

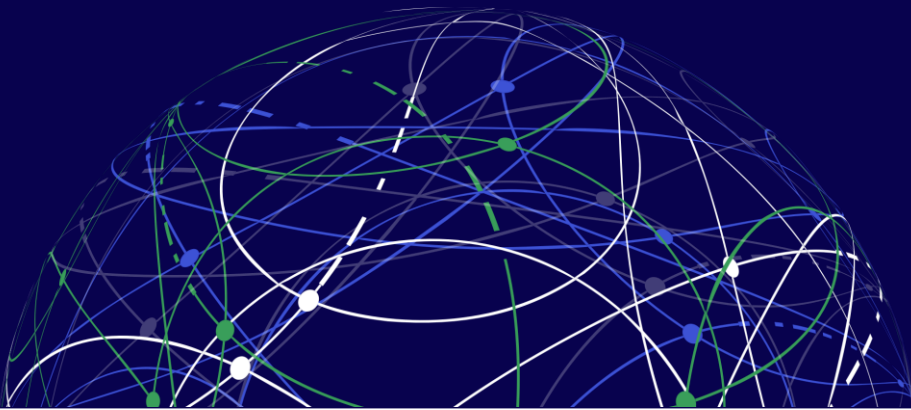
Privacy Enhancing Technologies Whitepaper

Developed by Centre of Excellence – Data Sharing and Cloud

July 2023

Authors:

Constantijn Molengraaf, Daniel Kapitan, Pepijn Groen, Yekaterina Travkina



About Authors



Pepijn Groen
Programme lead for CoE-DSC
& Manager, INNOPAY, Netherlands

Pepijn Groen is a programme lead of CoE-DSC. He is well versed in multi-stakeholder projects that aim to design, develop and implement innovative product and service concepts in open, collaborative ecosystems. In this capacity he has facilitated various industry collaborations in developing innovative services in such areas as public transport payments, asset-based finance, energy and mobility.



Daniel Kapitan
Expert from CoE-DSC community
& Eindhoven AI Systems Institute (EIASI) Fellow at TU/e

Daniel Kapitan is an EIASI Fellow with over 20 years of hands-on experience as a data scientist, business architect and strategic advisor in industry. He is responsible for the data science & professional education program at Eindhoven University of Technology. Through combining his work as an advisor, lecturer, and applied researcher, he aims to make a positive contribution towards data solidarity and civic AI.



Constantijn Molengraaf
Expert for CoE-DSC program
& Consultant, INNOPAY, Netherlands

Constantijn Molengraaf is a consultant on Digital Ecosystems and Trust Frameworks with expertise in deep tech (Privacy Enhancing Tech, Quantum computing). He has facilitated multiple use cases in the CoE-DSC and developed various CoE-DSC use case tools like the Use Case Implementation Guide. His mission is to align the use of data, technology, and AI with societal goals.



Yekaterina Travkina
Expert for CoE-DSC program
& Consultant, INNOPAY, Netherlands

Yekaterina Travkina is a consultant specialising in research on trust and governance for Data Spaces, and use of AI and Privacy Enhancing Technologies for data sharing. She has supported various CoE-DSC use cases in healthcare, finance, and manufacturing. She is part of CoE-DSC harmonisation efforts for Data Spaces development. Her goal is to drive sustainable change and equitable development.

Abstract

This whitepaper provides decision-makers with insights on the benefits of Privacy Enhancing Technologies (PETs) for data collaborations. With recent growth and development of data sharing, public and private organisations can realise new economic and societal value potential. However, data collaboration participants often face barriers for data sharing in form of privacy, commercial and reputational risks. PETs can play a role for reducing these barriers and increasing trust in data collaborations where data cannot be shared directly, since PETs allow to generate insights without disclosing the underlying data. The paper focuses on the most important PETs and their benefits for respective use cases. It also covers challenges that need to be overcome for large-scale adoption of PETs and lastly, shows tangible steps for fostering implementation of these technologies in organisations.

About Contributors

This whitepaper was reviewed and improved with the help of experts from various fields. Below you can find the list of all contributors who were part of the process. Authors would like to thank these experts for their contribution and knowledge.

