NIH Genomic Data Sharing

X01 Data Analysis Collaboration Workshop

June 16, 2017

NIH Genomic Data Sharing Policy (NOT-OD-14-124)

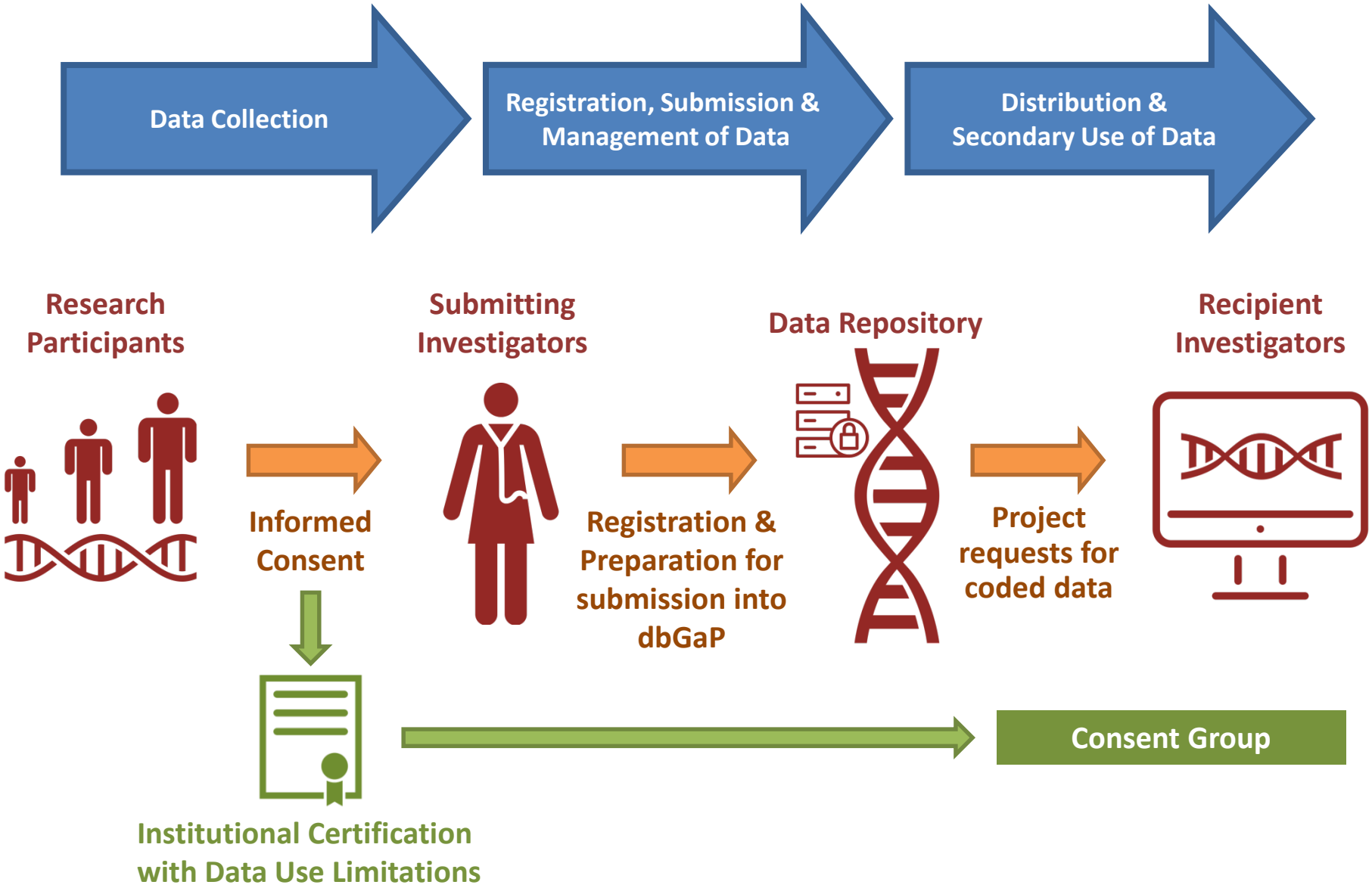
- Effective **January 25, 2015**
- Replaces the NIH GWAS Data Sharing Policy (NOT-OD-07-088)
- **Guiding Principle**: The greatest public benefit will be realized if data are made available
  - Under terms and conditions consistent with the informed consent provided by individual participants
  - In a timely manner to the largest possible number of investigators

