GUIDELINES FOR CATEGORIZATION AND SECURITY OF RESEARCH DATA & INFORMATION
(Final)

UNIVERSITY of GUELPH

CHANGING LIVES
IMPROVING LIFE

POLICY TITLE:	**Guidelines for Categorization and Security of Research Data & Information**
POLICY #:	CIO-ITSecurity – 03.2
Initial Draft By - Position / Date:	Genevieve Gauthier, Privacy Officer; Gerrit Bos IT Security Officer / Sep.10, 2012
APPROVED BY / DATE:	**Rebecca Graham, CIO; John Livernois, AVP Research Services / April 25, 2014**
Last Revised Date:	**Final Draft – December 08, 2013**
Next Review Date:	One year after Approval (May 2015)
BRIEF DESCRIPTION:	This document describes guidelines for categorization and security of research data. It is intended to be able to serve as a blueprint for a more general data classification policy or guideline. It is expected to be upgraded to a policy during the May 2015 review.

CONTENTS

1. INTRODUCTION AND PURPOSE

Confidential information[1] collected and used during the course of research must be protected from loss, destruction or unauthorized access. The following information security categorization is offered as a guideline for researchers on the treatment of Research Data/Information ("research data"). Information security categories are defined below, along with their application to the research domain. Information is categorized

[1] Confidentiality may arise from law, policy, practice, contract or agreement and includes but is not limited to personal or third party data, or information provided implicitly or explicitly in confidence.

according to the level of risk[2] to research participants, researchers, sponsors of research and the University, should data security be breached.

Users of these guidelines should note that it is both the content of the information, as well as the expectation[3] of those to whom the information relates or from whom the information was obtained that determines the category of information and level of associated risk of harm to individuals, sponsors of research and/or the institution from unauthorized disclosure.

2. SCOPE AND APPLICATION

These guidelines apply to all users of research data (faculty, staff, students and visiting scholars) as well as any third-party agents who are authorized to access research data within the custody or control of the University of Guelph[4]. The use of these guidelines will assist Research Ethics Boards, researchers, users and recipients of research data in handling and organizing data based on access and security needs.[5]

3. DEFINITIONS

Cloud Environment: A configuration where multiple distributed networked computers store information.

Electronic Data: Information that is stored, transmitted or read in an electronic format such as a file on a drive or device or information in a database.

Encryption: Involves transforming information into an unintelligible format that can only be made legible using an authorized key or password[6].

Financial Information: Information about an individual's or an organization's financial matters, such as income, expenses, banking and credit information.

Hardcopy Data: Is information that is stored and read in a physical format such as a paper file or a book.

Harm: Anything that has a negative effect on the welfare of an individual or organization; the nature of the harm may be social, behavioural, psychological, physical, or financial[7].

Identifiable Individual: An individual who could reasonably be identified, directly or indirectly through personal information; or an individual who can be reasonably identified by linking previously separate groups of information.

Personal Health Information (PHI): Information about an identifiable individual and related to their health or health care history including but not limited to medical history, details of visits to health-care practitioners, and test results.

Personal Information /Personally Identifiable Information: Recorded information about an identifiable individual including but not limited to the individual's name, age, race, sex, address, SIN or other identifying numbers, financial history, education and/or employment history, personal opinions, and more[8]. Personal

[2] As noted in the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, 2010 (TCPS2); Chapter 5 "Privacy and Confidentiality"*.

[3] Expectation as to how the information will be used, based on notice provided to participants or the circumstances under which the information is provided

[4] See Appendix for detail on data safeguards for third parties

[5] Users should note the application of security controls as outlined herein does not ensure access to data outside the scope of these guidelines

[6] Refer to the *University of Guelph End-Point Encryption Policy*. Also see the *Appendix* in this document.

[7] As referenced in the *TCPS2*

[8] In the province of Ontario and for the purposes of this guideline, "personal information" is as defined in the *Freedom of Information and Protection of Privacy Act (FIPPA)*

information ranges in sensitivity, becoming more sensitive in accordance with the risk and/or harm that may ensue as a result of unauthorized release or disclosure of the information.

Research Data/Information: Information collected, obtained and used during the course of research. Includes original data, previously existing data sets (secondary use), as well as the analysis, results, or dissemination resulting from the research process.

***Security*:** Enables the protection of assets or property, including information or data, through the application of physical, technical, and/or administrative safeguards.

