STEMPA
2025

DATA SHARING GUIDEBOOK

The Navigation Map for Sharing Research Data

Contents

Executive Summary

Public healthcare data in Singapore is governed by stringent legal, ethical, and institutional requirements that safeguard patient confidentiality and maintain public trust. Navigating these requirements can be challenging for research teams, particularly in multi-institution collaborations.

The **STEMPA Data Sharing Guidebook** serves as a clear and practical resource for responsible, compliant, and efficient data sharing.

It consolidates key governance principles, and illustrative considerations relevant to the preparation, anonymisation, transfer, storage, access, and disposal of public healthcare data for research purposes.

The guidebook's most significant contributions lie in three areas:

- Guidance on anonymisation and de-identification: Clear instructions aligned with MOH requirements, including the treatment of identifiers, residual risk mitigation, and practical application in research settings.
- Clarification of the approval workflow: An overview of how public healthcare institutions typically assess data-sharing requests, including circumstances under which lower-risk data may be subject to more streamlined review pathways, subject to institutional assessment.
- Illustrative case studies: Practical, scenario based examples from public healthcare research that demonstrate how multi-institution collaborations handle data anonymisation, aggregation, and governance considerations in practice.

Glossary

A	**Anonymisation**	A data treatment process that removes both direct and quasi-identifiers to minimise the re-identification risk.
C	**Chief Information Office (CIO)**	Institutional function that reviews DSRA, endorses security plans, and oversees IT risk management for data sharing projects.
D	**Data Access Provider (DAP)**	The data owner/provider within an institution that can verify the availability of requested datasets and support anonymisation or data sharing.
	Data Advisory Committee (DAC)	Committee in programmes/trust infrastructures that governs access to programme data/samples, authorship/acknowledgement, and publication review.
	Data Exchange Office (DXO)	National Healthcare Group department that evaluates data requests for compliance and routes them to the relevant governing offices/authorities.
	Data Protection Impact Assessment (DPIA)	Phase-1 assessment of the data request process: classification, legal/regulatory compliance, anonymisation sufficiency, managerial controls.
	Data Protection Officer (DPO)	Designated institutional function/personnel responsible for classification reviews, DPIA endorsement, incident escalation, and compliance with PDPA/MOH.
	Data Security Risk Assessment (DSRA)	Phase-2 technical assessment of infrastructure, transfer, and security risks; takes longer for identifiable data; anonymised data will experience an expedited process/skipped steps.
	Data Transfer Agreement (DTA)	Contract that defines purpose, scope, security, access, retention, overseas transfer obligations and incident reporting.
	De-Identification	Light processing of data coupled with safeguards to reduce, but not eliminate, the risk of re-identification; not all de-identified data qualifies as anonymised.
	Delta-Presence	An assessment of the risk that specific sensitive attributes – such as medical diagnoses or physical features – could result in re-identification, especially in small datasets.
E	**Encryption**	Approved cryptographic standards mentioned for secure transfer: ECC≥384-bit, RSA≥2048-bit, AES≥128/256-bit, SAFER SK-128.

F

Fast-Lane Approval — Expedited approval (typically ~6–8 weeks end-to-end), granted when data is anonymised and management controls meet MOH standards.

G

Genomic Data — Sequencing data (e.g. SNPs, STRs) and other biological modalities; treated with special anonymisation and managerial controls in PHIs.

Group Information Security Office (GISO) — Cluster-level security office that may conduct additional reviews for identifiable data and extended DSRA checks.

H

HealthTech Instruction Manual (HIM) — Regulatory framework for public healthcare institutions specifying data classification and IT/security requirements. Includes Data Access & Distribution (HIM-DA&D) and ICT Security Policy (HIM-ISP).

Health Insurance Portability and Accountability Act (HIPAA) Safe Harbor — The United States' standard listing of 18 identifiers to remove for de-identification of PHI; broadly aligned with MOH's 14 identifiers.

Human Biomedical Research Act (HBRA) — Singapore law (2015) regulating identifiable health data use in human biomedical research, ethics oversight and consent requirements.

I

Information Security RCST Framework — MOH framework complementary to ISF that governs security classifications (e.g. "Restricted [Classified]") and associated controls.

Information Sensitivity Framework (ISF) — MOH framework that defines sensitivity levels (e.g. "Unrestricted, Restricted Sensitive", Normal/High) used for classification and incident reporting.

