Confidentiality and Data Security Guidelines for Electronic Data
University of Kentucky (UK) Institutional Review Board (IRB)

Current research practices routinely involve electronic data in a variety of ways. Traditionally provisions for the confidential handling of research materials referenced keeping them in locked file cabinets and in locked offices. These provisions still have application when the research materials are in hard copy form. However, there is growing complexity in how to protect research data that are collected and maintained electronically. This guideline applies to the major types of electronic data collection and data maintenance, recognizing that advances in device design, software, and university systems are constantly changing. However, certain basic protection provisions are described which are expected of investigators.

With the anticipated increase in use of electronic devices including portable devices and drives, as well as web-based survey tools, data security needs greater attention on the part of investigators. The wide range of diversity in studies, methods, and electronic data devices means that investigators need to evaluate confidentiality and data security when electronic data is collected and/or stored.

The guideline applies to all studies involving electronic data that are participant-identified and that include information of a personal or health nature. This can include even low to minimal risk studies **if** the information is personal or health related. Thus, personally identified low-risk surveys on smoking and drug use would be included. Studies that would be excluded would be those with no identifying information or, if identified, cover topics that are non-personal such as marketing surveys about product use or political interests. In this guideline two terms are used for personal information: (1) personal health information or PHI consistent with HIPAA concepts; and (2) personal identifying information or PII for other than HIPAA-related studies – see the definitions section below for clarification.

In reviewing the guidelines below, investigators should also consult their departmental IT staff for assistance in applying the recommended data security steps. Departments vary in the ways in which they oversee and advise investigators about specific security procedures.

A set of basic definitions is at the end of this document.

I. Data security guidance for investigators who collect, use and store electronic data

1. The recommended electronic devices for entering and storing human subjects data are secure servers or stand-alone PCs that have encryption software for all PHI or other identifying data.
    a. Stand-alone PCs can be used for data storage of de-identified data without encryption, but with password protections over the use of the PC.
    b. Server-based PHI or other identified human subjects data should be behind a firewall and be encrypted.
    c. Anonymous data or de-identified data that cannot be tracked back to a person, using cue information in the data set matched to other data sources, can be stored on servers without encryption, but still would require authorized password access.

2**. Laptop data collection devices**. Laptops may be issued by sponsors for specialized projects. These laptops likely have a high degree of security built in. However, sponsor provided laptop devices should meet the same criteria as stated below. Laptops can be approved for data collection of human subject data when the following are provided:
    a. The device uses software that encrypts all personal health information or other identifying information.
    b. The data are formatted such that PHI or other identifying data are in separate files or tables from any clinical or research information about the persons.
    c. All files are password protected in addition to the laptop having a password.
    d. Laptops can be used for anonymous data collection without encryption.
    e. Laptops can be used for storing and analyzing de-identified data on human subjects.

3. **Jump drives** are only to be used under the following conditions:
    a. The jump drive uses files that have software to automatically encrypt all personal health information or other identifying information or the entire jump drive is encrypted.
    b. The data are formatted such that PHI or other identifying data are in separate files or tables from any clinical or research information about the persons.
    c. All files are password protected.
    d. Jump drives can be used for storing and analyzing de-identified data on human subjects.

4. **Web-based data entry**.
    a. Web-based PHI or other identifying data should be into a secure web server (https) and the server should encrypt any PHI or other identifiers upon submission. The server should be behind a firewall.
    b. Web-based anonymous or de-identified data need not be encrypted. Firewall protections advised but not essential.
    c. Sponsor web-based data sets may require the use of security tokens to access files. Security tokens decrypt files and open them for investigators to do data entry. These devices are an acceptable device for web-based data entry.

5. **PDAs, I-PODs and Blackberry devices**.
    a. The device uses software that encrypts all personal health information or other identifying information.
    b. The data are formatted such that PHI or other identifying data are in separate files or tables from any clinical or research information about the persons.

6. **CDs and DVDs**.
    a. PHI, PII or other identifying data should not be stored on CDs or DVDs unless the entire CD or DVD is encrypted.
    b. De-identified human subject data can be stored on CDs or DVDs in open format.

7. **Email**
    PHI or PII or other identifying data should not be contained in email communications that are sent outside of the UK Medical Center (MC) firewall.

