Policy Name: Guidelines for Availability of Data Sets	**Policy Number:** 15.1
Section: Data Sharing – 15	**Date Revised:** June 1, 2021

15 Data sharing

Each Alliance study has a formal protocol document, which includes a statement of the objectives of the study. Patient consent and authorization are obtained to collect the individual patient data required for addressing the study objectives. These data are transmitted from the treating or enrolling institution to the Alliance Statistics and Data Management Center (SDMC), where these data are reviewed, processed and stored in the Alliance database. Not all information submitted becomes part of the electronic database; for example, only some information on supporting documents such as operative and pathology reports may be entered into the database. The electronic database is used as the basis for analyses of Alliance studies, with the analyses performed by the staff at the Alliance SDMC.

The procedures described here do not cover requests – from the National Cancer Institute (NCI), the Food and Drug Administration (FDA), or other federal agencies – for information required by federal regulations or by the terms of the grant awards from federal agencies (e.g., Cancer Therapy Evaluation Program (CTEP), Division of Cancer Prevention (DCP) and Division of Cancer Control and Population Sciences (DCCPS), NCI, and National Institutes of Health (NIH) to the Alliance. Such requests will be honored as expeditiously as possible.

This policy covers requests for existing data, not requests for collection of additional data. Requests for individual-level genomic or other high-dimensional data not used in the primary publication (see section 15.4) may be subject to other NCI and NIH regulations.

Data requested by an investigator can include images and/or data generated from Alliance laboratory correlative studies. However, requests for use of biospecimens are covered by a separate evaluation and review procedure described in section 11.

The sharing of data with industry is further described in section 13. However, in cases where industry requests data from studies in which it has not participated, it would follow the procedure indicated in this section.

15.1 Guidelines for availability of data sets

For phase III studies, it is anticipated that individual-level de-identified data sets that would be sufficient to reproduce results provided in a publication (i.e., published manuscript) containing the primary study analysis, will be available via the NCTN Data Archive generally within six months of publication of the manuscript. It is anticipated that data sets containing patient-level entry data of all baseline variables summarized in the publication will be available within 12 to 15 months after the publication of the primary analysis. The NCTN Data Archive has its own requesting procedures: https://nctn-data-archive.nci.nih.gov. Some data may also be available via Project Data Sphere:

https://www.projectdatasphere.org/projectdatasphere/html/home. If the desired data are not contained within the NCTN Data Archive or Project Data Sphere, these data will be available to individuals via the requesting procedures described in section 15.2.

For non-phase III studies, a patient data set containing the variables analyzed in the primary results paper will be available upon request (subject to restrictions in sections 15.3 and 15.4). This process could take several months, based on the type of request and workload amount/priorities of the SDMC. The release of data may also be constrained in cases where the sample sizes are too small to reliably de-identify data

Data sets from the following types of publications will be available via the NCTN Data archive: publications reporting QOL or toxicity results from one or more phase III trials; publications reporting updated phase III trial data, e.g., updated survival data; and publications reporting the results of a biomarker analysis from a phase III trial (including subset analyses based on a biomarker), or meta-analyses of multiple trials (some of which are phase III). If desired data are not available in the NCTN Data archive for publications that are not presenting the primary analysis of the trial, patient data sets containing the variables analyzed in the manuscript will be available upon request (subject to restrictions in sections 15.3 and 15.4). This process could take several months depending on workload and prioritization within the SDMC

Release of data collected in a clinical trial conducted under a binding collaborative agreement between CTEP and a pharmaceutical/biotechnology company must be in compliance with the terms of the binding collaborative agreement and must be approved by CTEP and the company. Release of data is also subject to the terms of any contracts between the Alliance and other entities, which cover any of the requested data. These two considerations could, in some instances, delay the release of data to requesting investigators.

15.2 Request procedures

While most analyses of Alliance studies are performed at the Alliance SDMC, the Alliance also makes research data available to other investigators, as required by the policies of the NIH. An investigator who wishes to use individual patient data from one or more of the Alliance studies that are not available through NCTN Data Archive or Project Data Sphere must make a formal request to the Alliance Chicago Office.

The Alliance requires the investigator to fill out a formal request form, available on the Alliance website. The Alliance also requires the investigator to sign a data release specifying who will have access to the individual patient data and specifying that it will not be shared with others outside this specified set of individuals unless first approved by the Alliance.

There will be no scientific review of requests for data. If the Alliance is unable to fulfill a request, the Alliance will inform the investigator(s) of the reason the request cannot be fulfilled. In most cases it is likely the investigator(s) will be able to amend the request to comply with the procedures. If the Alliance believes the request will not be amendable, the Alliance will inform the investigator of the appeals process outlined in section 15.6, and also notify the lead chief of the Clinical Investigations Branch (CIB) of CTEP in the Division of Cancer Treatment and Diagnosis (DCTD) at the NCI, the lead NCTN program director, and the DCP or NCORP Director, as appropriate. Release of data is subject to the disclaimer in section 15.5.

15.3 Regulatory considerations

All research use of data collected on human subjects from network group studies led by the Alliance Central Protocol Operations Program and Alliance SDMC is subject to applicable Office of Human Research Protections (OHRP) regulations and to applicable regulations of the Privacy Rule of the Health Insurance Portability and Accountability Act (HIPAA). Generally, patients consent to have their data stored and possibly shared for the purpose of future research, with protections for privacy. The standard policy is to provide data that have been rendered fully anonymous, de-identified, or coded. IRB approval from an investigator's institution may be required to fulfill requests for non-de-identified data.

