

Summary of Database of Genotype and Phenotype (dbGaP) Policy Compliance Violations (38 Cases, 2007 to July 1, 2018¹)

Brief Summary	Policy Expectation(s)	Actions Taken and/or Preventative Measures Implemented
Data Submission Incidents		
A submitted dataset included data from a research participant who had withdrawn from the study.	Submission of data to NIH and subsequent sharing for research purposes should be consistent with the wishes of the research participants from whom the data were obtained.	The mistake was discovered by the institution that submitted study data to dbGaP. NIH requested, and confirmed, that the 49 Approved Users ² who downloaded the data destroyed their copy of the dataset and all resulting analyses. NIH removed the participant's data from the dataset. The corrected version of the dataset was reposted.
An investigator erroneously submitted data from 77 research participants who did not consent to broad sharing of their data.	Submission of data to NIH and subsequent sharing for research purposes should be consistent with the informed consent of the study participants from whom the data were obtained.	NIH staff worked with the investigator to ensure that all affected data were removed from dbGaP. NIH requested, and confirmed, that 64 Approved Users who downloaded the data destroyed their copy of the dataset and all resulting analyses. The submitting investigator reported the incident to the Institutional Review Board, and the institution took corrective actions, such as developing institutional standard operating procedures regarding genomic data sharing and submission to the NIH. The corrected version of the dataset was reposted.

¹ As of July 1, 2018, there were 42,292 approved Data Access Requests.

² An Approved User is a user approved by the relevant Data Access Committee to access one or more datasets for a specified period of time and only for the purposes outlined in the Principal Investigator (PI)'s approved Research Use Statement. The Information Technology (IT) Director indicated on the Data Access Request, as well as any staff members and trainees under the direct supervision of the PI are also Approved Users and must abide by the terms laid out in the NIH Data Use Certificate Agreement.

Brief Summary	Policy Expectation(s)	Actions Taken and/or Preventative Measures Implemented
<p>An investigator submitted a dbGaP dataset without sufficiently random identifier codes for individual participants, and the dataset was released for access before the submitting investigators realized the error.</p>	<p>Data submitted to dbGaP should be de-identified and coded using a random, unique code.</p>	<p>NIH requested, and confirmed, that Approved Users who downloaded the data destroyed their copy of the dataset and all resulting analyses. As a preventive measure, the institution that submitted the dataset implemented an additional review in their system. The institution created the appropriate random codes for the dataset, and this version replaced the previous version in dbGaP.</p>
<p>NIH discovered that it was possible to deduce human genome sequence from microbial sequence data posted in the unrestricted access Sequence Read Archive database. The microbial sequence data was collected from human specimens, and the human DNA sequence data had not been thoroughly removed at the data cleaning stage.</p>	<p>Human DNA sequence data intended only for controlled-access distribution through dbGaP should not be available in an unrestricted access database.</p>	<p>NIH requested, and confirmed, that 35 Approved Users who downloaded the data destroyed their copy of the dataset and all resulting analyses. NIH worked with the submitting institutions to ensure that the affected files were removed from the unrestricted access Sequence Read Archive, re-filtered with improved data cleaning software to remove detectable human sequence, and reposted. Data that could be used to reconstruct any remaining human sequence was also removed.</p>

Brief Summary	Policy Expectation(s)	Actions Taken and/or Preventative Measures Implemented
<p>An investigator who submitted data to NIH inadvertently reversed the Data Use Limitations for two datasets. The dataset restricted to “Disease-specific” research use was labeled for “General Research Use,” and the “General Research Use” dataset was labeled for “Disease-specific” research use.</p>	<p>Submitting institutions certify that the appropriate research uses of the data are delineated.</p> <p>Research with the requested datasets must be consistent with the Data Use Limitations as delineated by the submitting institution.</p>	<p>NIH requested, and confirmed, that the five Approved Users who downloaded the data destroyed their copy of the datasets and all resulting analyses, and they confirmed that the data had been used in accordance with the correct Data Use Limitations. The submitting investigator implemented an additional mechanism for checking study data configuration prior to submission to NIH. The corrected version of the dataset was reposted.</p>
<p>While uploading human sequence data to controlled-access, the submitting investigator was erroneously able to upload the data to the unrestricted access Sequence Read Archive database because an NIH software filter that would have prevented the upload was inactive at the time.</p>	<p>Human DNA sequence data intended only for controlled-access distribution through dbGaP should not be available in an unrestricted access database.</p>	<p>NIH removed the data from the unrestricted access Sequence Read Archive and reactivated the Sequence Read Archive filter software. The European Bioinformatics Institute and DNA Data Bank of Japan also removed these data from their unrestricted access databases upon NIH request.</p>
<p>An investigator erroneously submitted data from 29 research participants who did not consent to broad sharing of their data.</p>	<p>Submission of data to NIH and subsequent sharing for research purposes should be consistent with the informed consent of the study participants from whom the data were obtained.</p>	<p>NIH staff worked with the investigator to ensure that all affected data were removed from dbGaP. NIH requested, and confirmed, that the five Approved Users who downloaded the data destroyed their copy of the dataset and all resulting analyses. The corrected version of the dataset was reposted.</p>
<p>Research or Data Access Incidents</p>		