***Sensitive Information/Data*:** Includes but is not limited to personally identifiable information[9], and must be defined according to context and the expectation[10] of the individual or entity to whom the information relates. For example, the names and addresses of subscribers to a College mailing list would generally not be considered sensitive information. However, the names and addresses of subscribers to a specific research interest group may be considered sensitive. Federal and provincial legislation, as well as contractual obligations and agreements may also specify data elements that require protection from unauthorized creation, access, modification and/or deletion.[11]

4. INFORMATION CATEGORIZATION

The categorization system presented in these guidelines is comprised of components that work together to assist the researcher in assessing the data to determine the appropriate security controls for use, storage and destruction of research data. *Table A* outlines the definitions and examples (i.e. research application) of each of the three types or categories of data.

TABLE A—INFORMATION CATEGORIES

INFORMATION CATEGORY	DEFINITION	RESEARCH EXAMPLES *(non-exhaustive)*
Public (Type 1)	• Information deemed to be public by government legislation and/or University policy • Information in the public domain	• All information has been gathered from within the public domain (e.g. publicly available data published by Statistics Canada) • Information is of a non-personal nature (and for which there is no expectation or obligation of confidentiality)

[9] As defined in the *University of Guelph End-point Encryption Policy*.
[10] An understanding of how personal information will be used should be expressly defined for the individual to whom the information relates through a *Notice of Collection.*
[11] As defined in the *University of Guelph End-point Encryption Policy*.

INFORMATION CATEGORY	DEFINITION	RESEARCH EXAMPLES *(non-exhaustive)*
Internal /Private (Type 2)	• Information not approved for circulation outside of the University of Guelph • Loss would inconvenience the researcher, organization or sponsor of research • Unauthorized disclosure could result in moderate risk to an individual, sponsors of research, the University or its affiliates, but is unlikely to result in financial loss or serious damage to credibility	• Data are anonymous and collected without identifiers (e.g. anonymous surveys) • Data are anonymized[12] and all information that could be used directly or indirectly to identify an individual has been permanently removed or modified. No code exists for future identification or re-linkage • Data are coded with direct identifiers removed and replaced with a code[13] • Information about an identifiable individual, the disclosure of which would not reasonably be expected to cause material harm to the individual, but for which confidentiality has been assured • Identifiable information that is generally available to the public (e.g. name and address only) • Information about an individual where the subject would not reasonably have an expectation of confidentiality (e.g. classroom presentation) • Information obtained from a sponsor or developed during the course of research that by law or contract is "confidential information", the disclosure of which would not reasonably be expected to cause the sponsor to incur financial loss or competitive disadvantage
Confidential /Sensitive (Type 3)	• Information is protected by government legislation (e.g. FIPPA, PIPEDA, PHIPA, Controlled Goods Regulations, etc.) • Information is not otherwise protected by legislation but is protected by contractual agreements or as part of participant's consent agreement • Information is only available to authorized persons • Unauthorized disclosure could result in a significant level of risk to an individual, sponsors of research, the University or affiliates and cause serious financial impact or damage to reputation of the University, affiliates or sponsors of research	• Information about an identifiable individual, that could be damaging to a person's reputation or lead to embarrassment if disclosed • Highly sensitive information about an identifiable individual that could cause a non-minimal risk of harm if disclosed • Information provided by a sponsor or developed during the course of research that by law or contract is "confidential information", the disclosure of which could cause the sponsor to incur financial loss or competitive disadvantage

[12] See also TCPS2 (Ch.5) definition of "anonymized information"

[13] Code/data keys for the purpose of identifying/tracking individuals in research should be categorized and protected in accordance with the level of security required if the data were not coded.

5. RISK LEVEL ASSESSMENT

In order to implement the proper information-security controls for research data, the researcher must assess the *level of risk*[14] associated with unauthorized release or disclosure of the research data as well as the *category/type*[15] of information. The level of risk is assessed according to the magnitude of harm and the probability that this harm will occur[16] should the research data be lost, stolen or accessed by unauthorized parties.

The magnitude of harm should be considered from the perspective of the research participants or sponsors of research and should take into account the nature and sensitivity of the information, as well as relevant legal or contractual obligations, or regulatory requirements. The probability of harm to participants or sponsors of research should be reasonably estimated. The impact on the researcher from unauthorized loss, release or disclosure of research data should also be considered, both from a reputational and regulatory perspective and as some data would be inconvenient, difficult, or impossible to replace if lost or compromised.

Table B indicates the minimum level of risk relative to the probability and magnitude of harm[17]. A designation of Low Risk should only be used for instances where the magnitude and probability of potential harm is no greater than that which the participants or research sponsors could reasonably be expected to encounter in, respectively, their daily lives or day-to-day business operations.