Know Your GPA

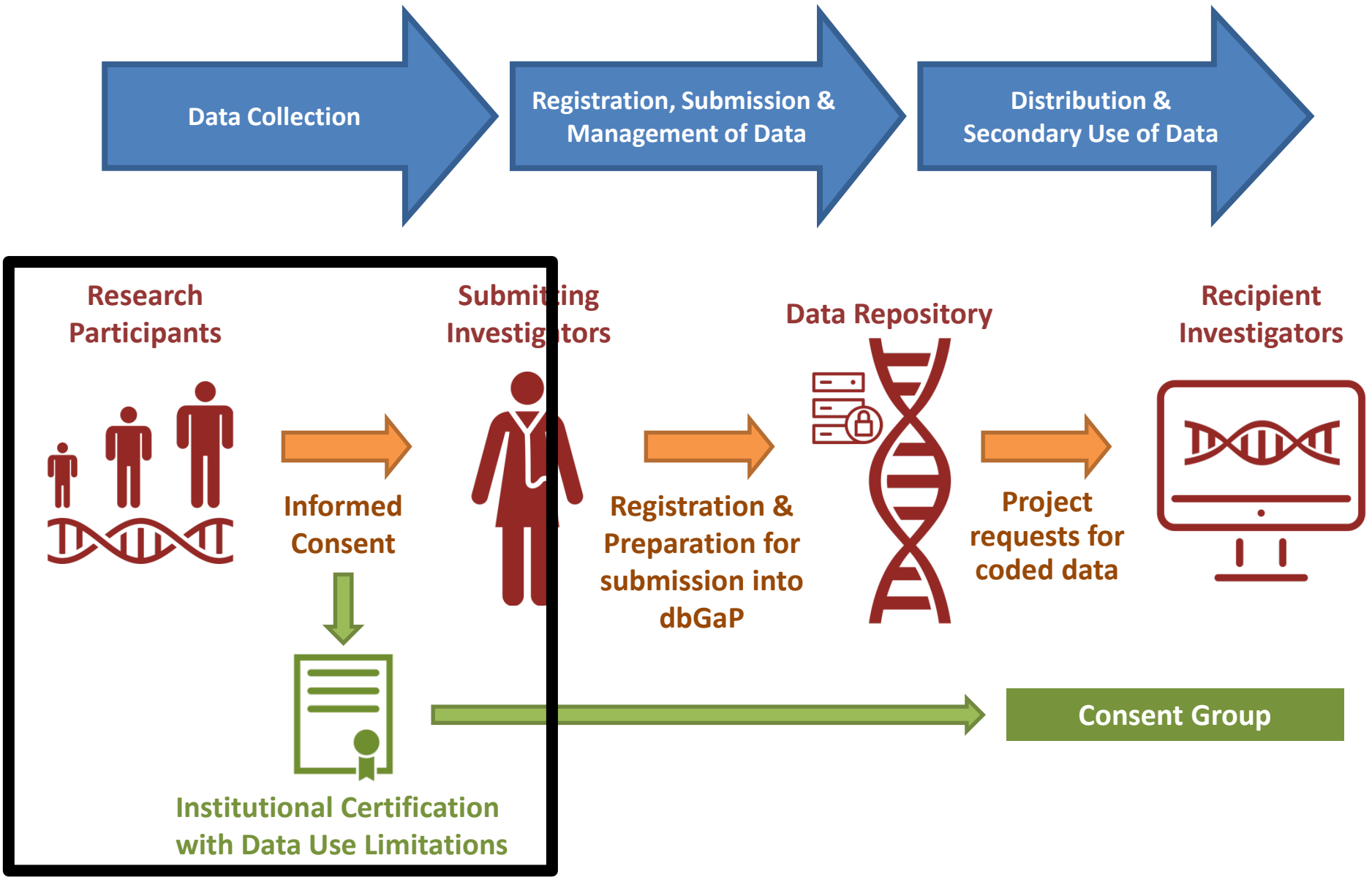
Each NIH Institute and Center (IC) has a Genomic Program Administrator (GPA) who serves as a point of contact for GDS Policy implementation within the IC.