Brief Summary	Policy Expectation(s)	Actions Taken and/or Preventative Measures Implemented
<p>Due to a dbGaP software configuration error, an Approved User was able to download an unapproved dataset.</p>	<p>Controlled-access data will be accessed by Approved Users only.</p>	<p>NIH requested, and confirmed, that the Approved User who downloaded the unapproved dataset destroyed their copy of the dataset and all resulting analyses. As a preventive measure, NIH conducted an audit of its system and changed how dbGaP’s software was configured.</p>
<p>An Approved User accessed data that was part of a study the Approved User helped conduct. The Data Access Committee failed to determine whether the Approved User had access to personal or identifying information for the research participants which may have meant that Institutional Review Board approval was required before accessing the data.</p>	<p>Approved Users who have access to personal identifying information for research participants in the original study at their institution or through their collaborators may be required to obtain Institutional Review Board approval in order to access data.</p>	<p>The Approved User’s access to the dataset was suspended until the Data Access Committee confirmed that the Approved User did not have access to personal identifying information. As a preventive measure, the Data Access Committee implemented multiple checkpoints in its review process. The NIH Data Use Certification agreement was updated to clarify when Institutional Review Board approval is required.</p>
<p>An Approved User was approved to access dbGaP datasets without including an Information Technology Director³ on the Data Access Request.</p>	<p>The Information Technology Director should be listed as a “Senior/Key Person” in the Data Access Request.</p>	<p>The Data Access Committee implemented additional checks in the Data Access Request review process, and NIH modified the dbGaP system to make the Information Technology Director requirement more apparent.</p>

³ “Information Technology (IT) director” is expected to be a senior IT official with the necessary expertise and authority to affirm the IT capacities at an academic institution, company, or other research entity. The IT Director is expected to have the authority and capacity to ensure that the NIH Security Best Practices for Controlled-Access Data Subject to the NIH GDS Policy and the institution’s IT security requirements and policies are followed by the Approved Users.

Brief Summary	Policy Expectation(s)	Actions Taken and/or Preventative Measures Implemented
<p>An Approved User conducted research that was not stated in the Data Access Request.</p>	<p>Approved Users are expected to limit research use to that described in the approved Data Access Request.</p>	<p>The Approved User's access to dbGaP datasets was suspended for three months. The Approved User may submit a new Data Access Request after the suspension period. NIH also recommended that the Institutional Signing Official⁴ receive training in oversight of Approved User projects. The institution sent a memo to all Approved Users stressing the importance of adhering to the terms of access, and the Institutional Signing Official will review all Research Use Statements and Data Use Limitations in the future and remind Approved Users that changes in the scope of their project requires submission of a revised statement.</p>
<p>An Approved User conducted research that was not stated in the Data Access Request.</p>	<p>Approved Users are expected to limit research use to that described in the approved Data Access Request.</p>	<p>In this case, the Approved User recognized the error and contacted NIH to report the mistake. NIH suspended the Approved User's access to the data for three months. The Approved User may submit a new Data Access Request after the suspension period.</p>

⁴ The Signing Official has institutional authority to legally bind the institution in grants administration matters. The label, "Signing Official," is used in conjunction with the NIH eRA Commons. The individual fulfilling this role may have any number of titles in the grantee organization, but is typically located in its Office of Sponsored Research or equivalent.