TABLE B—DATA RISK ASSESSMENT

		MAGNITUDE OF HARM		
		MINIMAL	***MODERATE***	***SUBSTANTIAL***
PROBABILITY OF HARM	***MINIMAL***	LOW RISK	LOW RISK	MEDIUM RISK
	MODERATE	LOW RISK	MEDIUM RISK	HIGH RISK
	SUBSTANTIAL	MEDIUM RISK	HIGH RISK	HIGH RISK

6. SECURITY CONTROLS

The Information Category (Table A) and Risk Level (Table B) of research data determine the Security Level required for the treatment of the research data. *Table C* includes examples (non-exhaustive) of the kind of research data that could be classified in each Information Category and at each Risk Level and is coded to indicate the appropriate Security Level for that research data.

Table D indicates the appropriate safeguards for the storage, transmission, and disposal of research data based on its designated Security Level.

[14] See Table B
[15] See Table A
[16] As noted in the *TCPS2*.
[17] "Magnitude of harm" refers to the seriousness or intensity of harm to an individual or organization if a particular risk should occur. "Probability of harm" refers to the likelihood of an individual or organization actually suffering the relevant harm.

TABLE C—INFORMATION CATEGORY/RISK CROSS-REFERENCE

	LOW RISK	**MEDIUM RISK**	**HIGH RISK**
Type 1 **PUBLIC**	Business contact information; information on public record **(S-I)**	Information shared in a group that is of a non-personal nature, where the expectation of privacy of the participants is low **(S-I)**	Compiled information from many public sources which would be time-consuming or costly to re-compile if lost or destroyed[18] **(S-II)**
Type 2 **INTERNAL/ PRIVATE**		De-identified data which would be difficult, though not impossible, to re-identify/link Information shared in a group that is of a moderately personal nature[19] **(S-II)**	Individual voice or video recordings that, even if not highly sensitive in content, would be impractical or impossible to replace if lost or destroyed[20] **(S-III)**
Type 3 **CONFIDENTIAL/ SENSITIVE**		Identified data about a highly sensitive topic that could cause embarrassment/psychological harm if released Student records including grades, opinion material, examples of work **(S-III)**	Identified data about a highly controversial topic that could put participants at risk if released SIN numbers, medical, criminal or employment history **(S-III)**

[18] That is, the risk of harm is to the researcher, as opposed to individuals who may have a risk of harm

[19] Depending on the expectation of privacy of the participants (e.g. opinions expressed regarding a non-controversial topic)

[20] And where there is an expectation of confidentiality

TABLE D—INFORMATION SECURITY LEVEL KEY AND REQUIREMENTS

	STORAGE *Holding of data in either electronic or hardcopy format*	**TRANSMISSION** *Transfer of data, usually refers to electronic format*	**DESTRUCTION** *Eradication of data so it may not be recovered*
SECURITY LEVEL-I (S-I)	No security controls required for data storage or transmission.		Files may be recycled or deleted.
SECURITY LEVEL-II (S-II)	Electronic files and/or data should be stored on a University-sponsored shared directory with controlled access. Electronic files and/or data should be encrypted when stored on portable or insecure devices[21]. Data should not be stored in a "cloud" environment unless hosted by the University of Guelph or supported by suitable agreements. Portable or insecure devices should be stored in a secured location (i.e. where access is limited) when not in use. Hardcopy files should be stored in a locked office or file cabinet.	Data should be transmitted via secure wired, wireless, cellular or other network (encrypted, secure wifi) Transmission by fax – machine should have limited access and only those authorized can view, or recipient must agree only authorized person will be present when material is received.	Electronic files and/or data should be formally removed and media should be reformatted. Hardcopy files should be shredded.

[21] Using the centrally provided encryption solution; refer to *University of Guelph End-point Encryption Policy* (http://www.uoguelph.ca/cio/content/end-point-encryption-policy); also see the *Appendix* for alternative solutions

	STORAGE *Holding of data in either electronic or hardcopy format*	**TRANSMISSION** *Transfer of data, usually refers to electronic format*	**DESTRUCTION** *Eradication of data so it may not be recovered*
SECURITY LEVEL-III (S-III)	Electronic files and/or data must be stored on a University-sponsored shared directory or stationary device (i.e. desktop computer or server) with controlled physical access. Electronic files and/or data must be encrypted when stored on portable or insecure devices.[22] Data must not be stored in a "cloud" environment unless hosted by the University of Guelph or supported by suitable agreements. Portable or insecure devices must be stored in a secured location when not in use. Hardcopy files must be stored in a locked office or file cabinet with controlled access (on University premises when possible to do so).	Data must be transmitted via secure wired, wireless, cellular or other network (SSL, SSH encrypted, secure wifi) Transmission by fax – machine should have limited access and only those authorized can view, or recipient must agree only authorized person will be present when material is received.	Electronic files and/or data and media must be degaussed (magnetic information wiped). Devices may be physically destroyed. Hardcopy files must be cross-cut shredded.

