NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

RSTF 2020 Roundtable Discussion: Managing Sensitive Data

Ms Goh Su Nee
Dr Amy Chou

23 Oct 2020

NTU Library

Office of Information, Knowledge and Library Services

Questions to consider

1. If a researcher unknowingly deposited sensitive data, would you be able to recognise it?
2. When a researcher claimed to have anonymised his/her dataset, would you know how to verify this? Have effective barriers been set up to ensure little/no risk of re-identification?
3. Should an anonymised dataset be given the same level of access as a non-sensitive dataset?

Types of sensitive research data

Does the data fall under any of the categories?

- Sensitive data is defined as information that needs to be protected against unauthorised access or unwarranted disclosure.
  - **Identifiable data**: Data that can be used to identify an individual, endangered species, object or location.
  - **Proprietary data**: Data that is internally generated and gives competitive advantage to its owner. This includes research data with commercialisation potential.
  - **Restricted or confidential data with contractual** (e.g. Research Collaboration Agreements, Non-Disclosure Agreements) **or legal obligations** (e.g. Official Secrets Act).

Source: NTU Library. Research Data Management: Sensitive data LibGuide.

Recognising sensitive research data

Does the data contain other sensitive information? Consider:

- Content/nature
  - Data of racial, ethnic, political, sexual, criminal origins
- Context
  - Sensitivity varies with time and across populations or subjects

Data anonymisation

If identifiers are present, can the identifiers be removed?

- Anonymisation
  - A state in which re-identification is no longer possible
  - Achieved by different methods or strategies
  - Removal of identifiers alone is insufficient (e.g. Massachusetts Group Insurance Commission Data Leak, AOL Search Data Leak)

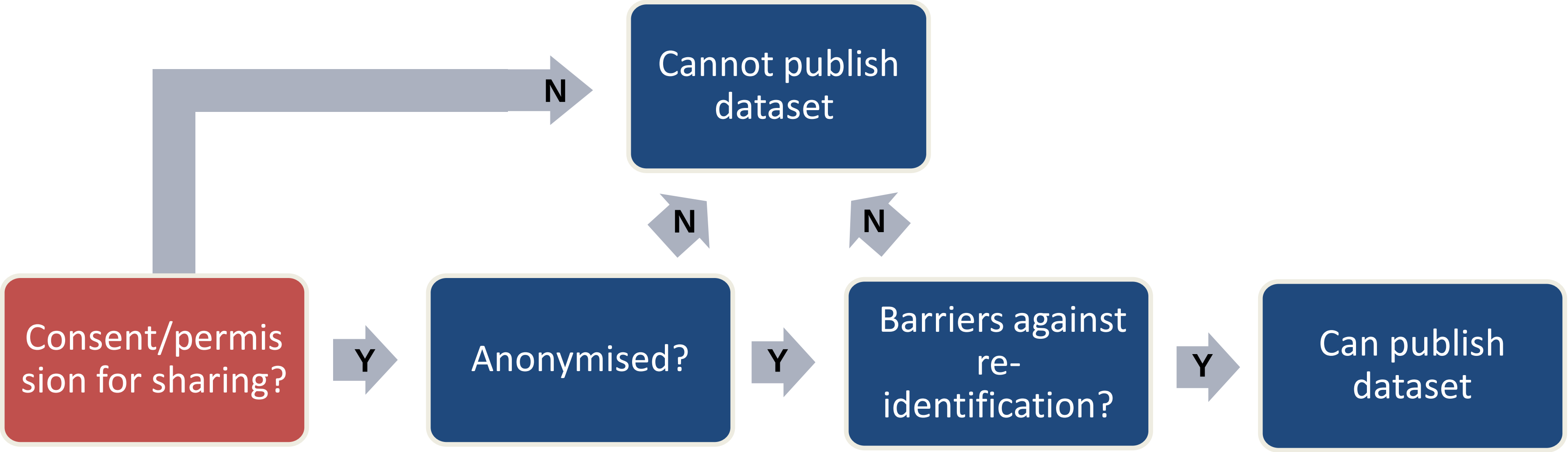
Sources: NTU Library. Data Anonymisation: What is it?; Kayaalp M. (2018) Modes of De-identification. AMIA Annu Symp Proc. 2017: 1044–1050.