Brief Summary	Policy Expectation(s)	Actions Taken and/or Preventative Measures Implemented
<p>Due to a coding mistake in the dbGaP system software that misconfigured a dataset, an Approved User who had been approved to access a certain subset of data was able to download data from another subset. The Approved User did not have approval to access the other dataset and should not have downloaded it.</p>	<p>Controlled-access data will be accessed only by Approved Users.</p>	<p>NIH requested, and confirmed, that the Approved User who downloaded the unapproved dataset destroyed their copy of the data and all resulting analyses. NIH corrected the coding mistake configuration error, and, as a preventive measure, implemented an additional configuration review protocol.</p>
<p>An Approved User who moved to a new institution downloaded data at the new institution without submitting a new Data Access Request.</p>	<p>Approved Users agree that if they change institutions during the access period, they will close their project before leaving, and submit a new Data Access Request from the new institution if they wish to continue their work.</p>	<p>NIH requested, and confirmed, destruction of the data. The Approved User was allowed to submit a new Data Access Request and was warned that a future violation could lead to access suspension.</p>
<p>NIH incorrectly configured a study file resulting in the reversal of the Data Use Limitations for two datasets. The dataset labeled for “Non-profit Research Use Only” was incorrectly labeled “General Research Use” and the “General Research Use” dataset was labeled for “Non-profit Research Use only.”</p>	<p>Research with the requested datasets must be consistent with the Data Use Limitations as delineated by the submitting institution.</p>	<p>NIH requested, and confirmed, that the two Approved Users who had downloaded the data destroyed their copy of the datasets and all resulting analyses, and they confirmed that the data had been used in accordance with the correct Data Use Limitations. NIH also adjusted the processing software and, as a preventive measure, implemented additional checks. The corrected version of the dataset was reposted.</p>
<p>An Approved User conducted research that was not stated in the approved Data Access Request.</p>	<p>Approved Users are expected to limit research use to that described in the approved Data Access Request.</p>	<p>NIH suspended the Approved User’s access to the data for three months. The Approved User may submit a new Data Access Request after the suspension period.</p>

Brief Summary	Policy Expectation(s)	Actions Taken and/or Preventative Measures Implemented
<p>An Approved User conducted research that was not stated in the approved Data Access Request.</p>	<p>Approved Users are expected to limit research use to that described in the approved Data Access Request.</p>	<p>NIH suspended the Approved User’s access to the data for three months. The Approved User may submit a new Data Access Request after the suspension period.</p>
<p>NIH incorrectly configured a study file resulting in a swap of participants between the group consented for “Disease-specific” research only and “General Research Use.”</p>	<p>Research with the requested datasets must be consistent with the Data Use Limitations as delineated by the submitting institution.</p>	<p>NIH requested, and confirmed, that the nine Approved Users who downloaded the data destroyed their copy of the datasets and all resulting analyses. The corrected version of the dataset was reposted.</p>
<p>A dbGaP audit identified compressed archive “MULTI” files distributed with several studies that were improperly configured by including individual-level phenotype data that was restricted to “Disease-specific” research use. The MULTI files are intended to distribute data files characterizing the entire study or analysis such as pedigrees, release notes, or analysis-specific sample attributes, where redistribution is not restricted by Data Use Limitations. The restricted individual-level phenotype data should not have been included in these MULTI files.</p>	<p>Research with the requested datasets must be consistent with the Data Use Limitations as delineated by the submitting institution.</p>	<p>NIH requested, and confirmed, that Approved Users who downloaded the data destroyed their copy of the datasets and all resulting analyses. dbGaP developed automated checks of data files to confirm that the contents of MULTI files are appropriate. In addition, submitting investigators will be asked to review and certify the final packaging of files before their release. The corrected version of the dataset was reposted.</p>

Brief Summary	Policy Expectation(s)	Actions Taken and/or Preventative Measures Implemented
<p>An Approved User uploaded controlled-access dbGaP data to a third party data sharing site, which was designed for the sharing of unrestricted access data in a password protected working group space. The controlled-access dbGaP data were downloaded by three unapproved users.</p>	<p>Controlled-access data will be accessed only by Approved Users.</p> <p>Approved Users must not distribute controlled-access data to any entity (e.g., a third party) not included in the Data Access Request.</p>	<p>The Approved User notified NIH of the inappropriate data sharing within 24 hours of becoming aware of it. The data were promptly removed from the data sharing site, and NIH confirmed that the three unapproved users who downloaded the data destroyed their copy of the dataset and all resulting analyses. NIH suspended the Approved User's access to the data for six months. The Approved User may submit a new Data Access Request after the suspension period.</p>
<p>An Approved User uploaded controlled-access dbGaP data to a third party cloud computing platform. Only Approved Users had access to the data, no unauthorized access occurred.</p>	<p>Approved Users must not distribute controlled-access data to any entity (e.g., a third party) not covered in the Data Access Request.</p> <p>Approved Users and institutions must adhere to dbGaP security best practices, which at the time of the incident did not support the use of cloud computing.</p>	<p>The Approved User notified NIH of the inappropriate data sharing within 24 hours of becoming aware of it. NIH requested that dbGaP controlled-access data be removed from the cloud, which was confirmed. NIH suspended the Approved User's access for two months. The Approved User's institution sent a notice out to all Approved Users about the importance of best practices and data security. The Approved User may submit a new Data Access Request after the suspension period.</p>
<p>An Approved User conducted research that was not stated in the Data Access Request. The research was inconsistent with the Data Use Limitations assigned to the datasets.</p>	<p>Approved Users are expected to limit research use to that described in the approved Data Access Request.</p> <p>Research with the requested datasets must be consistent with the Data Use Limitations as delineated by the submitting institution.</p>	<p>NIH suspended the Approved User's access to the data for six months. The Approved User may submit a new Data Access Request after the suspension period.</p>

