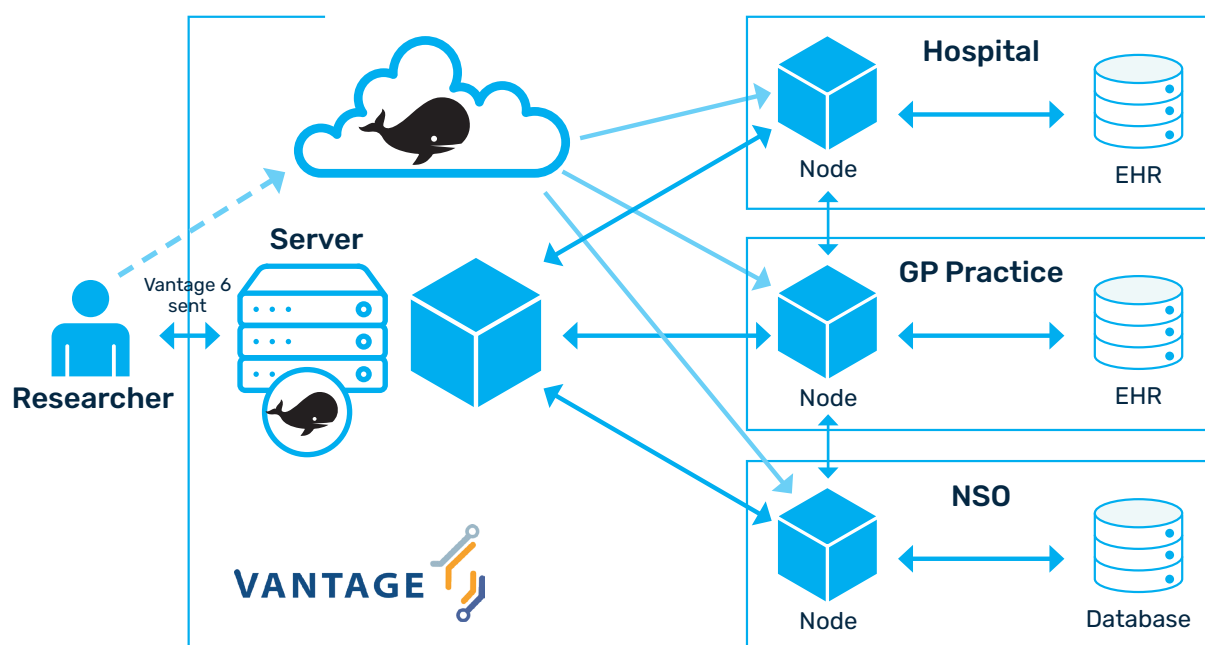


Figure 3.12: Architecture of federated learning infrastructure use

As shown in the figure, the input and compute parties each run predefined code supplied via approved Docker images. Only images that are approved by the local party may be executed on local data. This process is controlled via Vantage6, an open source infrastructure for federated learning. The various parties involved have the ability to review the Docker images independently. A central log is kept of the various transactions executed (e.g. which Docker image is executed as part of what workflow). The final output is inspected manually for potential privacy leaks before release outside of the cooperating organisations.

The main legal challenges are:

1. The research undertaken needs to adhere to current national and international legislations.
2. The research, as well as the later project phases, should not only be legally but also ethically justifiable/acceptable.
3. To enable the continuous research after the successful development of a prognostic model.

Other challenges include linking data of individuals scattered across different datasets and making different calculations (e.g. scalar product) from data distributed across different parties.

OUTCOMES AND LESSONS LEARNED

Legal agreements can help to bridge the gaps in the existing technology. In order to create a solid legal data governance framework suitable for the purposes of CARRIER, the following legal documents will at least have to be in place between the different parties involved:

- Consortium agreement
- Joint controller agreement
- Data Impact Privacy Assessment

Those agreements build the foundation for the federated learning procedure.

In the development phase of the project, the data governance framework is needed to support the information flows required for the federated learning.

In the second phase of the project, once the prognostic tool has been developed and is ready to be used, a data governance framework is needed which enables the tool to function accurately and effectively, whilst ensuring the privacy and data security of the patients using the tool is protected.

CASE STUDY 13: STATISTICS NETHERLANDS: MEASURING EFFECTIVENESS OF AN EHEALTH SOLUTION USING PRIVATE SET INTERSECTION

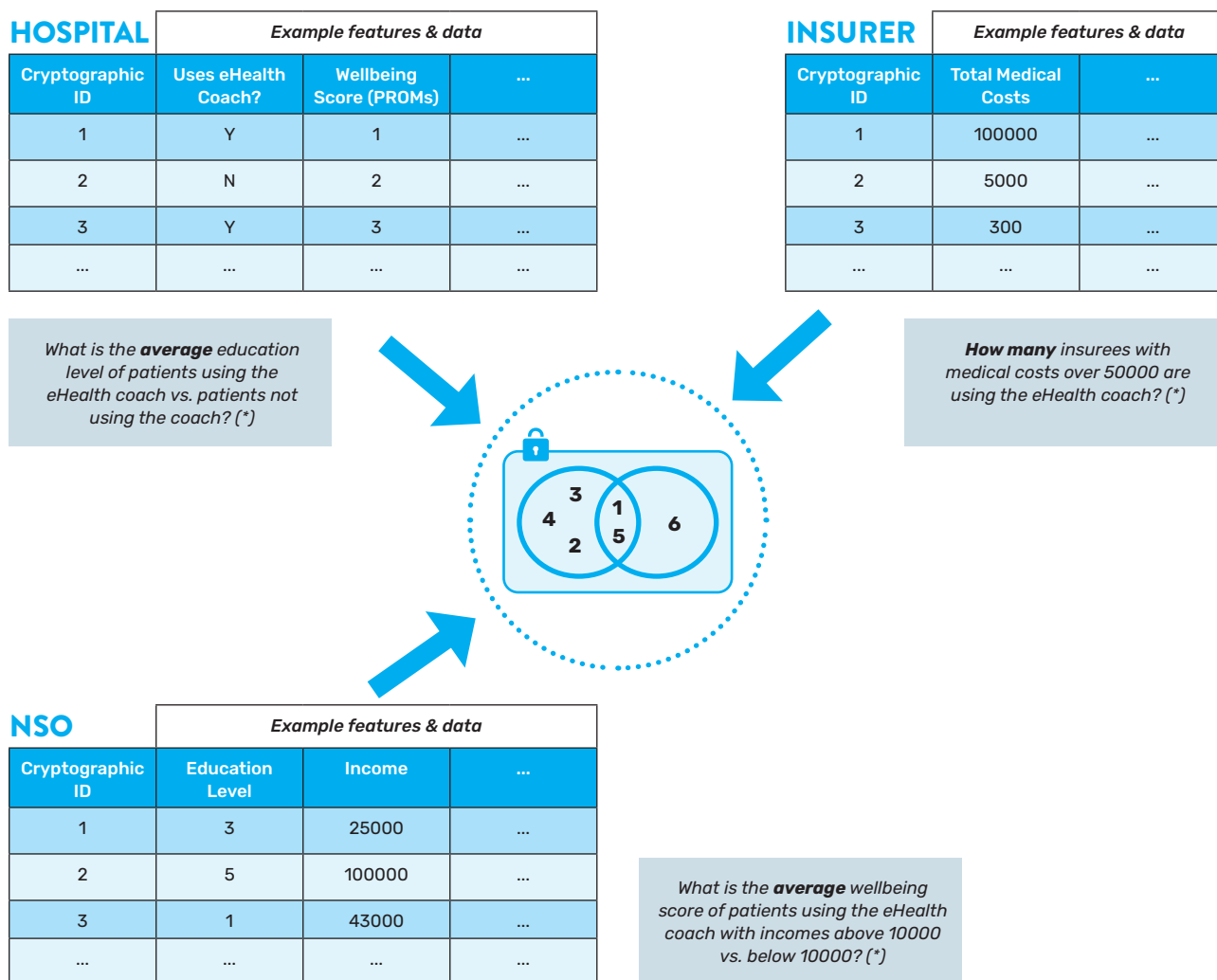
Purpose	To measure the effectiveness of a specific eHealth solution, without sharing patient information
Datasets	Medical data (Hospital), treatment cost data (Medical Insurance Company) and socio-economic data (NSO). Synthetic data was used in the initial Proof of Concept, and real data used for the Pilot phase.
PETs used	Homomorphic Encryption, Secret Sharing, Secure Multi Party Computation
Application	Private set intersection with analytics
Details of computation	Records are linked based on long term cryptographic IDs and local feature-based filtering, and aggregate statistics are derived, including sums, counts, averages, and standard deviations. Processing is on demand and query-based..
Parties and trust relationship	Three input parties who also act as compute parties in the sMPC network. A trustless model regarding input security is assumed. Aggregate output is cryptographically disclosed to the NSO first for SDC checks. Once cleared by the NSO, outputs can be disclosed to the other input parties.
Implementation status	Pilot
Resources	https://www.cbs.nl/nl-nl/corporate/2021/37/succesvolle-pilot-cbs-tno-cz-en-zuyderland-analyse-van-data-op-afstand https://www.cz.nl/over-cz/nieuws/praktijktest-succesvol https://www.tno.nl/nl/over-tno/nieuws/2021/9/eerste-gemeenschappelijke-analyse-zonder-onderlinge-datadeling/ https://gitlab.com/ppa-project https://medium.com/applied-mpc/secure-and-private-statistics-with-distributed-paillier-8a186410b5af https://www.linksight.nl/en/projects/ppa/

BACKGROUND

The project stems from the concept of ‘value-based care’ – delivery of the right care in the right place. A thorough understanding of the effectiveness of eHealth solutions will help to pinpoint the right care for the right patient. This means care that is tailored to the individual patient. However, the delivery of this care and determining the effectiveness of it requires data. This data is held by various different parties such as hospitals, general practitioners, health insurers and NSOs, but it may not be shared and mutually disclosed due to its highly sensitive nature. The usage of Privacy-Enhancing Technologies enables shared calculations while using cryptographic primitives to protect the data from mutual disclosure.

CASE STUDY DESCRIPTION

HIGH LEVEL FUNCTIONAL PERSPECTIVE



*For illustration purposes only, actual allowed queries are subject to implemented smart contract business rules

Figure 3.13: An example of Private Set Intersection with Analytics (PSI-A)

The PET techniques used are homomorphic encryption and secret sharing. The techniques to assure compliance are blockchain based smart contracts along with other techniques such as long term cryptographic IDs and TLS encryption. The approach to test and prove the effectiveness of the applied techniques and technologies was achieved through thorough tests with synthetic data (unit & feature testing) and record linking (match-%

using long term cryptographic IDs) before going live with real data. A governance board with privacy / data protection officers from each party was established to oversee compliance with the regulatory environment. All parties received the outcome (aggregated statistics) of a joint query-based analysis on combined data to create statistics and derive new insights for policymaking.



OUTCOMES AND LESSONS LEARNED

Many legal challenges related to the functional and technical features used to preserve medical and statistical confidentiality, and privacy in general had to be addressed.

The project took the the following legal measures as a joint effort:

- External independent legal review
- Review of specific applicable laws (Medical, Health Insurance, Statistical)
- Review of general applicable laws (GDPR, SSN)
- Development of a Data Protection Impact Assessment
- Creation of Data Processing Agreements
- Creation of an Agreement of Cooperation
- Review of relevant Intellectual Property / Patent / Export law.

The main functional challenges were:

- necessity of high-quality descriptive metadata
- need for synthetic test data
- special attention for common linking identifiers
- necessity to use Statistical Disclosure Control (SDC) to guarantee output privacy.

The overall organisational challenges when deploying a PET-based solution are the interdisciplinary involvement of multiple departments within each party such as methodology, statistics, legal, IT, communications etc.

Finally, transparent communication with the data owners (i.e. the patients) is very important to build trust in the project by explaining its purpose, clarifying its legal basis, and ensuring appropriate consent is given.

CASE STUDY 14: TWITTER AND OPENMINED: ADVANCING THIRD-PARTY AUDITS AND RESEARCH REPRODUCIBILITY OVER UNRELEASED DIGITAL ASSETS

Purpose	The primary purpose of this project is to evaluate the efficacy of PETs for algorithmic transparency. If successful, the goal is to enable researchers outside of Twitter to perform research on data and models within the firm using privacy-enhancing technologies (without having direct access to the underlying information being studied).
Datasets	The central datasets in the first project come from the paper Algorithmic amplification of politics on Twitter, in addition to synthetic reproductions of the private datasets therein for the purpose of development and testing. The largest synthetic dataset contains approximately 1 billion rows of data.
PETs used	Remote execution (sometimes called federated learning/analytics), differential privacy, and secure multi-party computation.
Details of computation	<p>A dataset is uploaded to a PySyft domain node and the data owner configures their domain node with a user account for a data scientist. A data scientist can then obtain a pointer to the uploaded dataset that lets them perform operations on the dataset as if it were a normal NumPy array though they are not able to see the results of their computations in the process.</p> <p>Once they have concluded their computation, they can see the results by using the adversarial differential privacy system, which adds statistical noise to their result. This process of adding noise also spends privacy budget (related), which is tracked at an individual data subject level.</p>
Parties and trust relationship	The current phase of the project is designed for an external trusted party to test the system by using PySyft to reproduce research results against a synthetic dataset. The next phase of the project is anticipated to replace this data with data from the paper for end-to-end testing.
Implementation status	Ongoing Proof of Concept
Resources	<p>Announcing our Partnership with Twitter to Advance Algorithmic Transparency Investing in privacy enhancing tech to advance transparency in ML Algorithmic Amplification of Politics on Twitter</p> <p>Christchurch Call Initiative on Algorithmic Outcomes https://www.christchurchcall.com/media-and-resources/news-and-updates/christchurch-call-initiative-on-algorithmic-outcomes/</p> <p>Christchurch Call Initiative on Algorithmic Outcomes https://www.amcham.co.nz/page-1334006/12928098</p>

BACKGROUND

Since 2016, Twitter users have been able to choose a preferred order for viewing their Home timeline from two options. The first option is to view Tweets from accounts the user has chosen to follow presented in reverse chronological order. The second option is to view Tweets that are algorithmically selected and ordered based on a personalization algorithm to prioritize content shown to each user based on the system design and how they interact with the algorithmic system, resulting in potential for older Tweets and those from accounts they do not follow to be prioritized in the Home timeline.

In October 2021, Twitter published learnings from an internal analysis of whether its recommendation algorithms amplify political content. The study analyzed millions of Tweets from elected officials in seven countries: Canada (House of Commons members), France (French National Assembly members), Germany (German Bundestag members), Japan (House of Representatives members), Spain (Congress of Deputies members), United Kingdom (House of Commons members), United States (official and personal accounts of House of Representatives and Senate members) from 1 April - 15 August 2020 and hundreds of millions of Tweets containing links to articles shared on Twitter during the same timeframe.