Experts from Knowledge Institutions:

- Daniel Worm, TNO
- Johan van Soest, BISS and Maastricht University
- Mark de Reuver, TU Delft
- Wirawan Agahari, TU Delft
- Yannis Velegarakis, Utrecht University

Representatives of PETs Providers:

- Edwin Kooistra and Vincent Campfens, BlueGen AI
- Maarten Everts and Pieter Verhagen, Linksight
- Toon Segers, Roseman Labs
- Wim Kees Janssen, Syntho

Other Experts:

- Miranda Graftdijk, Verbond van Verzekeraars
- Paul Fockens, Sustainable Rescue Foundation
- Theo Hooghiemstra, Hooghiemstra & Partners

Table of Contents

<i>Management summary</i>	5
<i>Privacy Enhancing Technologies Whitepaper</i>	6
1. Data sharing in collaborative ecosystems is key to capture new value	6
2. Privacy, reputation and commercial risks hinder progress of data sharing initiatives	6
3. Privacy-Enhancing Technologies (PETs) offer mechanisms to collaborate on data to generate new value, while preserving privacy	7
4. Use cases can benefit from three emergent groups of PETs: Synthetic Data, Federated Learning, Multi-Party Computation, and apply Differential Privacy to assess PETs use	8
5. PETs are still a relatively new technology, at least five challenges remain for large-scale adoption	10
6. Key next steps to use PETs in your organisation	10
<i>Appendix: selected use case descriptions</i>	12

Management summary

Various studies¹ show that data sharing is becoming increasingly important for realising economic and societal value from data and analytics. Business leaders in both private and public organisations are faced with the challenge to define strategic opportunities and layout tactical plans to put data sharing in practice as part of their digital transformation agenda.

This paper aims to provide decision-makers with insights on the benefits that Privacy Enhancing Technologies (PETs) have to offer, clarifying what it can - and cannot – bring to capture that value. The common denominator of PETs is that they allow specific insights to be extracted from shared data without actually disclosing the underlying data itself. As such, PETs allow organizations to capture the utility of the combined (shared) data, whilst at the same safeguarding privacy, commercial interests, and reputational interests of all those involved. Thus, PETs can play a role in increasing trust for participants in data collaborations where data cannot be shared directly, as the result allowing them to unlock new value from data².

The six key take aways of this paper are that:

1. Data sharing in collaborative ecosystems is key to develop new insights from data sets.
2. Privacy, commercial and reputational barriers hinder progress of data sharing initiatives.
3. PETs can accelerate use case development since barriers are significantly reduced and/or simplified due to PETs allowing participants to safeguard their data. This is observed in practice by data collaborations in various sectors (see p.7 and p.10 for practical examples).
4. Use cases can benefit from three emergent groups of PETs: Synthetic Data, Federated Learning, and Multi-Party Computation. Additionally, in specific settings, use cases can utilise Differential Privacy as a technical assessment measure of privacy parameters in algorithms³.
 - PETs typically apply to use cases where there is a need to gain statistical insights by analysing datasets from collaborating organisations, without disclosing underlying (sensitive) data. In some sectors this is referred to as a secondary use of data, meaning data is re-used for other purposes than for which it was originally collected (examples can include monitoring sector performance, contributing to sustainable and societal goals).
 - PETs are less applicable in cases where actual data sharing is required. For example, where party A and party B need to transact actual datapoints among each other in order to achieve a desired outcome (e.g. to complete invoicing, to approve order details, to dispatch delivery). This is referred to primary use of data, meaning that data is used for the same purpose as it is collected for.
5. PETs are still a relatively new technology, with several challenges that need to be addressed to enable large-scale adoption. Challenges include usability, interoperability, trust, legal and regulatory frameworks that enhance PETs potential.
 - Addressing these challenges will require collaboration between technical experts, policymakers, and potential PET users to develop agreements enabling the adoption of PETs on a large scale and capturing the value of data for society.
6. And lastly, to implement PETs in an organisation, it is helpful to experiment with demos and trial projects to learn what works in your context, raise internal awareness, build internal data capabilities, and collaborate with stakeholders and data sharing partners.

¹ OECD. (2019). [Economic and social benefits of data access and sharing](#)

² Ofe, H., Minnema, H., & de Reuver, M. (2022). The business value of privacy-preserving technologies: the case of multiparty computation in the telecom industry. *Digital Policy, Regulation and Governance*, 24(6), 541-557. <https://doi.org/10.1108/DPRG-10-2021-0132> and Attema, T. & Worm, D. (2021). [Technological breakthrough finally, a privacy-friendly way to harness data.](#)

³ Near, J., & Darais, D. (2022). [Differential Privacy: Future Work & Open Challenges](#). NIST

Privacy Enhancing Technologies Whitepaper

1. Data sharing in collaborative ecosystems is key to capture new value

Pursuing new business opportunities on data is high on the digital agenda of decision-makers and goes under the name of for example “data-driven innovation”, “data economy innovation”, “social data infrastructure” and “internet of FAIR” (Findable, Accessible, Interoperable, Reusable).