Effective barriers

After de-identification, are there still risks of disclosure?

- Risk of disclosure depends on de-identification technique (e.g. removal of identifiers, pseudonymization)
- Effective barriers must be implemented, such that keys to re-identification are held by Trusted Third Parties only

Flow chart for sensitive data curation advice



Consent/permission for sharing?

Anonymised?

Barriers against re-identification?

Can publish dataset

Cannot publish dataset

Handling sensitive research data

- How to protect sensitive data
  - Access authorised data only
  - Work with data responsibly
  - Store data in appropriate places
  - Manage devices containing sensitive data
  - Report a breach or compromise of sensitive data
  - Get help if needed

Sources: Safe Computing, University of Michigan. Protect Sensitive Data; University of Michigan Library. Best practices for research data management.

POLL

Institutional Sensitive Research Data Management Readiness

Policies & Guidelines

- Alignment with other policies & guidelines
- Roles & responsibilities
- Awareness

Tools & Environment

- Platforms
- Tools

Education & User Services

- User training
- Data curation support

NTU

NTU Research Data Policy[1]
NTU Research Data Classification[2]

Developing a data governance policy
Developing classification-based handling guidelines for research data

RESEARCH DATA LIFECYCLE APPROACH[3]

BEFORE	DURING	AFTER
NTU DMP Tool[4]	NTU Electronic lab notebook	DR-NTU (Data)[5]

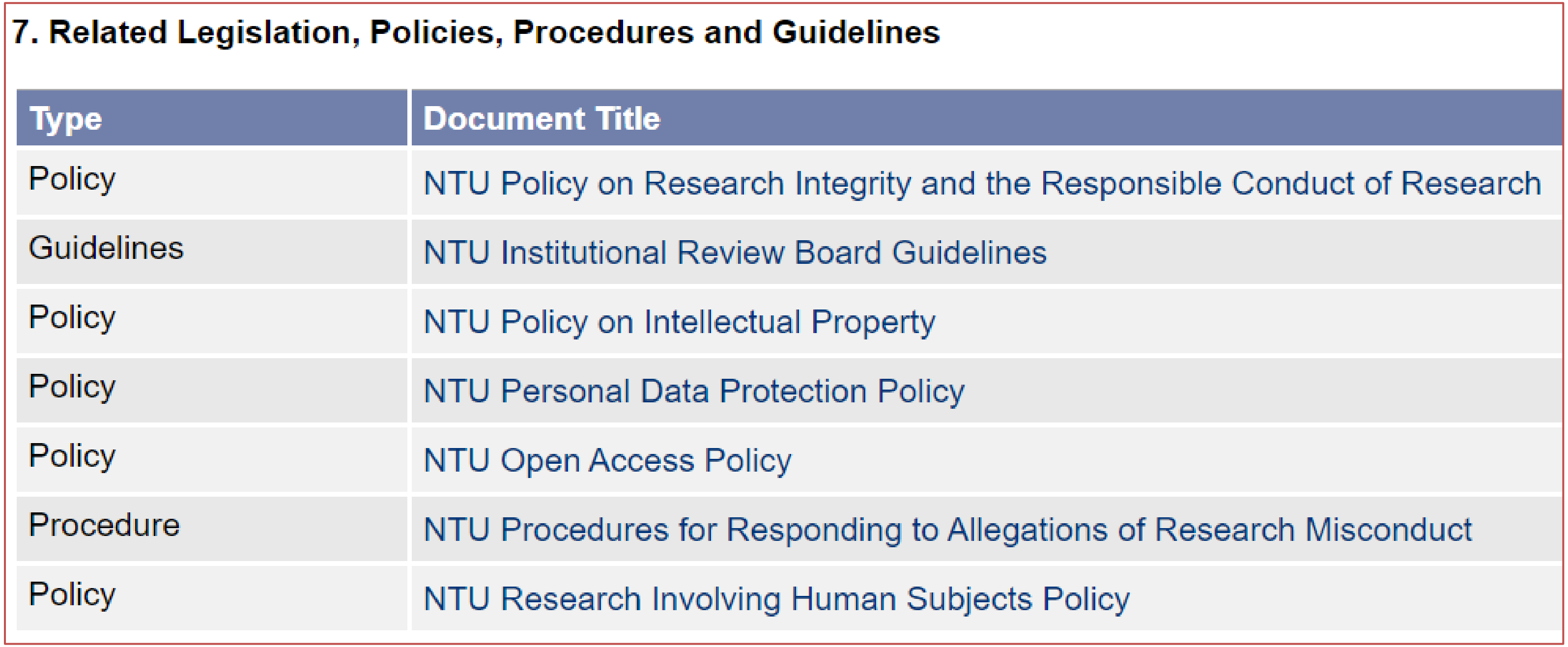
Exploring a separate centrally managed research data storage system for research data that cannot be shared
Exploring enterprise level encryption tools