https://osp.od.nih.gov/wp-content/uploads/IC_GPAs.pdf

dbGaP Data Sharing: The Big Picture



dbGaP Data Sharing: The Big Picture



Institutional Certification template

Extramural Institutional Certification*

For studies using data generated from cell lines created or clinical specimens collected after January 25, 2015

Date: [MM/DD/YYYY]
Name of GPA
Genomic Program Administrator
Select IC [, NIH, HHS
9000 Rockville Pike
Bethesda, MD 20892-7395

Re: Institutional Certification of _____ [NAME OF INSTITUTION] to Accompany Submission of the Dataset from _____ [ORIGINAL STUDY NAME[1]] for _____ [PROJECT TITLE FOR DATA TO BE SUBMITTED] to an NIH-designated data repository.

Dear ,
The submission of data to the NIH-designated data repository is being made with institutional approval from _____, along with appropriate institutional approvals from collaborating sites, as listed here:

[IF APPLICABLE ENTER COLLABORATING SITE NAMES HERE AND CLICK 'ADD TO LIST'] | *LIST OF COLLABORATING SITES*

Add to list >> | Clear list

The _____ hereby assures that submission of data from the study entitled _____ to an NIH-designated data repository meets the following expectations, as defined in the Genomic Data Sharing Policy:

- The data submission is consistent, as appropriate, with applicable national, tribal, and state laws and regulations as well as relevant institutional policies.[2]
- Any limitations on the research use of the data, as expressed in the informed consent documents, are delineated in the table on page 3.[3]
- The identities of research participants will not be disclosed to NIH-designated data repositories.
- An Institutional Review Board and/or Privacy Board, and/or equivalent body, as applicable, has reviewed the investigator's proposal for data submission and assures that:
  - The protocol for the collection of genomic and phenotypic data is consistent with 45 CFR Part 46.[4]
  - Data submission and subsequent data sharing for research purposes are consistent with the informed consent of study participants from whom the data were obtained;
  - Consideration was given to risks to individual participants and their families associated with data submitted to NIH-designated data repositories and subsequent sharing;
  - To the extent relevant and possible, consideration was given to risks to groups or populations associated with submitting data to NIH-designated data repositories and subsequent sharing; and
  - The investigator's plan for de-identifying datasets is consistent with the standards outlined in this Policy (see section IV.C.1.).

The data are to be made available through ☐ unrestricted[5] or ☒ controlled-access[6] (If unrestricted access is marked, the data use limitation table on page 2 does not need to be completed.)

The National Center for Biotechnology Information is authorized to upload the display of variant ☐ alleles and/or ☐ frequencies from this study in public variation archives (i.e., dbSNP and dbVar)[7].

* Certification must be provided for all sites contributing samples. If more than one site is contributing samples, the primary site may submit one Institutional Certification indicating that they are providing certification on behalf of all collaborating sites. Alternatively, each site providing samples may provide their own Institutional Certification.

