Protection of Identification

Protecting the Rights & Privacy of Human Subjects

**

There are two basic tools to protect from disclosure of sensitive data and subjects' identities: Restricting information in the dataset, and restricting access to the data. Thus, data intended for broader use should be free of identifiers that would permit linkages to the research participants and free of content that would create unacceptably high risks of subject identification.

Stripping a dataset of items that could identify individual participants is referred to by several different terms, such as data redaction, de-identification of data[1]., and anonymizing data. It is rarely sufficient to simply remove names, addresses, telephone numbers, Social Security Numbers, and the like. Deductive disclosure of individual subjects becomes more likely when there are unusual characteristics or the joint occurrence of several unusual variables. Samples drawn from small geographic areas, rare populations, and linked datasets can present particular challenges to the protection of subjects' identities.

https://grants.nih.gov/grants/policy/data_sharing/data_sharing_workbook.pdf

Access & Control Measures of Data

Measures used to minimize the risk of breaching the confidentiality of data include the following:**

- Mandatory agreements to maintain confidentiality
- Data encryption
- Electronic firewalls and locked storage facilities,
- Password authentication of users
- Audit trails
- Disaster prevention and recovery plans
- Security measures for backup tapes.

Institutions and investigators should work closely to develop and update plans and procedures to protect the security of data.

https://grants.nih.gov/grants/policy/data_sharing/data_sharing_workbook.pdf

Modes of Access & Control of Data Sharing

Addressing NIH Data Sharing Policies

Lisa Federer, Research Informationist
UCLA Louise M. Darling Biomedical Library

**

From Flickr– Username: ichibod
From Wikimedia– Username: Dake

under the auspices of the PI

From Flickr– Username: Novartis AG

in a data archive

From Flickr– Username: loop_oh

in a data enclave

Mode Data Sharing – Data Archives

Data Archives

There are many archives for data. Many data archives facilitate the sharing of data using Web-based platforms. A searchable list of Websites for archives is available through the University of California at San Diego at http://odwin.ucsd.edu/idata/.

Most journals now expect that DNA and amino acid sequences that appear in articles will be submitted to a sequence database before publication. The **National Center for Biotechnology Information (NCBI)**, National Library of Medicine (NLM), NIH, was established in 1988 as a national resource for molecular biology information. NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information with the goal of improving understanding of molecular processes affecting human health and disease. NCBI provides timely and accurate processing and biological review of new entries and updates to existing entries, and is ready to assist authors who have new data to submit. For more information about submitting and downloading data, see the NCBI Website at http://www.ncbi.nlm.nih.gov/Genbank/index.html

The National Center for Chronic Disease Prevention and Health Promotion at CDC operates the **Youth Risk Behavior Surveillance System (YRBSS)**. This system provides data on six health risk behaviors among youth: unintentional injuries and violence, tobacco use, alcohol and other drug use, sexual behaviors, dietary behaviors, and physical activity. The YRBSS is composed of several surveys of different populations of youth, but focuses on national, State, and local school-based surveys of students in grades 9 through 12.

https://grants.nih.gov/grants/policy/data_sharing/data_sharing_workbook.pdf

Mode Data Sharing – Data Enclaves

Data Enclaves

Some data can be shared only under the most controlled conditions. If, for example, there is any risk of subject identification, the investigator may ask that users submit requests for specific analyses or come to the investigator's site to run analyses under supervision. Data enclaves were designed to deal with such situations.

One such enclave is the **Research Data Center at the CDC's National Center for Health Statistics (NCHS)**. The Research Data Center supports use of several NCHS restricted-use datasets through the Internet and within the Data Center itself. Additional information on the Research Data Center is available at http://www.cdc.gov/nchs/r&d/rdc.htm.

