DIVERSITY PROGRAM CONSORTIUM DATA SHARING AGREEMENT APPROVED 09/06/2019

Enhancing the Diversity of the NIH-funded Workforce Overview

The National Institutes of Health (NIH) recognizes the need to diversify the scientific workforce by enhancing the participation of individuals from groups identified as underrepresented in the biomedical, clinical, behavioral and social sciences (collectively termed "biomedical") research workforce. With that in mind, in 2012, it convened the NIH Advisory Committee to the Director (ACD) Working Group on Diversity in the Biomedical Research Workforce to explore strategies to attract, prepare and sustain the interest of individuals in the scientific workforce, including those from underrepresented groups (NIH's Interest in Diversity). In response to the Working Group's recommendations, which were endorsed by the ACD, the NIH established the Common Fund Program "Enhancing the Diversity of the NIH-funded Workforce," also known as the Diversity Program Consortium (DPC). In Phase I, this program allowed for the formation of a national collaborative consisting of three integrated initiatives: (1) Building Infrastructure Leading to Diversity (BUILD), (2) the National Research Mentoring Network (NRMN), and (3) the Coordination & Evaluation Center (CEC). In Phase II, new awardees joined the consortium through the DPC Dissemination and Translation Awards (DPC DaTA).

In partnership with the NIH, DPC awardees employ approaches to strengthen institutional capacity to engage and prepare individuals, including those from underrepresented groups, for successful careers in biomedical research. The interventions focus on infrastructure, faculty development, student engagement, research training and mentorship across the career pathway. A primary goal of the DPC is to provide robust evidence on effective ways to enhance diversity by engaging and sustaining the interest of individuals in the biomedical research workforce and to encourage the dissemination of successful diversity enhancing interventions to a wide variety of institutions across the United States.

Table of Contents

1. PURPOSE

To facilitate the evaluation of the program, a core set of data across the DPC awardee sites is collected at intervals as outlined in the Consortium-Wide Evaluation Plan (CWEP). In addition, individual awardees collect and store site-level data for evaluation and research purposes to meet the goals of the program. This Data Sharing Agreement (DSA), developed in conjunction with DPC's awardees and the Executive Steering Committee (ESC), describes the requirements for data collection, integrity, storage, security, confidentiality, use, sharing, ownership, rights, and responsibilities. The DSA may be modified to meet the evolving needs of the consortium. A revised DSA must be approved by the ESC, which includes representation from each awardee and the NIH.

2. PERIOD OF POLICY

The original DSA developed and implemented during Phase I of the DPC remained in effect from March 16, 2016 until this revised DSA was approved on 09/06/2019. To ensure compliance, the DSA is incorporated into the terms and conditions of each award. This DSA supersedes the original DSA and should be referenced to address topics related to DPC data going forward. This DSA will remain effective until June 30, 2029, 5 years after the end of the funding period.

3. GOVERNANCE, AUTHORITIES, DATA RIGHTS & COMPLIANCE

The CEC is responsible for implementation of the consortium-wide evaluation activities in collaboration with the awardees, the ESC, and the NIH. The NIH will be responsible for oversight and adherence of the DPC awardees to this DSA. These terms apply to each DPC awardee in Phase I (RFA-RM-13-017, RFA-RM-13-016, RFA-RM-13-015) and Phase II (RFA-RM-18-004, RFA-RM-18-003, RFA-RM-18-002, RFA-RM-18-006, RFA-RM-18-005; RFA-RM-19-003). In addition to the DPC awardees, all non-consortium parties granted access to DPC data (see below for access conditions) will adhere to this DSA. Failure to abide by the terms and conditions of this DSA, including data security/disclosure provisions, may result in (i) denial of further access to the DPC data, (ii) denial of access to NIH-funded resources, and (iii) federal or state penalties. Liability will be aligned with data ownership and rights.

The NIH is committed to protecting the rights and privacy of those whose information is collected during the conduct of its funded research, and awardees will be responsible for compliance with this DSA. This DSA is made under the NIH's authority to conduct and fund research; to provide training/training assistance; to collect information as to the practical application of such research and training activities; to assemble accurate data to evaluate research priorities and scientific opportunities; and to maintain records in connection with these or other agency functions (42 U.S.C. §§ 241 and 282, and 44 U.S.C. § 3101). This DSA incorporates by reference the NIH's data sharing policies for research and the Family Educational Rights and Privacy Act (FERPA) guidelines on use or disclosure of student educational records in the conduct of research, when applicable.