Sincerely,

Authorized Institutional Official:[9]
Name: _____ Title: _____
Signature: _____ Date: _____

Investigator:
Name: _____ Title: _____
Signature: _____ Date: _____

All Kids First sequence data (e.g., BAMs, VCFs) are "controlled-access"

[1] For the submission of data derived from cell lines or clinical specimens lacking research consent that were created or collected before the effective date of this Policy, the Institutional Certification needs to address only this item.
[2] For guidance on clearly communicating inappropriate data uses, see NIH Points to Consider in Drafting Effective Data Use Limitation Statements, http://gwas.nih.gov/pdf/NIH_PTC_in_Drafting_DUL_Statements.pdf
[3] Aggregate-level data include summary statistics from the research study, such as allele frequencies or effect sizes and p-values for test of association. If "yes" is checked, your aggregate-level data will be included in the Compilation of Aggregate Genomic Data, a collection of analyses across many dbGaP studies that can be accessed with a single Data Access Request.
[4] The Single Nucleotide Polymorphism Database (dbSNP) is a public archive for genetic variation (apparently neutral polymorphisms, polymorphisms corresponding to known phenotypes, and regions of no variation) within and across species. The Database of Genomic Structural Variation (dbVar) is a collection of genomic structural variation data, typically 50 nucleotides in length or longer, for different organisms. For more information, see: http://www.ncbi.nlm.nih.gov/SNP/ and http://www.ncbi.nlm.nih.gov/dbvar/.
[5] 45 CFR Part 46. Protection of Human Subjects. See http://www.gpo.gov/fdsys/pkg/CFR-2011-title45-vol1/xml/CFR-2011-title45-vol1-part46.xml.
[6] As noted earlier, for studies using data or specimens collected before the effective date of this Policy, the IRB, privacy board, or equivalent body should review informed consent materials to ensure that data submission is not inconsistent with the informed consent provided by the research participants.
[7] Data made publicly available to anyone
[8] Data made available for secondary research only after investigators have obtained approval from NIH to use the requested data for a particular project.
[9] A senior official at an institution who is authorized to enter the institution into a legally binding contract and sign on behalf of an investigator who has submitted data to NIH, e.g., Dean, Vice President for Research.

Institutional Certification template

Extramural Institutional Certification

For guidance on drafting data use limitations, please refer to the NIH Points to Consider in Drafting Effective Data Use Limitation Statements found at: http://gds.nih.gov/pdf/nih_ptc_in_drafting_dul_statements.pdf. Data use limitations are developed based on the original informed consent from the participant.

Data Use Limitations (will be used in dbGaP to create Consent Groups)

General Research Use	GRU	Use of the data is limited only by the terms of the Data Use Certification:these data will be added to the dbGaP Collection.
Health/Medical/Biomedical	HMB	Use of the data is limited to health/medical/biomedical purposes, does not include the study of population origins or ancestry.
Disease-specific [list disease]	DS	Use of the data must be related to the specified disease.
Other		[ENTER CUSTOMIZED TEXT, IF APPLICABLE]

Data Use Limitation Modifiers

IRB approval required	IRB	Requestor must provide documentation of local IRB approval.
Publication required	PUB	Requestor agrees to make results of studies using the data available to the larger scientific community.
Collaboration required	COL	Requestor must provide a letter of collaboration with the primary study investigator(s).
Not-for-profit use only	NPU	Use of the data is limited to not-for-profit organizations.
Methods	MDS	Use of the data includes methods development research (e.g., development of software or algorithms)
Genetic studies only	GSO	Use of the data is limited to genetic studies only.