One of the datasets that can be used at the NCHS Research Data Center is a periodic survey called the National Survey of Family Growth (NSFG). Data from this survey provide an accurate statistical picture of family life, marriage and divorce, contraception, sexual experience, pregnancy, and infertility. Information concerning the NSFG is available at the NCSH Website at http://www.cdc.gov/nchs/nsfg.htm.

https://grants.nih.gov/grants/policy/data_sharing/data_sharing_workbook.pdf

Use of De-identified Data

Attachment A: Human Subjects Research Implications of "Big Data" Studies **

In any event, institutions must certify to the NIH that the de-identified data, which have been collected under such a broad consent, may be used either on an unrestricted basis or on a "controlled access" basis. Under "controlled access," those researchers who seek to access and use the data must agree to a number of terms and conditions, including:

- Using the data only for the approved research;
- Protecting data confidentiality;
- Following, as appropriate, all applicable national, tribal, and state laws and regulations, as well as relevant institutional policies and procedures for handling genomic data;
- Not attempting to identify individual participants from whom the data were obtained;
- Not selling any of the data obtained from NIH-designated data repositories;
- Not sharing any of the data obtained from controlled-access NIH-designated data repositories with individuals other than those listed in the data access request;

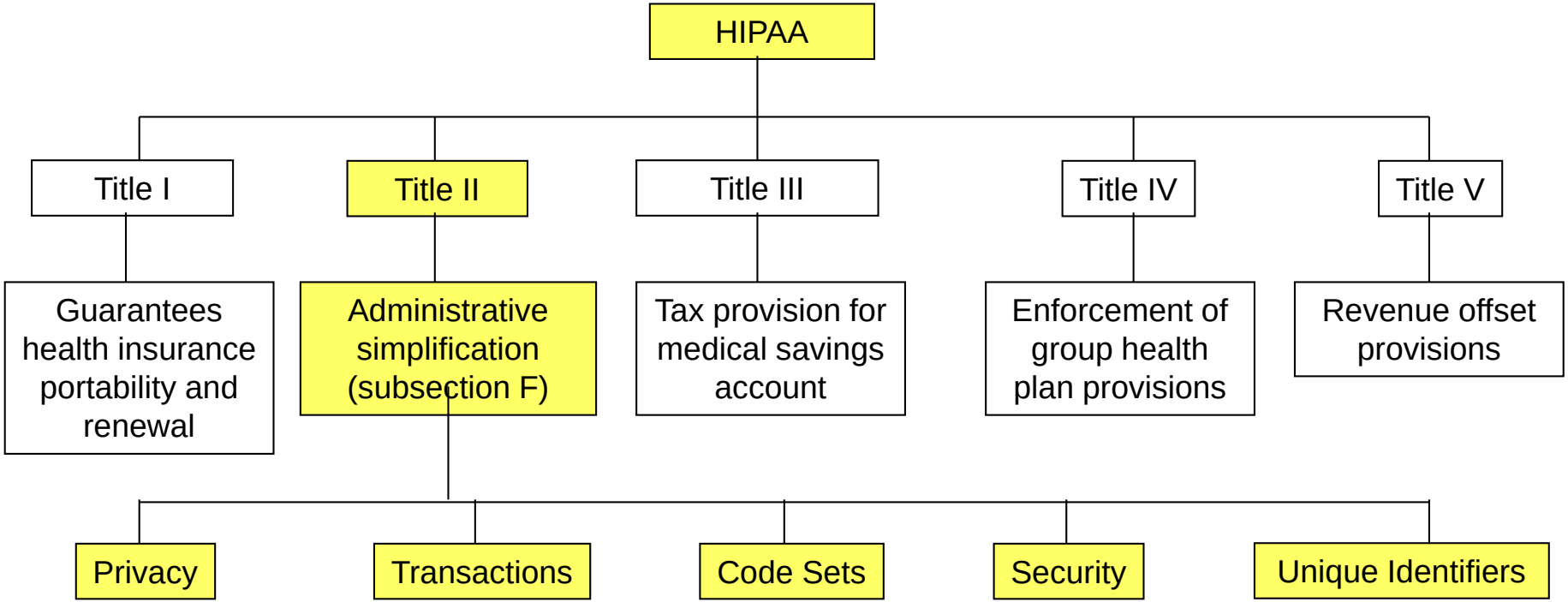
Use of De-identified Data

Attachment A: Human Subjects Research Implications of "Big Data" Studies

In any event, institutions must certify to the NIH that the de-identified data, which have been collected under such a broad consent, may be used either on an unrestricted basis or on a "controlled access" basis. Under "controlled access," those researchers who seek to access and use the data must agree to a number of terms and conditions, including:

