NIH Policy Manual

3016 - Intramural Research Program Human Data Sharing (HDS) Policy

Issuing Office: OD/OIR **Phone:** (301) 496-1921

Approving Official(s): CIO

Release Date: 7/31/2015 ?

Transmittal Notice

1. **Explanation of Material Transmitted:** This Manual Chapter describes policy for sharing of and secondary research with human data in the NIH Intramural Research Program (IRP). All NIH-owned or jointly-owned data, obtained from humans are covered.
2. **Filing Instructions:**

   Remove: n/a
   Insert: Manual Chapter 3016, dated 07/31/2015

PLEASE NOTE: For information on:

- Content of this chapter, contact the issuing office listed above.
- NIH Manual System, contact the Division of Management Support, OMA on 301-496-2832, or enter this URL: http://oma.nih.gov/public/MS/manualchapters/Pages/default.aspx

A. Purpose

Research often involves sharing and secondary use of data. Data sharing advances the NIH mission by facilitating the validation of research results, and allowing the strength of analyses to be increased by combining datasets, providing access to unique data that cannot be readily replicated, informing future research, increasing the return on investment of scientific research, and accelerating the translation of research results into knowledge, products, and procedures to improve the public health.

NIH's commitment to data sharing is longstanding. As a valuable resource, data must be managed in a way that promotes their responsible and fair distribution. This Policy describes how NIH's commitment to data sharing is implemented in the NIH IRP, the duties of intramural researchers, and the types or authorizations, agreements and documentation that may be needed for sharing human data from intramural laboratories and intramural collections for secondary research. This Policy is complementary to other applicable NIH policies [1].

These other policies include, as applicable, the *Standards for Clinical Research in the NIH Intramural Research Program (*
http://www.cc.nih.gov/ccc/patientcare/pdf/cc_research_standards.pdf), the *Guidelines for the Conduct of Research in the*
Intramural Research Program
(https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical_conduct/guidelines-conduct_research-6_11_07.pdf), NIH HRPP SOP 5 – *Research Activities with Human Specimens and* Data, the *NIH Genomic Data Sharing Policy*
(http://gds.nih.gov/03policy2.html) and the *Guidelines for Human Biospecimen Storage and Tracking within the NIH Intramural Research Program*
(https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical_conduct/guidelines-biospecimen.pdf). They also include, as applicable, the Office of Management and Budget Open Data Policy, which requires that data be generated, collected and made available in a manner that supports subsequent use and analysis. See OMB, Memorandum M-13-13, *Open Data Policy-Managing Information as an Asset* (May 9, 2013), available at https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf.

B. Scope

This Policy applies to all human data in the NIH Intramural Research Program (IRP), including the NIH Clinical Center as well as NIH Institutes and Centers. Included in this scope are human data, some of which is "human subjects" data, needed to validate research findings in scholarly publications whether or not the human data is identifiable to individual subjects, and regardless of where the data originated. Data from sources outside of NIH that are housed at NIH are covered unless NIH has no ownership interest, e.g., NIH obtained the data under a confidentiality agreement precluding ownership and the data is not used for Research by IRP investigators. The Policy applies to all research activities resulting in data collection, including natural history studies and interventional clinical trials, as well as epidemiologic research. It governs the sharing of data among intramural investigators as well as the sharing of intramural Data from intramural investigators to investigators outside of the NIH.

C. Background

NIH's mission is to seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability. In addition to funding biomedical and behavioral research into the causes, diagnosis, prevention, and cure of human diseases and the training of basic and clinical researchers capable of carrying out such research, NIH also is responsible for expanding the knowledge base in basic, medical, and associated sciences and for ensuring a continued high return on the public investment in research.Data sharing facilitates expedited translation of research results into knowledge, products, and procedures to improve human health. NIH has longstanding policies to make data from the research activities that it funds widely available for secondary research in a timely manner.[2]

The collection and storage of human data, including human subjects data, in research studies is a complex task that is integrated into the medical and research practices of each component of the IRP. The NIH Clinical Center's Clinical Research Information System (CRIS) collects clinical data as part of the clinical research process for patients of the Clinical Center. In addition, each Institute stores clinical research Data in a variety of management systems and repositories. Another system, the Biomedical Translational Research Information System (BTRIS) serves as a central repository for IRP clinical data.