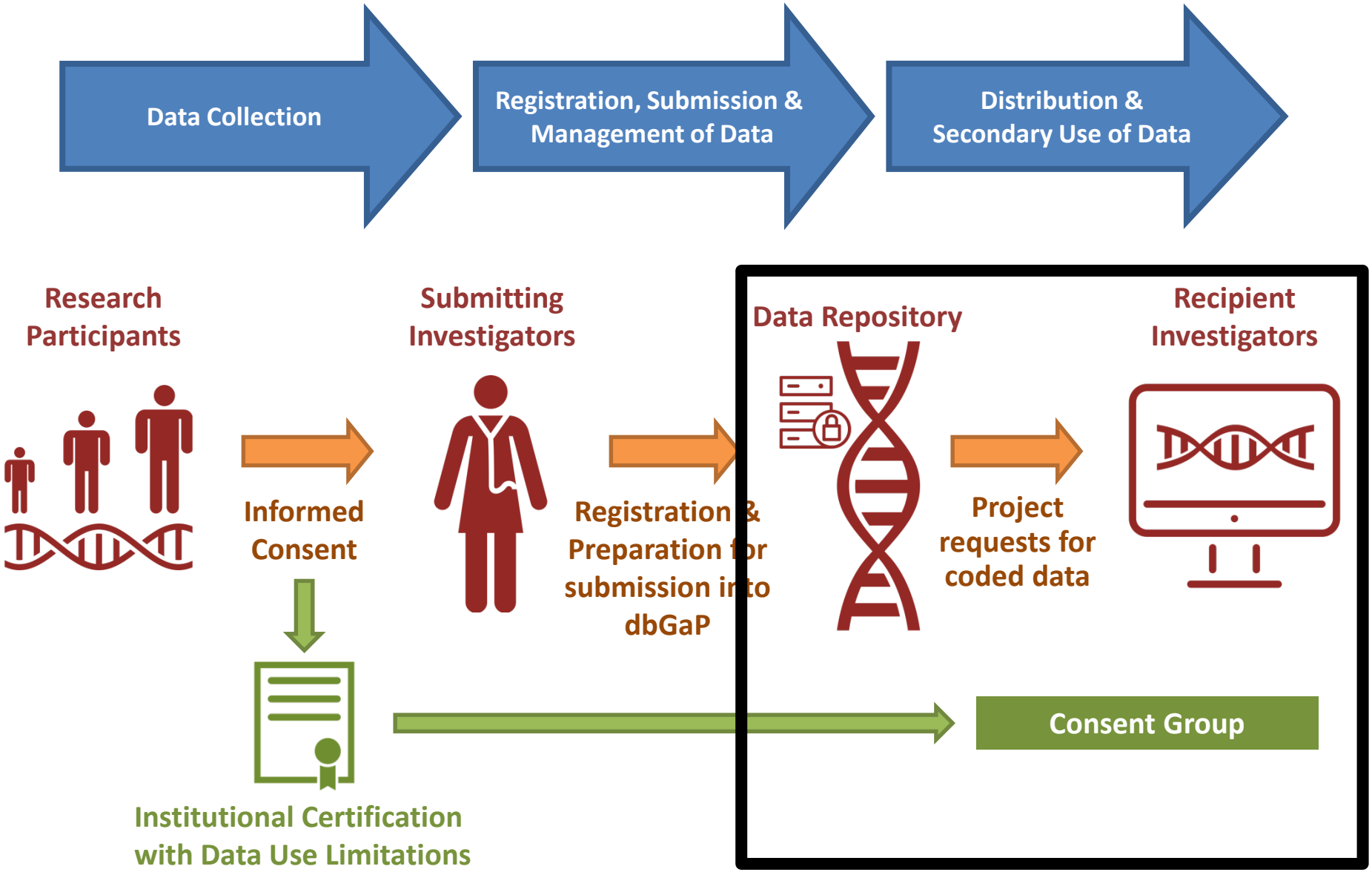
Using the tables above, please indicate in the form below the consent group(s) for each collaborating study site. Use one row per consent group.