In practice, organisations now view data as a strategic asset and increasingly become aware that data represents a key asset both for economic growth and societal benefits. The data an organisation collects contains many insights, but when the data from one organisation is combined with data from other organisations, new value creation can be unlocked as new insights are generated⁴.

Sharing data with others offer opportunities for business model innovation, new revenue streams and process efficiency. The significance of data sharing to macro-economic effects is also widely recognised in various studies on both national and international level⁵. Next to economic growth, data is a strategic asset for achieving societal goals like realising affordable and accessible healthcare, aiding green energy transition, innovating sustainable mobility, achieving supply chain efficiency for circular economy, climate change damage prevention and more⁶.

To capture the value of data sharing, organisations need to collaborate with each other on data. Collaboration requires trust, efficient deployment and use of jointly defined standards, rules and agreements, and ensuring that data is findable, accessible, reusable and interoperable (FAIR). Data sharing is gaining traction as new digital collaboration models arise. Examples include ecosystem partnerships and collaborative models such as data spaces that focus on data sharing for specific themes (e.g., human trafficking), for specific sectors (e.g. logistics, manufacturing, health care) or across sectors. On Dutch national level, a wide variety of data spaces are already up or in development across various sectors such as energy (MFF BAS), mortgage (HDN), health (HealthRI, MedMij), construction (DSGO) and the manufacturing industry (SCSN).

2. Privacy, reputation and commercial risks hinder progress of data sharing initiatives

Even though organisations see value in data sharing initiatives, there are typical barriers related to privacy, commercial and reputational interests of those providing the data (see Table 1).




Barriers	1. Privacy Barrier	2. Commercial Barrier	3. Reputational Barrier
			
Description	In some circumstances it is difficult for organisations to directly share data that is sensitive and involves Personal Identifiable Information (PII). For example, data sharing requires legal basis under GDPR, in addition to other sectoral laws that may apply and require compliance from data providers (e.g., duty of confidentiality and professional secrecy).	Organisations are hesitant to share data with others because data constitutes commercial value and is considered a strategic asset. For non-commercial parties, such as non-profits and charitable organisations, this also holds true since data constitutes a strategic asset to show performance and help attract donors to their initiatives. ⁷	Organisations are not willing to share data because they fear that once their data is shared it will be re-used for other purposes than it was intended for, leading to reputational damage. This typically is the case when there are no sufficient instruments in place to mitigate liability risks (e.g., weak identity assurance)
Example context	Typical for initiatives where sensitive personal data is involved. E.g., medical records of the patient, travel patterns of an individual etc.	Typical for initiatives where organisations collaborate on data that are also in competition with each other. E.g., mobility providers, healthcare providers, insurance providers etc.	For all kinds of initiatives where significant liability risks exist and there are little trust mechanisms in place to mitigate these.

Table 1. Barriers that hinder progress of data sharing initiatives

⁴ OECD. (2019). [Economic and social benefits of data access and sharing](#)

⁵ McKinsey (2021) Discussion paper. [Financial data unbound: The value of open data for individuals and institutions](#); World Economic Forum (2020) Whitepaper. [Share to Gain: Unlocking Data Value in Manufacturing](#);

⁶ The World Bank. (2021). [World Development Report 2021: DATA FOR BETTER LIVES](#)

⁷ Mayer, D. J. & Fischer, R. L. (2023). Exploring data use in non-profit organisations. <https://doi.org/10.1016/j.evalprogplan.2022.102197>

The respective barriers need to be reduced through mechanisms which bring assurance and trust among the participants on business, legal, operational, technical, and functional levels⁸. Specific mechanisms required are dependent on the context of the initiative. This paper focuses on privacy enhancing technologies (PETs) as a part of technical solution for initiatives where data cannot be shared directly for generating insights.