Using and sharing human data requires special attention. If the data set includes human subject data, i.e., private, individually-identifiable information of living persons as defined in Federal regulations for the protection of human subjects[3], then review or approval from an Institutional Review Board (IRB) or the NIH Office of Human Subjects Research Protection (OHSRP) is required. If human data without private, individually-identifiable information are used, then prospective review should be sought from the OHSRP, e.g., to determine what additional procedures may be required, if any. In addition, because the data may be subject to the Privacy Act of 1974, a data sharing agreement consistent with that law may be needed. Regardless of whether data are identified, coded, or completely anonymous ("unlinked"), OHSRP may be consulted to help determine whether the data may be shared or used for secondary research purposes.

For example: *Final NIH Statement on Sharing Research Data*, February 26, 2003 available at http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html; *NIH Intramural Policy on Large Database Sharing*, April 5, 2002 available athttp://sourcebook.od.nih.gov/ethic-conduct/large-db-sharing.htm; *NIH Policy on Sharing of Model Organisms for Biomedical Research*, May 7, 2004 available at http://grants.nih.gov/grants/guide/notice-files/NOT-OD-04-042.html*; Reaffirmation and Extension of NHGRI Rapid Data Release Policies: Large-scale Sequencing and Other Community Resource Projects*, February 2003 available at https://www.genome.gov/10506537; *NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS)* available at http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html; and *NIH Genomic Data Sharing Policy*, available at http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html.

See 45 CFR 46.

D. Definitions

Definitions [4]:

1. **Coded:** Data for which (a) identifying information (such as name or social security number) that would enable the investigator readily to ascertain the identity of the individual to whom the private information pertains has been replaced with a number, letter, symbol, or combination thereof (i.e., the code); and (b) a key to decipher the code exists, enabling linkage of the identifying information to the private information.

2. **Data:** The recorded factual material commonly accepted in the scientific community as necessary to validate research findings, including data sets, used to support scholarly publications, but does not include laboratory notebooks, preliminary analyses, drafts of scientific papers, invention disclosures or patent applications, plans for future research, peer reviews, communications with colleagues, or physical objects such as laboratory specimens.
3. **Derivatives:** Materials produced by the modification of human specimens or isolation of a component(s) of specimen samples originally obtained from humans. Examples include human cell lines, recombinant DNA clones of human genes, and isolated infectious agents from humans.
4. **Human Data:** Data that are obtained directly from humans as well as derivatives of such data.
5. **Human Subject:** A living individual about whom an investigator (whether professional or student) conducting research obtains:

    a. Data through intervention or interaction with the individual, or
    b. Identifiable private information. (45 CFR 46.102)

6. **Identified:** Data that are still attached to a readily available subject identifier such as name, social security number, study number, hospital number, medical record number, address, telephone number, etc., such that the identity of the subject can be ascertained.
7. **Intervention:** Physical procedures by which data are gathered (e.g., venipuncture) and/or manipulations of the subject or the subject's environment that are performed for research purposes.
8. **Interaction:** Communication or interpersonal contact between investigator and subject.
9. **Private information:** 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 provided for specific purposes by an individual, which the individual can reasonably expect will not be made public (for example, a medical record). In order for obtaining the information to constitute research involving human subjects, private information must be *individually identifiable* (i.e., the identity of the subject is or may readily be ascertained by the investigator or may be associated with the information). (45 CFR 46.102)
10. **Individually Identifiable:** The identity of the subject is or may readily be ascertained by the investigator or associated with the information.(45 CFR 46.102)
11. **Publicly Accessible:** Data are available to qualified researchers outside of the NIH. It may include either data that are openly accessible and available for any use or data that are accessed in a controlled manner to protect appropriately certain interests, for example, the privacy of research subjects, intellectual property or security.
12. **Research:** A systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge. Activities that meet this definition constitute research for purposes of this Policy, whether or not they are conducted or supported under a program considered to be research for other purposes. For example, some demonstration and service programs may include research for purposes of this Policy.

13. **Research Repository:**A collection of data maintained for the overall purpose of sharing data for current and future research. Such research may be general and cover a wide array of diseases and topics, or may be more specific, focusing on a single disease or topic.
14. **Secondary Research:** Research use of data for other than the original purpose(s) for which the Data were initially collected through interaction or intervention with living individuals.
15. **Specimens:** A quantity of tissue, blood, urine or other biologically-derived material. A specimen can include everything from subcellular structures (DNA) to cells, tissue (bone, muscle, connective tissue), organs (e.g. liver, bladder, heart, kidney), blood, gametes (sperm and ova), and waste (urine, feces, sweat, hair and nail clippings, shed epithelial cells, and placenta).
16. **Unlinked:** Data that were initially collected with identifiers but, before research use, have been irreversibly stripped of all identifiers by use of an arbitrary or random alphanumeric code and the key to the code is destroyed, thus making impossible for anyone to link the samples to the sources. This does not preclude linkage with existing clinical, pathological, and demographic information so long as all individual identifiers are removed prior to distribution or receipt.