Collaborating Site Name	Data Use Limitation	Data Use Limitation Modifiers
Eg: Cold Cohort Study	*Health/Medical/Biomedical*	IRB ☐ PUB ☐ COL ☐ NPU ☐ MDS ☐ GSO ☐
Eg: Cold Cohort Study	*Disease Specific Research* [*Lung Cancer*]	IRB ☒ PUB ☐ COL ☐ NPU ☒ MDS ☐ GSO ☐
	Select consent group title	IRB ☐ PUB ☐ COL ☐ NPU ☐ MDS GSO
	Select consent group title	IRB ☐ PUB ☐ COL ☐ NPU ☐ MDS GSO
	Select consent group title	IRB ☐ PUB ☐ COL ☐ NPU ☐ MDS GSO
	Select consent group title	IRB ☐ PUB ☐ COL ☐ NPU ☐ MDS GSO

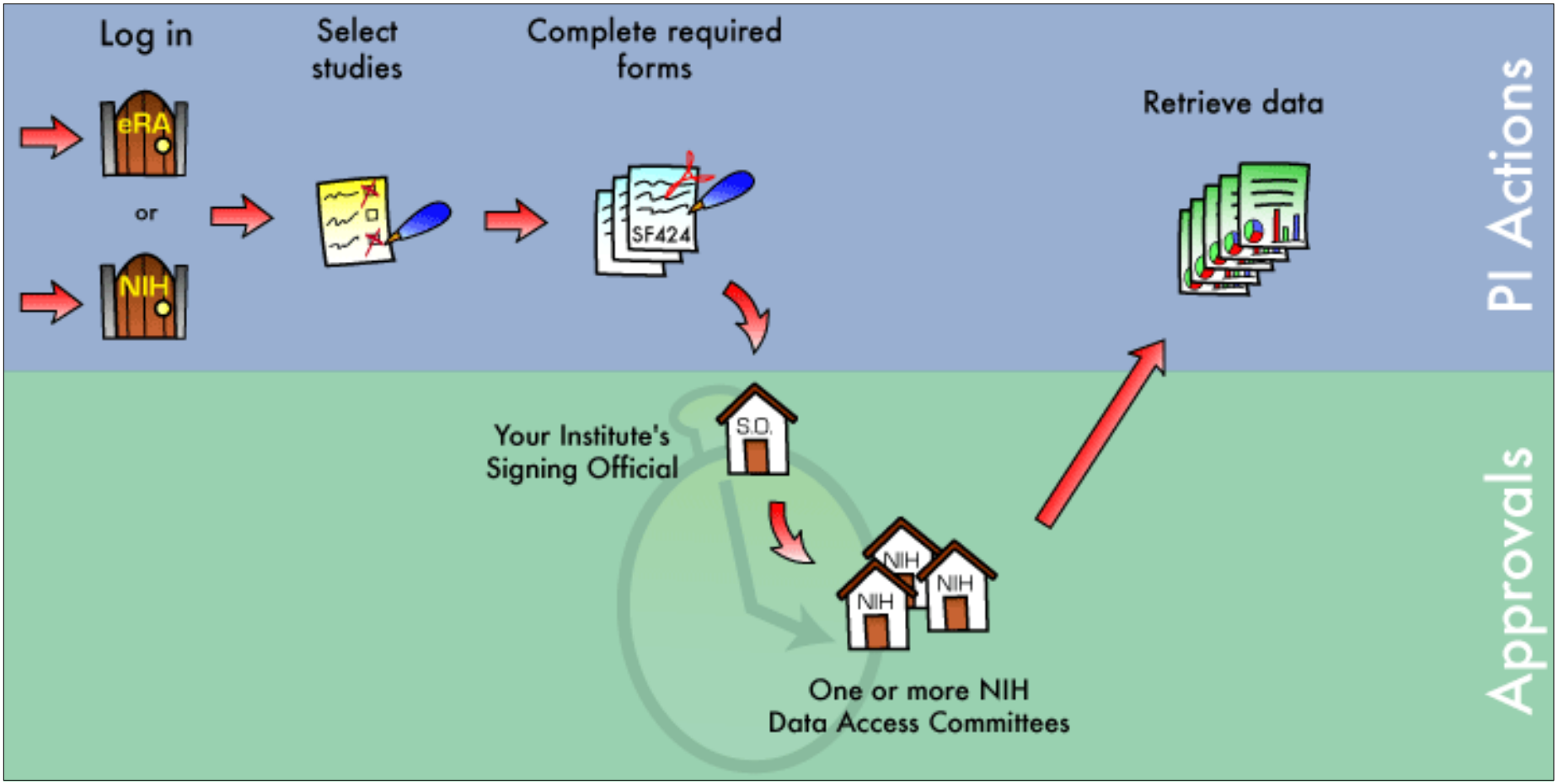
Data Use Certification (DUC)