Brief Summary	Policy Expectation(s)	Actions Taken and/or Preventative Measures Implemented
An Approved User sent unencrypted controlled-access datasets to other Approved Users via email.	Approved Users and institutions must adhere to the NIH Security Best Practices for Controlled-Access Data Subject to the NIH Genomic Data Sharing Policy.	NIH suspended the Approved User's access to the data for one month. The Approved User may submit a new Data Access Request after the suspension period.
An Approved User posted controlled-access participant attribute information on a public database website.	Controlled-access data will be accessed only by Approved Users.	NIH suspended the Approved User's access to the data for three months. The Approved User may submit a new Data Access Request after the suspension period.
An Approved User shared controlled-access data with a journal's editorial staff, who were not approved by NIH to access these data. The journal requested the data to accompany peer review of the approved user's manuscript.	Controlled-access data will be accessed only by Approved Users.	The journal and peer reviewers destroyed the data per NIH request.
An Approved User conducted research that expanded the scope of work in the Data Access Request prior to receiving Data Access Committee approval to expand the scope of work.	Approved Users are expected to limit research use to that described in the approved Data Access Request.	NIH suspended the Approved User's access to the data for three months. The Approved User may submit a new Data Access Request after the suspension period.
A Data Access Committee inadvertently approved a Data Access Request that was inconsistent with the Data Use Limitations assigned to the datasets.	Research with the requested datasets must be consistent with the Data Use Limitations as delineated by the submitting institution.	The Approved User destroyed the data per NIH request. As a preventive measure, the Data Access Committee implemented additional checkpoints in its review process.

Brief Summary	Policy Expectation(s)	Actions Taken and/or Preventative Measures Implemented
<p>NIH incorrectly configured a study file resulting in the reversal of the Data Use Limitations for two datasets. The dataset labeled for “Non-profit Research Use Only” was incorrectly labeled “General Research Use” and the “General Research Use” dataset was labeled for “Non-profit Research Use Only.” Despite the incorrect configuration, research conducted with these data was consistent with the Data Use Limitations.</p>	<p>Research with the requested datasets must be consistent with the Data Use Limitations as delineated by the submitting institution.</p>	<p>NIH requested, and confirmed, that Approved Users who downloaded the data destroyed their copy of the datasets and all resulting analyses. The corrected version of the dataset was reposted.</p>
<p>An Approved User conducted research that was not stated in the Data Access Request. The research was inconsistent with the Data Use Limitations assigned to the datasets. Results of the research were included in a manuscript posted on a pre-print server.</p>	<p>Approved Users are expected to limit research use to that described in the approved Data Access Request.</p> <p>Research with the requested datasets must be consistent with the Data Use Limitations as delineated by the submitting institution.</p>	<p>NIH suspended the Approved User's access to the data for six months. NIH confirmed that the Approved User who downloaded the data destroyed their copy of the dataset and all resulting analyses including the analyses from the manuscript that was submitted to a journal. The Approved user may submit a new Data Access Request after the suspension period; however, NIH recommended that a senior Institutional Official review the Approved User's Research Use Statement and dataset Data Use Limitations in the future.</p>
<p>An NIH processing error resulted in the generation of incorrect data file identifiers used to organize the data into the appropriate consent groups.</p>	<p>Submission of data to NIH and subsequent sharing for research purposes should be consistent with the informed consent of the study participants from whom the data were obtained.</p>	<p>NIH requested, and confirmed, that five Approved Users who downloaded the data destroyed their copy of the dataset and all resulting analyses. The corrected version was reposted.</p>