As delineated in the Notices of Grant Awards, each *"awardee will retain custody of and have primary rights to site-level data and software developed under these awards, subject to Government rights of access consistent with current DHHS, PHS, and NIH policies. All evaluation-related consortium data will be shared with the NIH upon request and at the conclusion of the award."* During the funding period, the CEC has responsibility for management and oversight of the consortium-wide evaluation BUILD data and long-term NRMN follow-up data as delineated in this agreement, and the individual awardees retain ownership over the use of their site-level data.

4. DATA COLLECTION APPROVALS, CATEGORIES & QUALITY

4A. Data Collection Approvals/Clearance

The CWEP data collection instruments, processes, and schedules for data collection developed on behalf of the DPC adhere to those approved by the UCLA-CEC Institutional Review Board (IRB), or other relevant IRB governing the consortium-wide data collection. All non-CWEP site-level data collection instruments and procedures must adhere to the governance of the local site IRB.

4B. Data Categories

Consortium-wide and site-level data are required to determine the effectiveness of DPC training, mentoring, and research-capacity building interventions on outcomes. The data categories include the following:

1. **Consortium-Wide Evaluation Plan Data:** Data collected to complete the ESC-approved CWEP. The CWEP includes the scheduled collection of both qualitative and quantitative data elements to measure psychosocial factors as well as outcomes. The CWEP is divided into the following broad categories: (1) student/mentee, (2) faculty/mentor, and (3) institutional/site as outlined in the logic models and the associated DPC Hallmarks of Success. Consortium-wide data are collected at defined intervals and include participant rosters for BUILD activities, survey responses, institutional records, and transcripts from CEC case studies.
    a. **BUILD Participant Data.** BUILD awardees submit participant rosters on an ongoing basis through the CEC Tracker, a tool developed by the CEC and utilized by the BUILD awardees to upload, collect, store, and manage consortium participant data. The CEC Tracker assigns each participant a unique nine-digit numeric identification number. This allows the CEC to maintain longitudinal data regarding exposure to BUILD activities as individuals progress through their careers. The CEC Tracker allows authorized site administrators to add identifying elements to the CEC Tracker to assist with longitudinal tracking (e.g., site-level identification numbers). The CEC conducts quality review and risk assessment of the data.
    b. **NRMN Participant Data Phase I.** NRMN data managers from across the NRMN Cores submitted participant rosters on an ongoing basis to the Data Management System (DMS), a tool developed by the UCLA Computer Technology Research Lab (CTRL) and utilized by the NRMN awardees. These data are assembled into a master participant spreadsheet and shared with the CEC for use in conducting the longitudinal follow-up surveys.
    c. **Survey Data.** The CEC administers the consortium-wide surveys. DPC awardees (BUILD and NRMN Phase I) work with the CEC to ensure robust participation. The survey data are cleaned, documented, and de-identified under the management and oversight of the CEC.
    d. **BUILD Institutional Record Data.** Institutional Record (IR) data is essential for accurate tracking of student persistence and graduation, as well as faculty accomplishments. CWEP IR data includes (1) de-identified data for introductory science and mathematics courses, and (2) identifiable data for students and faculty who have provided consent through surveys.
    e. **Transcripts from CEC Case Studies:** The CEC periodically conducts visits to awardee sites to gather qualitative data. The data are coded and curated under the management and oversight of the CEC.

2. **Site-Level Data – BUILD, NRMN, DPC DaTA:** Data elements collected by individual awardees to evaluate the impact of site-level variables on outcomes. Site-level data includes (1) site-specific data collected by the CEC and (2) non-CWEP data collected and stored by individual awardees. The CEC-equivalent to "site-specific" data are any data being used in analyses that form a component of the NIH-mandated CEC evaluation of the Diversity Program Consortium. CEC case studies data also fall within the category of "CEC-specific data" and cannot be shared due to the inability to appropriately de-identify those data.