The study found that Tweets from elected officials are algorithmically amplified when compared to political content on the reverse chronological timeline. Algorithmic amplification was found to be an individualized effect (i.e., similar users received different results) and Tweets posted from accounts on the political right received more algorithmic amplification than those posted by accounts on the political left in all countries but Germany. News outlets were categorized based on media bias rating from two independent organizations. The study found that right-leaning news outlets received greater algorithmic amplification compared to left-leaning news outlets.

Twitter's ML Ethics, Transparency, and Accountability (META) team aims to discover whether the algorithmic amplification identified in the study results from preferential treatment in the algorithm's design rather than representing user interactions in order to reduce adverse impacts. In addition to sharing aggregate data with researchers in order to reproduce the study, META would like to provide researchers with access to the raw data from which the aggregates were calculated. But, privacy concerns have previously prevented sharing raw data, which limits reproducibility and the benefits of having many researchers examine these important issues from multiple perspectives and approaches.

CASE STUDY DESCRIPTION

On 20 January 2022, Twitter announced a partnership with OpenMined, an open-source nonprofit organization, to use PETs to replicate the findings of the political amplification study using synthetic data based on the original data.

Most differential privacy work occurs in a "trusted curator" setting, where the data scientist has access to the raw (and often sensitive) data, and determines for themselves how much noise is sufficient to add in order to protect privacy, before publishing. This assumes a lot of expertise and trust.

In contrast, OpenMined has been building its differential privacy system in an adversarial setting, where differential privacy mechanisms are designed to protect data from an adversary that is studying it. In practice, this means

that any output party is required to remain within a privacy budget (low trust). Additionally, the use of remote execution environments and Tensor pointers allows output parties to use the data without having intimate knowledge of sophisticated differential privacy mechanisms and how and when to apply them.

In addition, OpenMined's differential privacy system allows for privacy budgets to be stored at an individual data subject level. This ensures that an output party can be further limited as to how much information they can learn about any individual in a dataset. The ability to store and track privacy budgets at an individual level also allows for much tighter privacy loss [\[source\]](#) compared to traditional differential privacy, which adopts a pessimistic approach by considering only the worst case estimate over all data subjects and all possible values of their data, for every single analysis.

However, to the best of our knowledge, no large-scale demonstrations of this kind of differential privacy system (adversarial, and with a large number of individual privacy budgets) have been conducted. The partnership between OpenMined and Twitter would be the first to attempt to show an adversarial differential privacy system with millions of individual privacy budgets for different data subjects.

A future aim of the project is to enable researchers to conduct studies with actual data rather than being limited to the data currently available via the public Twitter API. Therefore, this project serves as a first step towards implementing PETs to enable researchers to conduct research using the same data that Twitter uses in their own internal analyses to improve accountability while preserving privacy.

Initiative on Algorithmic Outcomes, a partnership between New Zealand, the United States, Twitter, Microsoft and OpenMined to develop and test a differential privacy system to enable privacy preserving research across multiple online platforms. The pilot will serve as a “proof of function” regarding the use of PETs to facilitate open and transparent, multi-stakeholder research to enable better understanding of algorithmic outcomes, particularly the role of algorithms in content discovery and amplification, while preserving the privacy of individuals. The pilot will also demonstrate that the underlying techniques can be scaled to meet real-world legal, policy and other requirements.

OUTCOMES AND LESSONS LEARNED

It is possible to have a differential privacy system that tracks millions of individual privacy budgets.

Differential privacy in an adversarial setting is possible, and displays several benefits:

- The output party (a data scientist in this case) can work with a Tensor pointer using the exact same functions and methods as if they were using their statistical analysis framework of choice (NumPy, PyTorch, etc)
- The output party is not forced to know how differential privacy works to get the insights they want.
- Output parties are constrained by their own privacy budget as to how much they can learn about a given dataset, thereby limiting any adversarial party’s ability to do damage.

This case study demonstrated that it is possible to perform queries on a dataset on a PySyft domain node and get results without ever viewing the private or sensitive dataset being queried.

Consequently, privacy-enhancing methods from this study are being further developed through an international pilot initiative to facilitate research on real world data across multiple platforms.

In September, 2022, in conjunction with the UN General Assembly and Christchurch Call leaders summit, New Zealand Prime Minister Jacinda Ardern and French President Emmanuel Macron announced the Christchurch

CASE STUDY 15: UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE: TRIALLING APPROACHES TO PRIVACY-PRESERVING FEDERATED MACHINE LEARNING

Purpose	To privately train a neural network model on isolated lifestyle data collected by smart devices.
Datasets	Publicly available dataset on human activity recognition with smart devices' accelerometer and gyroscope data. The data was split into four subsets, one for each participating statistical office, for the purpose of the experiments.
PETs used	Federated Learning, in combination with additive Homomorphic Encryption, Differential Privacy
Application	Development of a machine learning model.
Details of computation	A neural network is trained via a federated learning approach across the four data partitions. A simulated environment was built to enable both privacy-preserving training and inference to be tested.
Parties and trust relationship	Multiple input parties (Public organisations, i.e. NSO). A not-fully-trusted central authority acting as an aggregator.
Implementation status	Proof of Concept
Resources	Private ML Track (unece.org)

BACKGROUND

As part of the High-Level Group for Modernization of Official Statistics (HLG-MOS), the project on Input Privacy Preservation Techniques (IPP) has the goal to investigate modern and collaborative approaches to methods and tools on privacy-enhancing technologies that offer protection of the input data. Private Machine Learning (ML) is one of the tracks in the IPP project with the main objective to investigate best practices and open source tools for distributed and collaborative ML among multiple organisations in a low trust environment.

CASE STUDY DESCRIPTION

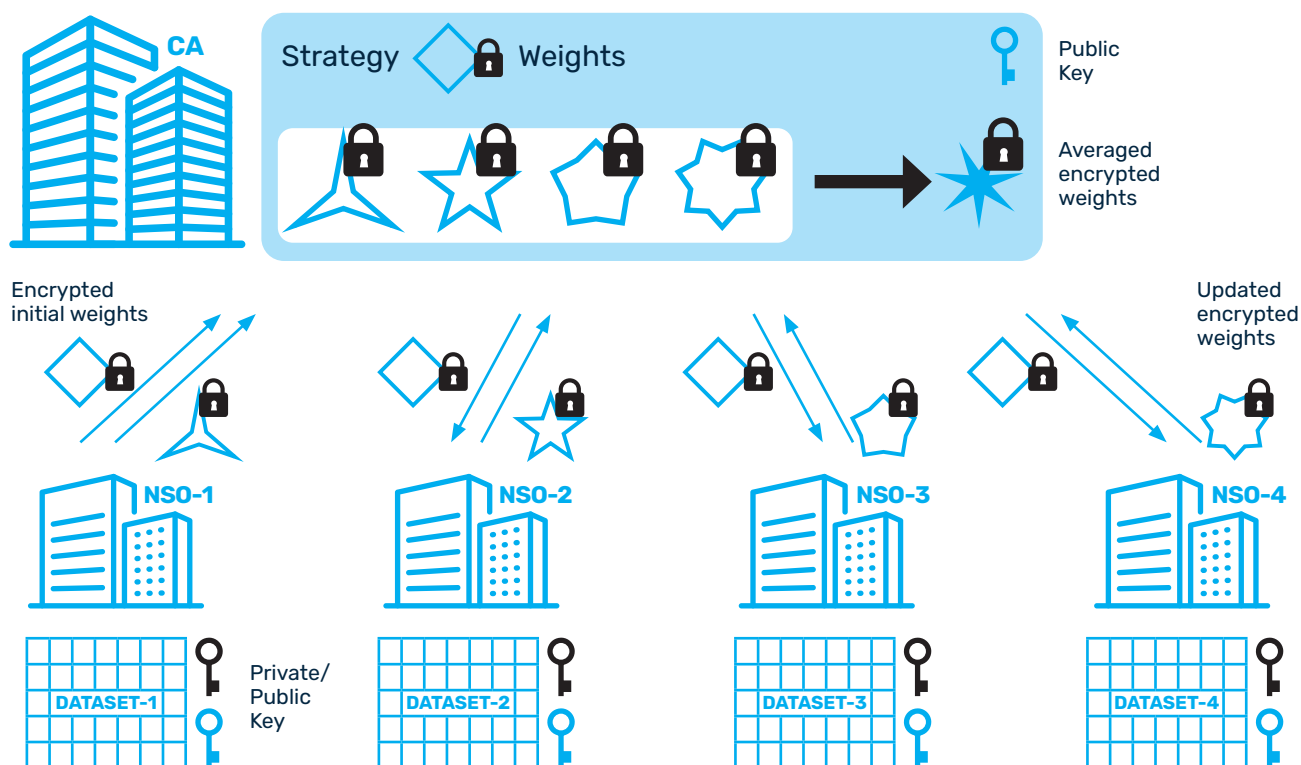


Figure 3.14: An overview of a distributed privacy preserving machine learning involving four National Statistical Offices (NSO) and a Central Authority (CA). A combination of Federated Learning and additive Homomorphic Encryption is used to train a neural network to classify human activities based on publicly available accelerometer data collected from smart devices.

In the first set of experiments, multiple National Statistical Offices (NSO) from Canada, Netherlands, Italy and the UK collaborated in building a simulated environment to validate the concept of multi-party privacy preserving machine learning (PPML) for both training and inference. A distributed and containerized PPML architecture was built utilising Federated Learning in combination with other privacy-enhancing technologies, such as additive Homomorphic Encryption, to train a neural network model on isolated lifestyle data collected by smart and wearable devices. A simulated environment was created by using open source tools and libraries to recognize and classify human activities into multiple categories based on publicly available accelerometer data collected from smart and wearable devices.

OUTCOMES AND LESSONS LEARNED

Preliminary results of our experiments in a simulated environment proves the feasibility of distributed and federated analytics among organisations while protecting the privacy of isolated data sources. We have built a community of Statistical Offices in the area of privacy-enhancing technologies with links to open source community, industry and academia. Moreover, there is a direct link to sustainability, when it comes to collaboration among NSOs, namely novel ways of collaboration, driven by privacy requirements and technological constraints. However, in real scenarios, prior agreements among participating agencies on a standard data format and preprocessing steps on a case-by-case basis seem to be necessary before deployment of distributed ML on sensitive data.

CASE STUDY 16: UNITED NATIONS PET LAB: INTERNATIONAL TRADE

Purpose	Enable multiple national statistical offices (NSOs) to perform reconciliation and joint analysis on independently collected trade datasets.
Datasets	The datasets involved were originally from the UN Comtrade Datasets and are now being extended to integrate third-party data sources.
PETs used	Differential Privacy, Secure Enclaves, Secure Multi Party Computation
Details of computation	Each NSO maintains an independent record of their international trade, such as imports, exports, re-imports, re-exports, and so on, at varying levels of granularity. The computations aimed to identify erroneous recordings of trade between pairs of countries and enable broader international trade analysis.
Parties and trust relationship	Multiple input parties (each NSO involved, including Statistics Canada, US Census, UK ONS, Statistics Netherlands, ISTAT Italy) with shared outputs. There is no assumption of trust between parties.
Implementation status	Proof of Concept (ongoing)
Resources	What is the UN PET Lab and Why is it Important? The Economists write-up on the PET Lab

BACKGROUND

International trade information is an important data source used to better understand the flow of commodities in and out of each country, measure the level of competitiveness of a country, and track economic growth. However, these figures have been typically tracked and maintained at a national level, and as such ambiguities and errors can appear when comparing recorded levels of trade on an international level.

The United Nations has aided in these challenges for a long time, in particular via the Comtrade portal, which publicly shares international trade statistics of each country on a monthly basis following the [Harmonised System of commodity categorization](#) (H2 through H6). Comtrade data provides each country's imports, exports, re-imports, and re-exports. As such, there is a unique pairing of data in opposite directions for each pairing of countries. For example, the United States will have recorded its exports of maize to Canada while Canada will have recorded their imports of maize from the United States. In theory, these numbers should match, although there are a number of reasons why this might not be the case. Through-trade is one such example - i.e. when a country receives goods as an intermediate stop rather than a final destination. However, even when through-trade is taken into account, disparities can still occur.

Ultimately, having a better understanding of global trade can be immensely beneficial in understanding global economic development, globalization, enabling the enforcement of trade restrictions, and reducing global money laundering to name but a few examples.

CASE STUDY DESCRIPTION

The goal of the project was to use privacy-enhancing technologies to share more granular information between countries and to enable the linkage of additional heterogeneous data sources. Comtrade data and the Harmonised System of classification undergo classical data disclosure controls prior to publication, thus minimizing the abilities of analysts to understand where and why disparities occur.

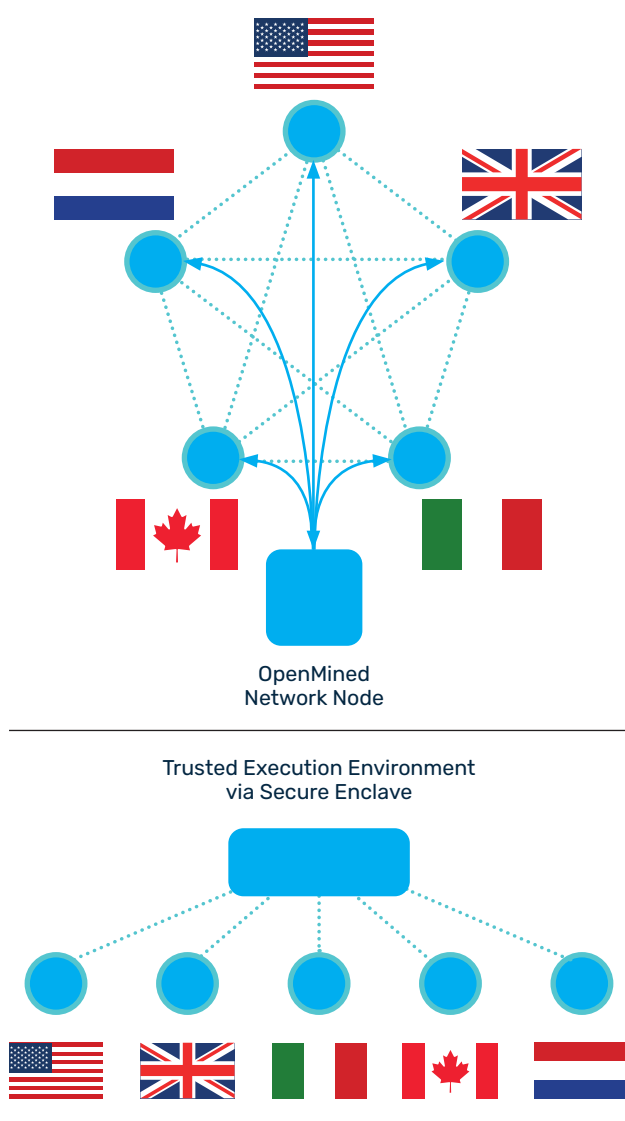


Figure 3.15: On the top, we visualize the peer-to-peer network which leverages sMPC and differential privacy. On the bottom, we see the secure enclave approach which centralizes the computation in a trusted execution environment.