II. Web Surveys

There is increasing interest in using web-based survey tools for research involving human subjects. There are two major forms of web-based surveys: (1) investigator-devised and programmed tools that are housed on university servers under the control of the investigator; and (2) independent proprietary survey programs that incorporate investigators' own measures but are data that reside on servers owned by the survey company. Investigators need to have clear assurances about the protections that are afforded by the independent proprietary providers, whereas those that are developed by the investigator can have total control in-house. While proprietary vendors of web-based survey tools are generally ethical, investigators should obtain information about the tool's security and privacy protections, including learning whether user IP addresses are captured and saved during completion of the surveys. Most vendors will 'scrub' IP addresses from the data at the investigators request. Some vendors also 'scrub' them upon submission of the completed form, but this should be clarified. Despite their stated privacy policies, many vendors, especially those who promote freeware, do in fact share IP addresses with their consortium of investors, and thus absolute anonymity cannot be guaranteed to survey respondents.

With either approach, the window of greatest vulnerability for data is during the time the program is open and being used by the participant. This is the point at which hacking could discover identity or other personal information. This is not unique to web-based research, but includes any period when a user is online. Investigators need to be assured that when any PHI or PII are being collected in web-based tools that once the data are transmitted, they are encrypted. Several things need to be considered with web-

based data collection. The informed consent forms or scripts used in lieu of consent forms will need to clarify the kinds of protections that are available to the web-using participant. Just as consent forms described research materials being in locked file cabinets when the data are in hard copy, the consent form should describe the specific web-based data security being used. If proprietary vendors are being used to collect the data, and if breach of confidentiality could put respondents at risk due to the nature of the survey questions, consent forms should spell out this possibility to potential respondents.

The following checklist has been developed to assist investigators in assessing their data security and protections. The list can provide guidance on what to include in protocol narratives and in consent forms.

Investigator checklist for data protections

TOPIC	Check Yes or No	GUIDANCE/RECOMMENDATIONS
DATA COLLECTION		
1. Will your study use personally identifiable information (PII) or personal health information (PHI) about participants?	___**Yes** ___ **No**	**If Yes then:** Generally, the IRB requires that research data be kept apart from PII or PHI. Separate tables or separate files should be used to maintain confidentiality of individual records. *Note this protection in the consent form*.
Will the PII or PHI be kept in separate files from research data on participants?	___**Yes** ___**No**	**If Yes then:** The crosswalk table shows the real PII or PHI along with a research record number. See Exhibit A below.
Will a research record number be used instead of PII or PHI in research data tables?	___**Yes** ___ **No**	**If Yes then:** The tables containing crosswalks between PII/PHI and research record numbers must be kept separately from any tables used for research. **Separate any PII or PHI from other research data. Create crosswalk, if needed, between participant record number and PII/PHI and ensure that the crosswalk file is separate from both the research data and the PII /PHI file. (See examples of table structures in Exhibit A below).**
2. Will the PII or PHI be obtained from existing electronic data systems (electronic medical records, institutional data sets?	__ **Yes** __ **No**	**If Yes then:** Ensure that applicable HIPAA or other authorizations have been approved for data access into a new electronic data table. **On receiving PII or PHI, separate these identifiers from other health or research data. Create crosswalk, if needed, between participant record number and PII/PHI and ensure that the crosswalk file is separate from both the research data and the PII /PHI file. (See example of table structures in Exhibit A below).**
3. Will data from participants be	__ **Yes**	**If Yes then:**

entered directly into electronic devices during research surveys or procedures?	__ **No**	Ensure that encryption is installed such that PII/PHI are always encrypted as they are entered, saved or submitted. The ideal is for the encryption to occur during entry, but it is acceptable to do this at "save" or "submit" functions (for VPN or other uploading). *Note this in the consent form*. An example of language for this might be: "All personal identifying information is encrypted as it is typed into the laptop." Or "All personal identifiers are encrypted when the data are uploaded."
4. Will data be loaded into a server system using a virtual private network (VPN)?	__ **Yes** __ **No**	**If Yes then:** The VPN-accessing server must be behind a firewall and all identifiers must be encrypted before being deposited in the VPN accessed or uploaded file or the data must be moved immediately into an encrypted file.
IF YES, THEN:		
Will you be using portable devices for data collection? (This includes Laptops, IPODs, IPADs, PDAs, Androids, Blackberries, etc.)	__ **Yes** ___**No**	**If Yes then:** Ensure that encryption is installed such that PII/PHI are always encrypted as they are entered, as they are saved or submitted. The ideal is for the encryption to occur during entry, but is acceptable at save or submit functions (for VPN or FTP or other uploading). There are two ways of encrypting data for portable devices: (1) Encrypting the entire laptop so that a password is needed to even open any operation of the device; and (2) encryption only of the specific file being used for research.
Are portable devices supplied by the sponsor?	__ **Yes** ___**No**	**If Yes then:** Ensure that the device comes with security keys or built-in encryption either for the specific files or for the entire device. Some sponsor-supplied devices do not use PII or PHI (research record number only) and rely on separate methods for transmitting these data. If PII or PHI are entered into or are stored on the device ensure that the keys are in place and/ or that if PII/PHI data are encrypted.
Are you planning to put data on small portable storage devices such as jump drives?	__ **Yes** ___**No**	**If Yes then:** **Ensure that PII or PHI are not stored on jump drives.** Other research data can be stored on jump drives as long as there is no way that the data could be traceable to a participant identity.
Is the device owned by the PI or by the university rather than being supplied	__ **Yes** __ **No**	**If Yes then:** Assume responsibility for maintaining two separate data tables (two files) and ensure that the PII or PHI

