OJPHI

A Virtual Data Repository Stimulates Data Sharing in a Consortium

Suzanne Siminski[1*], Soyeon Kim[1], Adel Ahmed[1], Jake Currie[1,] Alex Benns[1], Amy Ragsdale[2], Marjan Javanbakht[2], Pamina M. Gorbach[2], and the C3PNO Cohort Investigators.

[1]Frontier Science Foundation, Amherst, NY and Brookline, MA
[2]University of California Los Angeles, Los Angeles, CA

Abstract

Research data may have substantial impact beyond the original study objectives. The Collaborating Consortium of Cohorts Producing NIDA Opportunities (C3PNO) facilitates the combination of data and access to specimens from nine NIDA-funded cohorts in a virtual data repository (VDR).

Unique challenges were addressed to create the VDR. An initial set of common data elements was agreed upon, selected based on their importance for a wide range of research proposals. Data were mapped to a common set of values. Bioethics consultations resulted in the development of various controls and procedures to protect against inadvertent disclosure of personally identifiable information. Standard operating procedures govern the evaluation of proposed concepts, and specimen and data use agreements ensure proper data handling and storage.

Data from eight cohorts have been loaded into a relational database with tables capturing substance use, available specimens, and other participant data. A total of 6,177 participants were seen at a study visit within the past six months and are considered under active follow-up for C3PNO cohort participation as of the third data transfer, which occurred in January 2020. A total of 70,391 biospecimens of various types are available for these participants to test approved scientific hypotheses. Sociodemographic and clinical data accompany these samples.

The VDR is a web-based interactive, searchable database available in the public domain, accessed at www.c3pno.org. The VDR are available to inform both consortium and external investigators interested in submitting concept sheets to address novel scientific questions to address high priority research on HIV/AIDS in the context of substance use.

Keywords: common data elements, data repository

Abbreviations: National Institute on Drug Abuse (NIDA), Collaborating Consortium of Cohorts Producing NIDA Opportunities (C3PNO), human immunodeficiency virus (HIV), acquired immunodeficiency syndrome (AIDS), injecting drug users (IDU), virtual data repository (VDR)

Correspondence: siminski@frontierscience.org*

DOI: 10.5210/ojphi.v13i3.10878

Introduction

Research data are undeniably a valuable resource, requiring considerable effort, time, and funding to produce. By increasing access to research data, the "impact, efficiency, and effectiveness of scientific activities and funding opportunities" are also increased [1]. Affirming its commitment to access to data generated through support of public funds, the National Institutes of Health (NIH) issued a statement requiring a plan for data sharing for investigator-initiated applications [2]. The rationale for the requirement is that

> *[s]haring data reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research, makes possible the testing of new or alternative hypotheses and methods of analysis, supports studies on data collection methods and measurement, facilitates the education of new researchers, enables the exploration of topics not envisioned by the initial investigators, and permits the creation of new datasets when data from multiple sources are combined.*[3]

In order to leverage investments in publically-funded research such as ongoing cohorts that address high priority research on HIV in the context of substance use, the National Institute on Drug Abuse (NIDA) issued RFA-DA-17-019 to solicit cooperative agreement applications "to establish a virtual repository, and facilitate the leadership of the cohorts steering committee (SC), consisting of representatives from the NIDA-funded cohorts and NIDA staff" in order to "provide a strong resource platform for current and future collaborative efforts with other investigators to address emerging questions related to HIV pathogenesis, prevention, and treatment in the context of substance abuse, as well as to foster the creativity and efficiency of investigator–initiated research projects." [3] The goal is not only to optimize collaborations among cohort investigators but also to provide better access to data and specimens for researchers external to the cohorts.

The Collaborating Consortium of Cohorts Producing NIDA Opportunities (C3PNO) [4] was funded to facilitate broader access to rich data and biological specimens from the nine NIDA-funded cohorts described in Table 1. An important activity of the consortium has been the creation of a virtual data repository (VDR). A VDR is an online repository of data. A VDR facilitates information sharing within and beyond narrowly defined research communities. By access through an online interactive platform, a VDR can allow a user to define search criteria and obtain a summary of the number of participants meeting the criteria. A VDR may include information such as concise data descriptions displayed in the form of a master data catalog, provide details on the populations who contributed the data, and facilitate disseminating information to the interested users of the data and users of well-characterized banked specimens. By making a VDR available in the public domain, users are able to triage the appropriateness of the data for their purposes without having to make a formal request to each of the study teams for information.

By coordinating efforts and providing access to data and specimens across cohorts, data may have substantial value beyond that of addressing the original research studies' objectives. The most obvious value of combining data across cohorts is the greater sample size it affords, thereby increasing power to test hypotheses, which is particularly important in smaller populations or when events of interest are rare, for example, HIV seroconversion. In addition, research consistent with

NIDA's mission may be outside the expertise of the cohorts' investigators and providing broad access to data and specimens to researchers from other disciplines, including new investigators, may facilitate the establishment of new knowledge bases. Because the specimens come from well-characterized individuals, they are particularly valuable for assessing biomarkers.

