# National Institutes of Health
# Points to Consider in Drafting
# Effective Data Use Limitation Statements

## I. Introduction

A Data Use Limitation Statement (DUL) is a brief written description of the research purpose(s) for which data submitted to National Institutes of Health (NIH) genomic data repositories (for example, the database for Genotypes and Phenotypes, or dbGaP) should or should not be distributed. NIH Data Access Committees (DACs) refer to DULs when they evaluate Project Requests submitted by investigators seeking access to genomic data, to help determine whether the purpose(s) for which the data are being requested are consistent with the purpose(s) for which the Submitting Institution authorized the data to be used (that is, generally speaking, for purpose(s) that are not inconsistent with the terms of the informed consent under which the data were collected). **Drafting clear, understandable DULs is thus critical to ensure that data are made available for the full range of appropriate uses, but not for unauthorized ones.**

This guidance document lists points to consider when drafting DULs for data to be deposited in dbGaP. This guidance is for use by:

- Investigators preparing to submit data to NIH who are responsible for drafting DULs to describe how the data should or should not be used;
- Institutional Officials and Institutional Review Boards (IRBs) charged with reviewing and approving investigators' plans for the deposition of data in genomic databases; and
- NIH program staff responsible for overseeing the process of data deposition in NIH genomic databases.

## II. General Considerations

DULs that authorize broad use are optimal from the standpoint of promoting the rapid advancement of science by enabling broad use of the data. Such DULs are usually also easiest for NIH DACs to interpret. However, a DUL should be written no more broadly than the informed consent form that was signed by the participants in the original study (See "Points to Consider for IRBs and Investigators"). Any limitations on use of the data included in the informed consent form, or otherwise stipulated by the submitting institution, should be clearly stated. The DUL should provide enough information to enable DACs to determine easily if proposed research use statements for secondary research are consistent with the research purposes permitted by the submitting institution (i.e., consistent with the informed consent form and the process by which the data were initially collected).

Where the study involves an archived ("legacy" or "retrospective") dataset, the approach to drafting a DUL depends on how the consent form (or consent forms) for the study was (or were) written. Some studies—especially studies that have been underway for a long time—involve more than a single version of a consent form. In such cases, separate DULs may need to be drafted for different "Consent Groups," depending on any limitations stated in the consent forms used to enroll in the study at different points in time. Some general principles that apply to drafting DULs for studies that involve retrospective datasets include:

- If the consent form stated that the data would not be shared with other investigators, the data are not appropriate for deposition in NIH genomic repositories, unless a new consent is obtained from the research participants. In such cases, the investigator may wish to consider, in consultation with the IRB and the institution, whether it is appropriate and feasible to seek a new consent for broad data sharing. It may also be appropriate to consider submitting a request to the funding IC for an exception to the expectation for genomic data deposition to NIH databases.

- If the consent form was silent on the matter of data sharing, the investigator's institution must determine, in consultation with the IRB, whether or not the data are appropriate for deposition in dbGaP or other NIH repositories—and if data may be deposited, what if any limitations should be placed on its use. In making this determination, the institution should be guided by the original purpose(s) of the research, as described in the consent form and other information about the study and study population.

- If the consent form anticipated the possibility of data sharing, but stated that the data would be used only for the study of a particular disease, the investigator may wish to consider, in consultation with the IRB and the institution, obtaining a new consent from participants for broad research purposes. NIH may in some cases consider providing funding to obtain a new consent in the context of other funding priorities. However, in some cases, obtaining a new consent (or even initial consent) may be impossible or inappropriate; in such cases, the data may need to be deposited with a "disease-specific" DUL (see Section III below).

- If the consent form is written in a way that addresses whether or not *samples* may be used in the future (and for what research purposes), but is silent about future uses of *data*, drafting of the DUL should take into consideration (in consultation with the IRB and the institution) what a reasonable person signing the consent form in the circumstances may have understood he or she was agreeing to when signing the consent. In some cases, research participants may have significantly greater concerns about what might happen in the future to the samples they provide (for example, concerns about "cloning") than about what might happen to the data generated from the samples. On the other hand, some research participants may have greater concerns about what might happen to the data, which, unlike samples that may be depleted eventually, could remain "out there" forever.