The starting point of this was to initially work with safe, publicly available data from Comtrade, and as systems emerged to connect and analyze the data, more sensitive data may be included. This balanced the goals of proving the usefulness of PETs to securely link and perform analysis on trade data, whilst reducing project risk and minimizing the time to kick off the experimentation.


A secondary - but important - goal of the project was to understand the pros and cons of different privacy-enhancing technologies that can be used in such a setting. The experiments used two very different architectural approaches to address the problem.

The first approach used secure multi-party computation (sMPC) and differential privacy via a peer-to-peer federated data network, provided by OpenMined. In this setting, each party involved set up a node that housed their sensitive data and made requests to calculate the total value of the goods traded (imported/exported) across all the parties involved, without any party having to disclose the amount of any particular good imported or exported. The data queries did not require manual approval from a data compliance officer from any of the parties involved.

The second approach used enclave technology in combination with differential privacy. Each party was able to connect to the enclave via a secure proxy on their local device. The proxy makes an initial handshake with the enclave, authenticating the client and receiving the attestation document of the enclave which in turn guarantees the software running inside the enclave. Through this handshake, a symmetric key is also shared between the client and the enclave which enables bilateral secure communication thereafter.

While the enclave-based sMPC framework, provided by Oblivious, is reusable and highly generic, the software running inside the enclave was written especially for this case study. There were two core parts to the software. The first was the data science element which joined data from each party and applied various forms of aggregate queries. To ensure no low-level information was shared between parties, differential privacy was applied to outputs. Specifically, this was done in collaboration with the OpenDP (Harvard) and SmartNoise (Microsoft & Harvard) projects.

The second element of the software packaged the outputs into a formal PDF report for the purpose of upstream sharing. To ensure that the PDF would not be modified



throughout its life cycle, it was digitally self-signed from within the enclave and the public key of the signature was embedded into an attestation document from the enclave and used to watermark the document. This encapsulation of the attestation document and the self-signed public key in the PDF allows upstream users to confirm who uploaded the data originally, the software used to process the data and that the document was not modified since its creation.

The parties engaged in this collaboration spanned the NSOs from the United States, Canada, UK, Netherlands, and Italy, with infrastructure and assistance from the United Nations Global Platform.

OUTCOMES AND LESSONS LEARNED

A second, and important outcome in the context of this document, is that there can very often be multiple privacy-enhancing technologies that can solve a specific challenge. The OpenMined PySyft framework and Oblivious enclaves offered different pros and cons which may be more suitable in different contexts. OpenMined's PySyft enabled users to be very flexible in terms of access control management, with requests for queries being adhoc and approved just-in-time but after the initial infrastructure was created and approved. On the contrary, the Oblivious sMPC framework placed the access management controls prior to the deployment of the enclave itself.

Such differences directly stem from the range of functionalities the frameworks offer. OpenMined's PySyft is limited in nature to deal with specific types of queries and arithmetic combinations thereof. Thus when agreeing to use the federated network users are in turn agreeing to the access control mechanism and the differential privacy and sMPC mechanisms built within.

On the contrary, the enclave-based computation can run any software that would run on a typical server. For this reason, the enclave-based solution could do advanced functionality like generating PDF documents with corresponding signatures. However, due to this wide-ranging flexibility, it is the internal software that requires bilateral approval.

CASE STUDY 17: UNITED STATES CENSUS BUREAU: DEPLOYING A DIFFERENTIALLY PRIVATE DISCLOSURE AVOIDANCE SYSTEM FOR THE 2020 US CENSUS

Purpose	To protect against the disclosure of sensitive information collected by the census
Datasets	Data from the 2020 US census
PETs used	Differential Privacy
Application	Statistical disclosure
Details of computation	The basic algorithm computes a series of cross-category statistics (queries) on the raw data and then adds noise to each cell based on its computed disclosure risk and a privacy-loss budget.
Parties and trust relationship	A single input party (US Bureau of Census) and multiple output parties (public or private entities). Open access. There is no assumption of trust among various parties.
Implementation status	Production (2020 results were released in August 2021)
Resources	<p>[1] Abowd, J., Ashmead, R., Simson, G., Kifer, D., Leclerc, P., Machanavajjhala, A., & Sexton, W. (2019). Census topdown: Differentially private data, incremental schemas, and consistency with public knowledge. US Census Bureau.</p> <p>[2] https://github.com/usensusbureau/DAS_2020_Redistricting_Production_Code</p> <p>[3] https://www2.census.gov/library/publications/decennial/2020/2020-census-disclosure-avoidance-handbook.pdf</p>

BACKGROUND

Each decade, the US Census Bureau is required by law to conduct a “counting the whole number of persons in each State.” The results of this count determine the number of congressional seats within each state, are used by states to define district boundaries, and determine the allocation of billions of dollars of federal funding. In addition to this accounting, the US Bureau also conducts over 130 different surveys of the US population each year to measure socio-economic, public health and educational indicators.

Under the Census Act, all published statistics must protect the information and identity of citizens as defined by the Privacy Act of 1974. Historically, these requirements have resulted in limited release of Census data and possibly inadequately protected results.

Figure 5.16A shows the history of different disclosure control methods that the Census has used in the past. The advent of big data re-identification techniques has made many of the sensitive attributes counted by US Census, especially those collected on small geographic cells, inferable for subpopulations in the US. At the same time, users of census data have pushed for more accurate, more transparently created and more available statistics. Since 2010, the Census Bureau has sought to revamp its disclosure avoidance system to address both sets of concerns.



Figure 3.16: The history of disclosure control processes employed by the US Census

CASE STUDY DESCRIPTION

Starting in 2020, US Census has employed differential privacy techniques to improve both the privacy protections and the usability of census data. The agency has employed a two-phase procedure that applies differential privacy noise injection and a post-processing phase that enforces constraints on published results.

The census has implemented a version of differential privacy called TDA (TopDown Algorithm) [1]. The basic algorithm computes a series of cross-category statistics (queries) on the raw data and then adds noise to each cell based on its computed disclosure risk and a privacy-loss

budget. This is down from the largest units of geography (national) down to the smallest (census blocks) as shown in Figure 2. Allocation of budget across queries is handled by an iterative optimization process that reduces risk across potential queries. Census has made the code and process documentation for these algorithms publicly available [2][3].

In addition to adding noise, a series of post-processing steps are employed to ensure that certain invariants and constraints are enforced (e.g. total numbers add appropriately for published cells, certain key statistics are always noise free, non-negative population counts etc.).

DATA PROTECTION PROCESS

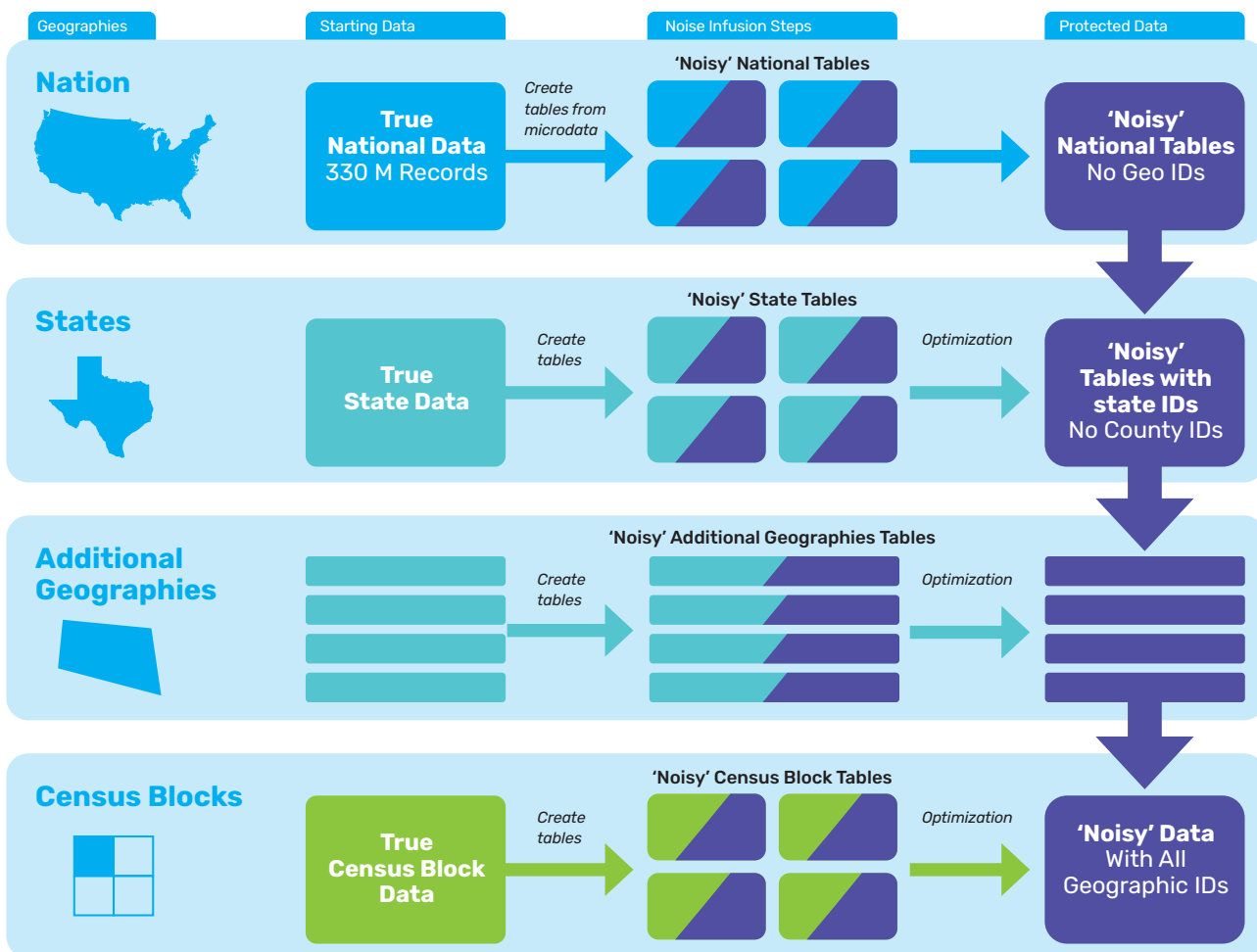


Figure 3.17: Census top-down noisification approach

OUTCOMES AND LESSONS LEARNED

The 2020 decennial census results were released with DAS controls as described above for the first time. Both state and federal agencies in the United States have since made use of these results to redraw congressional district lines and allocate funding.

While most of these activities have proceeded using differentially injected noise and without issue, some researchers and policymakers have expressed concerns about high levels of noise in small-cells and sparsely populated areas causing usability issues with the officially released numbers. A community of privacy academics has asked that the Census Bureau release noise injection meta-statistics to help users understand when and where their analyses are (or are not) meaningful.

In addition to usage issues, the current release of census numbers were legally challenged by the State of Alabama (in *Alabama v. Raimondo* (Dept. of Commerce)). In part, Alabama claimed that the inaccuracy of official statistics would result in an inability to comply with redistricting regulations such as the voting rights act. As of late June 2021, this case has been dismissed by a federal appellate court and the plaintiffs have indicated that they do not intend to appeal to the supreme court.

CASE STUDY 18: UNITED STATES DEPARTMENT OF EDUCATION: ANALYSING STUDENT FINANCIAL AID DATA USING PRIVACY-PRESERVING RECORD LINKAGE

Purpose	To compute statistics on average student loan and grant data across the US for 30 categories of undergraduate students
Datasets	Real student financial records for financial aid loans and grants
PETs used	Secure Multi Party Computation
Application	Private Set Intersection with Analytics
Details of computation	Record linkage between two parties, based on a common key, and computation of average loan and grant values for 30 student categories
Parties and trust relationship	Two parties - both are input, compute and output parties - who partially trust each other
Implementation status	Pilot
Resources	A Federal Government Privacy-Preserving Technology Demonstration

BACKGROUND

This case study applies Private Set Intersection with Computation (a form of secure multi-party computation) to real-world sensitive data for the US Department of Education. In our setting there are two parties, each a distinct organization within the Department of Education: the National Postsecondary Student Aid Study group (NPSAS), and the National Student Loan Data System (NSLDS).

Our case study reproduces a portion of the annual NPSAS survey of higher education financial aid - an annual statistical report on the average undergraduate financial aid in the US for the academic year. In particular, this case study focuses on statistics in the NPSAS report regarding undergraduate educational grants and loans by the US Government.

PRIVACY-PRESERVING STATISTICS FOR US DEPT. OF EDUCATION

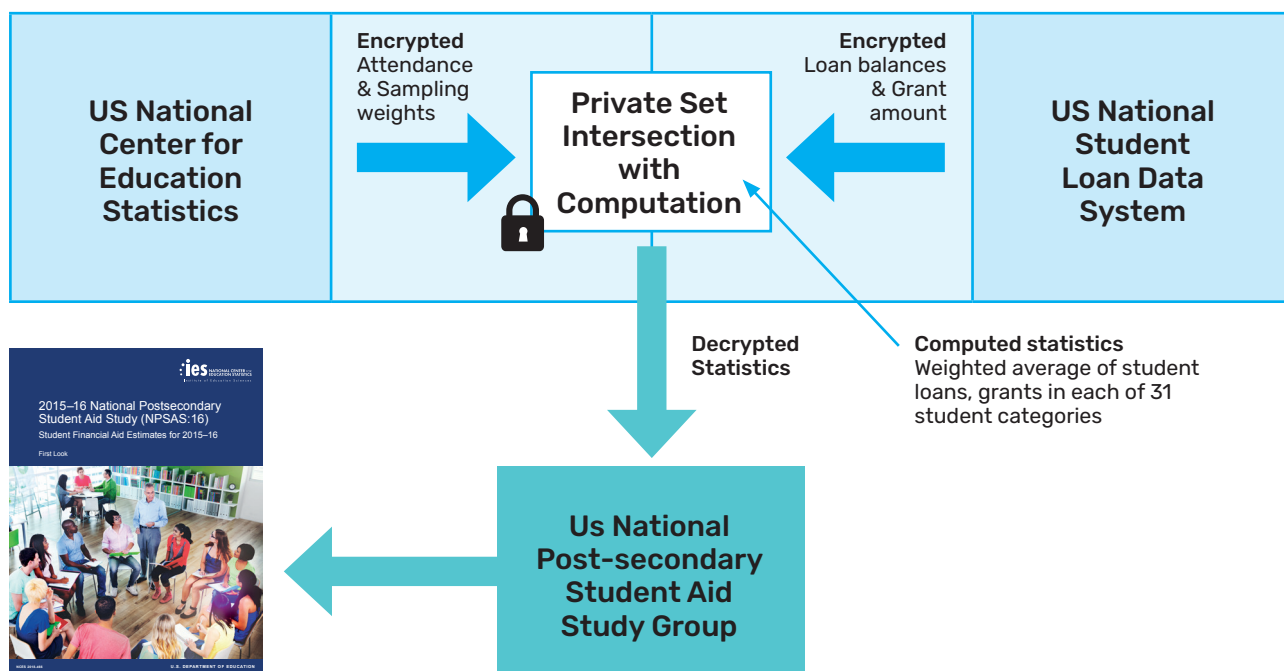


Figure 3.18: Privacy-Preserving statistics for US Department of Education

CASE STUDY DESCRIPTION

Today, NSLDS must share sensitive student financial information with NPSAS: the fact that a student receives grants or loans from the Government, detailed financial information about those grants and loans, and additional personal information. In this setting, NPSAS and NSLDS have a common key – the student Social Security number, which is also considered sensitive private information. As shown in the Figure above, to avoid disclosing any of this sensitive data while successfully and efficiently providing the same statistics, our prototype performs the same data linkage and statistical analysis without sharing that sensitive information between the agencies by using secure multi-party computation. This “zero-trust” approach relies on computing the necessary statistics *while the data remains encrypted*, and then decrypting only the results of the analysis.

