UNIVERSITY OF TORONTO

Data Classification Standard: Research guide

Level	Definition	Explanation and examples
Level 4	Non-public data that the University has designated as level 4. Level 4 data requires substantially greater protection measures than confidential data.	**Highly sensitive** research data[1], requiring **stronger security controls**, whose unauthorized access, disclosure, or loss poses **significant** financial, reputational, legal or physical risk to the data subject, researcher, University, etc. **Examples (not exhaustive):** • Personal health information (PHI). • Research data subject to export controls or the Controlled Goods Program. • Personal data from the European Union **classified** as "extra sensitive" under the General Data Protection Regulation (GDPR). • Information that, if disclosed, could place **data subjects or researchers** at risk of foreseeable physical, psychological, social, financial or legal harm. • Research data with confirmed dual-use potential. • Research data requiring **stronger** security controls by partners, funding agencies, the Research Ethics Board (REB), legislation or regulations.
Level 3	Non-public data that contains personal information (as defined by Freedom of Information and Protection of Privacy Act [FIPPA] for which appropriate permission to disclose has not been received) and other data that the University has designated as being level 3.	**Sensitive** research data, requiring **strong security controls**, whose unauthorized access, disclosure or loss poses **some** (non-minimal) financial, reputational or legal risk to the data subject, researcher, University, etc. **Examples (not exhaustive):** • Administrative records or data used for research purposes whose original data classification was level 3 (e.g., education/student records, employee records, other FIPPA-covered data). • Potentially identifiable information related to human subject data, including (de-identified) genomic data that can be re-identified using publicly available data. • Personal data from the EU **not classified** as "extra sensitive" under GDPR. • Collections of variables or indirectly identifiable information that, when merged, becomes sensitive. • Research data requiring **strong** security controls by partners, funding agencies, REB, legislation or regulations.

UNIVERSITY OF TORONTO

Level	Definition	Explanation and examples
Level 2	Data the University has not chosen to make public but has not been designated by the University as being in another level.	**Non-public but non-sensitive** research data; most active research data is at least level 2 prior to publication. **Examples (not exhaustive):** • Most active and/or unpublished research and intellectual property that is not already classified as level 3 or 4. • Published research data under embargo. • Research data which is REB-exempt and/or has no contractual obligations for additional protections. • Anonymous information (e.g., survey) where no identifiers were collected. • Anonymized, de-identified or coded information, which is not PHI-related, where all directly identifiable information has been obfuscated, and the risk of (unauthorized) re-identification is low or very low. ○ **Note:** The code/data keys for the purposes of re-linkage are classified at the same level as the original, uncoded data.
Level 1	Data available for broad or general open view.	**Publicly available**. **Examples (not exhaustive):** • Publicly available data or datasets. • Published research data not subject to embargo or beyond embargo period. • Open-source software source code. • Identifiable information which the data subject explicitly consented to make publicly available or has no expectation for privacy.

[1] University of Toronto Institutional Research Data Management Strategy (utoronto.ca)

UNIVERSITY OF TORONTO

Data classification decision tool for research

My research data includes…	Data classification	
Personal health information (PHI).[2]	**Level 4**	
Data subject to export controls or the Controlled Goods Program.[3]	**Level 4**	
Other sensitive research areas and data, including but not limited to[4]: • Confirmed dual-use (military, intelligence or dual military/civilian applications) potentiality. • Biological agent and toxin biosecurity, including security sensitive biological agents (SSBA). • National security/strategic implications.	**Level 4**	
Personal data classified as "extra sensitive" or similar under General Data Protection Regulation (GDPR) or equivalent privacy legislation.	**Level 4**	
Identifiable human subjects' data: • Directly identifiable information. • De-identified data that can be re-identified or linked using publicly available data. • Collections/constellations of variables or indirectly identifiable information that, when merged, becomes sensitive.	**Level 3**	
Administrative records or data used for research purposes: • Student records. • Employee records. • General-purpose emails and business records.	**Level 3**	
Data classified as confidential[5] or sensitive by partners (data use agreements), funding agencies, research ethics boards, legislation, regulations or the researcher. The following is a non-exhaustive list of considerations that can help parse out the degree of risk or potentiality of harm present within one's research data: • Vulnerability of the individual or community from which the data originates. • Social and cultural norms, wherein disclosure of controversial or stigmatized behaviour would be concerning or harmful to the individual's wellbeing. • Local laws and geopolitical situations, wherein disclosure of information would be concerning or harmful to the individual's wellbeing. • Likelihood that nation state, criminal or other malicious groups or individuals might want to steal, halt, destroy or alter research data. • The financial, reputational/social, psychological, behavioural, legal, and/or physical risk, impact or harm that an unauthorized disclosure might cause to the data subject, community or researcher. • The volume of data stored, wherein the scale of information which could be affected by a possible unauthorized disclosure requires additional security controls to limit risk. • The data subjects' ability to provide consent to the use of their data for research purposes.	**Level 3**	**Level 4**

UNIVERSITY OF TORONTO

My research data includes…	Data classification
Non-identifiable human subjects' data (non-PHI): • De-identified information (e.g., anonymized and/or coded information). ○ **Note:** The code or data keys for purposes of re-linkage are classified at the same level as the original, uncoded data. • Anonymous information where no identifiers were collected.	**Level 2**
Most active and/or unpublished research and intellectual property, by default (unless otherwise classified).	**Level 2**
Published research data under embargo by publisher or other body.	**Level 2**
Published research not subject to embargo or beyond embargo period.	**Level 1**
Publicly available data and datasets.	**Level 1**
Unpublished research and intellectual property (not otherwise classified), which the Principal Investigator (PI) wishes to be made generally accessible.	**Level 1**

[2] Personal Health Information Protection Act, 2004, S.O. 2004, c. 3, Sched. A (ontario.ca)
[3] Export Controls & Controlled Goods Program (utoronto.ca)
[4] National Security Guidelines for Research Partnerships (science.gc.ca)
[5] University of Toronto Institutional Research Data Management Strategy (utoronto.ca)