3. Privacy-Enhancing Technologies (PETs) offer mechanisms to collaborate on data to generate new value, while preserving privacy

PETs have emerged to tackle aforementioned challenges by sharing data without disclosing underlying information to an accepted degree⁹. At first glance, this may seem like an unsurmountable paradox: how can we share data without disclosing it? Advancements in privacy enhancing technologies (PETs) have made it possible to do exactly that. Intuitively, one can think of PETs to create insights without needing access to the raw underlying data. The data utility, defined as the usefulness of data in providing insights, is maintained by PETs such that specific insights can be extracted without disclosing the underlying data. Protecting the underlying data is very common for example when using HTTPS or VPNs to only allow specific recipients to use the data and to avoid the misuse of the data (given that applied control mechanisms are trusted). PETs take this a step further by not only restricting *who* can access the data but also *what* can be done with the data. To achieve that the emerging PETs solutions strive to ensure the balance between preserving data utility whilst safeguarding privacy*, as illustrated in Figure 1.

In open data collaborations, data is pooled together for an analysis, for which Trusted Third Parties (TTPs) are often involved. Relying on the TTP for centralised storing and analysing of the data in the open (i.e., without privacy enhancements) creates vulnerabilities and a single point of failure. While in privacy enhanced collaborations, PETs are applied to the data already at a source as well as when doing integrated analysis to avoid revealing sensitive information in the output of computations. Here data is not shared directly, and data utility is preserved to the required level of analysis adjusted to a specific use case.

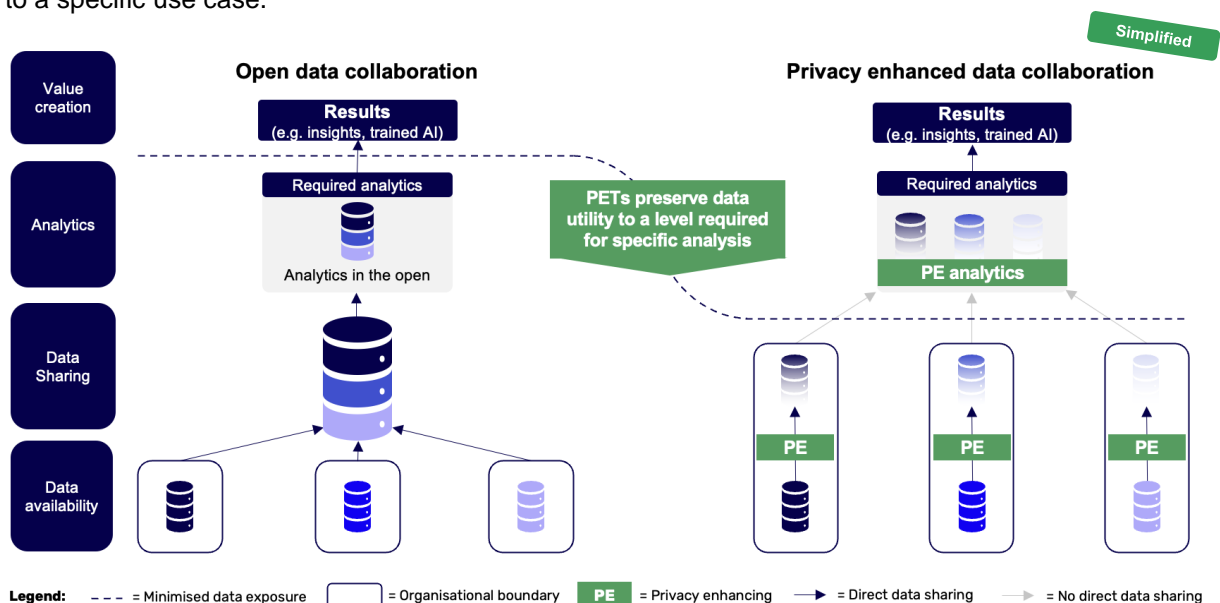


Figure 1. Differences between open and privacy enhanced data collaborations

* The definitions of privacy depend on the respective field - legal, technical, ethical¹⁰. In this paper we cover privacy that concerns safeguarding information of individuals, which is formally known as information privacy or data privacy¹¹. This entails controlled use, collection, and distribution of personal identifiable information (PII) as per data subject's consent.