OUTCOMES AND LESSONS LEARNED

Our experiments produced accurate results for average US Federal Pell Grants, Subsidized Federal Direct Loans, Unsubsidized Federal Direct Loans, and all Federal Direct Loans across institution type, attendance pattern, and income level that were comparable to the ground truth computations performed without privacy preservation. The experiments yielded total computation times and network traffic costs that were reasonable compared to those of the typical methods used to produce these statistics. The technique offers cryptographically proven security at levels comparable to those typically used for encrypting data today. This prototype demonstrated that sMPC technology can assure confidentiality of sensitive information while enabling practical, performant analysis of combined data held by diverse organizations.

CHAPTER 3. INTERNATIONAL COLLABORATION AND CASE STUDIES

BIBLIOGRAPHY

Bureau, US Census (2020). *DAS 2020 Redistricting Production Code Release*. url: https://github.com/uscensusbureau/DAS_2020_Redistricting_Production_Code. Accessed 2022-04-22.

4. STANDARDS

4.1 INTRODUCTION

This chapter provides an overview of standards-making activities relevant to the processing of data sets in general. In collating the content for this section, we have reviewed updates to the standards listed in the last edition and identified several new standards, including standards under development.

There has been a significant increase in standards-related activity relevant to privacy-enhancing technologies (PETs) and data in Artificial Intelligence (AI), and more specifically, Machine Learning (ML), since the publication of the UN Handbook on Privacy-Preserving Computation Techniques.¹ In the case of AI/ML, it can be observed that, in contrast to earlier approaches to standardization, which sought to draw together practice and experience collected over a period of time to benefit from hindsight, the driver now is foresight, with a view to prevention of perceived potential harms (known-knowns and known-unknowns).

Because of this expansion of activity dealing with PETs and, more broadly, the context in which they may be applied, the discussion has been split into two parts. The first identifies directly relevant essential standards, with sections on encryption and security techniques,

where there are current and emerging standards that address the PETs covered in Chapter 2. We recommend the Key Standards section below as a 'must read' for readers concerned specifically with PETs. The second part considers indirectly related standards that could affect the environment - technical and organizational - in which PETs may be deployed, with subtopics on cloud computing, big data, governance, artificial intelligence and data quality amongst others. The Related Standards section is provided as a 'should read' for readers who want to explore the 'bigger picture.' The inclusion criterion for Related Standards has intentionally been relaxed to facilitate broad rather than focused coverage. Consequently, not all the subtopics may have equal relevance for a given reader.

The textual description of a standard reproduces with permission some or all of the published description of the document appearing on the standards organization's public web pages. The definitive text may be viewed by following the link provided. Unless otherwise specified, the link access date is 2022-01-25

¹ Archer et al., *UN Handbook on PPTs* (2023).

4.2 KEY STANDARDS

The standards material in this section pertains directly to PETs, covering encryption and security techniques and then highlighting a couple of isolated standards, one on data de-identification terminology and the other on trusted execution environments.

ENCRYPTION

[ISO/IEC 18033-1](#) Information technology security techniques – Encryption algorithms – Part 1: **Encryption algorithms**. This introduces a set of standards (18033 Parts 1-5) covering asymmetric, block, stream and identity-based ciphers. Subsequently, Part 6 (Homomorphic encryption, see next) was added and published in 2019, but Part 1 has not been updated to reflect this.

[ISO/IEC 18033-6](#) Information technology security techniques – Encryption algorithms – Part 6: **Homomorphic encryption** is a standard on homomorphic encryption (HE) schemes. An HE scheme aims to allow operations directly on encrypted data, which is achieved by representing the plaintext as elements of a group rather than as conventional computer data. HE mechanisms are characterized by the operations they support, typically addition and multiplication in a given group. The standard describes two HE schemes (Exponential ElGamal and Paillier) and the processes for generating parameters and keys, encryption, decryption, and operating on encrypted data. As the title of this standard indicates, it is one part of a set of standards related to encryption techniques. Outside formal standardization activities, the **Homomorphic Encryption Standardization**^a consortium is open to industry, government, and academia participants. The initiative attempts to build broad community agreement on security levels, encryption parameters, encryption schemes, core library API, and eventually, the programming model to drive adoption of this technology.

SECURITY TECHNIQUES

There is a substantial portfolio of security and privacy techniques standards, covering both technology and management. ISO/IEC standards are increasingly multi-part documents. We summarize relevant families, highlighting parts that could directly affect PETs.

ISO/IEC 19592 SECRET SHARING

ISO/IEC 19592 consists of two parts in a series. [ISO/IEC 19592-1](#) focuses on the general secret sharing model and the related terminology. It introduces properties that secret sharing schemes could have, e.g. the crucial homomorphic property for several secure multi-party computation (sMPC) systems. [ISO/IEC 19592-2](#) considers specific schemes. It starts with the classic ones like Shamir and replicated secret sharing. All schemes are systematically described using the terms and properties from Part 1. There were original plans to have more parts for this standard that would describe sMPC paradigms, but there has been no published progress since the publication of the previously cited handbook.²

ISO/IEC 27000 INFORMATION SECURITY MANAGEMENT

ISO/IEC 27000 is part of the ISO Management System Standards (MSS)^b series with a specific focus on Information Security Management System (ISMS) to provide repeatable steps to help organizations improve their performance to achieve their goals and objectives with a continuous cycle of self-evaluation, correction and improvement of operations and processes.

ISO/IEC 27000 consists of 19 parts, with each part dedicated to a specific focus. ISO/IEC 2700[1,6,9] specify requirements for systems, for bodies providing audit and certification, and for sector-specific application of requirements. ISO/IEC 2700[2,3,4,5,7,8,] and ISO/IEC 270[13,14,16,21] describe general guidelines. ISO/IEC 270[10,11,17,18,19] and ISO/IEC 27799 describe sector-specific guidelines. The standards selected below seem to have the most relevance to PETs:

^a Homomorphic encryption: <http://HomomorphicEncryption.org>, Accessed 2018-07-02.

² op.cit. Archer et al., p.109

^b See <https://www.iso.org/management-system-standards.html> for more information about this series.

1. [ISO/IEC 27001](#) Information technology – Security techniques – Information security management systems – **Requirements:** useful for internal and external assessment of the organization’s ability to meet its own security requirements;
2. [ISO/IEC 27002](#) Information technology – Security techniques – Information security management systems – **Code of Practice:** guidance on developing industry- and organization-specific guidelines for controls, including policies, processes, procedures, organizational structures and software and hardware functions and how these controls need to be established, implemented, monitored, reviewed and improved, as appropriate;
3. [ISO/IEC TS 27006-2:2021](#) Information technology – Security techniques – **Requirements for bodies providing audit and certification of information security management systems – Part 2: Privacy information management systems** This document specifies requirements and provides guidance for bodies providing audit and certification of a privacy information management system (PIMS) according to ISO/IEC 27701 in combination with ISO/IEC 27001, in addition to the requirements contained within ISO/IEC 27006 and ISO/IEC 27701. It is primarily intended to support the accreditation of certification bodies providing PIMS certification;
4. [ISO/IEC 27010](#) Information technology – Security techniques – **Information security management for inter-sector and inter-organizational communications Information technology:** demonstrates the application of 27002 to the kind of domain identified;
5. [ISO/IEC 27017](#) Information technology – Security techniques – **Code of practice for information security controls based on ISO/IEC 27002 for cloud services:** provides guidelines for information security controls applicable to the provision and use of cloud services by providing: additional implementation guidance for relevant controls specified in ISO/IEC 27002; additional controls with implementation guidance that specifically relate to cloud services. This document provides guidance for both cloud service providers and cloud service customers; Initiatives vary

substantially in scale and impact. Objectives falling under the heading of “privacy” will depend on culture, societal expectations and jurisdiction. This document is intended to provide scalable guidance that can be applied to all initiatives. Since guidance specific to all circumstances cannot be prescriptive, the guidance in this document should be interpreted with respect to individual circumstances;

6. [ISO/IEC 27018](#) Information technology – Security techniques – **Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors.** Although aimed at providers, this is also relevant for procurement, because both parties need to meet the requirements of applicable legislation and regulations covering PII protection. The requirements and how they are divided between the cloud service provider and its customers vary according to legal jurisdiction, and the contract terms between the cloud service provider and the customer. A public cloud service provider processing PII is a “PII processor” the customer can range from a natural person (“PII principal”) processing their own PII in the cloud, to an organization (“PII controller”) processing PII relating to many PII principals. This discussion also clearly has legal implications and should be read in conjunction with Chapter 7 (Legal and Regulatory Issues), the document on legal issues arising from PETs³ and, where appropriate, with advice from qualified legal and regulatory expertise;
7. [ISO/IEC 27550](#) Information technology – Security techniques – **Privacy engineering for system life cycle processes:** provides privacy engineering guidelines to help organizations integrate recent advances in privacy engineering into system life cycle processes. It describes the relationship between privacy engineering and other engineering viewpoints (system and security engineering, risk management); privacy engineering activities in key engineering processes such as knowledge and risk management, requirement analysis, and architecture design. The audience includes all involved in developing, implementing or operating systems that need privacy consideration and managers in organizations responsible for privacy, development, product management, marketing, and operations. Content includes:

³ Varia, *Legal Issues arising from PETs* (2023).

- a) how privacy engineering supports system and security engineering, information risk management, knowledge management etc.;
 - b) conceptual principles such as privacy-by-design and privacy-by-default, important design goals noted in GDPR and elsewhere;
 - c) processes for identifying, evaluating and treating privacy risks in the course of IT systems design;
 - d) how IT systems can be engineered to support and satisfy the OECD privacy principles, which form the basis of most privacy laws and regulations;
- 8. ISO/IEC 27551** Information security, cybersecurity and privacy protection – **Requirements for attribute-based unlinkable entity authentication.** This document addresses collection limitations. Attribute-based unlinkable entity authentication (ABUEA) allows PII principals to establish the authenticity of a selected subset of their identity attributes without revealing a larger subset. Particular focus is put on unlinkability and a metric that measures the strength of this property in implementations of ABUEA is introduced. This document focuses on cases where a third party attests to at least one attribute. This document also identifies security properties to be met to achieve various protections and unlinkable properties;
- 9. ISO/IEC TS 27570:2021** Privacy protection – **Privacy guidelines for smart cities** The document provides guidance on: smart city ecosystem privacy protection; how standards can be used at a global level and at an organizational level for the benefit of citizens; and processes for smart city ecosystem privacy protection;
- 10. ISO/IEC 27701:2019** Security techniques – **Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management – Requirements and guidelines** This document specifies requirements and provides guidance for establishing, implementing, maintaining and continually improving a Privacy Information Management System (PIMS) in the form of an extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy management within the context of the organization. This document specifies PIMS-related requirements and provides guidance for PII controllers and PII processors holding responsibility and accountability for PII processing;

ISO/IEC 29100 SECURITY TECHNIQUES

ISO/IEC 29100 is another multipart standard providing a high-level framework for protecting personally identifiable information (PII) in computer systems. As with the earlier discussion of PII in cloud computing, this is related to legal and regulatory issues. The standards are not a substitute for the advice of qualified legal and regulatory experts.

- 1. ISO/IEC 29100** Information technology – Security techniques – **Privacy framework.** This document aims to help organizations define privacy safeguarding requirements in respect of PII by: setting out terminologies for privacy, identifying actors and their roles in processing PII, outlining privacy safeguarding requirements and connecting with established privacy principles;
- 2. ISO/IEC 29101** Information technology – Security techniques – **Privacy architecture framework.** This document supersedes the 2013 publication cited in the publication of the previously cited handbook.⁴ This is one of the oldest standards efforts that handles secure computing. It presents architectural views for information systems that process personal data and show how Privacy-Enhancing Technologies such as secure computing, secret sharing, and also anonymisation, pseudonymisation, query restrictions and more may be deployed to protect PII. It also considers appropriate actions and considerations during the PII processing life cycle, which identifies the phases: collection, transfer, use, storage and disposal;
- 3. ISO/IEC 29134:2017** Information technology – Security techniques – **Guidelines for privacy impact assessment.** A privacy impact assessment (PIA) is an instrument for assessing the potential impacts on privacy of a process, information system, programme, software module, device or other initiatives that process PII and, in consultation with stakeholders, for taking actions as necessary to treat privacy risk. A PIA report may include documentation about measures taken for risk treatment, for example, measures arising from the use of the ISMS in ISO/IEC 27001. A PIA is a process that begins at the earliest possible stages of an initiative when there are still opportunities to influence its outcome and thereby ensure privacy by design. It continues until, and even after, the project has been deployed;

