

Interim Guidance on Human Subjects Research Data Security

Revised 20 August 2008

This documentation is offered as interim guidance as the University revises both the current IT and HIPAA security policies and procedures. Research data is susceptible to alteration, loss and theft that poses both a risk to subjects as well as a risk of loss of valuable data for the research project itself. The increased use of portable devices and media to store research data has led to an increased potential for loss and theft of large data sets. To protect subject confidentiality, ensure the integrity and availability of research data, and to comply with state & federal regulatory requirements and University policy, the Yale IRBs and ITS Information Security Office require that investigators use the attached Data Classification Table and Security Matrix as described below.

Information and information system Classification:

The level of security necessary for a given data set depends on the nature of the data, with more sensitive data requiring more stringent security controls. The level of security is also dependent on the storage media and on the configuration of the computing device where the information is created, accessed, transmitted or received. Table 1 (below) provides examples of data types in a three tiered information classification scheme. The list is not intended to be all inclusive but provides examples of data in each category. Once the classification of the data has been determined, Table 2 should be referenced to determine what steps are necessary to provide reasonable data security depending on the computing system, media storage or transmission requirements.

Approaches to Data Security:

Coded/de-identified/anonymous data sets

The current best practice involves de-identifying the data sets. If no identifiers, as defined under HIPAA (see <http://hipaa.yale.edu/policies/forms/Form5039-DiIdentificationChecklist.pdf>), are needed then no further steps are necessary as a data loss would not disclose individually identifiable data. When identifiers are needed, the demographic information can be maintained as a separate data file with a code which links the demographics to the study data itself. The two files are then stored in separate locations such as the de-identified data on a laptop and the code stored in a secure location (e.g., an appropriately secured Yale server – an *above-threshold* ePHI system).

Encryption

By far, the most vulnerable data storage is on mobile devices and portable storage media, including laptops, PDA's, flash drives, CD's, etc. When individually identifiable data must be maintained on these devices and coding the data as described above is not feasible, the device should be equipped with encryption software such as [PGP](#). Encryption poses certain risks in and of itself including potential data loss through loss of the encryption key and data loss during installation (see 'University Endorsed Encryption Implementations' <http://www.yale.edu/ppdev/Procedures/its/endorsedencryption/EndorsementEncryptionImplementation.pdf>). It is recommended that encryption software be installed through Yale ITS so that appropriate steps can be taken to minimize these inadvertent data losses. Note that some systems, such as Linux do not currently have an

encryption solution that meets va.gov requirements and thus use of encryption of VA data on operating systems other than Windows or Mac OS X is not recommended. Encryption of data on servers or other file sharing systems is not currently supported at this time.

Above-Risk-Threshold Systems

The risk to the integrity and availability (as opposed to simply the risk of confidentiality), is increased if 1) the computing system creates, accesses, transmits or receives primary source data or 2) the system is configured to allow access by multiple people. Data can be stored on appropriately secured systems or servers on the Yale network that are registered as *above-threshold* ePHI systems (hipaa.yale.edu/security/sysadmin/). Access to the system or server should be restricted to appropriate research staff through unique ID and password combinations along with other security requirements. Files stored on the server can be further restricted with a separate application and/or database level ID-password requirement. Strong passwords should be implemented (see <http://www.yale.edu/ppdev/Guides/its/passwords.pdf>).

Paper Records

Study data maintained as paper documents should be stored under lock and key, preferably in a locked cabinet and/or in a locked room.

Data Transfer

Data exchange with other research personnel or collaborators located off-site can lead to data being under reduced security en route. Options to minimize data loss/disclosure include sending the data using the ITS 'Yale File Transfer Facility' (www.yale.edu/its/email/transfer.html), using secure file transfer software (SSH, sCP, sFTP), encrypting the data via email or on a disc, and sending it through a courier service such as Fedex. Consult Information Security staff for appropriate options for data exchange via email, web form or other modes of data and file exchange.

Allowing non-Yale personnel or collaborators access Yale systems

Sharing data with non-Yale research personnel or collaborators by authorizing them to access computers on the Yale network requires special security measures. Documentation of authorization, authentication, and the security of the remote systems sharing Yale data is key to good compliance practice. In most cases, a written agreement describing the data privacy and security obligations of the non-Yale user should be obtained prior to providing access to University systems.

