Policy Name: Guidelines for Availability of Data Sets	**Policy Number:** 15.1
Section: Data Sharing – 15	**Date Revised:** March 15, 2013

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 Center (SDC), where the 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 the analysis of Alliance studies, with the analyses performed by the staff at the Alliance SDC.

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], 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.

The data requested by an investigator can include 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.

15.1 Guidelines for availability of data sets

For phase 3 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 to individuals via the requesting procedures described in section 15.2 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.

For non-phase 3 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 SDC. Since these studies could be quite small, the release of data may also be constrained by the ability to de-identify data.

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 SDC.

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.

Policy Name: Request Procedures	**Policy Number:** 15.2
Section: Data Sharing – 15	**Date Revised:** March 15, 2013

15.2 Request procedures

While most analyses of Alliance studies are performed at the Alliance SDC, 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 must make a formal request to the Alliance Chicago Office.

The Alliance requires documentation, which includes a brief description of the project, as well as documentation of Institutional Review Board (IRB) approval or exemption from the institution of the requesting investigator (see section 15.3). The Alliance also requires the investigator to sign a data use agreement specifying who will have access to the individual patient data and specifying that it will not be shared with other 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 and the lead NCTN program director. Release of the 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 SDC 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 have only consented to have their health information used for the objectives of the clinical trial in which they participated. Use of the data for other research projects is allowed only if an IRB has determined that use of the data in the project meets the minimal risk criteria for conducting the research without the patient's consent, if the use of the data in the project is exempt from consent requirements, or if the project does not constitute human subjects research. The required level of review or approval will generally depend on the degree to which the data have been rendered fully anonymous, de-identified, or coded.

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 Genomic data sharing

15.4.1 NIH data sharing policies

In accordance with NIH data sharing policies, genomics data generated from Alliance studies are deposited into the database on Genotypes and Phenotypes (dbGaP). The Alliance Bioinformatics Unit is responsible for this process.

NIH data sharing policies are evolving as a consequence of the cost of investments in large-scale data generation in high-throughput genotyping and sequencing of samples collected in NIH-funded research studies. There is strong consensus that making data and complete results of studies broadly available to the scientific community helps to insure that the investments made in data collection and genomic (and other " –omic" studies requiring substantial investment of resources and generating large-scale data) studies provides the greatest benefit to stakeholders including NIH, taxpayers and the scientific community.

At the same time, there is appropriate concern for maintaining the privacy of patients participating in such studies, and respecting the consent procedures within existing studies. Unique challenges to full compliance with data sharing policies arise in the context of clinical trials because some of the outcomes measured within clinical trials are available relatively early in the trial, while some of the key outcomes are not available until the primary endpoint(s) of the trial are met. Because high-throughput genomic data can be generated very rapidly and may be appropriately applied to outcomes available early in the trial, the desire to facilitate data sharing and fully comply with NIH data sharing policies will inevitably collide with long-standing practices in clinical trials research that have traditionally precluded sharing of data from a study until it is completed.

15.4.2 Alliance genomics studies

Alliance genomics studies are typically conducted as companions to Alliance clinical trials. To ensure that the data sharing process addresses the concerns of all parties involved, a steering committee consisting of the relevant committee and study chairs and statisticians will be formed for each genomic study. Any steps to be taken with respect to data sharing will be reviewed by the relevant steering committee and proceed only upon approval.

It should be noted that as NIH policies regarding high-throughput genomic data sharing continue to evolve, it is expected that the corresponding Alliance policies will of necessity evolve as well.

15.4.2.1 Genotype data

De-identified (coded) high throughput genotype data (including intermediate files and/or information useful for copy number variation analysis) will be made available to public repositories (such as dbGaP) upon completion of quality control studies. The Alliance statistician associated with analysis of the trial and genotype data will determine when quality control studies have been completed, and will prepare data for submission. Publications by others that make use of only Alliance genotype 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 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 analysis of the trial and genotype data will determine when the standard Statistics and Data 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 DSMB monitored study, will be released without a formal approval from the DSMB.

15.4.2.3 Results databases

As considerable time may elapse between submission of genotype and submission of phenotype data, the Alliance will develop results databases (see example at http://www.pgscore.org) to serve results of genotype - phenotype association studies for phenotypes that have not yet been deposited. For example, genome-wide association studies conducted on intermediate phenotypes (e.g., pharmacogenetic phenotypes, or surrogate

outcomes) may be available through the web site before final phenotype data are deposited in dbGaP because of the length of time required to obtain and quality check full outcome information. The use of results databases permitting extensive queries will improve access of the scientific community to results of the studies and serve as the necessary intermediate between completion of initial genetic studies (which may involve intermediate data) and completion of the clinical trial. Results databases may be made publicly available upon completion of Alliance-approved analyses that have undergone review of the steering committee and, in the case of studies that have not yet met the primary endpoint, been approved by the DSMB.

15.5 Release conditions and disclaimer

A simple, formal data use agreement 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.

Sometimes the data requested for analysis 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, the 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 SDC to perform the abstraction. Some funding for clerical support may still be required. Likewise, when data requested require data sets not available in easily obtained electronic format, especially for older trials, the Alliance may require funding for support to create the data set in an electronic format.

In releasing the 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 the 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 manuscript arising from the project associated with the data request must be sent to the Alliance; however, approval of the manuscript is not a condition for use of the data.

Policy Name: Appeals Process	**Policy Number:** 15.6
Section: Data Sharing – 15	**Date Revised:** March 15, 2013

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 program director, CTEP associate director or his/her designee, 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 program director.

Policy Name: Fees	**Policy Number:** 15.7
Section: Data Sharing – 15	**Date Revised:** March 15, 2013

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 SDC funded under the NCTN Program. Fees will not be charged for the release of non-complex electronic data sets. For complex data sets where substantial work is involved, fees may be charged for preparing and documenting the data set. Any fees will be limited to the actual time, effort, and materials required for preparing and documenting the data set.