⁴ op.cit. Archer et al., p.109

4. [ISO/IEC 29184:2020](#) Information technology – **Online privacy notices and consent.** The broader availability of communication infrastructures and improved information processing capability have enabled much wider-ranging collection and analysis of personal information. While these technological improvements may offer consumer and business benefits, consumers are becoming increasingly “privacy aware” and questioning the privacy impact of the collection and use of PII. This can be due to an inadequate explanation of how PII is processed, stored, maintained and managed. This document specifies controls and associated additional information for organizations: (a) to provide the basis for presenting clear, easily understood information about PII collection and processing and (b) to obtain consent from PII principals in a fair, demonstrable, transparent, unambiguous and revocable (withdrawable) manner. This document details the implementation of two privacy principles from ISO/IEC 29100 (i.e., Principle 1: Consent and choice, Principle 7: Openness, transparency and notice);
5. [ISO/IEC 29190:2015](#) Information technology – Security techniques – **Privacy capability assessment model.** This document provides organizations with high-level guidance about how to assess the level of their ability (capability) to manage privacy-related processes. It sets out an approach for assessing the efficiency and effectiveness of privacy-related processes used by organizations. This includes (a) decision-support information for formulating and executing a privacy strategy and for operations and line-of-business staff, (b) consideration of the range of “privacy stakeholders” who might have very different requirements, driven by legal and regulatory compliance requirements, and (c) by inter-related “good practice” provisions from a variety of internal and external sources. The document provides guidance on how to set up a capability assessment program within an organization, with an iterative and incremental process of improvement, with the aim of self-assessment against a capability assessment model; metrics against key performance indicators; outputs from privacy process management audits and management practices for input into improving capability;

OTHER KEY STANDARDS

1. [ISO/IEC 20889](#) **Privacy-enhancing data de-identification terminology and classification of techniques.** This standard is closely related to the 29100 family in that it addresses how to use de-identification techniques on PII while maintaining compliance with the principles put forward in 29100. The standard notes that it is oriented toward tabular data and the techniques may not be applicable to data as free-form text, image, audio or video. The aim here is to mitigate in these specific areas re-identification, but no guarantees are offered. If stronger guarantees are required, alternative approaches to the processing as outlined in Chapter 4 (Methodologies and Approaches) may be more appropriate. For longevity, the standard classifies techniques rather than the detail of implementation, but there is much useful information in the informative annexes;
2. [ISO/IEC 24760-1:2019](#) IT Security and Privacy – **A framework for identity management – Part 1: Terminology and concepts.** This document defines terms for identity management, and specifies core concepts of identity and identity management and their relationships. It is applicable to any information system that processes identity information;
3. [IEEE P2830](#): **Standard for Technical Framework and Requirements of Trusted Execution Environment based Shared Machine Learning.** TEEs are discussed in detail in Chapter 4 (Methodologies and Approaches). At a generic level this standard specifies functional components, workflows, security requirements, technical requirements, and protocols. This is supported by and draws on various use cases with characteristics for which TEE is applicable. The sources to aggregate are subject to various constraints on their combination. The purpose of TEEs is to support shared ML, where (encrypted) data are shared with a trusted third party for the learning computation. The purpose of the standard is to provide a verifiable basis for trust and security.

4.3 RELATED STANDARDS

There are numerous other standard activities that, while not directly connected with PETs, do affect the environment in which PETs may be deployed or have something to say about the concerns that PETs aims to address. This section summarizes relevant standards in Cloud Computing, Big Data, Governance, Artificial Intelligence and Data Quality.

CLOUD COMPUTING

1. [ISO/IEC 17789](#) Information technology – Cloud computing – **Reference architecture**. This standard was reconfirmed in 2021. The relevance for PETs is that clause 8.5.9 refers to the protection of PII. Everything that has been written under the 27000 series about PII is applicable to cloud computing. In addition, cloud provisioning will almost certainly involve the transfer of data, which counts as “processing” under the General Data Protection Regulation (GDPR) and other similar legislation. The (informative) appendix to the standard (A.4.2) discusses the idea of a privacy impact audit.
2. [ISO/IEC 19944-1](#) Cloud computing and distributed platforms Data flow, data categories and data use – **Part 1: Fundamentals**. This document, along with [ISO/IEC 19944-2](#) Cloud computing and distributed platforms – Data flow, data categories and data use – **Part 2: Guidance on application and extensibility** replaces ISO/IEC 19944:2017. 19944-1 extends the vocabulary and reference architecture in ISO/IEC 17788 and ISO/IEC 17789 to describe an ecosystem involving devices using cloud services; it describes the various types of data flowing within the devices and cloud computing ecosystem, the impact of connected devices on the data that flow within the cloud computing ecosystem, flows of data between cloud services, cloud service customers and cloud service users, provides foundational concepts, including a data taxonomy, and identifies the categories of data that flow across the cloud service customer devices and cloud services. Notable changes compared to ISO/IEC 19944:2017 include how to handle organizational data, the introduction of the concept of data facets and new data use categories, particularly the use of data associated with artificial intelligence systems.

BIG DATA

ISO/IEC provides well-defined big data standards. ISO/IEC 20546 focuses on big data overview and vocabulary, followed by ISO/IEC 20547 five-part standard on a big data reference architecture. Of these, ISO/IEC 20547-4 is the most relevant to security and privacy:

1. [ISO/IEC 20546](#) Information technology – Big data – **Overview and vocabulary**: This document provides an overview of big data’s key concepts, along with a set of terms and definitions. It gives a terminological foundation for big data-related standards;
2. [ISO/IEC TR 20547-1](#) Information technology – Big data reference architecture – **Part 1: Framework and application process**: This document provides a framework to describe a big data architecture and implementation, a process for mapping a specific problem set/use case to the architecture and evaluating that mapping;
3. [ISO/IEC TR 20547-2](#) Information technology – Big data reference architecture – **Part 2: Use cases and derived requirements**: This document provides a collection of big data use cases and decomposes those use cases into technical considerations that big data architects and system implementers can consider;
4. [ISO/IEC 20547-3](#) Information technology – Big data reference architecture – **Part 3: Reference architecture**: This document describes the reference architecture in terms of User and Functional views;
5. [ISO/IEC 20547-4](#) Information technology – Big data reference architecture – **Part 4: Security and privacy**: This document describes the security and privacy aspects unique to big data;
6. [ISO/IEC TR 20547-5](#) Information technology – Big data reference architecture – **Part 5: Standards roadmap**: This document provides a list of standards and their relationship to the reference architecture that architects and implementers can consider as part of the design and implementation of their system.

There are also two related British Standards:

1. **BS 10102-1** Big data Part 1: **Guidance on data-driven organizations**: This document gives guidance on realizing value from data, including big data, such as gaining insights, informing strategies, enhancing reputation, and improving compliance, efficiency and performance;
2. **BS 10102-2** Big data Part 2: **Guidance on data-intensive projects**: This document provides guidance on good practice for implementing data-intensive projects to realize value, including: defining project objectives and project type; project roles and responsibilities; data project management methodology; defining the approach to governance and compliance (see BS 10102-1, Clause 6); operating governance and compliance within a framework; working with partners, suppliers, technology providers, consumers and other third parties; and project closure – review against project objectives, communication and lessons learned.

GOVERNANCE

Standards related to governance, both for IT in general and security and privacy in particular, are fairly mature. There are no standards yet that directly address the governance issues raised by the use of AI (but see the [Artificial Intelligence](#) section and [Section 4.4 Standards under Development](#)). The primary group of standards is the ISO/IEC 38500 series:

1. **ISO/IEC 38500** Information technology – **Governance of IT for the organization**: This document describes principles and provides definitions and a model for governing bodies to evaluate, direct, and monitor the use of information technology (IT) in their organizations. It is a high-level, principles-based advisory standard, whose aim is to provide broad guidance on the role of a governing body and encouragement to use appropriate standards to underpin the governance of IT.
2. **ISO/IEC TS 38501** Information technology – Governance of IT – **Implementation guide**: This document provides more specific methodological governance guidance in IT, with the goals of assuring that the risks associated with IT are appropriately managed and ensuring the maximization of IT investment value.

3. **ISO/IEC TR 38502** Information technology – Governance of IT – **Framework and model**: provides support in clarifying and distinguishing between governance and management in respect of IT, and a model that illustrates the relationship between governance and management, to identify the responsibilities associated with each.
4. **ISO/IEC 38505-1** Information technology – Governance of IT – Governance of data – Part 1: **Application of ISO/IEC 38500 to the governance of data** and **ISO/IEC TR 38505-2** Information technology – Governance of IT – Governance of data – Part 2: **Implications of ISO/IEC 38505-1 for data management**: jointly provide (Part 1) principles, definitions and a model for governing bodies to use when evaluating, directing and monitoring the handling and use of data in their organizations and (Part 2) what the governing body of an organization expects and requires from the data management team to be assured that the governing principles of IT can be implemented and are being upheld for data and its use by the organization. Together they show how the strategy can inform data policy, processes and controls and how to design the controls and processes to monitor the implementation of the strategy so that the governing body can be assured of the performance of and conformance to the strategy.
5. Governance also receives attention in the ISO/IEC 27000 series with **ISO/IEC 27014** Information security, cybersecurity and privacy protection – **Governance of information security**. This provides guidance for either a governing body, or top management, on concepts, objectives and processes for the governance of information security.

ARTIFICIAL INTELLIGENCE

There is much standardization activity around artificial intelligence taking place at ISO/IEC (Joint Technical Committee/Sub Committee 42), CEN/CENELEC (Joint Technical Committee 21) and the IEEE Standards Association, but currently there is relatively little that has completed the standardization cycle, beyond that focusing on Big Data (five publications, see [Big Data section](#)) and three of the several technical reports discussed below. There are five working groups under ISO/IEC JTC 1/SC 42, which currently address Foundational standards, Data, Trustworthiness, Uses cases and applications, and Computational approaches and computation characteristics of AI systems, respectively.

Although [Section 4.4](#) discusses standards in development, several of the AI-related standards are listed here for coherence because some of those published refer to those not yet published, which latter are denoted “under development”:

1. [ISO/IEC 22989](#) Information technology – Artificial intelligence – **Artificial intelligence concepts and terminology**: defines standardized concepts and terminology to (i) help in the description and understanding of artificial intelligence technology; (ii) permit the comparison of different technologies in regard to properties such as trustworthiness, robustness, resilience, reliability, accuracy, safety, security, and privacy; (iii) support stakeholders in identifying appropriate solutions and analyzing market offerings;
2. [ISO/IEC 23053](#) Information technology – Artificial intelligence – **Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)**: proposes a framework for describing AI system components and their functions in an accepted standard way. Hence it aims to be a basis for future standards addressing implementation and use of AI systems, ways to improve transparency, and allocation of responsibility across AI system components;
3. [ISO/IEC TR 24372](#) Information technology – Artificial intelligence – **Overview of computational approaches for AI systems**: provides a summary of the computational methods and approaches within AI systems. It draws on ISO/IEC 22989 (concepts and terminology), ISO/IEC 23053 (framework for systems using ML) and ISO/IEC TR 24030 (AI use cases). This document describes the characteristics of an AI system and its computational approaches. The illustration of computational approaches in AI systems includes machine learning and non-machine learning methods, puts forward a general taxonomy of computational approaches including knowledge-driven and data-driven, and discusses selected algorithms, basic theories and techniques, main characteristics and typical applications;
4. [ISO/IEC TR 24028](#) Information technology – Artificial intelligence – **Overview of trustworthiness in artificial intelligence**: provides an analysis of the factors that can impact the trustworthiness of systems providing or using AI. The document discusses (i) some existing approaches to help trustworthiness in technical systems and discusses their potential application to AI systems; (ii) how to mitigate AI system

vulnerabilities relating to trustworthiness; (iii) how to improve the trustworthiness of AI systems;

5. [ISO/IEC TR 24029-1](#) Information technology – Artificial Intelligence – **Assessment of the robustness of neural networks – Part 1: Overview**: views robustness as a crucial property that poses new challenges in the context of AI systems – specifically neural networks – where some risks are specifically tied this property, and understanding these risks can be key to the decision to adopt AI. This document provides an overview of some approaches for assessing these risks;
6. [ISO/IEC TR 24030](#) Information technology – Artificial intelligence – **Use cases**: takes 132 use cases, collected through an open call and then analyzed, to identify AI applications, deployment models and application domains. The drivers for this work were to (i) provide input to and reference for AI standardization work; (ii) share the collected use cases to foster collaboration; (iii) reach out to new stakeholders interested in AI applicability; (iv) support the translation of science and technology through the identification of general-purpose requirements.

DATA QUALITY

The ISO 8000 series sets out frameworks for data quality for different kinds of data and can be used in conjunction with or separately from the ISO/IEC 9000 series. The series of 18 published parts encompasses data governance, data quality management, data quality assessment, quality of master data and quality of industrial information. From these, the following are most relevant for NSOs:

1. [ISO 8000-1](#) Data quality – Part 1: **Overview**;
2. [ISO 8000-2](#) Data quality – Part 2: **Vocabulary**. As the title indicates, its content sets out the terms and definitions used in the rest of the series, addressing topics including quality, measurement, syntax and semantics and data governance;
3. [ISO 8000-61](#) Data quality – Part 61: **Data quality management: Process reference model** specifies the processes required for data quality management. The processes are used as a reference to enhance data quality and assess process capability or organizational maturity for data quality management.

See also the summary of [ISO/IEC 5259-x, Artificial intelligence – Data quality for analytics and machine learning](#) in the Standards under Development section.

4.4 STANDARDS UNDER DEVELOPMENT

The majority of standards making activity is taking place in the context of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), specifically through ISO/IEC JTC 1/SC 42 Artificial intelligence, where JTC 1 is the Joint Technical Committee for Information technology and SC 42 is the sub-committee for Artificial intelligence. SC 42 currently has five working groups: Foundational standards (WG1), Data (WG2), Trustworthiness (WG3), Uses cases and applications (WG4) and Computational approaches and computational characteristics of AI systems (WG5). SC 42 relevant liaisons for the context of this document are with Information security, cybersecurity and privacy protection (SC 27) and Cloud computing and distributed platforms (SC 38). Additionally, there is a Joint Working Group between SC 42 and SC 40 (IT service management and IT governance) on Governance implications of AI.

At the European level, CEN-CENELEC coordinates standards adoption for its member countries, which comprise the member bodies of the 27 European Union countries, the United Kingdom, the Republic of Northern Macedonia, Serbia and Turkey, plus the three countries of the European Free Trade Association (Iceland, Norway and Switzerland). All CEN and CENELEC members commit to adopt identical European Standards and withdraw national conflicting standards. Many such standards are straightforward adoptions of ISO standards, but there is some separate activity, coordinated by the recently formed Joint Technical Committee on Artificial Intelligence, where there is deemed to be a distinctive European issue that is not being addressed at ISO. Its approved remit states: "CEN-CLC/JTC 21 will proceed with the identification and adoption of international standards already available or under development from other organizations like ISO/IEC JTC 1 and its subcommittees, such as SC 42 Artificial Intelligence. Furthermore, CEN-CLC/JTC 21 will focus on producing standardization deliverables that address European market and societal needs, as well as underpinning EU legislation, policies, principles, and values."

The IEEE provides the other major standards activity arena, much of which stems from the IEEE's Ethically-Aligned

Design initiative,⁵ spawning the P7000 series of standards and working groups and a variety of other more niche activities. These working groups are coordinated by the (recently formed) Artificial Intelligence sub-committee of the IEEE Standards Association's Standards Board.