3. **Site-Level Third-Party Data:** Data collected from BUILD awardee partner institutions or NRMN sub-awardees, which can include both consortium-wide and site-level data elements. Third-party institutions retain sole ownership of the use of their site-level data unless the data part of the CWEP (see above). Third-party data are subject to the terms of this DSA for all CWEP data, unless otherwise agreed upon in writing between an awardee and the third party that predates any DPC DSA. If such third-party agreement does not allow for the sharing of data as described in this or any other DPC DSA, the awardee shall attempt to secure permission for the sharing of third-party data consistent with the objectives of the DPC.

4C. Consortium-Wide Data Submissions and Quality Review

DPC awardees are expected to work collaboratively with the CEC to meet consortium-defined standards of data completeness and quality. The DPC process for submission of CWEP data includes the CEC providing a description of data to be submitted, the submission timeline, and access to the secure consortium data repository for the transfer of data. The process includes quality assurance activities to be completed by awardees, as well as quality review, risk assessment,

and de-identification to be completed by the CEC. Quality review includes confirmation of identifiers for linking with other consortium-wide data, assessment of valid values, explanations for missing data, and completion of logic or skip patterns. Disclosure risk assessment includes review for sensitive and infrequent (rare) data points that could be used to identify individuals. For both quality reviews and disclosure risk assessments, the CEC works with each institution to resolve any outstanding issues with data quality.

5. DATA SECURITY

To protect the rights and privacy of individuals whose information is collected, all parties under this DSA must agree to adhere to the highest standards for data transfer, storage, and access. The NIH and the DPC recognize that data security in multisite and collaborative research requires compliance with organizational policy, IRB guidelines, as well as local, state and federal laws and regulations to safeguard participant privacy and ensure data protection. To transfer data, awardees must use a secure file transfer service over an encrypted connection. Physical and/or electronic data are to be maintained securely and retained for up to 5 years following the end of the program funding period.

1. **CWEP Data.** All CWEP data management and storage systems operate behind a firewall on a private Internet Protocol (IP) space (i.e., accessible to defined IP addresses and inaccessible to regular internet traffic). The physical security of the data management and storage is maintained at the UCLA CTRL with multiple levels of security compliant with Health Insurance Portability and Accountability Act (HIPAA) data security standards. The UCLA CTRL maintains and manages restricted information that identifies participants (e.g., name, address, student/faculty institutional ID number). Access to files is protected via account and password protocols. Regular review of protocols ensures state-of-the-art network security.
   a. **BUILD Participant Data.** Access to the CEC Tracker requires authentication with a virtual private network (VPN) appliance in addition to CEC Tracker web application account verification. Because of the confidential nature of the data, the participant lists are not available for consortium or third-party use. Sites have ongoing password protected access to their own de-identified tracker data. Identifiable participant information is only provided to authorized educational officials at individual awardee institutions and is subject to their local IRB governance.
   b. **NRMN Phase I Participant Data.** Because of the confidential nature of the data, the participant lists are not available for consortium or third-party use. Identifiable participant information is only provided to authorized educational officials at individual awardee institutions and is subject to their local IRB governance.
   c. **Survey Data.** The CTRL administers each on-line CWEP survey. De-identified CWEP survey data is provided through a secured DPC online repository. Identifiable site-level survey data is only provided to authorized educational officials at individual awardee institutions and is subject to their local IRB governance.
   d. **IR Data**. Awardees must use a secure file transfer service over an encrypted connection to transfer CWEP IR data. Identifiable CWEP site-level IR data is only provided to authorized educational officials at individual awardee institution and is subject to their local IRB governance.
   e. **Case Study Data.** Because of the sensitive and identifiable nature of the data, the case study data is secured as described above and is not available for consortium or third-party use.

2. **Site-Level Data.** All DPC awardees must implement, maintain, and use IRB approved administrative, technical, and physical security measures to preserve the confidentiality and integrity of the data. Storage must be in a secure and locked location and all electronic data collected must be maintained in a password-protected directory behind the institutional firewall with access granted only to approved staff or officials.

