WONG PARTNERSHIP

Privacy Enhancing Technologies

Personal Data Protection: New Federated Learning Guide, Revised Synthetic Data Guide and New PET Sandbox Use Cases

On 20 July 2026, the Personal Data Protection Commission (**PDPC**) published new guidance on Privacy Enhancing Technologies (**PETs**), which help organisations derive insights from personal data in a secure way without exposing personal information.

The guidance comprises:

(a) A new Guide on Federated Learning (**FL Guide**), launched at the inaugural Singapore Data Festival;

(b) A revised Guide on Synthetic Data Generation (**SDG Guide**); and

(c) New use cases under the PET Sandbox operated by the Infocomm Media Development Authority (**IMDA**).

The FL Guide aims to help organisations understand Federated Learning (**FL**) techniques and use cases, and navigate practical considerations for exploring and implementing FL. It recommends an adoption roadmap (**Adoption Roadmap**) covering suitability and readiness assessments as well as key design considerations.

The FL Guide and SDG Guide are offered as resources within the PET Sandbox, which provides a safe environment for organisations to explore PETs on pilot use cases and gain regulatory clarity. The Guides are not intended to be authoritative statements of the law or substitutes for legal or other professional advice. They are living documents that will be updated from time to time to ensure their recommendations remain relevant.

FL Guide

What FL is

FL enables artificial intelligence (**AI**) model training across multiple data sources without requiring organisations to share or centralise their underlying data. By keeping input data local and sharing only model updates, FL allows organisations to collaborate securely, unlock the value of diverse datasets, and produce better AI models. FL is one of two complementary capabilities of the federated approach, the other being Federated Analytics, which enables distributed statistical analysis without centralising raw data and can serve as a precursor to FL. The FL Guide assumes that adopting organisations already meet baseline standards for data protection, cybersecurity, incident response and general AI governance. It focuses on the unique risks and controls introduced by FL's distributed and multi-party architecture.

When FL is suitable

The FL Guide advises that FL should be considered where it addresses genuine constraints or strategic needs. Where data can be feasibly centralised, or simpler PETs can meet the requirements, those conventional approaches are typically more efficient. The FL Guide identifies conditions where FL provides unique value, including:

(a) **Data cannot be centralised**: There is a clear and specific reason why data cannot or should not be centralised, such as regulatory restrictions, contractual obligations or operational constraints.

(b) **Data must remain at source**: Data is generated and must be used where it lives, such as on edge devices, within secure systems, or across national boundaries where cross-border data transfer would introduce unacceptable latency, security exposure or data integrity risks.

(c) **Data is a competitive asset**: Participants hold data that is competitively sensitive or represents a core business asset, such as proprietary trading patterns or clinical trial results.

(d) **Collaboration produces better models**: Training on diverse, cross-organisational data produces models that are statistically more robust, better generalised and higher performing than any model trained on a single participant's dataset alone.

For illustration, the largest FL effort to date connected 71 sites across six continents to train a consensus model on 6,314 patients for rare cancer boundary detection. Participating sites shared only model parameter updates, so that raw magnetic resonance imaging (MRI) scans never left the local institutions. The global consensus model achieved a 33% improvement in delineating the "tumour core" and a 23% improvement in the complete tumour extent compared to a model trained on publicly available data.

FL Adoption Roadmap

The FL Guide recommends a phased Adoption Roadmap that transitions from strategic alignment to operational planning and technical execution, spanning three core stages:

(a) Assess Suitability, which establishes strategic alignment;

(b) Assess Readiness, which focuses on operational planning; and

(c) Design and Configure, which delivers the technical execution.

Organisations should navigate these stages while keeping risk management as the central, connecting concern underpinning the entire roadmap.

Managing FL risks

While FL reduces data exposure risks by keeping raw datasets local during training, its distributed, multi-party architecture introduces distinct risks and coordination challenges that go beyond those of traditional single-organisation AI systems. There are four key risk areas:

(a) Private information leakage through model updates;

(b) Corruption of the global model through malicious or faulty participant contributions;

(c) Disruption of the training process through denial-of-service or Byzantine attacks; and

(d) Supply chain compromise through tampered software or model artefacts.

No single safeguard addresses all these concerns. Effective risk management requires layering multiple controls from the outset, running through the entire FL adoption roadmap.

SDG Guide

In its accompanying media release, the PDPC stated that it has also revised the SDG Guide, first issued in 2024, with the latest industry research, and that, unlike FL (which enables collaboration whilst keeping data stored locally), synthetic data (**SD**) creates artificial datasets mirroring the statistical properties of real data without containing individuals' personal information. The updated guide expands on new SDG methods and best practices to produce synthetic data and prevent re-identification, and includes new case studies on organisations that have successfully generated synthetic data for practical applications.

The SDG Guide describes SD as artificial data generated using a purpose-built mathematical model, algorithm or AI/Machine Learning model trained on a source dataset to mimic the characteristics and structure of the source data. Good quality SD can largely retain the statistical properties and patterns of the source data, so that analysis of SD can produce results similar to those yielded with source data. SD is not inherently risk-free, due to re-identification risks as personal information can still be leaked, owing to the SD's resemblance to the source data.

