P&P: GA 106 Version No: 2.1 Effective Date:10/23/2015	RESEARCH DATA SECURITY: PROTECTING HUMAN SUBJECTS' IDENTIFIABLE DATA	Supersedes: Security of Research Subjects' Personally Identifiable Data Held by Researchers 9/2005

1. POLICY

People who volunteer to participate as subjects in research do so with the understanding that the researcher(s) will protect their identity and the information obtained about them from inadvertent or inappropriate disclosure. The principle that the Institutional Review Board (IRB) upholds in assessing the benefits and risks of the research is expressed in the Belmont Report as "beneficence"— to maximize possible benefits and minimize possible harms to human subjects. As benefits and risks may be reflected in protections for privacy and confidentiality, all human subject research protocols must have in place an acceptable and documented procedure for the protection of identifiable and/or confidential information before the protocol will be approved, amended, granted continuing approval, or determined exempt from IRB review.

The purpose of this policy is to delineate the requirements for appropriate data security measures to protect the identity of and/or confidential information obtained about individuals who participate as subjects in research.

1.1 Definitions

1.1.1 A *human research data set* constitutes a body of informational elements, facts, and statistics about a living individual obtained for research purposes. This includes information collected by an investigator through intervention/interaction with the individual or identifiable private information obtained without intervention/interaction with the individual.

1.1.2 *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 which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (e.g., a medical or school record).

1.1.3 *Identifiable Information* means information that can be linked to specific individuals either directly or indirectly through coding systems, or when characteristics of the information are such that by their nature a reasonably knowledgeable and determined person could ascertain the identities of individuals.

1.1.4 Personal identifiers are any data elements that singly or in combination could be used to identify an individual, such as a social security number, name, street address, demographic information (e.g., combining gender, race, job, and

location), student identification numbers, or other identifiers (e.g., hospital patient numbers). Email addresses are often personally identifiable but not in all cases. Other data elements, such as Internet IP addresses, have varying degrees of potential for identifying individuals, depending on context. These elements require consideration as to whether they should be treated as personal identifiers.

1.1.5 *A de-identified data set* refers to data that has been stripped of all elements or combinations of elements (including, but not limited to, personal identifiers and coding systems) that might enable a reasonably knowledgeable and determined person to deduce the identity of the subject. For example, while not directly identifiable, a dataset may include enough information to identify an individual if elements in the dataset are combined.

1.1.6 *A coded data set* refers to data that has been stripped of identifiers and assigned an identity code (typically a randomly generated number) which is associated with and unique to each specific individual; the code can be used to link data elements to the identity-only data set. This identity code should not offer any clue as to the identity of an individual.

1.1.7 An *identity-only data set* contains any and all personal identifiers absolutely necessary for future conduct of the research and the key to the identity code that can be used to link or merge personal identifiers with the coded set.

1.1.8 *Secure location* refers to a place (room, file cabinet, etc.) for storing a removable medium, device, computer, or equipment wherein reside data sets with personal identifiers to which only the principal investigator has access through lock and key. (Either physical or electronic keys are acceptable). Access may be provided to other parties with a legitimate need in the context of the research, consistent with the policies below and as disclosed in the research protocol.

1.1.9 *Secure data encryption* refers to the algorithmic transformation of a data set to an unrecognizable form from which the original data set or any part thereof can be recovered only with knowledge of a secret decryption key of suitable length, and using a suitable algorithm. (Refer to the Human Research Data Risk Assessment Matrix for additional information).

1.2 Specific Policies

The level of security necessary is relative to the risk posed to the subject should personally identifiable information be inadvertently disclosed or released as a result of malfeasance. 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. The IRB has the authority to decide if the security plan to protect subjects' confidentiality or anonymity appears acceptable. For data that retains identifiers, the protocol must describe adequate administrative, physical, and technical safeguards. When a study involves greater than minimal risk, investigators are encouraged to consult with appropriate information

technology and security experts such as their system administrators to develop appropriate data security plans.
Specifically, investigators should:

1.2.1 Collect the minimum identity data needed. Identifiers should only be collected if they serve a legitimate purpose in the context of the research.

1.2.2 De-identify data as soon as possible after collection and/or separate data elements into a coded data set and an identity-only data set. Coded data and identity-only data should always be stored separately in a secure location. Raw identity data should be destroyed whenever possible in accordance with UCOP Record Retention Policies (see below).

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.

1.2.3 Secure data encryption must be used if identifiable information is: (1) stored on a networked computer or device, either on campus or off-campus; (2) transmitted over a network; and/or (3) stored on a removable medium (e.g., laptop computer or a USB flash drive).

1.2.4 Limit access to personally identifiable information. The opportunity for human error should be reduced through: a) limiting the number of people (both users and administrators) with access to the data and ensuring their expertise and trustworthiness; and/or b) using automatic (embedded) security measures (such as storing data on non-volatile medium only in secure data-encrypted form) that are professionally installed and administered. If this 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.

1.2.5 When identifiable information is stored in personal or university-owned or -maintained computer, investigators are strongly encouraged to ensure that this 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.

1.3 Related Policies

- Minimum Security Standards for Electronic Information and Security of Networked Devices, among other policies, can be found on the Berkeley Security web page.
- UCOP Records Retention Policies
- Public Requests for Research Records

2. SCOPE

This policy applies to all human subject research reviewed by CPHS and conducted by or under the auspices of University of California Berkeley (UCB) faculty, graduate students, postdoctoral scholars, other affiliated researchers (investigators) or research conducted using UCB resources. This policy also applies to research conducted under an inter-institutional agreement (IIA) for which UCB is the reviewing IRB. The pertinent information or data containing personally identifiable information may be (or has been) collected or stored in any form such as electronic, digital, paper, audio or video tape. This information or data may be stored within computers or equipment that is privately owned, university-owned or -maintained or reside on removable electronic media, in either case located on university premises or elsewhere.

3. RESPONSIBILITY

Investigators are responsible for:

- Disclosing the nature of the confidential data they collect in their study protocol so that the IRB can assess the data security risk;
- Preparing data security plans and procedures in accordance with the appropriate security category requirements;
- Implementing and monitoring the data security plans and procedures over the course of the project.

Information Services and Technology (IST), or other centralized campus technology experts, can assist investigators to implement appropriate security measures for their research.

The UC Berkeley Research Data Management (RDM) program provides consulting services and guidance to UCB researchers who have research data questions. The RDM consulting network can be reached by sending an email to researchdata@berkeley.edu.

OPHS staff are responsible for checking that the confidentiality measures described in the protocol are consistent with this policy.

The IRB has a regulatory responsibility to ensure the adequacy of an Investigator's provisions to maintain confidentiality of the data in human subjects research.

4. APPLICABLE REGULATIONS

4.1 Regulations

4.1.1 The Office for Human Research Protections (OHRP) currently does not specify data security protections but instead requires IRBs to determine, when appropriate, that there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data.

4.1.2 The California State Committee for the Protection of Human Subjects (CPHS) mandates specific data security requirements for researchers who are requesting personally identifiable data from state agencies. These requirements include certification

from PIs and the campus Chief Information Officer that applicable data security controls are implemented effectively. Any UC Berkeley investigator requesting personally identifiable data from a state agency should contact UC Berkeley's Information Security and Policy (ISP) for more information on how to comply with California State CPHS data security requirements.

Helpful Links:

California State CPHS Data Security Assessment Service

California State Committee for the Protection of Human Subjects (CPHS)

California State CPHS Data Security Requirements

California State CPHS Sample Data Security Letter