- Data management planning workshops[6]
- DR-NTU (Data) workshops[6]
- Sensitive data management online guide[7]
- Data sharing curation advice

Developing in-house e-learning modules for faculty and students that will include sensitive data management

[1]NTU Research Data Policy: https://libguides.ntu.edu.sg/rdm/researchdatalifecycle
[2]NTU Research Data Classification: https://libguides.ntu.edu.sg/rdm/dataclassification
[3]Research data lifecycle: https://libguides.ntu.edu.sg/rdm/researchdatalifecycle
[4]NTU DMP Tool: https://libguides.ntu.edu.sg/rdm/dmp
[5]DR-NTU (Data): https://researchdata.ntu.edu.sg/
[6]Workshops: https://libguides.ntu.edu.sg/rdm/workshops/
[7]Sensitive data management online guide: https://libguides.ntu.edu.sg/rdm/workshops/

NTU Research Data Policy – related polices & guidelines



7. Related Legislation, Policies, Procedures and Guidelines

Type	Document Title
Policy	NTU Policy on Research Integrity and the Responsible Conduct of Research
Guidelines	NTU Institutional Review Board Guidelines
Policy	NTU Policy on Intellectual Property
Policy	NTU Personal Data Protection Policy
Policy	NTU Open Access Policy
Procedure	NTU Procedures for Responding to Allegations of Research Misconduct
Policy	NTU Research Involving Human Subjects Policy

[1] NTU Research Data Policy at https://libguides.ntu.edu.sg/rdm/researchdatalifecycle

NTU Research Data Classification

	Unclassified/Open (Level 1)	**Restricted (Level 2)**	**Confidential (Level 3)**	**Highly Confidential (Level 4)**
Sensitivity	*Low or no sensitivity*	*Moderate*	*Moderate high*	*High*
Examples of research data	- Public domain information - Research data in publications made available on OA platforms - Published research data in open repositories - Published IP and related information	- Manuscript drafts - In-progress or unpublished research data (not classified under Level 3 or 4) - Unpublished IP and related information	- Identifiable personal data# - De-identified data which can be made identifiable through various means* - Research data governed by research contracts or relevant legal agreements	- Research data under Official Secrets Act, or government data classified as Secret