6. DATA USE & SHARING

6A. Data Dissemination

The long-term impact of the DPC will be in the broad dissemination of evidence-based effective DPC training, mentoring, and research-capacity building strategies. All DPC awardees are expected to disseminate outcomes to the wider community. Two types of dissemination products include evaluation outcomes and hypothesis-based research findings. *Evaluation outcomes* represent the results with respect to the Hallmarks of Success and will inform the community about the overall effectiveness of the training, mentoring, and research-capacity interventions. *Hypothesis-based research* is

based in a theoretical framework, tests models or hypotheses, and delineates findings that will inform the biomedical community about what factors influence the outcomes.

During Phase II, the DPC is expected to develop and implement a Dissemination Strategic Plan for both Phase I and Phase II data describing (1) major consortium-wide evaluation themes and hypothesis-driven research areas, (2) the types of dissemination products (e.g., data briefs, presentations, publications), (3) the data required for the analyses, (4) the expertise needed to produce rigorous products, and (5) a realistic timeline for producing the dissemination products, taking into consideration the time required for outcomes. The Dissemination Strategic Plan will be developed by consortium members and approved by the ESC. Additionally, each awardee is responsible for implementing their site-level dissemination plans.

6B. Data Access Approval for Consortium Members

To promote synergies and reduce redundancies, the DPC developed an approval process for DPC awardees to access data for disseminating consortium products. The approval process is described in the Publications and Presentations (P&P) Policy and is managed by the Publications and Presentations sub-Committee (PPsC) of the Executive Steering Committee. The process requires either a notification or an application as described below.

Notification
- **Consortium-Wide Evaluation Outcomes:** The CEC is responsible for disseminating consortium-wide evaluation outcomes. These ESC approved DPC evaluation products follow the site-level notification process for tracking outlined in the P&P Policy.
- **Site-Level Evaluation Outcomes or Hypothesis-Driven Research Findings Using Site-Level Data:** Awardees retain ownership of site-level data and are responsible for the dissemination of site-level evaluation and research findings. Dissemination of the findings using site-level data follows the notification process for tracking outlined in the P&P Policy.

Application:
- **Hypothesis-Driven Research Findings Using Consortium-Wide Data:** All products based on hypothesis-driven research findings using consortium-wide data must go through the approval process described in the P&P Policy. Requests for access to de-identified site-level data requires sponsorship by the Principal Investigator(s) of the site(s). The PPsC will review and approve all meritorious requests. After approval, the CEC prepares and releases the data through a secure data management system. If a consortium product changes in scope, as determined by the PPsC, the authors must obtain a second approval.

6C. Data Access Approval for non-Consortium Parties

During the funded period of the program, access to de-identified data for parties outside of the DPC requires submission of a *Data Request Form* describing the proposed use of the data and identifying an institutional sponsor who is a member of the DPC. Requests for access to de-identified site-level data from an outside party requires sponsorship by the Principal Investigator(s) of that site. Documentation of Human Subjects Ethics training and IRB approval or exemption must be provided with the data request. Outside parties must agree to use the requested data only for the approved use. All such requests will be reviewed by the ESC according to the P&P Policy. In addition, within 24 months of the termination of Phase II funding, de-identified CWEP data will be made available for public use, consistent with the NIH data sharing agreement for NIH funded research, through an open access portal on the CEC public website.

6D. Unauthorized Use of Data

DPC awardees agree not to misuse or disclose DPC data except as permitted under this DSA and/or required by law. Unauthorized use or disclosure of data must be reported to an NIH Program Official within one day of discovery and include both corrective actions to prevent future unauthorized use/disclosure, as well as efforts to mitigate any adverse effect of the unauthorized use/disclosure. The NIH Program Official will implement review procedures within the NIH.

7. DISPUTES

As per the Cooperative Agreement Terms and Conditions, "any disagreements that may arise in scientific or programmatic matters (within the scope of the award) between award recipients and the NIH may be brought to Dispute Resolution. A Dispute Resolution Panel composed of three members will be convened. The three members will be a designee of the Steering Committee chosen without NIH staff voting, one NIH designee, and a third designee with expertise in the relevant area who is chosen by the other two. In the case of individual disagreement, the first member may be chosen by the individual awardee. This special dispute resolution procedure does not alter the awardee's right to appeal an adverse action that is otherwise appealable in accordance with PHS regulation 42 CFR Part 50, Subpart D and DHHS regulation 45 CFR Part 16."