- Agreeing to the listing of a summary of approved research uses in dbGaP along with the investigator's name and organizational affiliation;
- Agreeing to report any violation of the GDS Policy to the appropriate DAC(s) as soon as it is discovered;
- Reporting research progress using controlled-access datasets through annual access renewal requests or project close-out reports;
- Acknowledging in all oral or written presentations, disclosures, or publications the contributing investigator(s) who conducted the original study, the funding organization(s) that supported the work, the specific dataset(s) and applicable accession number(s), and the NIH-designated data repositories through which the investigator accessed any data.

Health Insurance Portability and Accountability Act (HIPAA), 1996



Source: John Butterworth, Ethics and Clinical Research

Health Insurance Portability and Accountability Act (HIPAA)

What is protected health information (PHI)?

According to the US Department of Health and Human Services, **protected health information (PHI)** is individually identifiable information (see below for definition) that is:

1. except as provided in item 2 of this definition,
    i. transmitted by electronic media;
    ii. maintained in electronic media; or
    iii. transmitted or maintained in any other form or medium (includes paper and oral communication).
2. Protected health information excludes individually identifiable health information:
    i. in education records covered by the Family Educational Rights and Privacy Act, as amended, 20 U.S.C. 1232g;
    ii. in records described at 20 U.S.C. 1232g(a)(4)(B)(iv);
    iii. in employment records held by a covered entity (see below for definition) in its role as employer; and
    iv. regarding a person who has been deceased for more than 50 years.

Health Insurance Portability and Accountability Act (HIPAA)

Electronic protected health information (ePHI)

Electronic protected health information (ePHI) is any protected health information (PHI) tha created, stored, transmitted, or received electronically.

Electronic protected health information includes any medium used to store, transmit, or receive PHI electronically. The following and any future technologies used for accessing, transmitting, or receiving PHI electronically are covered by the HIPAA Security Rule:

- Media containing data at rest (storage)
  - Personal computers with internal hard drives used at work, home, or traveling
  - External portable hard drives, including iPods and similar devices
  - Magnetic tape
  - Removable storage devices, such as USB memory sticks, CDs, DVDs, and floppy disks
  - PDAs and smartphones
- Data in transit, via wireless, Ethernet, modem, DSL, or cable network connections
  - Email
  - File transfer

Health Insurance Portability and Accountability Act (HIPAA)

Individually identifiable health information is information that is a subset of health information, including demographic information collected from an individual, and

1. is created, or received by a health care provider, health plan, or health care clearing house; and

2. relates to past, present, or future physical or mental health conditions of an individual; the provision of health care to the individual; or past, present, or future payment for health care to an individual, and

    i. that identifies the individual; or

    ii. with respect to which there is a reasonable basis to believe the information can be used to identify the individual.

Individually identifiable health information (i.e., PHI) is subject to state and federal privacy and security rules including, but not limited to, the Health Insurance Portability and Accountability Act (HIPAA).

HIPAA & De-identification

[1] Under the HIPAA Privacy Rule, de-identification of a dataset means removing the following variables: names; geographic information (including city, state, and zipcode); elements of dates such as those for birth, hospital admission and discharge, death; telephone numbers; fax numbers; electronic mail addresses; Social Security Number; medical record and prescription numbers; health plan beneficiary number; account numbers; certificate or license number; any vehicle identifier or serial number, including license plate number; any device identifier or serial number; Web Universal Resource Locator (URL); Internet Protocol (IP) address number; any biometric identifiers, including finger or voice prints; full face photographic images or any comparable images; and any other unique identifying number, characteristic, or code consisting of any segments of the previously listed identifiers. **