ISO/IEC STANDARDS UNDER DEVELOPMENT

At ISO/IEC, there are relevant standards at various stages of development. It normally takes at least three years to pass through all the processes to publish an ISO international standard. In the following paragraphs, we offer a non-exhaustive list of potentially relevant ISO/IEC standards under development. Where information is published on-line at the time of writing, a URL is provided.

- 1. ISO/IEC 5259-x Artificial intelligence – Data quality for analytics and machine learning – Parts 1-5.** ISO/IEC JTC 1/SC 42/WG2 is developing this collection of standards, covering Overview, terminology and examples (5259-1), Data quality measures (5259-2), Data quality management requirements and guidelines (5259-3), Data quality process framework (5259-4) and Data quality governance framework (5259-5). The aim of the series is to provide tools and methods to assess and improve the quality of data used for analytics and Machine Learning, drawing on the ISO 8000 series (Data quality and Enterprise Master Data) with ISO/IEC 25012 (Data quality model) and ISO/IEC 25024 (Measurement of data quality) from the ISO/IEC 25000 series on software and data quality.
- 2. ISO/IEC 8183 Information technology – Artificial intelligence – Data life cycle framework.** This document defines the key stages in application neutral and AI technology-neutral terms of the data life cycle related to using data in AI systems. As such, it constitutes a bridge between the data life cycle for conventional software and the specialized life cycles for big data (ISO/IEC 20547), machine learning, data quality for ML (ISO/IEC 5259-x), etc. to facilitate alignment and comparability across a range of standards.

⁵ IEEE, *Ethically Aligned Design* (2019).

3. [ISO/IEC 27555](#) Information security, cybersecurity and privacy protection – **Guidelines on personally identifiable information deletion.** Many functional processes and IT applications use personally identifiable information (PII) which is subject to various compliance provisions relating to privacy. Thus, organizations need to ensure that PII is not retained for longer than is necessary, and is deleted at the appropriate time. This also may require organizations to fulfill PII principals’ rights such as right to obtain the erasure (to be forgotten). ISO/IEC 29100 defines principles of “data minimization” and “use, retention and disclosure limitation” for personally identifiable information (PII) which can be enforced using deletion as a security control.
4. [ISO/IEC 27556](#) Information technology – **User centric framework for the handling of personally identifiable information (PII) based on privacy preferences.** This standard will lay out a “user-centric framework” to handle personal information in a controlled manner in accordance with the privacy-by-design and other requirements of applicable privacy laws and regulations. It will outline a mechanism for organizations handling personal data to comply with the data subject’s privacy requirements, while sharing and collaborating on processing the data. The architecture will be generic, to avoid specifying the content and format of privacy preference information, but sufficient to inform the design and implementation of IT systems handling personal information and communicating it between organizations, while managing the privacy preferences of data subjects. The standard builds on ISO/IEC 29100 “Privacy framework”.
5. [ISO/IEC 27557](#) Information Technology – **Organizational privacy risk management.** This standard will guide organizations in managing privacy risks that could impact the organization and/or individuals (data subjects) as an integral part of the organization’s overall risk management. The standard will distinguish information risks (with the potential to harm the organization directly) from privacy risks (with the potential to harm individuals directly and the organization indirectly), emphasizing the difference in the respective risk management activities.
6. [ISO/IEC 27559](#) **Privacy-enhancing data de-identification framework.** This standard will provide a non-prescriptive framework for identifying and mitigating privacy-related risks such as re-identification *etc.* during the life cycle of de-identified data. Organizations can use the standard to properly de-identify (anonymise) data, build trust with data subjects and meet compliance requirements.
7. [ISO/IEC 24745](#) Information security, cybersecurity and privacy protection – **Biometric information protection.** Provides requirements and guidelines for the secure and privacy-compliant management and processing of biometric information. This will shortly replace the 2011 edition of the standard.
8. [ISO/IEC PWI^c 6102](#) **Guidance on illustrative processes for a privacy information management system.** Determine if SC 27 needs a standard for “Guidance on processes of a privacy information management system” as part of the ISO /IEC 27000-family. Consider the following: ISO/IEC 27001 and ISO/IEC 27003, ISO/IEC 27701 (a.k.a. DIS 27552), ISO Handbook “The integrated use of management system standards”, ISO/IEC 33004, 2nd WD of ISO/IEC 27022, ISO/IEC PWI 6089 Impact of AI on security and privacy.
9. [ISO/IEC WD 27565](#) **Guidelines on privacy preservation based on zero knowledge proofs.** This document provides guidelines on using zero knowledge proofs (ZKP) to improve privacy by reducing the risks associated with the sharing or transmission of personal data between organisations and users by minimizing the information shared. It will include several ZKP functional requirements relevant to a range of different business use cases, then describes how different ZKP models can be used to meet those functional requirements securely.
10. [TR ISO/IEC 27563](#) **Impact of security and privacy in artificial intelligence use cases.** This document provides information on how to assess the impact of security and privacy in AI use cases, covering in particular those published in [ISO/IEC TR 24030 \(Information technology – Artificial Intelligence \(AI\) – use cases\)](#)

^c Preliminary Work Item

We list several other ongoing activities without further detail, in some cases because projects have only recently been initiated (e.g. PWIs), for the sake of raising awareness that such matters are being addressed:

1. [ISO/IEC WD 27006-2](#) – Information security, cybersecurity and privacy protection – Requirements for bodies providing audit and certification of information security management systems – Part 2: Privacy information management systems;
2. [ISO/IEC 27553-1](#) Information technology – Security techniques – Security and Privacy requirements for authentication using biometrics on mobile devices - Part 1: Local Modes;
3. [ISO/IEC 27553-2](#) Information technology – Security techniques – Security and Privacy requirements for authentication using biometrics on mobile devices – Part 2: Remote modes;
4. [ISO/IEC WD TS 27561](#) Information technology - Security Techniques - Privacy operationalisation model and method for engineering;
5. [ISO/IEC WD 27562](#) Information technology – Security techniques – Privacy guidelines for fintech services
6. [ISO/IEC 29100:2011/Amd 1:2018](#) Information technology – Security techniques – Privacy framework – Amendment 1: Clarifications.

IEEE STANDARDS UNDER DEVELOPMENT

To initiate the establishment of an IEEE Standards Association working group, a proposal (a project authorization request, or PAR) to set the group up is submitted to the relevant committee of the IEEE standards organization, such as the AI sub-committee. If the PAR is approved, along with the nominations for chair and vice chair, the project typically has three years in which to deliver a draft. The outcome of a PAR can be a standard (e.g. P7001) or a guide (P2894) or a recommended practice (P2842). Membership in a working group is open to any individual. Please note that although the listed activities are “in development” at the time of writing, some of those listed below may have become draft standards or approved standards when this is read.

The IEEE’s P7000 series is working on a family of standards addressing ethically-aligned design, from which the following are most relevant to PETs:

1. [P7000-2021](#) - IEEE Standard Model Process for Addressing Ethical Concerns during System Design establishes a process model by which engineers and technologists can address ethical considerations throughout the various stages of system initiation, analysis and design. Expected process requirements include management and engineering view of new IT product development, computer ethics and IT system design, value-sensitive design, and stakeholder involvement in ethical IT system design;
2. [P7001](#) - Transparency of Autonomous Systems This document describes measurable, testable levels of transparency, so that autonomous systems can be objectively assessed and levels of compliance determined;
3. [P7002](#) - Personal Data Privacy Process This document specifies how to manage privacy issues for systems or software that collect personal data. It will do so by defining requirements that cover corporate data collection policies and quality assurance. It also includes a use case and data model for organizations developing applications involving personal information;
4. [P7003](#) - Algorithmic Bias Considerations This document provides developers of algorithms for autonomous or intelligent systems with protocols to avoid negative bias in their code. Bias could include the use of subjective or incorrect interpretations of data like mistaking correlation with causation. The project offers specific steps to take for eliminating issues of negative bias in the creation of algorithms;
5. [P7012](#) - Standards Project for Machine Readable Personal Privacy Terms This document aims to provide individuals with means to proffer their own terms respecting personal privacy, in ways that can be read, acknowledged, and agreed to by machines operated by others in the networked world. In a more formal sense, the purpose of the standard is to enable individuals to operate as first parties in agreements with others—mostly companies—operating as second parties.

Additionally, but not exhaustively, the following IEEE working groups are currently active:

^d Working Draft

1. **P2830 - Standard for Technical Framework and Requirements of Shared Machine Learning** This document defines a framework and architecture for machine learning in which a model is trained using encrypted data that has been aggregated from multiple sources and is processed by a third-party trusted execution environment. A distinctive feature of this technique is the essential use of a third party trusted execution environment for computations. The standard specifies functional components, workflows, security requirements, technical requirements, and protocols. In “shared machine learning” the data are shared but are encrypted and given to a trusted third party to train a model that is then shared. This standard will provide a verifiable basis for trust and security. A draft standard was published in May 2021;
2. **P2841 Framework and Process for Deep Learning Evaluation** This document defines best practices for developing and implementing deep learning algorithms and defines a framework and criteria for evaluating algorithm reliability and quality of the resulting software systems;
3. **P2842 Recommended Practice for Secure Multi-party Computation** This document provides a technical framework for Secure Multi-Party Computation, including specifying: An overview of Secure Multi-Party Computation, A technical framework of Secure Multi-Party Computation, Security levels of Secure Multi-Party Computation, Use cases based on Secure Multi-Party Computation.
4. **P2863 Recommended Practice for Organizational Governance of Artificial Intelligence** This document specifies governance criteria such as safety, transparency, accountability, responsibility and minimizing bias, and process steps for effective implementation, performance auditing, training and compliance in the development or use of artificial intelligence within organizations;
5. **P2986 Recommended Practice for Privacy and Security for Federated Machine Learning** This

document provides a recommended practice of privacy and security safeguarding for Federated Machine Learning, including security and privacy principles, defense mechanism against non-malicious failures and adversarial attacks towards a Federated Machine Learning system. This document also defines an assessment framework with the extent that defense mechanisms can achieve under various settings;

6. **P3652.1 Guide for Architectural Framework and Application of Federated Machine Learning** This document defines a machine learning framework that allows a collective model to be constructed from data that is distributed across data owners. This guide provides a blueprint for data usage and model building across organizations while meeting applicable privacy, security and regulatory requirements. It defines the architectural framework and application guidelines for federated machine learning, including: 1) description and definition of federated learning, 2) the types of federated learning and the application scenarios to which each type applies, 3) performance evaluation of federated learning, and 4) associated regulatory requirements.

NATIONAL STANDARDS IN DEVELOPMENT

BS EN 17529. Data protection and privacy by design and by default. This document provides the component and subsystem developers with an early formalized process for identification of privacy objects and requirements, as well as the necessary guidance on associated assessment. It further provides support for understanding the cascaded liability and obligation of manufacturers and service providers (Reference to GDPR and as applicable reference to Article 23, as well as to rules applicable to governmental applications). This document is intended for the use by manufacturers, suppliers, hard- and software developers, and system integrators providing products and services for the use by as data controller, and for the use by controllers when selecting products and services for data processing.

4.5 SUMMARY

In conclusion, we observe that, as noted at the outset of this chapter, there is much practice to draw upon in extant standards, but the notable volume of preemptive activity taking place in current standards development,

where the focus is more on the finer technical details, has the potential for greater relevance and impact in the coming few years.

CHAPTER 4. STANDARDS

BIBLIOGRAPHY

Archer, David W., Borja de Balle Pigem, Dan Bogdanov, Mark Craddock, Adria Gascon, Ronald Jansen, Matjaž Jug, Kim Laine, Robert McLellan, Olga Ohrimenko, Mariana Raykova, Andrew Trask and Simon Wardley (2023). *UN Handbook on Privacy-Preserving Computation Techniques*. arXiv publication: 2023, originally published 2019. doi: [10.48550/ARXIV.2301.06167](https://doi.org/10.48550/ARXIV.2301.06167).

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (2019). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition*. url: <https://standards.ieee.org/content/dam/ieeestandards/standards/web/documents/other/ead1e.pdf>.

Varia, Mayank (2023). *Legal Issues arising from Privacy Enhancing Technologies*. In preparation.

5. LEGAL AND REGULATORY ISSUES

5.1 INTRODUCTION

There is increasing awareness of PETs across governmental, commercial and private organizations. The security and privacy properties they offer clearly connect with the values that are increasingly being embedded in legislative and regulatory frameworks; however, because PETs are new and do not map cleanly onto existing laws and regulations, it can be problematic to determine whether they are acceptable to use in any specific scenario, and indeed this very issue imposes a substantial barrier to the adoption of PETs today. This chapter offers a short introduction to some of the issues, underlining the importance of timely incorporation of legal advice in a project, followed by reiterating the main issues addressed in the Legal Task Team¹ document on the legal issues raised by the use of PETs.

There are five key messages to take away from this section:

1. Involving legal experts early in any project is strongly advised;
2. Specific PETs are typically not mandated by legislation (see Articles 13 and 14 of the proposal for a regulation on European statistics on population and housing²), but PETs do enable compliance with legal requirements such as regarding “data minimization” or “data protection by design and by default”, and specific PETs might be recommended or required by a particular regulator for certain use cases;
3. The use of PETs must be consistent with existing laws, policies, and ideally cultural norms, and PETs can open up new opportunities and affordances within this social structure;
4. Any activity involving data from more than one jurisdiction will be more complicated;
5. Different laws and jurisdictions may take different views on the adequacy of a PET for a given use case, so we encourage regulators to publish guidance about the use of PETs.

In terms of law and regulation, it is critical that any statistics-related project must involve appropriate legal and regulatory experts, either internal or external, when the project starts, so that they can provide input on an ongoing basis from the outset, as part of both the governance and the risk management activities as well as project requirements and design. In any statistics-related project, this expertise is needed to provide input on the combination of privacy and statistics, especially in the case of NSOs, since each NSO has to comply not only with its regional and national privacy and data protection laws, but also any regional and national legislation in relation to statistics. Trying to “DIY” law and regulation, then seeking expert input only towards the end of a project or programme, when it may be too late to change the project’s key features or implementation, could be a very costly or even impossible exercise, leading to (possibly avoidable) legal risks for the project, the organization conducting or participating in it and the organizations using its outputs.

For projects that involve data processing organizations or personal data (or both) from different countries, the legal and regulatory implications that must be taken into account also include the cross-border aspects, and possible differences between the legal and regulatory regimes of different countries or even different states within the same country (e.g. Germany or the USA).

¹ Varia, Legal Issues arising from PETs (2023).

² European Union, European statistics on population and housing (Proposal) (2023).