Brief Summary	Policy Expectation(s)	Actions Taken and/or Preventative Measures Implemented
<p>A Data Access Committee inadvertently approved a Data Access Request and the erroneously Approved User downloaded the data.</p>	<p>Controlled-access data will be accessed only by Approved Users.</p>	<p>NIH requested, and confirmed, that the erroneously Approved User who downloaded the data destroyed their copy of the dataset and all resulting analyses. As a preventive measure, the Data Access Committee implemented additional checkpoints in its review process.</p>
Data Security Incidents		
<p>A dbGaP software bug caused a disapproval decision for a Data Access Request to be recorded as an approval, and the Approved User downloaded the data.</p>	<p>Controlled-access data will be accessed only by Approved Users.</p>	<p>NIH confirmed that the Approved User who downloaded the data destroyed their copy of the dataset and all resulting analyses. NIH corrected the software problem, and implemented other audit review procedures as a preventative measure.</p>
<p>The security system at an Approved User’s institution was discovered (by the Approved User) to be vulnerable to security breaches. The institution corrected the problem before a breach occurred.</p>	<p>Approved Users and institutions must adhere to dbGaP security standards. Approved Users agree to notify NIH of security breaches within 24 hours.</p>	<p>The institution undertook a thorough analysis of the problem. The institution implemented additional security measures to protect machines used for analyzing controlled-access dbGaP data.</p>
Publication Embargo Incidents		
<p>An Approved User’s trainee submitted an abstract for a scientific meeting prior to the expiration of the embargo date of a dbGaP dataset. The violation was discovered prior to the meeting when the trainee contacted NIH staff to clarify how to acknowledge the dataset in presentations.</p>	<p>Approved Users will not make presentations or submit manuscripts for publication before the embargo date(s) of dbGaP datasets.</p>	<p>The NIH requested, and confirmed, withdrawal of the abstract. As a preventive measure, the Approved User reminded trainees about NIH expectations for research using dbGaP datasets. NIH improved the visibility of the embargo date information on the study page and in the NIH Data Use Certification agreement.</p>

Brief Summary	Policy Expectation(s)	Actions Taken and/or Preventative Measures Implemented
<p>An Approved User's manuscript was published prior to the dataset's publication embargo date.</p>	<p>Approved Users will not make presentations or submit manuscripts for publication before the embargo date(s) of controlled-access dbGaP datasets.</p>	<p>NIH suspended the Approved User's access to all dbGaP data for six months. NIH requested, and confirmed, destruction of their downloaded data. NIH requested a report from the investigator's institution describing the circumstances surrounding the violation. NIH quickly published an editorial in the same journal emphasizing the importance of the embargo policy. The Approved User may submit a new Data Access Request after the suspension period.</p>
<p>A submitting investigator erroneously set the embargo period for two related dbGaP datasets at zero months, the default setting in the system, instead of 12 months. NIH was notified by the submitting investigator that an Approved User had submitted a journal manuscript that included analysis of the datasets within the 1-year embargo period.</p>	<p>Approved Users will not make presentations or submit manuscripts for publication before the embargo date(s) of controlled-access dbGaP datasets.</p>	<p>The investigator who submitted the data reviewed the manuscript, and provided permission for the embargo date to be waived in this one case. NIH corrected the embargo dates and set the default embargo period in the study registration system for 12 months from the date the study was released. All Approved Users were notified by NIH of the correct embargo date.</p>
<p>An Approved User's manuscript was published prior to the dataset's publication embargo date.</p>	<p>Approved Users will not make presentations or submit manuscripts for publication before the embargo date(s) of controlled-access dbGaP datasets.</p>	<p>NIH suspended the Approved User's and collaborator's access to controlled-access dbGaP data for six months, and confirmed destruction of all downloaded data and subsequent analyses. NIH requested that the Approved User inform the journal of the embargo violation. The Approved User may submit a new Data Access Request after the suspension period.</p>
<p>An Approved User presented findings at a conference prior to the dataset's publication embargo date.</p>	<p>Approved Users will not make presentations or submit manuscripts for publication before the embargo date(s) of</p>	<p>NIH suspended the Approved User's access to controlled-access dbGaP data for six months, and confirmed destruction of all</p>

Brief Summary	Policy Expectation(s)	Actions Taken and/or Preventative Measures Implemented
	controlled-access dbGaP datasets.	downloaded data and subsequent analyses. The Approved User may submit a new Data Access Request after the suspension period.