---

[4]Definitions from 45 CFR 46 where applicable and from other NIH policies, such as the *Policy for the Transfer of Materials from NIH Intramural Laboratories*, United States Public Health Service Technology Transfer Policy Manual Chapter 500A (2014), found at: http://www.ott.nih.gov/sites/default/files/documents/policy/pdfs/500-A-Policy_12042014.pdf.

E. Policy

To further advance and accelerate research to benefit the public health, data developed in the NIH IRP (NIH-owned or jointly-owned) should be collected in a manner that permits and promotes the broadest sharing possible. Data sharing may be complicated or limited, in certain cases, by agreements with outside collaborators, e.g., Cooperative Research and Development Agreements (CRADAs) or Clinical Trial Agreements, by IRB rules or by other constraints. NIH IRP investigators should share data broadly for secondary research purposes, in all cases consistent with applicable laws, regulations, and policies.

1. **Data Sharing Plans**

   All IRP-supported investigators must develop an appropriate Data Sharing Plan (DSP) for each research activity involving human data. The DSPs will describe the data to be produced in the proposed research, a commitment to the principle of data sharing, and a willingness to share, at a minimum, the data underlying any publications resulting from the research or an explanation of why sharing is not possible. It is understood that, in some cases, limitations like contractual obligations may constrain data sharing and these should be described. If other NIH policies for data sharing apply to the data, e.g., for genomic data, the DSP should indicate that those policies will be followed.[5]

Additionally, each DSP should include, as applicable:

a. Provisions for protecting the subjects' privacy and the confidentiality of the data, including, if necessary, any limitations on secondary research with the data based on the original informed consent (if known) or other limits. If data are not appropriate for sharing on an individual level basis even with individually-identifiable information removed, the DSP should include an explanation as well as an alternative for data sharing such as aggregate data sharing. Anticipated sharing and use limitations should be included;
b. A timeline for making the data publicly accessible, in general, no later than the time of publication of the main findings;
c. Any intellectual property issues (e.g., patent filings) or contractual obligations that would preclude sharing and secondary research with the data; and,
d. An estimate of costs and resources that may be needed to support the proposed DSP (e.g., preparation of data for submission to an intramural or public repository).

Scientific Directors, or their delegates, will review and approve each DSP, prior to the start of research. Data sharing plans may be amended. Scientific Directors will determine where copies of DSPs will be maintained.

2. **Data Collection and Participant Consent for Human Data**

All IRP-supported clinical investigators are expected to develop protocols and consent processes/forms to enable broad data sharing for secondary research consistent with this Policy.

At minimum, the protocol in which specimens and/or data are to be collected from humans and the individual participant consent process/form should identify which entities are expected to have access to data for the proposed (primary) research, including researchers at institutions outside of the IRP, if any. Additionally, because it is expected that data will be shared for secondary research – for example, through a central research repository or individual repositories – collection protocols and consent processes/forms should describe these plans, including anticipated plans to remove individually-identifiable information from the data before sharing.[6]

The consent process/form may not condition enrollment in a research activity on agreement to participate in secondary research, including by contributing to a research repository, unless it is intrinsic to the design of the proposed (primary) research. For example, if a prospective human subject does not agree to the use of his or her data in a research repository, the individual should not be denied entry into the research unless the primary purpose of the research involves creation of a repository and sharing data for future research.

The OHSRP, IRB staff and the Department of Bioethics may be consulted for assistance in developing specific consent language. One possible example of language

for future use and broad data sharing, to be included within a consent form explaining the circumstances under which data will be shared includes:

> *Example:* To advance science, it is helpful for researchers to share information they get from studying human samples. They do this by putting it into one or more scientific databases, where it is stored along with information from other studies. Researchers can then study the information combined from many studies to learn even more about health and disease.
>
> If you agree to take part in this study, some of your genetic and health information will be placed into one or more scientific databases after it has been stripped of identifiers such as name, address or account number, so that it may be used for future research on any topic and shared broadly for research purposes. A researcher who wants to study the information must apply to the database and be approved. Researchers with an approved study may be able to see and use your information, along with that from many other people. We do not expect any direct benefits for you from research resulting in the use of your data and information. When genomic data is shared, even when access is limited to approved users, confidentiality cannot be guaranteed because it may be possible to re-identify the data. De-identified data could be used to discriminate against or stigmatize you, your family or other groups to which you belong. However, state and federal laws provide some protections against genetic discrimination. If you have any questions, please ask the Principal Investigator.
>
> You may stop participating in this study and withdraw permission for your individual data, specimens and health information to be used for additional or future research at any time. If you choose, you may request to have your [*data/biological materials/interview transcript*] destroyed. However, it may not be possible to withdraw or delete materials or data once they have been shared with other researchers.

Consent documents must also include the name and contact information of an NIH staff person to address participant questions. For more information on the principles and practices to consider, see "Guidelines for Human Biospecimen Storage and Tracking with the NIH IRP" (2013). See also the section on Data Sharing Plans, above.

3. **Data Access and Distribution for Secondary Research**

Data are available to IRP-supported investigators and their collaborators outside of the IRP for secondary research in various ways.

Data should be shared in a fair manner and without bias to qualified recipients. Generally, individually-identifiable information about humans should not be distributed and used for secondary research. Identified data should be shared only in rare circumstances in which the research cannot reasonably be accomplished without certain identifying information and only after IRB review or OHSRP clearance, as applicable.

When identifiable information or coded information with access to identifiable information is shared, specific human subject and privacy protections apply and documentation is needed. For more information see HRPP SOP 5 – *NIH Research Activities with Human Data/Specimens*.

Generally, human subject and privacy protections apply to secondary research as follows:

a. **"Individually identifiable" or "Coded" Data when the IRP investigator can identify the subjects** requires prospective and continuing IRB review and approval - unless determined to be exempt from IRB review by OHSRP. Continuing IRB review and approval is required as long as research analyses are ongoing. Data use are subject to the Privacy Act of 1974 and records must be kept in accord with it, e.g., written agreement for transfer to outside collaborators.
b. **"Coded" Data when the IRP investigator cannot identify the subjects** may be exempt from or not subject to the requirement for IRB review and approval. The NIH OHSRP determines whether a research activity is exempt or not subject to IRB requirements. IRP-supported investigators must submit a formal request to the OHSRP. Data use may or may not be subject to the Privacy Act.
c. **"Unlinked" Data** is excluded from the requirement for IRB review and approval. The NIH OHSRP is authorized to determine whether a research activity is excluded. IRP investigators must submit a formal request to OHSRP. Data use is not subject to the Privacy Act.

When data are obtained from and shared within the NIH, transfers should be recorded and appropriate approvals obtained and documented for Human Subject compliance. There is no requirement under this Policy to use a specific data use agreement or obtain administrative approval from the principal investigator for the transfer. However, individual research repositories, such as the NIH database of Genotypes and Phenotypes (dbGaP) or the NIH BTRIS may have additional application and documentation requirements.

When data are shared outside of the NIH, transfers should be recorded and appropriate approvals obtained and documented for human subject and Privacy Act compliance. Additional documentation may be needed. At minimum written documentation between the NIH and the recipient of the data must be made to memorialize what is being shared and under what conditions. Individual research repositories, such as dbGaP or BTRIS, may have additional application and documentation requirements. For further information see Guidelines for the Conduct of Research in the Intramural Research Program (2007), HRPP SOP 20 Requirements for Collaborative Research, and HRPP SOP 20A Obtaining a Reliance (Authorization) Agreement at the NIH.

For additional information on these requirements, please consult the OHSRP, applicable IC Privacy Act office, or the Office of the General Counsel.

4. **Deposits to Research Repositories**

   All IRP investigators are encouraged to deposit data in publicly accessible research repositories for sharing to the extent feasible and appropriate. Data to be shared in publicly accessible research repositories should be deposited after data cleaning and quality control, within any applicable required timeframes.[7] In addition, data made available for secondary research by IRP-supported investigators through publicly accessible research repositories shall be handled in a manner consistent with applicable laws, regulations and policies.
5. **Effective Date**

   This Policy is effective on October 1, 2015. Proposed studies beginning scientific review on or after that date must comply with this Policy.

   For research that is ongoing or has begun scientific review before the effective date of this Policy, IRP-supported investigators are expected to share data for secondary research to the maximum extent possible, consistent with the informed consent provided by the research participant.