Guidance on these matters can be found in the OHRP document "Guidance on Research Involving Coded Private Information or Biological Specimens" located at http://www.hhs.gov/ohrp/policy/cdebiol.html. Information is also available on the NIH website (http://privacyruleandresearch.nih.gov/clin_research.asp) for Clinical Research and the HIPAA Privacy Rule. The criteria for de-identification of data under HIPAA are given in the Code of Federal Regulations, Part 46, Section 164.514. It is possible to conduct most projects using coded data (as described in the OHRP Guidance) that meet the criteria for a limited data set that can be released under a data use agreement (as described in Part 46 of the CFR, Section 164.512 and in the NIH HIPAA guidance documents), without obtaining additional patient consent or authorization.

15.4 Bioinformatic data sharing

15.4.1 NIH data sharing policies

Within the Alliance, 'bioinformatics data' is understood to encompass large scale genomic and molecular data from high-dimensional assays for which the number of markers considerably exceeds the number of patients. Assay types used most often include microarrays and next generation sequencing. In accordance with NIH data sharing policies, bioinformatic data generated from Alliance studies are deposited into the database on Genotypes and Phenotypes (dbGaP) or other appropriate public data repository.... The study team statisticians and bioinformaticists, commonly but not always within the Alliance Computational Genomics and Bioinformatics (CGB) Unit, are responsible for this process. It is expected that the corresponding bioinformatics data sharing policies will, of necessity, evolve as NIH policies regarding large scale data evolve.

15.4.2 Alliance bioinformatics studies

Alliance bioinformatics studies are typically conducted as substudies to Alliance clinical trials.

15.4.2.1 Bioinformatic data

De-identified (coded) high throughput genotype data and other large-scale 'omic data (which may include primary analysis files and/or intermediate files) will be made available to public repositories (such as dbGaP) according to NIH policies. The study team associated with analyses of trial and genomic data will determine when quality control studies have been completed, and will prepare data for submission. The Alliance SDMC CGB Unit is not responsible for data deposits/sharing in situations in which they are not members of the study analysis team. Publications by others that make use of only Alliance bioinformatic data (for example, as control data for other studies) may be published at any time after submission.

15.4.2.2 Phenotype data

Phenotype data used in the bioinformatics studies will be submitted at the completion of the trial once all data have been subject to quality and integrity checks. All phenotype data that are part of the

Alliance electronic database, have been checked for quality and integrity, and are used in genetic studies will be deposited. The Alliance statistician associated with analyses of trial and bioinformatic data will determine when the standard Statistics and Data Management Center quality control processes have been completed and will prepare data for submission. Publications by others making use of Alliance phenotype data (with or without genotype data) will be embargoed until after publication of the primary paper reporting the primary endpoint results of the clinical trial. As in the case of any Alliance data sharing request, no phenotype data on a Data and Safety Monitoring Board (DSMB) monitored study, will be released without a formal approval from the DSMB.

15.5 Release conditions and disclaimer

A simple, formal data release form specifying who will have access to the individual patient data (and specifying that it will not be shared with others outside this specified set of individuals), as well as covering the release conditions described below and the regulatory considerations described in sections 15.3 and 15.4 above, is required.

It is anticipated that most data requests can be provided as non-complex data sets in electronic form.

In releasing data, the Alliance makes no representations and extends no warranties of any kind, either expressed or implied. There are no expressed or implied warranties of merchantability or fitness for a particular purpose, or that the use of data will not infringe any patent, copyright, trademark, or other proprietary rights. No indemnification for any loss, claim, damage, or liability is intended or provided.

Copies of any abstract or manuscript arising from the project associated with data requests must be sent to the Alliance prior to submission. Alliance requires review and approval of any abstract or manuscript arising from any Alliance-led analyses.

When abstracts or manuscripts are based on Alliance-led analyses using data shared from multiple studies, authorship will include all members of the research team involved in the current research. The Alliance encourages the inclusion of the original study teams in assignment of abstract or manuscript authorship credits. This applies to Alliance-led meta-analyses of data sets obtained from the Alliance, as well as to meta-analyses of Alliance data sets available through the NCTN Data Archive or Project Data Sphere. Full authorship guidelines and further publications guidelines are contained in section 10.4.3.2.

Policy Name: Appeals Process	**Policy Number:** 15.6
Section: Data Sharing – 15	**Date Revised:** June 1, 2021

15.6 Appeals process

If a request for data is denied, the applicant may appeal the decision. The appeal is reviewed by the Alliance Group Chair, the lead NCTN or NCORP Program Director (as applicable), CTEP or DCP Associate Director or his/her designee (as applicable), and an outside statistician (i.e., a statistician who is not a member of the Alliance). The outside statistician is named jointly by the Alliance Group Chair and the lead NCTN or NCORP Program Director.

15.7 Fees

Routine costs associated with preparing standard data sets are viewed by NCI as covered by grants for the Alliance Operations Center and Alliance SDMC funded under the NCTN Program, or the Alliance Research Base under the NCORP program. Fees will not be charged for the release of non-complex electronic data sets.

Sometimes data requested for analyses will not all be coded in the Alliance database but will be available from supplementary material that was submitted as part of the trial. In this case, data would need to be abstracted from the supplementary material. Data abstractions can only be performed if adequate funding to support the abstraction is available. Even if funding is available, the Alliance may not have staff available to perform the abstraction. In this situation, Alliance may consider inviting the investigator(s) to the Alliance SDMC to perform the abstraction. Some funding for clerical support may still be required.

For biometric requests, sometimes data requested will not have been generated during completion of the original work, and thus would require additional bioinformatic analyses to generate the requested data. In this case, raw files would be shared so that the requestor can generate the needed data at their expense. A fee may be required for additional analysis.

Likewise, when data requested require data sets not available in easily obtained electronic format, especially for legacy trials, the Alliance may require funding for support to create the data set in an electronic format. Any fees will be limited to the actual time, effort, and materials required for preparing and documenting the data set.