https://grants.nih.gov/grants/policy/data_sharing/data_sharing_workbook.pdf

PHI & PII

Table 1: Types of Identifiers

Protected Health Information (PHI)	Personal Identifying Information (PII)	Sensitive Information **
An individual's personal and health information that is created, received, or maintained by a health care provider or health plan and includes at least one of the 18 personal identifiers listed below in association with the health information: ○ Name ○ Street address ○ All elements of dates except year ○ Telephone number ○ Fax number ○ Email address ○ URL address ○ IP address ○ Social security number ○ Account numbers ○ License numbers ○ Medical record number ○ Health plan beneficiary # ○ Device identifiers and their serial numbers ○ Vehicle identifiers and serial number ○ Biometric identifiers (finger and voice prints) ○ Full face photos and other comparable images ○ Any other unique identifying number, code, or characteristic ***Limited Data Set*** - a limited data set can include the following identifiers: a unique number code, or characteristic that does not include any of the above listed identifiers, geographic data (without street address), and/or dates.	Information about an individual which includes any of the identifiers below: ○ Name ○ Street address ○ All elements of dates except year ○ Telephone number ○ Fax number ○ Email address ○ URL address ○ IP address ○ Social security number ○ Account number, credit or debit card number, in combination with any required security code, access code or password that would permit access to an individual's financial account ○ Driver's License numbers or California or other identification card number ○ Device identifiers and their serial numbers ○ Vehicle identifiers and serial number ○ Biometric identifiers (finger and voice prints) ○ Full face photos and other comparable images ○ Any other unique identifying number, code, or characteristic (e.g., student identification number)	An individual's first name (or first initial) and last name in combination with any of the following: ○ Social Security Number ○ Driver's License Number or California ID card number ○ Financial account information such as a credit card number ○ Medical Information

UCLA OHRPP
Office of the Human Research Protection Program

Guidance and Procedure: Collection, Use, Sharing and Secondary Analyses of Human Specimens and/or Data for Research Purposes
(last updated April 23, 2013)

Use of PHI in Research

Research Use and Disclosure of PHI Without Authorization:

Preparatory to Research

**

- Requires notification of the entity holding the PHI
- Researcher must provide representation that:
  - PHI is to be used solely to prepare a protocol or a similar purpose
  - PHI will not be removed from the covered entity
  - PHI is necessary for research
- May be used to develop hypothesis, protocol or characteristics of research cohort
- May not be summarized, used or presented as a research study without prior IRB approval
- May allow access to PHI to identify subjects for recruitment

Source: John Butterworth, Ethics and Clinical Research

Use of PHI in Research

Table 1 Outline of the major characteristics of the five categories of sample labeling

Category	*Link between subject identity and genetic data*	*Records identifiable for clinical monitoring*	*Actions possible if consent is withdrawn*	*Return of individual results to subject*	*Scope of subject privacy* **
Identified	Yes, directly	Yes	Sample can be destroyed Data can be deleted[a]	Possible	Similar to general healthcare confidentiality
Coded	Indirectly, via code numbers	Yes, via protocol-specified procedures	Sample can be destroyed Data can be deleted[a]	Possible	Standard for clinical research Conforms to ICH guidelines
De-Identified	Very indirectly via two levels of code numbers	Yes, via protocol-specified procedures	Sample can be destroyed and data can be deleted[a] via protocol-specified procedures	Possible	Offers added privacy over single coding, breached only via specified procedures
Anonymized	No. Key between first and second codes is deleted	No	Sample and data are not identifiable and cannot be destroyed once key is deleted	Not possible	Genetic data not linked to individuals, offering additional security
Anonymous	No	No	None	Not possible	Maximum

[a]Data can be deleted up to the time it is reported, but not thereafter.

Terminology for sample collection in clinical genetic studies

The Pharmacogenetics Working Group*

PII in Education Records (FERPA)

The Family Educational Rights and Privacy Act

Guidance for Reasonable Methods and Written Agreements

What is the Family Educational Rights and Privacy Act?