There are three common use case archetypes for SD:

(a) **Generating training datasets for AI models**:

  (i) Augmenting training datasets with synthetically generated labelled data can be more cost-effective, especially when the source datasets are sparse; and

  (ii) SD can be used to simulate rare events or augment under-represented groups in training AI models.

(b) **Data analysis and collaboration**:

  (i) SD can enable data sharing for analysis, especially in industries and sectors, e.g., healthcare, where source data can be sensitive; and

  (ii) Synthetic data can be used in data exploration, analysis and collaboration to provide a representative preview of the source data without exposing sensitive information. This enables stakeholders to gain assurance of the data quality before finalising any agreement or collaboration.

(c) **Software testing**: SD can be used instead of production data for software development and to help organisations avoid data breaches if the development environment is compromised.

To reduce the re-identification risks of synthetic tabular data, a five-step approach to generating SD is recommended: (a) know your data; (b) prepare your data; (c) generate SD; (d) assess re-identification risks; and (e) manage residual risks.

The SDG Guide explains that generally, the re-identification risk assessment for SD is an attack-based evaluation against four categories, namely singling out, linkability, attribute inference and membership inference. Where the assessed risk level is unacceptable, organisations should regenerate the SD and check it by: (a) removing outliers if such trends or insights are not necessary for business needs; (b) generalising granular data or adding noise; (c) performing data integrity checks by validating the data format, structures, and the like against the earlier documented data dictionary; (d) selecting relevant metrics that meet data objectives to measure data fidelity; and (e) selecting relevant performance metrics that meet data objectives to measure data utility. Residual re-identification risks may be mitigated through technical, contractual and governance measures.

New PET Sandbox Use Cases

IMDA is enhancing the PET Sandbox with new practical resources. To date, 11 organisations from sectors such as finance, healthcare, construction, transport, and advertising tech have successfully implemented PETs within the Sandbox, with more expected to join. Two new use cases illustrate PETs in practice:

(a) **Singapore General Hospital (SGH)**: SGH used a Trusted Execution Environment (**TEE**) solution to explore whether medical images can be analysed securely in the cloud. TEE is a secure, isolated, hardware-based environment that keeps sensitive data protected while it is being processed. By combining the computing power of cloud with the security of TEE, healthcare providers can process more medical images more quickly while protecting patients' sensitive medical information.

(b) **Ant International**: Ant International used Multi-Party Computing to verify transactions and evaluate risk control checks for payment requests and marketing discount eligibility checks. Sensitive data from collaborating merchant and wallet partners does not need to be collected or stored, while authorised parties can still perform risk controls.

If you would like information and/or assistance on the above or any other area of law, you may wish to contact the Partner at WongPartnership whom you normally work with or any of the following Partners:

LAM Chung Nian
Head – Intellectual Property, Technology & Data

Kylie PEH
Partner – Intellectual Property, Technology & Data

Connect with WongPartnership.

WPG MEMBERS AND OFFICES

- contactus@wongpartnership.com

SINGAPORE

-
WongPartnership LLP
12 Marina Boulevard Level 28
Marina Bay Financial Centre Tower 3
Singapore 018982
t +65 6416 8000
f +65 6532 5711/5722

CHINA

-
WongPartnership LLP
Shanghai Representative Office
Unit 1015 Link Square 1
222 Hubin Road
Shanghai 200021, PRC
t +86 21 6340 3131
f +86 21 6340 3315

INDONESIA

-
Makes & Partners Law Firm
Menara Batavia, 7th Floor
Jl. KH. Mas Mansyur Kav. 126
Jakarta 10220, Indonesia
t +62 21 574 7181
f +62 21 574 7180
w makeslaw.com

MALAYSIA

-
Cheang & Ariff
Advocates & Solicitors
Loke Mansion
273A, Jalan Medan Tuanku
50300 Kuala Lumpur
t +60 3 2691 0803
f +60 3 2693 4475
w cheangariff.com
-
Foong & Partners
Advocates & Solicitors
13-1, Menara 1MK, Kompleks 1 Mont' Kiara
No 1 Jalan Kiara, Mont' Kiara
50480 Kuala Lumpur, Malaysia
t +60 3 6419 0822
f +60 3 6419 0823
w foongpartners.com

MIDDLE EAST

-
Al Aidarous Advocates and Legal Consultants
Abdullah Al Mulla Building, Mezzanine Suite 02
39 Hameem Street (side street of Al Murroor Street)
Al Nahyan Camp Area
P.O. Box No. 71284
Abu Dhabi, UAE
t +971 2 6439 222
f +971 2 6349 229
w aidarous.com
-
Al Aidarous Advocates and Legal Consultants
Oberoi Centre, 13th Floor, Marasi Drive, Business Bay
P.O. Box No. 33299
Dubai, UAE
t +971 4 2828 000
f +971 4 2828 011

PHILIPPINES

-
Gruba Law
27/F 88 Corporate Center
141 Valero St., Salcedo Village
Makati City 1227, Philippines
t +63 2 889 6060
f +63 2 889 6066
w grubalaw.com