⁸ We refer to [Data Sharing Canvas](#) on trust framework for data sharing

⁹ [OECD Digital Economy Papers](#). (2023). Emerging Privacy Enhancing Technologies

¹⁰ Smith, H. J., et al. (2011) [Information Privacy Research: An Interdisciplinary Review](#)

¹¹ Sun, C. et al. (2021) [A systematic review on privacy-preserving distributed data mining](#)

4. Use cases can benefit from three emergent groups of PETs: Synthetic Data, Federated Learning, Multi-Party Computation, and apply Differential Privacy to assess PETs use

In this article¹², we focus on PETs that accelerate large scale data collaborations. In this category we consider Synthetic Data (SD), Federated Learning (FL), and Multi-party Computation (MPC)¹³. As a separate category, the article covers Differential Privacy (DP) as a mechanism to assess privacy impact by quantifying the privacy parameters of an algorithm¹⁴. Although, by itself DP doesn't enhance privacy, it is relevant, nevertheless, for making assessments in use cases where aggregated insights are generated from sensitive data. As such Differential Privacy serves as a technical parameter (rf. footnote¹⁵ for the distinction between security and privacy).

The table below provides an overview of the different PETs and privacy assessment measures in scope of this article.

	Privacy Enhancement Technologies			Privacy Assessment Measure
	Synthetic data	Federated Learning (FL)	Multi-Party Computation (MPC)	Differential Privacy (DP)
Illustration				
Description	Synthetic data tools transform a dataset into a new data set with similar statistical properties while removing privacy sensitive information from the original data set.	Federated learning trains one 'overarching' AI model by locally training 'sub' AI models, removing the need to share & combine the original data sets.	Family of encryption schemes for computation on encrypted data. E.g.: <ul style="list-style-type: none"> • Secret sharing • Circuit garbling • Homomorphic encryption 	DP provides a measure on how much personal data is exposed by a specific data analysis algorithm. It shows mathematical parameters on privacy.
Typical data context	Data provider with large data set provides access to synthesised data set to one or multiple data consumers.	Collective need for training AI model on sensitive data sets distributed over multiple organisations.	Collective need for creation of insights/training ML on sensitive data distributed over multiple organisations.	DP is useful when there is a need to assess privacy impact from creating aggregated insights or training AI models on large data sets.
Characteristics	Computation is performed 'in the clear' on synthetic data (i.e., no data encryption is needed, since the original data is not used), allowing for open exploration.	Computation is performed 'in the clear', i.e., data is not encrypted for the analysis, as it is conducted at the source and only model parameters are shared afterwards.	Computation is performed 'in the blind', i.e., data is encrypted, and only aggregated results and/or model parameters are shared afterwards.	DP applies randomisation compared to deterministic statistical techniques; thus, it helps to ensure privacy in the context of re-identification attacks.
Limitations / Risks	<ul style="list-style-type: none"> • Data utility is dependent on performance of synthesization algorithms to capture relevant statistical properties. • Remains to be seen if synthetic data can be used to train models for production or only useful for exploration. • Not suited for distributed data as multi-variate relations between datasets is difficult to mimic. 	<ul style="list-style-type: none"> • Not suitable for highly sensitive data as FL has no formal security proof (the trained sub models may still contain sensitive characteristics from the data set). • Semantic standardisation is required for efficient collaboration between data contributors. 	<ul style="list-style-type: none"> • Resource complexity (interactions, computational) depend on specific computation and implementation combination. • Semantic standardisation is required for efficient collaboration between data contributors. 	<ul style="list-style-type: none"> • There is no general application-agnostic recipe to choose values of the DP parameters. • Achieving high degree of privacy can result in a loss of accuracy. • DP only assesses the algorithm used.
<ul style="list-style-type: none"> • Solution providers (in NL) and open source solutions 	<ul style="list-style-type: none"> • BlueGen AI • Syntho • Syntric AI • Synthetic Data Vault • Smart Noise 	<ul style="list-style-type: none"> • BranchKey • Syft + Grid (from OpenMined) • Flower • TensorFlow Federated • IBM Federated Learning • OpenFL 	<ul style="list-style-type: none"> • Linksgint • Roseman Labs • ABY framework with EzPC compiler • SCAPI • the SCALE-MAMBA • swanky • Motion • JIFF • CryptTen 	<ul style="list-style-type: none"> • OpenDP/SmartNoise Core and SmartNoise SDK • Google DP • TensorFlow Privacy (DP-SGD) • Pytorch Opacus (DP-SGD) • IBM Diffprivlib • Diffpriv (R package)

Source: INNOPAY B.V. analysis

Legend: Raw data Data with minimised personal data Generate synthetic data AI model Encryption Value Measure of privacy parameter

Table 2. The overview of PETs and Privacy Assessment Measures in scope of this paper

¹² We refer to [UN PET Guide](#), [UK Royal Society report](#) and [OECD Digital Economy Papers](#) for more elaborate introduction of PETs

¹³ MPC denotes various cryptographic protocols including secret sharing, garbled circuits, and Homomorphic Encryption because of their similar added value (they enable multiple parties to compute on data which is encrypted), although their technical implementations would justify separating them. For detailed comparisons between the performance, security, privacy and technical implications of these protocols we refer to ².