Information Technology Office (ITO) — Institutional office that advises on secure transfer/storage, encryption standards, and technical risk mitigation.

Institutional Review Board (IRB) — Ethics review board that approves human-subject research protocols and amendments before data access/sharing requests proceed.

K

K-Anonymity — Risk test ensuring each record is indistinguishable from at least k others with respect to quasi-identifiers.

L

L-Diversity — Extension of k-anonymity, ensuring that sensitive attributes within an equivalence class have at least L diverse values.

M

Man-in-the-Middle (MITM) Attack — An interception attack risk during remote access/transfer; mitigated via secure networks and multi-factor authentication.

MOH — Singapore Ministry of Health; issues sector guidelines/circulars including the HealthTech Instruction Manual and Information Sensitivity Framework.

N	**Nosocomial Antimicrobial Resistance (AMR)**	Hospital-acquired transmission patterns of antimicrobial resistance studied via microbiology, clinical and genomic data.
O	**Ordering Unit (OU)**	Operational unit descriptor used in clinical/administrative records (e.g. Nursing OU Description, Doctor OU Description).
P	**Personal Data Protection Act (PDPA)**	Singapore law requiring legal basis and reasonable safeguards for the collection, use, and disclosure of personal data; governs consent and overseas transfer obligations.
	Pseudonymisation	Replacing direct identifiers with surrogate keys/codes to enable data linkage while reducing identifiability; still considered re-identifiable and typically requires stronger controls.
	Public Healthcare Institution (PHI)	Hospitals/clusters in Singapore's public healthcare system contributing/controlling health data used in research.
	Public Research Institution (PRI)	Government/public agency/institute that receives and analyses anonymised or de-identified data in collaborative projects.
Q	**Quasi-Identifiers**	Attributes (e.g. age band, gender, geography, timestamps) not directly identifying an individual but that can enable re-identification when combined.
R	**Re-Identification Key**	Secret mapping enabling linkage back to identities; typically held by TTP under strict managerial and contractual controls.
	Research Informatics and Computing (RIC)	Institutional team that reviews requests, supports anonymisation and data preparation, and performs technical evaluations for sharing.
S	**Short Tandem Repeats (STRs)**	DNA motifs used in genetic analyses; and in certain studies, are regarded as non-identifiers to reduce re-identification risk.
	Single Nucleotide Polymorphisms (SNPs)	Point DNA variations studied in genomics; recognised as non-identifiers in specific research contexts to reduce re-identification risk.
T	**Trusted Third Party (TTP)**	Institution-appointed party that generates/controls keys for pseudonymised data and may assist with anonymisation, risk assessments, and linkage workflows.

Introduction

Working with public healthcare data often requires navigating a range of legal, ethical, and institutional requirements across different organisations. Variability in institutional review requirements can create uncertainty for research teams as they prepare, review, and share data in collaborative or multi-institution projects.

This guidebook supports the broader **Shorten Time to Execute Multiparty Project Agreements (STEMPA)** initiative, which seeks to streamline collaborations by aligning key governance positions and reducing avoidable delays. As one of STEMPA's reference resources, the guidebook provides a common understanding of how public healthcare institutions typically approach datasharing governance and controls.

Intended for researchers, research offices, and data governance teams involved in the preparation, request, or management of public healthcare data, this guidebook outlines the typical data-sharing workflow. It highlights key considerations that may affect approval timelines, and provides practical guidance on preparing data for sharing and navigating institutional approval requirements. It also draws on established good practices across the public healthcare ecosystem to support consistent and effective data-sharing efforts.

For further technical or operational support, this guidebook includes contact details for the Data Protection Offices and Trusted Third Parties (TTPs) across public healthcare institutions.

Chapters are organised around the typical datasharing workflow:

Part 1: Legal and Regulatory Requirements in Data Sharing Between Public Health and Research Entities

Topic 1 Anonymisation and De-Identification of Healthcare Data – Key Standards and Considerations

Topic 2 What to consider when merging datasets and how this may affect project timelines.

Part 2: Processes of Preparing Data for Sharing, Getting Approval, and Managing Anonymised Data

Topic 3 The standard data approval process and what different stakeholders look out for.

Topic 4 How to manage data access, transfer, storage, and disposal in collaborative projects.

Topic 5 What to do if a data incident occurs, and how institutions typically respond.

This guide should be read in conjunction with any institutional resources or protocols you are required to adhere to.

PART 1

Legal and Regulatory Requirements in Data Sharing Between Public Health and Research Entities