- NIH-wide document outlining the conditions under which researchers may access data
  - DUC is generated for every study registered in dbGaP
- **Secondary users and their supporting Institution's Signing Official must agree to conditions in DUC before access will be granted**
  - Includes any Data Use Limitations pertinent to the requested dataset

dbGaP Data Sharing: The Big Picture



dbGaP Access Process (secondary user)



Data Use Limitations (DULs)

- DULs are delineated in the original informed consent language agreed to by the research participants who contribute tissue/samples to a study
- DULs are included in Institutional Certification and delineate appropriate research uses of the data, as well as research that is specifically excluded:
  - E.g. consent may have been broad and data available for "general research use," or consent may have been more specific and data restricted to "research related to cancer"
  - Once data is released for access in dbGaP, DULs/consent groups are used to determine which data access requests are approved (by the Data Access Committees)
- DULs have been standardized in dbGaP (can be customized if warranted due to consent language)

Standard dbGaP DULs & Modifiers

Consent Group Types/ DULs (Only ONE applies)

Name	Abbreviation
General Research Use	Use of the data is limited only by the terms of the model Data Use Certification.
Health / Medical / Biomedical	Use of this data is limited to health/medical/ biomedical purposes, does not include the study of population origins or ancestry.
Disease-specific [ ]	Use of the data must be related to [ ]
Other	[XX]

DULs & DUL modifiers are based on consent language, <u>not</u> PI preference!

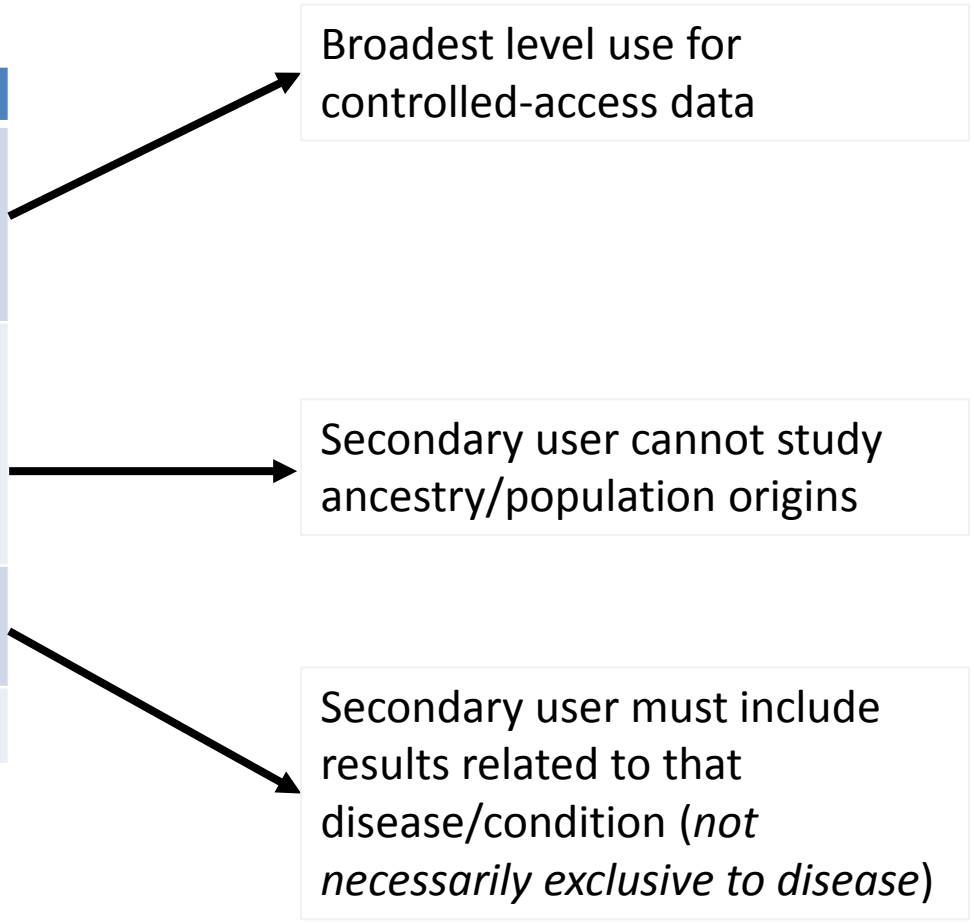
Consent Group Attributes/ DUL modifiers (ANY may apply)