The Family Educational Rights and Privacy Act (FERPA), 20 U.S.C. §1232g, is a Federal privacy law administered by the Family Policy Compliance Office (FPCO or Office) in the U.S. Department of Education (Department or we). FERPA and its implementing regulations in 34 CFR part 99 protect the privacy of students' education records and afford parents and eligible students (i.e., students who are 18 years of age or older or attend an institution of postsecondary education) certain rights to inspect and review education records, to seek to amend these records, and to consent to the disclosure of personally identifiable information from education records (PII from education records).

The general rule under FERPA is that PII from education records cannot be disclosed without written consent. However, FERPA includes several exceptions that permit the disclosure of PII from education records without consent. Two of these exceptions are discussed in this document – the studies exception and the audit or evaluation exception. The two exceptions contain specific, and slightly different, requirements, described more fully in the implementing regulations (34 CFR Part 99).

Data Confidentiality & Security Agreement (FERPA)

Confidentiality Agreement

Researcher agrees to fulfill their responsibility on this project in accordance with the following guidelines:

1. To comply in all respects with the provisions outlined in (a) the DCPS Process and Requirements to Conduct Research or Obtain Confidential Data, and (b)the MOA between my organization and/or myself and DCPS.
2. To comply in all respects with applicable provisions of the Family Educational Rights and Privacy Act (FERPA) (20 U.S.C. § 1232g; 34 CFR Part 99).
3. To maintain, use, disclose, and share data received pursuant to the MOA in a manner authorized by FERPA and any applicable federal and District of Columbia law or regulation.
4. To use data shared under the MOA with DCPS for no purpose other than the research project described in the MOA, and as authorized under 34 CFR §§ 99.31(a)(6). Nothing in the MOA shall be construed to authorize me/my organization to have access to DCPS data beyond that included in the scope of the MOA. I/my organization further agree not to share Confidential Data received under the MOA with or permit access to such data by any individual or entity other than the Parties named in the MOA, for any purpose, except as permitted by the MOA and applicable law. I/my organization shall put procedures in place to safeguard the confidentiality and integrity of Confidential Data, to place limitations on its use and to maintain compliance with applicable privacy laws. I understand that the MOA does not convey me/my organization ownership of any Confidential Data.
5. To obtain all necessary approvals from authorized officials of my organization prior to beginning the Project. I will also obtain informed consent from Project participants as described in the DCPS Process and Requirements to Conduct Research or Obtain Confidential Data.

Data Confidentiality & Security Agreement (FERPA)

Confidential Data Request Application
Data Confidentiality and Security Agreement Form

All principal and secondary investigators who will have access to the confidential data requested in the *Confidential Data Request Application* must sign this form and submit it with the Application.

I _____, as a principal or secondary investigator, agree to receive confidential data from the Michigan Department of Education (MDE) and/or the Center for Educational Performance and Information (CEPI), and to observe the following security provisions in transferring, storing, analyzing and reporting of the data.

1. Policy for data storage
   a. The location of ***all copies*** of the data must be carefully tracked
   b. The data must be stored where ***only*** the Confidential Data Request Application designed principal and secondary investigator(s) may access the data
   c. Data files ***must*** remain secure throughout the duration of data storage

2. Policy for data usage
   a. Data may be accessed ***only*** by the Confidential Data Request Application designed principal and secondary investigator(s)
   b. Data ***may not*** be shared with any other individuals outside those designed as the principal and secondary investigator(s) in the Confidential Data Request Application
   c. Data may be used ***only*** for analyses that respect privacy and confidentiality of all concerned parties including students, teachers, classrooms, schools, districts, intermediate school districts and the State of Michigan

3. Policy for data disposal
   a. The data ***must*** be destroyed in accordance with the date designated for destruction in the signed Confidential Data Request Application
   b. If an extension on the data destruction deadline is needed, the Research Collaborative Internal Review Board ***must*** be contacted, in writing, to approve an extension
   c. A certificate of destruction will be sent via US mail to the Research Collaborative Internal Review Board on the date of the data loan expiration

Data License Contract (FERPA)

The School District of Philadelphia
Standard Terms
for
Research Data License Agreements

The School District of Philadelphia (the "School District") has received and carefully considered your request for certain data held by the School District in connection with certain academic research or an evaluation you seek to carry out.