by a sponsor?		table is encrypted.
5. Are you using a university desktop PC or MAC for entering study data?	__ **Yes** ___**No**	**If Yes then:** Ensure that your device is behind the university firewall. Ensure that backups are to secure system servers or if an external hard drive is used for backups, ensure that it contains only encrypted PII or PHI.
Is the device a personally owned desktop PC or MAC?	__ **Yes** ___**No**	**If Yes then:** Do not have any PII or PHI stored on personal desktop devices. For all other research data on human subjects, ensure that a firewall is installed and turned on at all times.
6. Will web survey tools be used to collect data?	__ **Yes** __ **No**	**If Yes then:** If you have the capability and expertise, host the survey in-house on a secure university fire-walled , password-protected server
7. Will the survey be hosted on a commercial or independent proprietor's (external) server?	__ **Yes** ___**No**	**If Yes then:** Be sure to find out to what extent access to the server is limited, what protections are in place to protect the data against unauthorized access, and whether the data can be encrypted upon transmission.
Are survey questions of a sensitive nature such that a breach of confidentiality could put subjects at risk?	__ **Yes** __ **No**	**If Yes then:** Consent form should address the possibility of breach of confidentiality and that anonymity cannot be guaranteed.

EXHIBIT A – CROSSWALK EXAMPLE BETWEEN PII AND DATA TABLES

This table would show personal identifying information (PII) associated with the research record number and would need to be encrypted

PARTICIPANT CROSSWALK TABLE					
Participant ID Number	Participant Name	Address	Telephone number	SSN	DOB
10001	John Smith	403 Plum Street, Louisville, KY 40202	502-666-6666	555-55-5555	Dec-75
10002	Ophelia Doe	600 Sixth Street, Lexington, KY 40505	859-999-9999	666-66-6666	Nov-81
10003	Justin Tyme	100 Walnut Avenue, Novgorod, KY 40699	859-888-8888	111-11-1111	Oct-82
10004	Mary Laffer	26 Clown Avenue, Lexington, KY 40509	859-777-7777	999-99-9999	Sep-86

This table would actually contain the clinical or other research data and would not need to be encrypted

BASELINE DATA TABLE						
Participant number	gender	age	Variable 1	Variable 2	Variable 3	Variable 4
10001	M	35	2	2	5	11
10002	F	29	1	3	5	13
10003	M	28	2	3	4	15
10004	F	24	2	4	7	13

Definitions for electronic data collection:

Personal health information (PHI): This is defined by HIPAA law and includes personal identifiers that are associated with medical information other than patient/subject self-reported information that may pertain to health. Information/data from medical records are considered PHI.

Personal identifying information (PII): For the purposes of this policy, this includes information that identify a person including any or all of the following: (1) names; (2) social security numbers; (3) birth dates; (4) addresses; (5) IP addresses; (6) other data that could reasonably lead to discovering a personal identity.

Server: A server is a computer device with software that networks/links PCs and databases or web applications.

PC/Personal computer: A stand alone or networked computer as a desktop device.

Laptop: A portable computer that includes traditional laptops, netbooks and other portable computing devices that generally have full range PC capacities.

External drives: This includes everything from jump drives to external hard drives.

Compact disks and DVDs: Plastic disks for storing electronic data.

Security tokens: Jump drive-like devices that contain security codes or de-encryptions to allow access to secure web-based data sets.

PDAs, IPODs, IPADs Androids, and Blackberry-like devices: Portable devices that send and receive emails, text phone messages, or other communications and that include data entry and data storage capacity.

Virtual Private Networks (VPNs): The university allows access to selected drives and folders on university servers from remote locations using software provided by the UK phone system. With a VPN, a researcher can connect to files from off-site computers using either Ethernet or wireless connectivity. VPNs are password protected.

File Transfer Protocols (FTPs): FTPs are used to transfer data from off-site computers to main campus servers via web connections to specified server files.