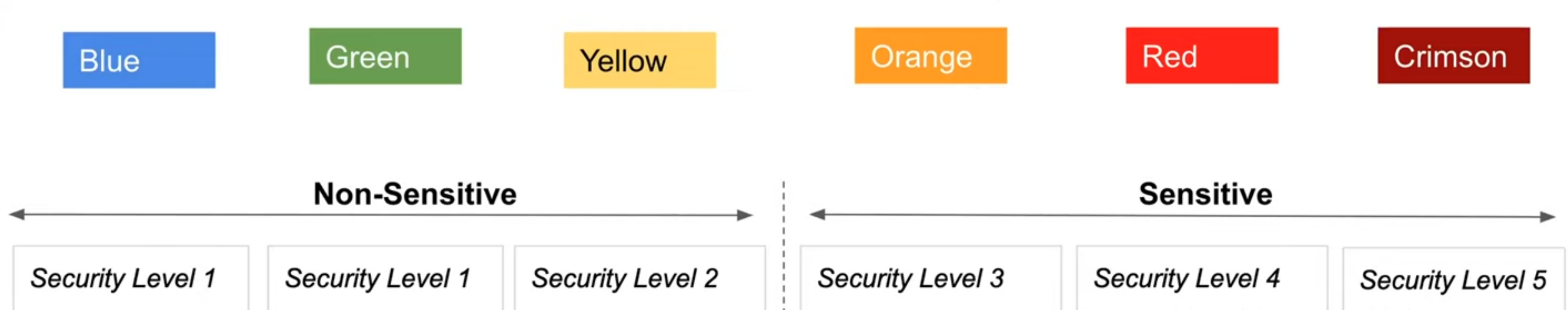
Each institution might have its own classification schemes – so do find out

#http://research.ntu.edu.sg/rieo/IRB/Guidelines/Pages/Personal-data.aspx

*Irreversibly de-identified personal data (i.e. linkages permanently deleted) can be re-classified as Level 2 (if unpublished) or Level 1 (if published).

Source: https://libguides.ntu.edu.sg/rdm/dataclassification

Dataverse’s upcoming functionalities



Non-Sensitive			Sensitive		
Blue	Green	Yellow	Orange	Red	Crimson
Security Level 1	*Security Level 1*	*Security Level 2*	*Security Level 3*	*Security Level 4*	*Security Level 5*
Publicly open	Publicly open	Restricted	Moderate sensitive	High sensitive	Maximum sensitive
	Publicly open Need to register to access	Restricted Need to be granted permissions, but non-sensitive	Moderate sensitive Requires Data Use Agreement (DUA), requires data enclave	Moderate sensitive Requires DUA, stricter security requirements and audits	Only metadata and no link to data, data stored outside network

Adapted from: Supporting Sensitive Data in Dataverse (Dataverse Community Meeting 2020), [video], last accessed 9 Oct 2020.

See also: DataSecurity Levels -Research Data Examples Quick ReferenceGuide, Harvard University, [webpage] created 22 Apr 2020, last accessed 12 Oct 2020.

Scenarios

As a repository manager, what would you do if someone deposited:

1. Interview transcripts involving human subjects.
2. Research data for PhD thesis or paper that has not been published in journals.
3. Research data bound by contractual or legal agreements that prevent open data sharing.

CLOSING

Useful resources

- MANTRA. The University of Edinburgh. Protecting sensitive data. [Online course]. CC-BY.
- University of Bristol. Sensitive research data bootcamp: Introduction. [Online course].
- Safe Computing. University of Michigan. Sensitive Data Guide to IT Services.
- Hiom, D., Gray, S., Steer, D., Merrett, K., Snow, K., & Beckles, Z. (2017). Introducing Safe Access to Sensitive Data at the University of Bristol. International Journal of Digital Curation 12(2), 26–36.
- ANDS Guides. Australian National Data Service. (2018). Publishing and sharing sensitive data.
- The President & Fellows of Harvard College. (2015). Sample Data Usage Agreement.
- Campbell-Dollaghan, K. (2018, Dec. 10). Sorry, your data can still be identified even if it's anonymized. [Magazine article] From Fast Company.