UNIVERSITY OF SOUTHERN MAINE
Office of Research Integrity & Outreach

Procedure #:	HIPAA-002; HRPP-020
AAHRPP:	Element II.3.D. & Element II.3.E.
Date Adopted:	04/11/2019
Last Updated:	04/16/2020
Prepared By:	Kelly J. Stevens, Regulatory Compliance Administrator
Reviewed By:	IRB Chair; IRB; ORIO; Cutler Institute
Procedure Title:	Anonymous and De-identified Data

1.0 Objective

1.1. To describe the policies and procedures in accordance with anonymous and de-identified data, the general process by which de-identified information is created, and the options available for performing de-identification for the protection of human subject research covered by the University of Southern Maine (USM) Human Research Protection Program (HRPP).

2.0 General Description

2.1. Federal regulations require IRBs to determine the adequacy of provisions to protect the privacy of subjects and to maintain the confidentiality of their data. The University of Southern Maine is committed to ensuring compliance with the privacy and security safeguards set forth under federal regulations to include the Department of Health and Human Services, Office of Human Research Protections, Food and Drug Administration as well as the Health Insurance Portability and Accountability Act (HIPAA).

2.2. The process of de-identification, by which identifiers are removed from health information, mitigates privacy risks to individuals and thereby supports the secondary use of data for comparative effectiveness studies, policy assessment, life sciences research, and other endeavors[1]. Because of the benefit of the use of health information even when it is not individually identifiable, 45 CFR§164.502(d) of the HIPAA Privacy Rule permits a covered entity or its business associate to create information that is not individually identifiable by following the de-identification standard and implementation specifications in 45 CFR 164.514(a)-(b).

[1] **Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule**

2.3. The Principal Investigator (PI) is responsible for ensuring that research data is secure when it is collected, stored, transmitted, or shared. All members of the research team should receive appropriate training about securing and safeguarding research data. This SOP also applies regardless of the source of funding for the research.

3.0 Definitions

3.1. Anonymous: The information obtained is recorded in such a manner that human subjects cannot be identified, directly or through identifiers linked to the subjects. [45 CFR 46.104(d)(2)(i)].

3.2. Coded Data: Identifying information (such as name) that would enable the investigator to readily ascertain the identity of the individual to whom the private information or specimens pertain has been replaced with a code (number, letter, symbol, or any combination) and a key to decipher the code exists, enabling linkage of the identifying information to the private information or specimens. If an individual electronic database file contains subject numbers as the only identifiers, that particular file has been de-identified. However, if an investigator maintains a code linking the subject numbers to identifiers, the research study itself is not completely de-identified (anonymous) as long as the investigators maintain the ability to identify an individual.

3.3. Covered Entity: means one of the following:

3.3.1. A health plan

3.3.2. A health care clearinghouse

3.3.3. A health care provider who transmits any health information in electronic form in connection with a covered transaction

3.4. Data Encryption: Encryption is the conversion of data into a form, through use of an algorithm, which cannot be easily understood by unauthorized people.

3.5. De-Identified Data: Data from which an investigator or others are not able to determine the identity of any particular individual. De-identified data cannot contain the following identifiers:

3.5.1. Names;

3.5.2. Street address (city, state and 5-digit zip codes are acceptable). GPS coordinates are considered identifiers unless geographic masking techniques are used that prevent re-identification;

3.5.3. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except

that such ages and elements may be aggregated into a single category of age 90 or older;

3.5.4. Telephone numbers;

3.5.5. Fax numbers;

3.5.6. Electronic mail addresses;

3.5.7. Social Security numbers;

3.5.8. Medical record numbers;

3.5.9. Health plan beneficiary numbers;

3.5.10. Account numbers;

3.5.11. Certificate/license numbers;

3.5.12. Vehicle identifiers and serial numbers, including license plate numbers;

3.5.13. Device identifiers and serial numbers;

3.5.14. Web Universal Resource Locators (URLs);

3.5.15. Internet Protocol (IP) address numbers;

3.5.16. Biometric identifiers, including finger and voice prints;

3.5.17. Full-face photographic images and any comparable images; and

3.5.18. Any other unique identifying number, characteristic, or code.
[45 CFR 164.514(b)(2)]

3.6. Identifiable Biospecimen: A biospecimen for which the identity of the subject is or may readily be ascertained by the investigator or associated with the biospecimen [45 CFR 46.102(e)(6)]