7. RELATED UNIVERSITY OF GUELPH GUIDELINES AND POLICIES

- Office of Registrarial Services Policy on the Release of Student Information
- Protection of Privacy and Access to Information at the University of Guelph
- End-Point Encryption Policy
- Acceptable Use Policy
- Enterprise Guideline – Management and Disposal of IT Assets
- REB Guideline 1-G-002 – Secondary Use of Data
- REB Guideline 1-G-018 – Establishing Risk Level
- REB Guideline 4-G-007 – Release of Anonymized or Anonymous Data
- REB Guideline 4-G-011 – Disposal of Research Data
- REB Guideline 4-G-016 – Data Retention
- REB Guideline 4-G-015 – Creating Databases of Personal Information

[22] Using the centrally provided encryption solution; refer to *University of Guelph End-point Encryption Policy* (http://www.uoguelph.ca/cio/content/end-point-encryption-policy)

8. SOURCES

- Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (2010): http://www.pre.ethics.gc.ca/pdf/eng/tcps2/TCPS_2_FINAL_Web.pdf
- Harvard University Information Security and Privacy: http://www.security.harvard.edu/
- University of Toronto Data Security Standards for Personally Identifiable and Other Confidential Data in Research: http://www.research.utoronto.ca/ethics/pdf/human/nonspecific/datasecurity.pdf
- Canadian Standards Association Privacy Code: http://www.csa.ca/cm/ca/en/privacy-code
- Carnegie Mellon Guidelines for Data Classification: http://www.cmu.edu/iso/governance/guidelines/data-classification.html
- Thompson Rivers University Information Classification Standard: http://www.tru.ca/its/infosecurity/Standards/Information_Classification_Standards.html
- Government of Alberta Privacy Taxonomy: http://www.ipc.on.ca/images/Resources/up-PPPP062.pdf
- Government of Alberta Information Security Classification: https://www.rimp.gov.ab.ca/publications/pdf/infosecurityclassification.pdf
- Federal Information Processing Standards Publication on Standards for Security Categorization of Federal Information and Information Systems: http://csrc.nist.gov/publications/fips/fips199/FIPS-PUB-199-final.pdf
- University of Calgary Information Security Classification Standard: http://www.ucalgary.ca/it_files/ea/Information%20Security%20Classification%20Standard.pdf
- University of Western Ontario Guidelines for Data Classification

APPENDIX – ADDITIONAL PRACTICAL CONSIDERATIONS

Encryption

The University of Guelph Encryption Service is a Whole Disk Encryption solution based on the SecureDoc product from WinMagic Inc., a Canadian company. Whole disk encryption ensures that all information is encrypted, including temporary files, memory files, and print files. The service logs activity, and thus provides independent proactive verification that a device is encrypted. The service includes "key escrow" service, so that if a key or password is lost or no longer available but the device is, the information can be recovered. If encryption is required (i.e. data "must" be encrypted in accordance with Security Level III requirements), the central encryption solution must be used.

If encryption is recommended or desired (i.e. data "should" be encrypted, in accordance with Security Level II requirements), but not required and the central encryption solution is not suitable, several alternatives are available. These include Microsoft Bitlocker, Apple FileVault, or TrueCrypt among others. It is important to note that these alternatives are not centrally supported, and may not have key escrow service or independent verification of encryption status.

Cloud Environments

On-campus alternatives to cloud environments include the Central File Service, central or departmental servers and purpose-bought storage systems with secure or limited access.

Cloud environment solutions from outside the institution may be considered provided the appropriate measures are in place which will provide the necessary level of data protection, and that such measures are supported by suitable agreements between the University and the vendors providing the cloud environment solutions. In such cases, data custodianship responsibilities will be clearly delineated in the agreements.

Third Parties

Consistent with responsibilities for the protection of research data that may arise out of law, contract or agreement, the University shall ensure appropriate measures are in place for the protection of research data within its custody and control, and offers a variety of solutions to achieve a satisfactory level of data protection.

In the case of research conducted in collaboration with third parties who are also users of research data, it may not be practicable to implement institutional solutions. In such circumstances, alternative solutions may be considered, provided the same level of protection shall apply and such measures are expressed as part of the research agreement or contract.

N:\Access & Privacy\Research-related\REB & Data Classification\Categorization & Security of Research Data_draft v2.07.docx