Name	Abbreviation
Publication required	Requestor agrees to make results of studies using the data available to the larger scientific community.
Not-for-profit use only	Use of the data is limited to not-for-profit organizations.
Methods	Use of the data includes methods development research (e.g., development of software or algorithms)
IRB approval required	Requestor must provide documentation of local IRB approval.
Collaboration required	Requestor must provide a letter of collaboration with the primary study investigator(s).
Genetic studies only	Use of the data is limited to genetic studies only.

Standard dbGaP DULs

Consent Group Types/ DULs
(Only ONE applies)

Name	Abbreviation
General Research Use	Use of the data is limited only by the terms of the model Data Use Certification.
Health / Medical / Biomedical	Use of this data is limited to health/medical/ biomedical purposes, does not include the study of population origins or ancestry.
Disease-specific [ ]	Use of the data must be related to [ ]
Other	[XX]



Standard dbGaP DUL Modifiers

Only non-profits can access this dataset

Secondary user must have local IRB approval before accessing dataset

Secondary user must have established a collaboration with PI who collected/submitted data before accessing dataset

DUL modifiers are based on consent language, not PI or IRB preference! "Preferences" can be listed in the DUC so that end users are aware (e.g. non-commercial use, publication collaboration)

Consent Group Attributes/ DUL modifiers (ANY may apply)

Name	Abbreviation
Publication required	Requestor agrees to make results of studies using the data available to the larger scientific community.
Not-for-profit use only	Use of the data is limited to not-for-profit organizations.
Methods	Use of the data includes methods development research (e.g., development of software or algorithms)
IRB approval required	Requestor must provide documentation of local IRB approval.
Collaboration required	Requestor must provide a letter of collaboration with the primary study investigator(s).
Genetic studies only	Use of the data is limited to genetic studies only.

Tips for Requesting Kids First Data

- End user must address the DULs in the Data Access Request
  - Example: For a study limited to research on scoliosis (disease specific DUL), you must address how your results will relate to scoliosis, even if you are comparing this dataset to data from a cancer cohort.
- Some Kids First cohorts have multiple "consent groups" (different datasets have different DULs). Therefore, you may not be granted access to the full dataset
  - Example: If you are studying ancestry in orofacial cleft, you may not be granted access to datasets that are "HMB", but you could have access to datasets that are "GRU"
- If you are not granted access to a particular dataset, you may be able to collaborate and use aggregate results

Going Forward

The Kids First Data Resource Center is tasked with making genomic data from various childhood cancer and birth defect cohorts broadly accessible to the research community.

- Projects with broad data sharing/use (GRU, minimal modifiers) will be prioritized for sequencing to facilitate collaborative research among the pediatric research community.
- If any DULs or modifiers were erroneously included, a revised Institutional Certification may be submitted to your GPA (after IRB approval) and the dbGaP registration will be corrected.
- New consent forms should allow broad sharing.