¹⁴ Near, J., & Darais, D. (2022). [Differential Privacy: Future Work & Open Challenges](#). NIST

¹⁵ When assessing PETs, security and privacy need to be differentiated, the first is about protection against data breaches while the second is about ensuring control over sensitive data of a data subject. Read more in Smith, H. J., et al. (2011) [Information Privacy Research: An Interdisciplinary Review](#)

Below is the list of selected use cases where PETs are applied (for CoE-DSC use cases refer to the table in the Appendix).

DUO provides synthetic datasets for education research in the Netherlands¹⁶

Starting in 2022, DUO (*Dienst Uitvoering Onderwijs*) - an education executive agency in the Netherlands – provides synthetic datasets for research purposes as part of the open data education project¹⁷. Registered researchers can make [requests](#) for synthetic data generated from original datasets. DUO collects various information on students in the Netherlands from primary to higher education, including obtained diplomas, programme completions and student financing. Such data, however, is privacy sensitive as it includes PII of students, and thus there is a barrier for sharing it with researchers directly. Therefore, to ensure data privacy, DUO generates synthetic datasets mimicking characteristics of the original dataset to be then used for training research models, and/or generating statistical insights to help improve education system.

Google uses Federated Learning to enhance next-word suggestions for virtual keyboards¹⁸

Google aimed to set up the keyboard next-word suggestions tailored to the users of smart devices like phones and tablets. Originally that would mean training an AI algorithm on pooled user data. However, this is not attainable due to privacy and connectivity barriers, as data from devices contains personal identifiable information of the end-users and these devices would need the bandwidth to share that data. To overcome these challenges, Google used a federated learning implementation which trains AI on-device in a decentralised manner without exporting end-users' data to servers. Such solution ensures users have privacy and control over their data, while the AI achieves better predictions, since training language models at a source allows to give personalised suggestions.

The Boston Women's Workforce Council (BWWC) uses MPC for periodic analysis of pay gaps¹⁹

Periodic analysis of gender and racial wage gaps is needed to stimulate equity roadmap development and provide accurate benchmarking. However, currently openly available sources for the analysis only include a self-reported US census data, while using the payroll data directly without MPC is not feasible as it contains personal information of employees. BWWC together with Hariri Institute since 2017²⁰ are implementing MPC to create aggregated insights and monitor pay gaps without revealing sensitive employee data. This implementation also allows to gain higher accuracy of benchmarks, compared to self-reported census data.

The CARRIER consortium applies Differential Privacy to ensure that their cardiovascular risk prediction models are secure and preserve privacy²¹

The CARRIER²² (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 (MUMC+, Zuyderland, Maastricht University/RNFM, ZorgTTP, and Statistics Netherlands CBS). Considering the reuse of gathered data, the project is heavily dependent on the legal basis on which the data was collected and other regulatory regimes impacting processing. Thus, one of the main challenges is the linkage of data sets owned by the different parties due to the risk of re-identification of subjects. This requires CARRIER to adhere to the highest standards of data security and privacy preserving measures. For that Differential Privacy is used to quantify parameters ensuring that developed algorithms are secure.

Note, that in practice the exemplified PETs can be used in combination with one another depending on the context of a use case.

¹⁶ Read more on DUO use case [here](#).

¹⁷ See more at: https://duo.nl/open_onderwijsdata/

¹⁸ Hard, A., et al. (2018). Federated Learning for Mobile Keyboard Prediction. <https://doi.org/10.48550/arXiv.1811.03604>

¹⁹ [BWWC on data privacy](#)

²⁰ More on BU Hariri Institute involvement can be found [here](#).

²¹ Scheenstra, B., et al. (2022). Digital Health Solutions to Reduce the Burden of Atherosclerotic Cardiovascular Disease Proposed by the CARRIER Consortium. *JMIR Cardio*. [doi:10.2196/37437](https://doi.org/10.2196/37437)

²² More on CARRIER project can be found [here](#), with involved project partners listed [here](#).