This agreement applies to, and represents an agreement by and among the Institution (the "Institution") named in the Data Request Form or research proposal submitted to the School District's Research Review Committee, the individual(s) named and Authorized Recipients, if any, named elsewhere therein, and the School District. "You" and "your" means the Institution, the Authorized Recipients and these individual(s).

You have described the purpose of your research or evaluation, your research project, the data you request from the School District, and how you anticipate publishing or distributing your research in the Data Request Form or research proposal submitted to the School District's Research Review Committee.

In its written response to your request, the School District has agreed to provide you with the data (the "Data") described therein, on the condition that you hereby agree to comply with the terms, conditions and limitations set forth in this agreement. Subject to the duties, conditions and limits set forth in this agreement, the School District hereby grants you a limited, nontransferable, revocable, non-exclusive license to use the Data solely for the purposes and solely in the manner set forth in this agreement. This limited, nontransferable, revocable, non-exclusive license does not permit you to use the Data for any other purpose or in any manner other than as expressly set forth in the Data Request Form or research proposal submitted to the Research Review Committee.

Privacy & Security of Data Sharing

Privacy and Security Workgroup: Summary of Big Data Public Hearings
Deven McGraw & Stan Crosley,
Health IT Policy Committee,
Department of Health & Human Services. February 9 2015

**

Key points:

1. Sometimes there is a need to use fully identifiable data
2. It is not possible to get informed consent for all uses
3. Impossible to notify individuals personally about all uses
4. Can't do universal opt-out because answers could be unreliable
5. There is likely a standard that could be developed that determines "clearly good/appropriate uses" and "clearly bad/inappropriate uses"

Focus on:

1. Minimum necessary amount of identifiable data (but offset by future use needs)
2. Good processes for approval and oversight
3. Uses of data stated publicly (transparency)
4. Number of individuals who have accessed to data minimized (distributed systems help accomplish this)

When we use identifiable data, we must store it in highly protected locations – "data enclaves"

Discussion ….

- Use of stored data / specimen (e.g., in biobank/databank, data storage, surveillance system) – who give consent? Data / specimen can be used only with having prior consent?
- Individual / Aggregated data can be used at what level (say, 506, NHSO,etc) with and with out consent?
- When should consent be required?
  - Based on type of use/disclosure
  - Based on higher “privacy” risk – for example, personal v. non-personal impact; level of sensitivity of the data?)
  - Based on identifiability of the data?
  - Based on commercial/profit use?
  - Disclosure outside of initial environment vs. internal uses – is this a worthwhile distinction for consent purposes?

Discussion ….

Privacy and Security Workgroup: Summary of Big Data Public Hearings
Deven McGraw & Stan Crosley,
Health IT Policy Committee,
Department of Health & Human Services.
February 9 2015

- HIPAA – PHI or PII for Thailand?

- Research uses in the HIPAA environment:

  Re-iterate/refine initial recommendation for ANPRM: for re-use of clinical or claims data to "contribute to generalizable knowledge," no need to obtain consent, as long as entity in control of data uses and fair information practices are implemented (for example, security, minimum necessary, etc.).

  – Any caveats to this? (what are higher risk vs. lower risk use cases? Personal v. non-personal impact?)
  – How to implement – through guidance on waivers or change to regulation?

Discussion ….

Privacy and Security Workgroup: Summary of Big Data Public Hearings
Deven McGraw & Stan Crosley,
Health IT Policy Committee,
Department of Health & Human Services. February 9 2015

- Consent issues **within** the HIPAA Framework

  To date discussions have focused on research uses and whether HIPAA and Common Rule requirements while building trust/protecting individuals

  – Low risk research – should this be acceptable **without consent and without an IRB waiver**? How does this get determined?

  – Consider role of transparency in place of consent when coupled with "appropriate use" definitions, particularly for observational research or health care operations

- Other comments on de-identification of data