Under the NIH *Genomic Data Sharing Policy*, for example, large scale human genomic data are expected to be submitted to an NIH designated repository to facilitate secondary use of the data. See http://gds.nih.gov/pdf/NIH_guidance_developing_GDS_plans.pdf.

These consent requirements apply generally, consistent with 45 CFR 46, to interactions with human subjects, not to data generated from de-identified specimens. In some cases, owing to other policy or legal mandates, they will apply to data generated from de-identified human specimens or material. For example, for genomic data, NIH requires express consent for future use and broad data sharing even when samples are de-identified or come from discarded clinical sources. See *NIH Genomic Data Sharing Policy* (2014), found at: http://gds.nih.gov/PDF/NIH_GDS_Policy.pdf. Therefore, it may be appropriate to include express consent for future use and broad sharing in some collection protocols and it may be impermissible to use some samples without such consent.

For example, for genomic data, NIH requires data to be deposited “in a timely manner.” See *NIH Genomic Data Sharing Policy* (2014), found at: http://gds.nih.gov/PDF/NIH_GDS_Policy.pdf.

F. Records Retention and Disposal

All records pertaining to this chapter must be retained and disposed of under the authority of NIH Manual 1743,"Keeping and Destroying Records," Appendix 1, "NIH Records Control Schedules"(as amended). These records must be maintained in accordance with current NIH Records Management and Federal guidelines. Contact your IC Records Liaison or the NIH

Records Officer for additional information.

G. Internal Controls

The purpose of this manual issuance is to create a policy and process to allow for maximum sharing of data and secondary research with data.

1. **Office Responsible for Reviewing Internal Controls Relative to this Chapter:** NIH Office of the Deputy Director for Intramural Research (DDIR) and individual IC Scientific Directors or delegates.
2. **Frequency of Review:** Ongoing; at least annually.
3. **Method of Review:** The IC Directors or Scientific Directors participate in the Annual Intramural Self-Assessment of Management Controls, through completion of a set of comprehensive checklists of questions. The Office of Intramural Research manages this process.
4. **Review Reports:** Review reports shall be sent to the NIH senior official responsible for the area covered by the Chapter, the Deputy Director for Intramural Research.

H. General References

1. *Standards for Clinical Research in the NIH Intramural Research Program* (2009) *(*http://www.cc.nih.gov/ccc/patientcare/pdf/cc_research_standards.pdf)
2. *Guidelines for the Conduct of Research in the Intramural Research Program* (2007) (https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical_conduct/guidelines-conduct_research-6_11_07.pdf)
3. *Guidelines for Human Biospecimen Storage and Tracking within the NIH Intramural Research Program* (2013) (https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical_conduct/guidelines-biospecimen.pdf)
4. *Policy for the Transfer of Materials from NIH Intramural Laboratories*, United States Public Health Service Technology Transfer Policy Manual Chapter 500A (2014) (http://www.ott.nih.gov/sites/default/files/documents/policy/pdfs/500-A-Policy_12042014.pdf)
5. NIH HRPP SOP 5 - *NIH Research Activities with Human Data/Specimens* (2013) (http://ohsr.od.nih.gov/OHSR/pnppublic.php)
6. NIH HRPP SOP 6 - *Determinations, Including Exemptions Made by the Office of Human Subjects Research Protections (OHSRP)* (2013) (http://ohsr.od.nih.gov/OHSR/pnppublic.php)
7. NIH HRPP SOP 20 - *Requirements for Collaborative Research* (2013) (http://ohsr.od.nih.gov/OHSR/pnppublic.php)
8. NIH HRPP SOP 20A - *Obtaining a Reliance (Authorization) Agreement at the NIH* (2013) (http://ohsr.od.nih.gov/OHSR/pnppublic.php)
9. *NIH Genomic Data Sharing Policy* (2014) (http://gds.nih.gov/03policy2.html)
10. NIH Manual 1743 – *Keeping and Destroying Records*
11. NIH Privacy Act Systems of Record Notice 09-25-0099, *Clinical Research Patient Medical Records*, HHS/NIH/CC

http://oma.od.nih.gov/public/ms/privacy/pafiles/0099.htm.
12. NIH Privacy Act Systems of Record Notice 09-25- 0200, *Clinical Basic and Population-Based Research Studies of the National Institutes of Health (NIH)*, HHS/NIH/OD,http://oma.od.nih.gov/public/ms/privacy/pafiles/0200.htm