3.7. Identifiable Private Information: Private information for which the identity of the subject is or may readily be ascertained by the investigator or associated with the information [45 CFR 46.102 (e)(5)]

3.8. **Individually Identifiable Health Information:** Information that is a subset of health information, including demographic information collected from an individual, and:

3.8.1. Is created or received by a health care provider, health plan, employer, or health care clearinghouse; and

3.8.2. Relates to the past, present, or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual; and

3.8.2.1. That identifies the individual; or
3.8.2.2. With respect to which there is a reasonable basis to believe the information can be used to identify the individual. [45 §160.103]

3.9. Limited Data Set: A limited data set is protected health information that excludes the following direct identifiers of the individual or of relatives, employers, or household members of the individual:

3.9.1. Names;

3.9.2. Postal address information, other than town or city, state, and zip code;

3.9.3. Telephone numbers;

3.9.4. Fax numbers;

3.9.5. Electronic mail addresses;

3.9.6. Social Security numbers;

3.9.7. Medical record numbers;

3.9.8. Health plan beneficiary numbers;

3.9.9. Account numbers;

3.9.10. Certificate/license numbers;

3.9.11. Vehicle identifiers and serial numbers, including license plate numbers;

3.9.12. Device identifiers and serial numbers;

3.9.13. Web Universal Resource Locators (URLs);

3.9.14. Internet Protocol (IP) address numbers;

3.9.15. Biometric identifiers, including finger and voice prints; and

3.9.16. Full-face photographic images and any comparable images.
[45 CFR 164.514 (e)(2)]

3.10. Private Information: Includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information that has been provided for specific purposes by an individual and that the individual can reasonably expect will not be made public [45 CFR 46.102 (e)(4)]

3.11. Protected Health Information (PHI) means individually identifiable health information:

3.11.1. Except as provided in (2) of this definition, that is:

3.11.1.1. Transmitted by electronic media;
3.11.1.2. Maintained in any medium described in the definition of electronic media; or
3.11.1.3. Transmitted or maintained in any other form or medium

3.11.2. PHI excludes individually identifiable health information in:

3.11.2.1. Education records covered by FERPA (20 U.S.C. 1232g);
3.11.2.2. Records on a student who is eighteen (18) years of age or older, or is attending an institution of postsecondary education, which are made or maintained by a physician, psychiatrist, psychologist, or other recognized professional or paraprofessional acting in his professional or paraprofessional capacity or assisting in that capacity, and which are made, maintained, or used only in connection with the provision of treatment to the student, and are not available to anyone other than persons providing such treatment, except that such records can be personally reviewed by a physician or other appropriate professional of the student's choice;
3.11.2.3. Employment records held by a Covered Entity in its role as employer; and
3.11.2.4. Regarding a person who has been deceased for more than fifty (50) years.

3.12. Sensitive Research Data: Data is considered sensitive when disclosure of identifying information could have adverse consequences for subjects or damage their financial standing, employability, insurability, or reputation.

3.13. Standard De-identification: Health information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information. [45 CFR §164.514 (a)]

4.0 Responsibility

4.1. Execution of this SOP: IRB Chair, IRB Members, Office of Research Integrity and Outreach (ORIO) Staff, Research Compliance Administrator (RCA), Principal Investigator (PI), Study Personnel (SP).

4.2. This SOP does not supersede more stringent requirements for the handling and storage of protected information that may be required by other UMS or USM policies and procedures, such as those governing staff of the Cutler Institute in its Cutler Institutes Protected Information Policies and Procedures manual.

5.0 Use of Public Databases

5.1. The use of a publically-available database for research analysis generally does not require IRB review. An IRB application is required for public databases when any of the following conditions apply:

5.1.1. Investigators seek to merge or enhance data sets in such a way that individuals might be identified.

5.1.2. Use of restricted data sets.

5.1.3. Use of databases, which require IRB approval prior to access.

5.1.4. If submitting a proposal to obtain the use of a data set is required, an application to the IRB for review and approval is also required.