End of Life Issues

Computing equipment or storage media that has been used to store research data must have any individually identifiable data removed prior to donation, disposal or recycling of the equipment. Simply deleting research files or reformatting disks does not adequately remove the data from the equipment. ITS can assist in ensuring that the equipment or media is appropriately "wiped" prior to disposal. Alternately, devices can be physically destroyed by smashing CD's or drilling holes in the hard drive to render any data inaccessible.

TABLE 1 – Interim Human Subjects Research Data Classification

- I. Restricted:** Protection of data is required by law (FERPA, HIPAA, PCI etc.,) or has been determined as such by the University (e.g., legal or contractual).
Examples:
- All identified health data, as defined in HIPAA,
 - All data originating at a Veterans Administration (VA) facility or collected in collaboration with the VA
 - Identifiable information collected, stored, processed, transmitted or used on behalf of HHS or any of its component organizations which is governed by a grant or contract specifying grantee or contractor security obligations
 - Any fully identifiable (name, street address, phone number, medical record number, social security number – real or scrambled) data on HIV status, mental health status, participant or family substance/alcohol abuse, genetic predisposition to disease, or criminal history unless data is considered available from public sources
 - Data sets including a participants full social security number, financial account number(s), insurance plan number(s), credit card numbers, bank account numbers
 - Data covered under a Certificate of Confidentiality
 - Student data subject to FERPA
- II. Sensitive:** University has a contractual obligation to protect the data or has been determined to be sensitive by the University (e.g., internal access only).
NOTE: assumption is that this data has been first reviewed and is not classified as 'restricted'; information that is not sensitive to disclose within the organization, but could be harmful if disclosed externally.
Access is limited to individuals who have a business need to know.
Examples:
- Linking codes used in studies collecting high sensitivity data when stored on separate device from coded data itself
 - Coded anonymous data sets when stored on separate device from linking code
 - Data from non-exempt studies which do not pose more than minimal risk to subjects.
 - Identified data obtained under an obligation to maintain confidentiality such as academic records, subject responses which could lead to social or emotional harms.
 - Fully identifiable data other than that noted in Section I above,
- III. Low Risk:** No regulatory or contractual requirements; information is authorized for release to public; information that if disclosed outside the

University; would not harm the organization, its faculty, staff, students, or business partners.

Examples:

- De-identified data with or without code when code is stored in accordance with moderate sensitivity data classification
- Data collected from studies deemed exempt by a Yale IRB
- Data consisting solely of publicly available information.

Table 2
Interim Data Security Requirements for Human Subjects Data
[See also Yale Policy 1600 Information and Information System Security
Responsibilities]

I. Low Risk Data

- a. All electronic devices are password protected with strong passwords (see <http://www.yale.edu/ppdev/Guides/its/passwords.pdf>)
- b. All portable media are physically secured when not in use either in a locked office or using lock-down cables.
- c. All paper and non-electronic copies of data are physically secured under lock and key when not in use.
- d. Servers must have access controls but may be departmentally maintained.
- e. Data may be transferred by unprotected e-mail.
- f. Electronic devices may be disposed of following deletion of files or disposal of documents in regular trash

II. Sensitive Data

- a. Requirements a through d for low risk data apply
- b. Servers, whether centrally or locally maintained, which contain PHI must be registered in the University System Inventory Data base (<http://hipaa.yale.edu/security/sysadmin/>)
- c. Desktops, devices and copies of data must be physically secured including locked offices and/or locked facilities with access restricted to study personnel and their guests.
- d. Data may be transferred only through password protected e-mail attachments or via a courier service for portable media such as CD's.
- e. Devices must undergo secure deletion of the disc at the end of life of the device or prior to recycling. Contact ITS for further information on secure deletion.
- f. Paper and other non-electronic copies must be shredded when no longer needed.
- g. Data should be routinely backed up and the back up copy physically secured.

III. Restricted Data

- a. Requirements a through c for low risk data apply
- b. Requirements b and d through f for sensitive data apply
- c. Servers may be centrally maintained by ITS or locally but must be registered in the University System Inventory Data base (<http://hipaa.yale.edu/security/sysadmin/>) and appropriately secured
- d. Portable devices (laptops, CD's, flashdrives, etc) must contain only de-identified data or be encrypted.

- e. Electronic devices must be set to automatically log-off and lock after 15 minutes of inactivity.
- f. Data may only be transferred electronically in an encrypted state such as encrypted e-mail attachments or encrypted CD's or through web-based secure file transfer.