5. PETs are still a relatively new technology, at least five challenges remain for large-scale adoption

Privacy enhancing technologies (PETs) have the potential to enable the development and adoption of data spaces while protecting individuals' privacy. While there has been significant progress in the development of PETs, there are still several challenges that need to be addressed to enable their large-scale adoption.

The five challenges for large-scale adoption:

- *Legal and regulatory frameworks:* PETs operate in a complex regulatory environment, with different countries and sectors having different privacy laws and regulations. For example, EU agency ENISA highlights that steps need to be taken to assess PETs readiness as tools aiding privacy protection in the legal context.²³ Since current regulations were not written with PETs in mind, [Privacy Enforcement Authorities](#) - like EDPB in the EU and ICO in the UK - are now working on incorporating PETs in data protection frameworks to stimulate the adoption of these technologies.²⁴ In addition, under GDPR data collaborations need to perform [DPIAs](#) (Data Protection Impact Assessments). This requires privacy and security officers²⁵ of organisations to be involved prior to implementing PETs in a data collaboration, which takes time and effort. For more on the legal aspects around the deployment of PETs see CoE-DSC and Pels Rijcken [whitepaper](#).
- *Usability:* Privacy enhancing technologies can be complex to implement and require specialised skills to deploy, manage and maintain. This can limit their adoption by organisations and individuals who lack the necessary technical expertise. PETs must be user-friendly and easy to understand to ensure that individuals can use them effectively.
- *Standardisation/quality of data:* to effectively integrate PETs in a data collaboration, participants need to ensure the quality of their datasets first. The data providers need to clean and prepare the data (e.g., making it machine readable and ensuring it satisfies agreed standards). These procedures require time, effort and expertise from participating organisations since data has to be taken care of at the source prior to using PETs.
- *Interoperability:* Different privacy enhancing technologies utilising heterogenous code are not always compatible with each other, making it difficult to create a comprehensive and integrated privacy protection framework. Interoperability issues can also hinder the exchange and sharing of data between different (PET driven) data spaces.
- *Trust:* The adoption of PETs depends on the trust that individuals and organisations have in the technology and the service providers implementing the technologies. PETs must be secure and reliable to gain trust, and their effectiveness must be validated through independent testing and evaluation.

Addressing these challenges will require collaboration between technical experts, policymakers, and potential PET users to develop the agreements that enables the adoption of PETs on a large scale.

6. Key next steps to use PETs in your organisation

When it comes to implementing PETs in your organisation, there are a few key steps that you should take:

- Experiment with demos and trial projects to see what works best for your specific needs. This may involve using decision tools such as the one developed by [TNO](#) or [VKA](#) to help guide your decision-making process. It is helpful to identify the problem or a challenge that data collaboration aims to solve (refer to [CoE-DSC Playbook](#) to structure your use case), and from there identify PETs fit for the context.
- Build your internal data capabilities. This means investing in the right technology, infrastructure, and operating model to support PETs and ensure that they are integrated seamlessly into your existing workflows. Also, improving internal capabilities involves training data related skills of employees and fostering creative and innovative environment for IT departments.²⁶

²³ ENISA - [Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies](#)

²⁴ [OECD Digital Economy Papers](#). (2023). Emerging Privacy Enhancing Technologies




















































²⁵ Read [here](#) on the role of Data Protection Officer (DPO)









²⁶ Agahari, W., Ofe, H., & de Reuver, M. (2022). It is not (only) about privacy: How multi-party computation redefines control, trust, and risk in data sharing. *Electronic markets*, 32(3), 1577-1602.

- Collaboration is also crucial when it comes to successful implementation of PETs. You'll need to work closely with stakeholders across departments and even outside of your organisation to form agreements on for example compensation models, liabilities, and service level agreements. It's essential to consider what additional agreements, processes, or tools may be necessary beyond PETs to ensure successful adoption and the BLOFT framework – Business, Legal, Operational, Functional, Technical – is a great starting point for this consideration.²⁷ Refer to [CoE-DSC tools and approach](#) as a guide for analysing BLOFT topics in your use case.
- Measure security and performance of applied algorithms as this will allow you to evaluate the effectiveness of your PETs and make any necessary adjustments to optimise the solution.