5.1.5. If the investigator seeks to obtain additional information from the database owner.

5.1.6. A data use agreement is involved.

6.0 Procedure

6.1. Determine if protected health information from a covered entity is identifiable, de-identified, coded, or anonymous. If an individual electronic database file contains subject numbers as the only identifiers, that particular file has been de-identified. However, if an investigator maintains a code linking the subject numbers to identifiers, the research study itself is not completely de- identified (anonymous), as long as the investigators maintain the ability to identify an individual. Anonymous means that no one should be able to link an individual person to the responses of that person, including the investigator.

6.2. If data collected is protected health information, the use of a Data Use Agreement is required; [45 CFR 164.514 (b)(2)(e)(2)].

6.3. Secure data encryption must be used if identifiable information is:

6.3.1. Stored on a networked computer or device, either on campus or off campus;

6.3.2. Transmitted over a network; and/or

6.3.3. Stored on a removable medium (e.g., laptop computer or a USB flash drive);

6.4. Determine if the level of security necessary is relative to the risk posed to the subject, should personally identifiable information be inadvertently disclosed or released. In an effort to ensure best practice, it is always desirable to have a high level of security rather than to risk operating at a minimal standard.

6.5. Not all research data sets can reasonably be de-identified (for example, in a video or audio-recorded interview, the subject may be readily identifiable). In this case, the original research data set must be considered personally identifiable and treated accordingly.

6.6. When a study involves greater than minimal risk, investigators are encouraged to consult with appropriate information administrators and technology and security experts within their system to develop appropriate data security plans.

6.6.1. Specifically, investigators should:

6.6.1.1. Collect the minimum identity data needed. Identifiers should only be collected if they serve a legitimate purpose in the context of the research.
6.6.1.2. De-identify data as soon as possible after collection and/or separate data elements into a data set and an identity-only data set.
6.6.1.3. Coded data and identity-only data should always be stored separately in a secure location.

6.7. Limit access to personally identifiable information. The opportunity for human error should be reduced through:

6.7.1. Limiting the number of people (both users and administrators) with access to the data and ensuring their expertise and trustworthiness; and/or

6.7.2. Using automatic (embedded) security measures (such as storing data in secure data-encrypted form) that are professionally installed and administered. If the computer is connected to the campus network or to the public Internet, the professional administrator of the computer shall ensure that it complies with all minimum standards for network and data security listed below.

6.8. When identifiable information is stored on a personal, business or a university-owned or maintained computer, investigators are strongly encouraged to ensure that the computer be professionally administered and managed. If this is not possible,

investigators should disclose such, and provide the IRB with a plan for how the sensitive data will otherwise be secured.

6.9. Secure data encryption must be used if identifiable information is:

6.9.1. Stored on a networked computer or device, either on campus or off-campus;

6.9.2. Transmitted over a network; and/or stored on a removable medium (e.g., laptop computer or a USB flash drive).

7.0 Roles and Responsibilities

7.1. The IRB has the authority to decide if the security plan to protect subjects' confidentiality or anonymity appears acceptable.

7.2. Principal Investigator and Research Team are responsible for:

7.2.1. Disclosing the nature of the confidential data they collect in their study protocol so that the IRB can assess the data security risk;

7.2.2. Preparing data security plans and procedures in accordance with the appropriate security category requirements;

7.2.3. Implementing and monitoring the data security plans and procedures over the course of the project;

7.2.4. Principal Investigators are responsible for ensuring data under the research team's control remains secure;

7.2.5. Using appropriate safeguards to maintain the confidentiality, integrity, and availability of data that is collected, used, shared and/or stored for research purposes, including Protected Health Information (PHI);

7.2.6. Identifying all on-site and off-site research personnel who have or need access to research data in any form and ensuring they employ appropriate safeguards and follow all university policies regarding access to data;

7.2.7. Ensuring that for human subject research, the IRB application appropriately explains the safeguards used to protect the data, including the data source (i.e., the types of records that are used to gather the data) and the data recording/collection method; and

7.2.8. Immediately reporting any suspected or known security breaches that compromise research data to the Research Compliance Administrator.

8.0 References

8.1. OHRP Guidance on Research Using Coded Private Information or Specimens (2008)
8.2. Accountability Act of 1996 (HIPAA) Privacy Rule
8.3. 45 CFR § 46.102
8.4. 45 CFR § 160.103
8.5. 45 CFR § 46.111
8.6. 45 CFR § 164.502
8.7. 45 CFR §164.514
8.8. FERPA 20 U.S.C. 1232g