## III. Special Considerations Relevant to "Disease-Specific" DULs

Where the data (or the sample(s) from which the data were derived) were collected with an informed consent allowing the data to be used for future studies relating to a particular disease, the IRB or institution may decide that it would be appropriate to have a disease-specific DUL, because more general research uses would exceed the scope of the original consent.

Where the underlying consent form authorizes use of the data for studies within a particular disease "range," writing an easily-interpretable DUL can be especially challenging. For example, the consent form for a schizophrenia study may state that the data would be made available only for future research on "schizophrenia and related conditions," or the consent form used to collect the data in a Type 1 diabetes study might state that the data would be made available only for future research on "Type 1 diabetes and its complications." If these studies are submitted to NIH with DULs that simply quote the language from the consent form (stating in the first case that use of the data is "limited to genetic studies of schizophrenia and related disorders" and in the second case that the use of the data is "limited to genetic research on "Type 1 diabetes and its complications") it can lead to questions for the DAC, such as

"Which disorders are 'related' to schizophrenia" and "Which conditions qualify as 'complications' of Type 1 diabetes?"

In such cases, the investigators requesting the datasets will be expected to justify clearly how their proposed research relates to, or is a complication of, the specified diseases, and the DAC will make reasonable efforts to evaluate the proposed justifications. However, to minimize the chance that the DAC will interpret the DUL in a way that is not consistent with the institution's determination of appropriate research use, it is advisable that the "disease-specific" limitations be drafted with precision. Any potentially ambiguous language relating to the permissible scope of research uses should be clarified (for example, by the listing of examples of the types of conditions that would or would not qualify).

Here are two examples that illustrate this:

> **BAD EXAMPLE:**
> *Use of the data is limited to genetic studies of schizophrenia and related conditions.*
>
> **GOOD EXAMPLE:**
> *Use of the data is limited to genetic studies of schizophrenia and related conditions. "Related conditions" includes conditions with evidence of genetic relationships to schizophrenia or schizoaffective disorder, such as acute psychoses, bipolar disorder, substance abuse disorders, MDD, or "Cluster A" personality disorders (schizotypal, schizoid, paranoid).*
>
> **BAD EXAMPLE:**
> *Use of the data is limited to genetic research on Type 1 diabetes and its complications.*
>
> **GOOD EXAMPLE:**
> *Use of the data is limited to genetic research on Type 1 diabetes and its complications. "Complications" include nephropathy, cardiovascular disease, retinopathy, neuropathy, and mortality. Phenotypes related to diabetes and its complications, such as body mass index, blood pressure, lipids, and hemoglobin A1c, may also be studied."*

Note that even a very carefully drafted DUL will often require a DAC to make some judgments. Here is an example:

> *Use of the data is limited to genetic studies of psychiatric health and related somatic conditions. "Psychiatric health" refers to DSM-IV or ICD-10 psychiatric disorders (for example, major depressive disorder, bipolar disorder, schizophrenia, attention-deficit hyperactivity disorder, autism, or substance use disorders). "Related somatic conditions" refers to general medical disorders whose risks have been shown to be elevated in individuals with psychiatric disorders (for example, cardiovascular disease, migraine, or Type 2 diabetes).*

In this example, the DAC will still have to use its judgment to decide how much evidence is needed to establish that risk for a particular medical disorder has been "shown" to be elevated in individuals with psychiatric disorders; the evidence base for making these determinations may also evolve over time. Nevertheless, in the example shown, the DAC would have been given some general guidance to assist its decision-making.

## IV. Some Example DULs and Their "Default" Interpretation

Below are examples of some commonly-used DULs and common interpretations of NIH DACs in applying them. As these examples show, very subtle differences in wording between DULs can lead to significant differences in the way the DULs could be interpreted. For a more detailed discussion on drafting DULs to modify these "default" interpretations (i.e., to minimize the likelihood that a DAC will interpret them more broadly or more narrowly than was intended), see Section V below.

- ***The data may be used for "general research purposes"***

  These data would generally be made available to any qualified investigator, irrespective of the specific research purpose for which the data are requested. For example, the data would be made available for:
  - research on any disease, even if the research is on a disease very different from the disease being studied in the original project;
  - methods development research which may have applications to many different diseases (e.g., development of software or algorithms;
  - research on non-disease traits (e.g., intelligence, behavioral or personality traits); and
  - research relating to population structure; this would nclude research that involves the determination of allele frequencies in different populations and that may thus have forensic/criminal justice applications or implications for the understanding of ancestral history (note that applications of these types may be objectionable to members of groups that have historically been the targets of discrimination).

