

## Compilation of Aggregate Genomic Data

NIH has implemented a process for requesting access to certain aggregate-level genomic data in the Database of Genotypes and Phenotypes (dbGaP) through a single access request. The aggregate datasets included in this process have been designated as appropriate for general research use (GRU) by submitting institutions, which indicates that there are no further limitations on secondary research use beyond those outlined in the [dbGaP Genomic User Code of Conduct](#).

### Compilation of Aggregate Genomic Data

The authorized GRU aggregate datasets are combined into a single study, the [Compilation of Aggregate Genomic Data](#). These data will continue to remain in controlled access with the same expectations as other controlled-access data, including data management practices and annual reporting on data use. This Compilation does not include individual-level data.

### Submission Process

The [NIH Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies \(GWAS\)](#) expects that all submissions to dbGaP will be accompanied by an Institutional Certification from a responsible institutional official certifying approval of the data submission. Investigators are encouraged to submit aggregate datasets along with individual-level genomic data. If data use limitations are specified for the individual-level data, the Institutional Review Board should also determine whether the data use limitations that pertain to individual-level data would also apply to aggregate data. The Institutional Certification should note the determination and state whether the aggregate data are appropriate for general research use. If the submitting institution certifies that aggregate data from a project can be included in the Compilation, then a study description for the aggregate data should be provided by the investigator in addition to a description for the individual data.

### Access Process

The steps that secondary researchers need to follow to access the Compilation are no different from the steps taken to request access to other data in controlled access. Requestors will submit a data access request as well as a [Data Use Certification agreement](#), co-signed by the requestor and his or her Institutional Signing Official, and agree to the [Code of Conduct](#). Consistent with other access requests, requestors will be expected to provide a Research Use Statement explaining their intended use of the data, which will be made publicly available on the dbGaP website.

NIH review of the requests for access to the Compilation of Aggregate Genomic Data will be carried out by a central Data Access Committee. Investigators approved for access to the Compilation will have access to the data for the standard one-year approval period, including any datasets added to the study during that one-year period. GRU-designated aggregate datasets will be added to the Compilation after the publication embargo for the original study has expired. As with controlled access to individual-level data, access renewal will be considered after the submission of a new access request and an updated annual report. Investigators who do not request a renewal must close out the project when the access period ends by submitting a close-out report that includes a confirmation that the downloaded data were destroyed.