5.2 THE LEGAL AND REGULATORY OUTLOOK

It is critical to consult qualified legal experts when considering or commencing specific projects or types of processing operations and even when making changes to them, as in practice the legal position depends very much on individual fact patterns of the specific project design or deployment scenario or both. Expert legal input is needed to determine:

1. Whether and to what extent the organizations running or participating in the project or using its outputs are required to comply with any or all of privacy, data protection and statistics legislation, and if so which laws in which countries/states, bearing in mind some laws only apply to certain types of organizations and/or types of activities, for example:
 - a. The US California Consumer Privacy Act (CCPA)³ affects “businesses” and “service providers”, and might not apply to non-profit/public sector bodies etc, whereas the EU General Data Protection Regulation (GDPR)⁴ imposes data protection obligations on “controllers” and “processors”, including non-profit or public bodies;
 - b. Under GDPR, certain processing of personal data for historical research or statistical purposes etc. may be exempt or subject to less stringent requirements, the details of which depend on the laws of the relevant EEA country or countries;
 - c. The Dutch Statistics Act⁵ prohibits publication of data used for statistical purposes on the level of an individual person, household or organization/institution, unless, in the case of data relating to a company or institution, there are good reasons to assume that the company or institution concerned will not have any objections to the publication.
2. What actions are required for compliance with those legal/regulatory requirements or at least to mitigate/reduce the legal risks for the project/organizations concerned, and whether any steps could be taken (e.g. “anonymisation” of personal data) so as to take a project out of scope of those requirements altogether, or at least so that more relaxed requirements will apply to it (e.g. pseudonymisation of personal data), and if so what and how;
3. Whether, what and how PETs can be used to assist with such compliance or reduce the compliance requirements or legal risks, and what alternative PETs if any could be used towards that end.

Note that other laws (e.g. on cybersecurity or intellectual property) may also be relevant to a project, but are not discussed in this document.

³ California State Legislature, California Consumer Privacy Act of 2018 [\[2018\]](#).

⁴ European Union, General Data Protection Regulation [\[2016\]](#).

⁵ Government of the Netherlands, Wet op het Centraal bureau voor de statistiek [\[2003\]](#).

5.3 CHALLENGES AND RISKS WHEN USING PETS

From the perspective of law/regulation, specific PETS could be required for compliance with laws/regulations but this is unlikely, as many laws are not so rigid or precise as to require or even identify particular named PETS or standards, e.g. a particular type of encryption. A PET that has been “approved” or favorably commented upon by a relevant regulator for one use case (e.g. encryption with restricted key access and also sMPC in the context of international transfers) may not be good enough for another use case (even if similar but involving e.g. different types of data or from different sources), or may need to be implemented differently for the other use case in order to ensure compliance in that scenario.

PETs involve a range of different techniques and as such support multiple use-cases. Legislators, public bodies and agencies have started looking into and issuing recommendations with implications on the use of PETs for particular use-cases. However it is important to remember that each use case is unique and one has to be very careful about the requirements and circumstances to which such recommendation might apply.

Notably, the European Data Protection Board issued “Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data”⁶ in the context of the transfer of personal data outside of the EU to third countries and article 46 of the GDPR. An identified use-case of effective measure there includes multi-party processing, that is processing of data by multiple parties located in different jurisdictions, which can be leveraging secure multi-party computation protocols. However, the document describes very concrete requirements that need to be satisfied for the measure to be effective. Some of these requirements impose specific technical constraints on the PET used, such as security against active adversaries and non-technical measures. Other requirements must be satisfied through social or legal mechanisms, e.g. no evidence of collusion between

different parties and no evidence of collaboration between public authorities in those processors’ jurisdictions, that would allow access to all sets of personal data held by the processors and reconstitution and exploitation of the personal data. Other identified measures include encryption to protect against access to the data by the public authorities in a third country. Such encryption methods need to conform to the state-of-the-art, follow guidelines on the key lengths, choice of protocols, such as those issued by ENISA, NIST and offer security against active and passive security, amongst others.

There has been an increased interest from public bodies and authorities around secure and private data sharing and data processing. The Information Commissioner’s Office in the UK has been carrying out a consultation and publishing guidance on privacy-enhancing technologies and anonymisation tools.⁷ Further important implications for privacy technologies for data sharing in the EU can be expected to come from the new European Strategy for data⁸ and the Digital Service Package including Data Governance Act,⁹ the Regulation on a European approach for Artificial Intelligence,¹⁰ the Digital Services Act and the Digital Markets Act¹¹ which are all currently being discussed and proceed.¹² A new European Data Innovation Board will be an important advisory structure created as part of it.

The above aims to provide an overview of some of the potential issues and risks that may arise at the interface between the law and PETs. The remaining sections go into more detail to look at the possibilities that PETs create (Affordances of PETs), the (legal) parties that may be affected (Actors), regulatory environment aspects (Regulatory environment), some regulatory factors to consider regarding the deployment of PETs (Consider the PET deployed), and lastly where PETs may fit in the data life cycle (Application across the data life cycle). These sections share content with the more detailed supplemental document on PETs and the law.

⁶ European Data Protection Board, Recommendations 01/2020 (2020).

⁷ Information Commissioner’s Office, Anonymisation, pseudonymisation and PETs (2021).

⁸ European Commission, A European Strategy for Data (2022).

⁹ European Union, Data Governance Act (2020).

¹⁰ European Union, Artificial Intelligence Act (2021).

¹¹ European Union, Digital Markets Act (2022).

¹² European Data Protection Board, Digital Services and Data Strategy (2021).

5.4 OPPORTUNITIES AND AFFORDANCES OF PETS

Deployed properly, PETs can be a vital component in helping to ensure that a project complies with relevant laws, regulations, policies, and cultural expectations in relation to privacy. For example even encryption, which may be classed as a PET aimed at preserving the confidentiality of data, is required by the GDPR only where its use is “appropriate” to the risks to the relevant individuals in the particular circumstances of the data processing concerned, taking certain factors into consideration (although the use of encryption at rest and in transit will very often indeed be appropriate, with stronger encryption, e.g. longer keys, having to be applied in the case of more sensitive/risky data). In general, by protecting privacy by default, PETs can contribute toward the legality of data processing and reduce risks of data loss or theft.

More than just a compliance mechanism, PETs also offer substantive new opportunities to add value in scenarios where data processing is desirable but restricted by statutory, regulatory, ethical, contractual, organizational, or competitive limits on data use, processing, and disclosure. The data privacy and security, including anti-disclosure protections, offered by PETs can empower people to contribute their personal information toward data analyses and empower information technology organizations to participate in the analysis by reducing risks to an acceptable level. The result is greater representation and data products that better reflect all of society, that in turn can improve social trust and belief in the resulting data or other products

To determine whether the use of a PET supports adherence to the law in a particular scenario, one must verify that all participants have satisfied all of their legal obligations, although full compliance may not always be possible and a risk-based approach must be adopted. The supplemental Legal Task Team document recommends a four-step process for doing so: (i) enumerate all actors involved in any aspect of data processing or technology development, (ii) determine under the scope of which laws each actor falls, and the types of affirmative requirements or limiting restrictions that these laws impose, (iii) analyze the extent to which the deployed PET is consistent with these requirements, and (iv) ensure that these questions are considered and revisited throughout the data life cycle.

ACTORS

In the context of PETs, from a legal perspective there are 5 main types of actors, particularly (but not only) under privacy/data protection laws. The discussion below uses the EU’s General Data Protection Regulation (GDPR) by way of illustration; similar (though not identical!) categorizations exist in other privacy laws. These actors are:

- 1. Lawmakers:** This class includes legislators and judges who enact/make or interpret the meaning of laws that affect PETs and their use..
- 2. Regulators:** National regulators will differ with the law and jurisdiction concerned. Examples in the case of the GDPR include EEA Member States’ national supervisory authorities (SAs) who are charged with monitoring compliance and enforcement; and the European Data Protection Board (EDPB), comprising national SAs collectively and the EDPS (covered below), which has been allocated certain functions and powers under the GDPR (e.g. to promote the consistent interpretation and application of the GDPR across the EEA)..
- 3. The protected class:** This is the category of legal entities intended to benefit from or be protected by the relevant laws, such as the individuals (“*data subjects*” under GDPR, “*consumers*” under the CCPA, i.e. California residents) whose “*personal data*” or “*personal information*” are protected, and who have certain legal rights regarding their personal data/information..
- 4. Those required to comply with certain legal obligations:** This group covers any actors who have obligations under the relevant laws and typically are subject to certain legal liability for non-compliance with those obligations.
- 5. Producers and/or providers of privacy-enhancing technologies or tools:** Suppliers of tools enabling users to employ PETs are not necessarily subject to obligations under the GDPR but, depending on the situation, they can be. For example, service providers who offer PETs as a service to their customers (e.g. via a cloud-based service) could be “processors” under GDPR.

REGULATORY ENVIRONMENT

There are a variety of regulations that influence subjects' rights about processing of their data, including laws that are explicitly about data protection as well as regulations about data security, minimization, fairness, accuracy, accountability, and more. An understanding of a regulatory environment may be achieved through consideration of the following aspects:

1. Determine whether an action falls within the *scope of applicability* of a particular regulatory regime. Some jurisdictions have omnibus general-purpose regulations, whereas other regulations only target a subset of data held, sector, type of organization or activities performed;
2. Consider how data processing impacts the actor's use of data. Some regulations may limit disclosure of derived data products to other actors, others may restrict the use of all data of a specific protected type, and some regulations might apply to data processing more generally;
3. Consider the *style of regulation*: some regulations require the actor to take an affirmative step like obtaining informed consent, others prohibit certain uses or disclosures of data, and yet others might impose a fiduciary duty of care to consider the best interests of other actors throughout the system life cycle;
4. Consider how PETs affect *additional subject rights* beyond restraints on processing, use, or disclosure. For instance, long-term access to certain data may either be necessary to comply with obligations toward accuracy and accountability, or prohibited to safeguard the data subject's privacy.

CONSIDERATIONS IN THE DEPLOYMENTS OF PETs

At this point, readers may be asking the question: "is the use of a PET compliant with the law?" Actually, this is usually *not* the right question to ask. Few laws offer a binary yes/no answer to this question due to a combination of (i) unawareness of the existence of PETs and (ii) a deliberate

technology-neutral design of the law, in order to support a spectrum of possible protection mechanisms and for the law to remain relevant and flexible as technology evolves.

Instead, a better approach is to examine, for each actor, how the PET influences the four regulatory environment criteria above. This involves determining whether a deployment of PETs is within the scope of a regulation, whether data processing implicates one or more regulations, whether all obligations on data processing are satisfied, and whether the deployed system is consistent with providing all other required data subject rights.

LEGAL ISSUES ACROSS THE DATA LIFE CYCLE

Data is core to the work of NSOs, and the purpose of this section is to set out the stages in a prototypical system life cycle, that captures the use of data for analytics, machine learning and more broadly for artificial intelligence, to provide a framework on which to hang cross-cutting legal issues. The list below aims to identify all the stages that might arise in the system life cycle in order to support discussion of when "processing", in the sense that the term is used in the GDPR, might occur. The considerations set out in detail in the Legal Task Team report on PETs are informed by the GDPR's notion of "processing", not to make this GDPR-specific, but because GDPR currently sets the highest legislative bar, and hence by adopting its definition of what constitutes processing, should cover the circumstances identified in other legislations. For example, the GDPR requires "Data protection by design and by default", for example see¹³ across the data life cycle although the emphasis is on upfront anticipation of issues. For the purposes of this discussion, we align legal and regulatory considerations with the following life cycle stages, which we map (see Figure 5.4.1) for comparison purposes on to the stages identified in the ISO AI system life cycle¹⁴ and the Generic Statistical Business Process Model (GSBPM)¹⁵:

1. **Idea conception and requirements establishment:** while no processing takes place in these stages, it is strongly recommended to involve appropriate legal experts as early as possible to aid in spotting potential

¹³ Information Commissioner's Office, Data protection by design and default (2022).

¹⁴ International Standards Organization, ISO/IEC 22989 AI concepts and terminology (2022).

¹⁵ United Nations Economic Commission for Europe, GSBPM v5.1 (2021).

issues sufficiently far ahead, particularly those which might enforce reconsideration of organization goals. The non-functional aspects of the requirements should record legal advice and legal forethought on the functional requirements, that is *what* the system is going to do;

- 2. Design and development:** this phase incorporates data planning and acquisition in the scoping of what primary data needs to be collected and what secondary data needs to be obtained. These both have associated legal considerations, such as what obligations different jurisdictions may put on the security and privacy of data collected for a particular purpose, forethought on system input and output privacy and any restrictions on acquisition or use of data acquired from third parties. Acquisition of third party data will probably involve processing in the legal sense through the transfer of partial or complete data sets into an organization's facilities, or arranging for remote access to data sets or setting up the network infrastructure for the capture of streaming data. As such due legal consideration is necessary.

Any preparation of data is also processing and again therefore needs due legal consideration;

- 3. Building a model:** data is used to construct a model and then to test the model; the data may be artificial, synthetic or actual; synthetic data [\[see Chapter 2, section 4\]](#) is normally derived from actual data and hence its creation constitutes processing of personal data; synthetic data may also create additional issues, such as the introduction of bias, or that the removal of outliers may render the data set useless, but retention may facilitate deidentification; the model may be constructed centrally or through a distributed process [\[see Chapter 2, section 5\]](#); in the latter case, it is the model that moves to the data, but this still constitutes processing;
- 4. Deployment:** the model from the previous stage is now part of a live system that is receiving live data; this makes clear that data is being processed, but also that the model is being used as part of some organizational function; this latter may create a circumstance for an explanation of a model decision;

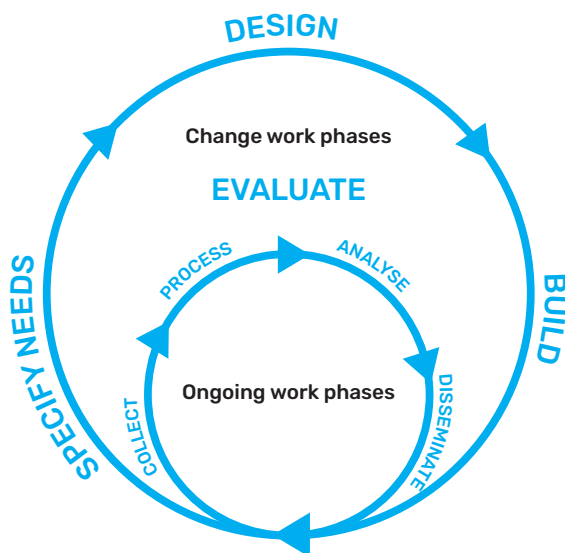


Figure 5.4.1: The Generic Statistical Business Process Model¹⁶

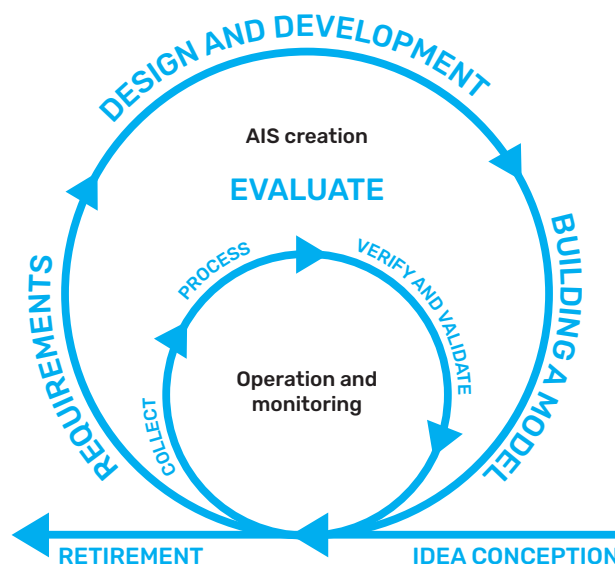


Figure 5.4.2: The AI system lifecycle¹⁷

¹⁶ op.cit. United Nations Economic Commission for Europe, [p.126](#).