- ***The data may be used for "biomedical research" (or "health research" or "health-related research")***

  These data would generally be made available to any qualified investigator, for:
  - research on any disease;
  - methods development research (which may have applications to many different diseases); and
  - research relating to population structure that may have implications for health.

  These data would *not* be made available for:
  - research on non-disease traits (e.g., intelligence, behavioral traits that have no clear relationship to disease); or
  - research relating to population structure that has an explicitly forensic focus.

  Note, however, that the distinctions here are sometimes difficult to draw. For example, these data would generally be made available for research on "learning disorders," because such studies can reasonably be interpreted as falling within the definition of "biomedical," "health," or "health-related" research—though the "line" between research on a "learning disorder" and "intelligence" can be murky. These data would also be made available for studies of behaviors that are health-related, such as for studies of alcohol or tobacco use. Note also that research relating to population structure that has implications for health also will unavoidably have some forensic applications or implications for the understanding of ancestral history, because of the information it may provide about allele frequencies in different populations.

- ***The data may be used for "genetic research"***

  These data would generally be made available to any qualified investigator, for:
  - research on any genetic disease (or any disease with a substantial genetic component);

- methods development research (which may have applications to many different genetic diseases);
- genetic research on non-disease traits (e.g., intelligence, behavioral traits); and
- research relating to population structure (including research that may have applications or implications for the understanding of ancestral history, because of the information it may provide about allele frequencies in different populations).

## V. Final Points to Consider

- Broadly written DULs (e.g., that authorize release of that data for "general research use," "biomedical research," "health-related research") have the broadest potential for the advancement of science and return of public benefit and are the most straightforward to administer. However, such DULs should only be used in cases where broad use is consistent with (i.e., is not precluded by) the original informed consent.

- In accord with the basic goal of the NIH genomics data sharing policy, the general approach of NIH DACs is to make data available to qualified investigators, regardless of whether the investigator is in the non-profit (academic, government) or profit sector, so long as the proposed Research Use Statement he or she submits is consistent with any limitations stated in the DUL. Thus, if there are any particular diseases, traits, or types of research that are *not* appropriate for data use, these should be listed specifically in the DUL. Similarly, if the data should *not* be made available to commercial entities due to constraints within the informed consent forms, this should be specifically stated in the DUL.

- In general, NIH DACs will approve data access to qualified investigators for research that will involve methods, software, or other tool development, even in instances of "disease-specific" DULs. This position was developed in consultation with the Participant and Data Protection Working Group of the Advisory Committee to the Director; the reasoning is that studies of this type can be expected to advance understanding of the specific disease, although presumably it will be useful beyond the targeted disease as well. Therefore, if the data should *not* be used for methods or tool development, this should be stated specifically in the DUL.

- NIH DACs are likely to approve data access to any qualified investigator, regardless of whether the research to be performed is specifically "genetic." Thus, if the data should *not* be made available for non-genetic research (i.e., research that will require use only of phenotype data), this should be stated specifically in the DUL.

- In most circumstances, NIH DACs will provide data access to any qualified investigator for studies focused on analyses of population structure; such studies will unavoidably have some forensic applications or implications for the understanding of ancestral history. The reasoning is that studies of this type may have important implications for understanding patterns of health and disease in populations. Such studies involve very little risk to individual participants, but some may have implications for groups. Thus, if the data should *not* be used for such purposes, this should be stated specifically in the DUL.

- A DUL should not contain extraneous, duplicative, or unnecessary information. In this regard, it may be helpful to review the language of the Data Use Certification (DUC) agreement that all investigators and their home institution sign when submitting a Project Request, in order to make sure that a particular issue is not already addressed through the policy itself. For example, it is unnecessary to state in a DUL that "Investigators must explicitly state their intention not to try to

use the data to identify an individual," because the terms and conditions within the DUC prohibit this.

- Clear and precise language is essential in drafting an effective DUL. Ambiguities in drafting should be carefully avoided. NIH program staff and the IRB should be consulted to ensure that the DUL is clearly written. NIH program staff may also consult with DAC Chairs or their IC GWAS Program Administrator for additional guidance on drafting DULs.