²⁷ Refer to [Data Sharing Canvas](#) - Section 1.4

Appendix: selected use case descriptions

Use cases	Description	Participants
Federated Learning for Health and Care (PoC by NLAIC)	HealthRI, LUMC, EMC explore distributive collaboration models (Algorithm to Data - A2D) with federated machine learning to ensure private AI data sharing based on FAIR principles for cancer medical research	    
Poverty Prevention in the Netherlands (ELSA)	ELSA lab explores the value and legal feasibility of using PETs for data collaboration between public institutions for early detection of poverty and debt	  
Mobility as a Service (MaaS) in the Netherlands	Roseman Labs, Publiek Vervoer and Transport Service Providers implement MPC on encrypted travellers' data to harmonise mobility offerings by matching the supply and demand and to contribute to sustainable mobility in the Netherlands	      
Monitoring Human Trafficking in the Netherlands	Roseman Labs, Pinsent Masons and Sustainable Rescue implement MPC on encrypted NGOs data to timely detect and aid victims of human trafficking across Netherlands	   
Monitoring Dutch Elderly Care	Linksight, DSW, Delft Municipality and Pieter van Foreest implement MPC for monitoring sector performance and measuring impact of Dutch care policies on elderly care provisions (i.e., WLZ, WMO, ZVW)	    
Privacy preserving entity resolution for AML and CFT	Knights Analytics and Roseman Labs implement MPC for entity resolution analysis on the data of Financial Institutions to detect money laundering and terrorist financing activities	  
GERDA (Integrated Regional Data Infrastructure in health domain)	Linksight, Population Health Data NL and 8RHK Gezond implement MPC for monitoring preventive healthcare provisions to improve Dutch Health care policy making (WMO, WLZ, ZVW)	  
CARRIER (Coronary ARtery disease: Risk estimations and Interventions for prevention and EaRly detection)	Maastricht and Limburg researchers and clinical practitioners, with SANANET and CBS implement MPC to aid in early detection and prevention of heart diseases through a personal health train data solution	      
The project privacy preserving analytics (PPA) on patients data	Linksight together with CZ, Zuyderland Medical Center and CBS analysed data from 4000 patients using MPC to determine the effectivity of eHealth app services for different patient groups in Limburg	    
Energy Use Case (grid monitoring)	Stedin, Roseman Labs implement MPC for monitoring energy generation and consumption to estimate grid capacity and develop congestion forecasts	 
Analysing 2-year survival rate of lung cancer patients (PHT use case)	Maastricht UMC led a consortium of 8 institutions in 5 countries to calculate the 2-year survival rate using federated learning with over 20,000 patient records. This approach enabled the study to be concluded in just 4 months, which without PETs would have taken much longer to execute.	      

Other use cases under Personal Health Train (PHT) project	For the list of personal health train use cases utilising Federated Learning and MPC see read more here .	
Federated energy management systems for building assets	BranchKey in collaboration with TNO developed Federated Machine Learning to predict and manage energy usage across buildings while ensuring privacy of sensitive data.	
Predictive maintenance for industrial marine equipment with FL	BranchKey developed Federated Machine Learning for marine equipment operators and manufactures to conduct timely maintenance of marine vessels while ensuring privacy of sensitive data.	
DUO synthetic data set for education research	DUO shares synthetic datasets with Dutch researchers to train research models, and help improve education system.	
CBS pilot to synthesise business registries	CBS together with Syntho did a Proof of Concept to synthesise a section of the General Business Register (ABR) in the Netherlands using a number of basic characteristics such as economic activity and size class.	
Churn prediction for Telecom customers using synthetic data	SAS and Syntho in collaboration with NL AI Coalition did a case study on generating a synthetic dataset from the original sensitive telecom data of 56 600 customers. Synthetic data was then used to train models for predicting customer churn in a privacy preserving way.	
Predicting deterioration and mortality for cancer research with synthetic data	Syntho in the SAS hackathon generated synthetic data from sensitive patients' datasets and used those synthetic datasets for making predictions for cancer research in a privacy preserving way.	
Predicting energy consumption with what-if scenarios using synthetic data	BlueGen AI generates synthetic data from smart meters datasets, which allows grid managers to make predictions of energy consumption and better manage supply and demand while ensuring privacy of original sensitive data from household meters.	
Financial Risk modeling and predicting faulty loans with synthetic data	BlueGen AI generates synthetic data from bank datasets of the customers and loan applicants. The synthetic data allows banks to train models for risk prediction and provide better loan conditions to customers while ensuring privacy of original sensitive clients' data.	