¹⁷ op.cit International Standards Organization, [p.126](#).

- 5. Operating and monitoring:** this stage incorporates the collection, processing, analysis of live data leading to outputs, in parallel with which is the monitoring function that includes continuous validation, firstly to support evaluation of ongoing adherence to requirements and secondly, in the case of deviation, to alert to the need for revision to account for data drift and model decay; such revisions may affect function over time in that the output at one time during operation is different from another; thus to satisfy a request for an explanation may require model archival^a to account for changes in behavior over time;
- 6. Retirement:** the data is processed to ensure secure deletion, archival or repurposing; legal obligations pertain to each in respect of the prevention of breach of privacy; the models in a system retain vestiges of the data used to construct them and as such each constitutes a representation of data and may be subject to the same legal obligations as data, in particular to prohibit unauthorized reverse engineering for the recovery of features that might identify individuals. If the risk of deidentification is present appropriate measures must be taken.

The processes making up the stages in the building, operation and monitoring and retirement may be the subject of audit by regulatory bodies or information requests from individuals and hence may need to be appropriately documented for both internal and external oversight. All the stages involve either forethought about or the actual processing of personal data and therefore might implicate a regulatory regime and subject an actor to the considerations listed above. For example, the GDPR requires, subject to derogations, that the data subjects have a right to be informed^b in clear and plain language about the way their personal data is being processed, and data subjects can, again subject to derogations, invoke

certain rights such as right of access by the data subject, rectification and erasure.

The discussion here is not intended to be either prescriptive or exhaustive, but aims to illustrate the sort of considerations for which organizational governance procedures might need to account. In practice, it is likely that only some of these stages will apply. For instance, in a situation where an organization already holds appropriate data and has experience of using a particular model building approach, then only the stages of idea conception, business requirements, deployment and retirement may be required, but such reuse still counts as processing for legal purposes.

^a For technical and organizational purposes the model may be treated as data, so archival must be done for a purpose and for a limited period of time. The GDPR is very clear that for public interest, scientific, historical research purposes and statistical purposes longer periods are allowed, but GDPR does not define "longer periods". Relevant laws on archiving by jurisdiction will have to be consulted to determine the correct action.

^b Under GDPR, and for illustrative purposes, this information should be at least: (i) the identity and the contact details of the controller and, where applicable, of the controller's representative; (ii) the contact details of the data protection officer, where applicable; (iii) the purposes of the processing for which the personal data are intended as well as the legal basis for the processing; (iv) where the processing is based on point (f) of Article 6(1), the legitimate interests pursued by the controller or by a third party; (v) the recipients or categories of recipients of the personal data, if any; (vi) where applicable, the fact that the controller intends to transfer personal data to a third country or international organization, the existence of appropriate (technical and organizational) measures and transparency about these.

5.5 CHALLENGES OF CROSSING JURISDICTIONS

The discussion above applies to data processing where only one legal jurisdiction is relevant. Support for PETs that cross borders is substantially more complex. Beyond the fact that the legal analysis of PETs must be conducted within each individual jurisdiction, one must also consider the laws that apply to cross-border data transfers.

The principles of data localisation and data sovereignty are at the very heart of the style of governance chosen. Albeit that there is no general definition for either data localisation or data sovereignty, the definitions of both principles most related to the application of privacy preserving techniques are as follows: *data localization* deals with the issue of a mandatory legal or administrative requirement directly or indirectly stipulating that data be

stored or processed, exclusively or non-exclusively, within a specified jurisdiction,¹⁸ while *data sovereignty* deals with the issue of the power over one's digital domain exercised by a State's or possibly a private organization.¹⁹

Whereas data localization focuses on the national aspect of keeping certain data within national borders, data sovereignty focuses on both the national and international aspects of the data stored. So when PETs are deployed in a cross-border context and when the governance therefore needs to be determined, the actual application of these principles by each country involved (laws, regulations etc.) will have to be assessed and compared in order to create the required foundation for said governance.

5.6 ADVICE TO REGULATORS

As PETs and software implementations mature and become more widely considered for adoption by governments, companies, and non-profit organizations, one important question will be whether and to what extent regulators conclude that their guarantees suffice to fulfill the legal obligations of the parties to a computation. This is a complex question because, as said above, adhering to the law rarely boils down to a simple yes/no question and often involves a risk analysis.

Regulatory agencies are no different; they must also perform a cost-benefit analysis when weighing whether to promote adoption of a new technology or express caution and risk-aversion toward a new, untested technology. Moreover, due to limited resources regulatory agencies sometimes focus on cases of egregious wrongdoing and defer to community guidance in promoting best practices rather than proactively pushing for a new technology.

That said, data privacy and data analysis are topics of great

interest nowadays, and overall legislatures and regulators are increasingly displaying openness toward the possibility of allowing and even encouraging the use of PETs. For example, financial regulators in the United Kingdom²⁰ and the United States²¹ have encouraged innovation and adoption^c of new technologies to identify cases of money laundering and other financial crimes, and the Information Commissioner's Office in the UK has consulted with health organisations to shape thinking on privacy-enhancing technologies.²²

In part, this shift is due to the successful tech transitions of PETs in unregulated contexts, and pilot projects to explore how PETs would operate within a government agency or regulated industry, both of which are discussed in detail in [Chapter 3 Case Studies]. Given the difficulties of being a first mover in an industry to adopt PETs, concrete guidance from regulatory agencies can be a critical step that spurs adoption of new technologies that can improve overall data privacy.

¹⁸ Svantesson, Data localisation trends and challenges (2020).

¹⁹ Fabiano, "Digital Sovereignty Between "Accountability" and the Value of Personal Data" (2020).

²⁰ Financial Conduct Authority, Financial Crime TechSprint (2019).

²¹ Federal Reserve Board, Money Laundering and Terrorist Financing (2018).

^c UK and US launch innovation prize challenges in privacy-enhancing technologies to tackle financial crime and public health emergencies. <https://www.gov.uk/government/news/uk-and-us-launch-innovation-prize-challenges-in-privacy-enhancing-technologies-to-tackle-financial-crime-and-public-health-emergencies>, accessed 2023-01-23.

²² Information Commissioner's Office, Health organisations and PETs (2022).

CHAPTER 5. LEGAL AND REGULATORY ISSUES

BIBLIOGRAPHY

California State Legislature (2018). *California Consumer Privacy Act of 2018*. url: http://leginfo.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5. Accessed 2022-06-22.

European Commission (2022). *A European Strategy for Data*. url: <https://digital-strategy.ec.europa.eu/en/policies/strategy-data>. Accessed 2022-06-22.

European Data Protection Board (2020). *Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data*. url: https://edpb.europa.eu/our-work-tools/ourdocuments/recommendations/recommendations-012020-measures-supplementtransfer_en. Accessed 2022-06-22.

European Data Protection Board (2021). *Statement on the Digital Services Package and Data Strategy*. url: https://edpb.europa.eu/our-work-tools/our-documents/statements/statementdigital-services-package-and-data-strategy_en. Accessed 2022-06-22.

European Union (2016). *General Data Protection Regulation*. url: <http://data.europa.eu/eli/reg/2016/679/oj>. Accessed 2023-01-04.

European Union (2020). *Data Governance Act*. url: <http://data.europa.eu/eli/reg/2022/868/oj>. Accessed 2023-01-04.

European Union (2021). *Artificial Intelligence Act (Proposal)*. url: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>. Accessed 2023-01-04.

European Union (2022). *Digital Markets Act*. url: <http://data.europa.eu/eli/reg/2022/1925/oj>. Accessed 2023-01-04.

European Union (2023). *European statistics on population and housing (Proposal)*. url: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2023:31:FIN>. Accessed 2023-01-23.

Fabiano, Nicola (2020). "Digital Sovereignty Between "Accountability" and the Value of Personal Data". In: *Advances in Science, Technology and Engineering Systems Journal* 5.3, pp. 270–274. doi: [10.25046/aj050335](https://doi.org/10.25046/aj050335).

Federal Reserve Board (2018). *Joint Statement on Innovative Efforts to Combat Money Laundering and Terrorist Financing*. Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, Financial Crimes Enforcement Network, National Credit Union Administration, and Office of the Comptroller of the Currency. url: <https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20181203a1.pdf>. Accessed 2022-05-23.

Financial Conduct Authority (2019). *2019 Global AML and Financial Crime TechSprint*. url: <https://www.fca.org.uk/events/techsprints/2019-global-amland-financial-crime-techsprint>. Accessed 2022-04-22.

Government of the Netherlands (2003). *Wet op het Centraal bureau voor de statistiek*. url: <https://wetten.overheid.nl/jci1.3:c:BWBR0015926&z=2022-03-02&g=2022-03-02>. Accessed 2022-09-07. An unofficial English translation of the Statistics Netherlands Act is available via: <https://www.cbs.nl/en-gb/about-us/organisation>. Accessed 2023-01-04.

Information Commissioner's Office (2021). *ICO call for views: Anonymisation, pseudonymisation and privacy enhancing technologies guidance*. url: <https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/icocall-for-views-anonymisation-pseudonymisation-and-privacy-enhancingtechnologies-guidance/>. Accessed 2022-06-22.

Information Commissioner's Office (2022a). *Data protection by design and default*. Information Commissioner's Office. url: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-by-design-and-default/>. Accessed 2022-06-22.

Information Commissioner's Office (2022b). *ICO consults health organisations to shape thinking on privacy-enhancing technologies*. url: <https://ico.org.uk/about-theico/media-centre/news-and-blogs/2022/02/ico-consults-healthorganisations-to-shape-thinking-on-privacy-enhancing-technologies/>. Accessed 2022-06-22.

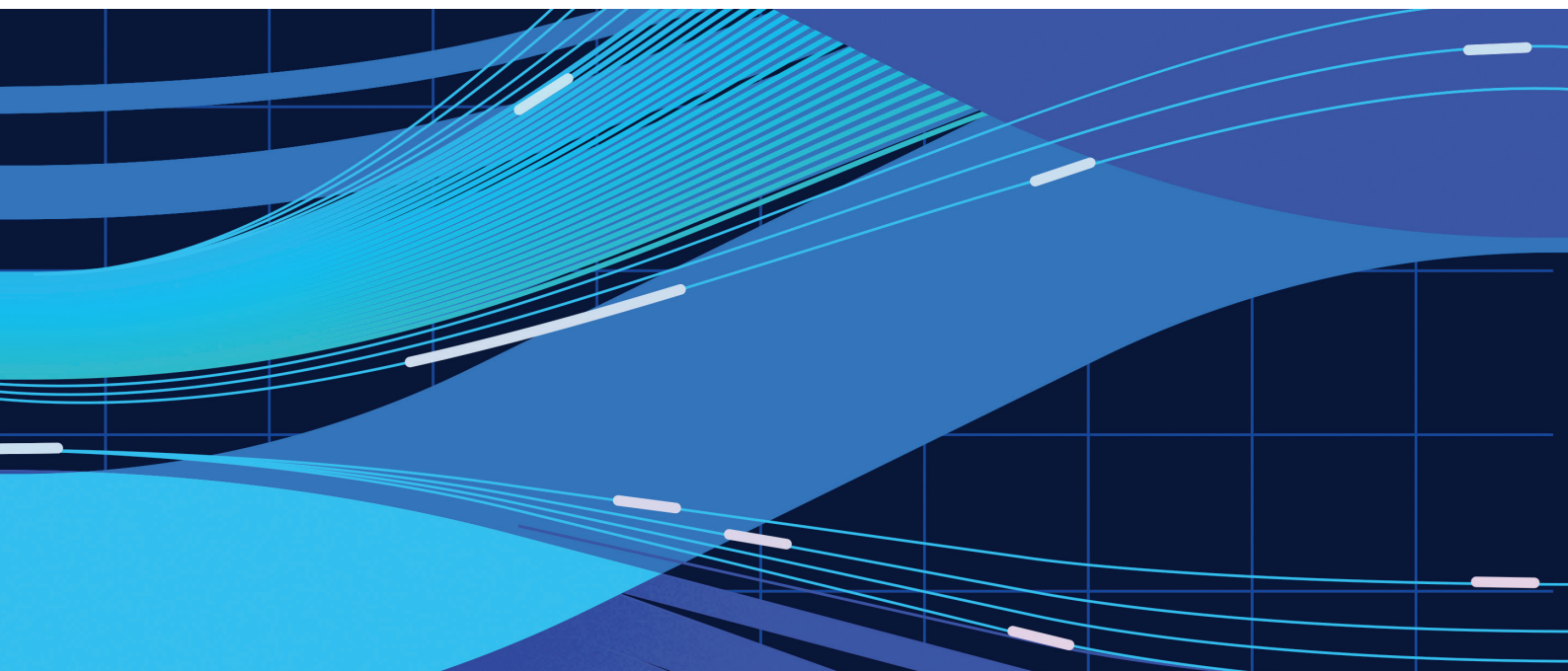
International Standards Organization (July 2022). *ISO/IEC 22989:2022 Information technology – Artificial intelligence – Artificial intelligence concepts and terminology*. url: <https://www.iso.org/standard/74296.html>. Accessed 2022-09-16.

Svantesson, Dan (2020). *Data localisation trends and challenges*. OECD Digital Economy Papers No. 301. OECD Publishing, Paris. doi: [10.1787/7fbaed62-en](https://doi.org/10.1787/7fbaed62-en).

United Nations Economic Commission for Europe (2021). *Generic Statistical Business Process Model (GSBPM v5.1)*. url: <https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1>. Accessed 2022-06-22.

Varia, Mayank (2023). *Legal Issues arising from Privacy Enhancing Technologies*. In preparation.





THE PET GUIDE

THE UNITED NATIONS GUIDE ON
PRIVACY-ENHANCING TECHNOLOGIES
FOR OFFICIAL STATISTICS.



**United
Nations**

Department of
Economic and
Social Affairs

Statistics Division,
UN Department of Economic
and Social Affairs.
unstats.un.org/bigdata
bigdata@un.org