It is worthwhile to note that the nine cohorts above were established independently, span nearly 20 years in dates of inception, and are funded separately. Each cohort has its own leadership structure and addresses different NIDA priorities. Each cohort has distinct study aims, research objectives, study populations, specimens, and data, and each stores data in separate formats. To take advantage of C3PNO cohorts' unique opportunity to study the intersection of substance use and HIV to answer questions in key populations, otherwise not possible with a single cohort's data, requires that certain challenges be overcome. We address these challenges by marshalling a multidisciplinary team with expertise in data curation and mapping, epidemiology, bioinformatics, data standardization, and data linking.

Table 1: Overview of Cohorts

Cohort	Population	Year Started	No. of Study Participants
ACCESS	HIV-positive PWID	2005	1100
ALIVE	PWID	1988	1500
Heart Study	HIV-positive African Americans	2004	1400
HYM	Young Men of Color Who Have Sex with Men	2015	450
JHHCC	HIV-positive persons in receiving care through Johns Hopkins HIV/AIDS Services ambulatory clinics	1989	1100
MASH	Hispanic persons in South Florida	2002	1400
mSTUDY	Latino and African-American/black men MSM at UCLA Vine Street Clinic	2013	500
RADAR	Young MSM	2014	1100
V-DUS	HIV-negative PWID	1996	3500

**Abbreviations*: No., Number; HIV, human immunodeficiency virus; PWID, persons who inject drugs; AIDS, acquired immunodeficiency syndrome; MSM, men who have sex with men; UCLA, University of California Los Angeles.

The Cohorts and Coordinating Center

The consortium of cohorts are described elsewhere [4]. Briefly, the cohorts geographically span the United States and Canada and include: the AIDS Care Cohort to Evaluate Access to Survival Services (ACCESS) Study (Vancouver, Canada); AIDS Linked to the Intravenous Experience (ALIVE) Study (Baltimore, MD); the Heart Study (Baltimore, MD); the Healthy Young Men's (HYM) Study (Los Angeles, CA); the Johns Hopkins HIV Clinical Care Cohort (JHHCC) (Baltimore, MD); the Miami Adult HIV (MASH) Study (Miami, FL); mSTUDY (Los Angeles,

CA); RADAR (Chicago, IL); and the Vancouver Drug Users Study (V-DUS) (Vancouver, Canada). The cohorts follow HIV-positive and high risk HIV-negative persons, people who use drugs including persons who inject drugs (PWID), and men who have sex with men (MSM) in either community or clinical settings. They include participants in and out of HIV care and substance use treatment. Some cohorts focus specifically on adolescents and young adults. All follow participants longitudinally. The cohorts collect extensive demographic, behavioral, clinical, and laboratory assay-based information on study participants. The differing study objectives of the cohorts result in heterogeneous populations and thus various types of data collected, ultimately leading to a lack of common standards across cohorts. Even when data domains align, instrument selection may be tailored to the population or to best serve the original studies' aims. C3PNO, among other functions, has developed and maintains a VDR which is described herein.

Methods

Common Data Elements (CDEs)

The C3PNO Steering Committee, composed of the cohorts' principal investigators, NIDA scientific staff, and a Scientific Advisory Board, developed an initial list of high-priority CDEs, which were selected based on their perceived importance for defining study populations, risk factors, potential confounders, or outcomes for assessing the feasibility of a wide range of research proposals addressing high priority research domains. Because some cohorts have been in existence for decades while others were established in the last five years and because some data elements are captured repeatedly and can vary with time (e.g., participant age, CD4+ counts, HIV/HCV viral loads, substance use patterns, sexual behaviors), the initial focus of this effort was on the most recent full visit for a participant. As such, not all participants currently or ever enrolled in the cohorts are represented in the available repository data, but it does ensure that all the data represent an up-to-date snapshot of cohort participants. The first data submission transfer in May 2018 included general demographic, socioeconomic, substance use, behavioral, and HIV related data from cohort participants. The first submission allowed the consortium's data staff to learn about cohorts' data management systems, obtain data dictionaries and/or data catalogues, establish data transfer agreements, and comply with any institutional regulatory requirements. With the first transfer, we also established processes for data transfer, data mapping, and conversion to a common format. This process allowed us to establish rapport between the coordinating center staff and cohort PIs and data managers. Multiple discussions, including an initial on-site, face-to-face meeting, followed by emails and teleconferences assisted in resolving any pending issues. Not all CDEs are collected by all nine cohorts, and data elements collected by multiple cohorts were not standardized across the cohorts. Some cohorts collected a data element with finer granularity than others. Initially data mapping to a common format was performed centrally by the C3PNO coordinating center, and mapping and final tabulations of CDEs were reviewed with data managers and cohort PIs for accuracy. In the case that additional data were submitted by the cohort data center, only data that are part of the CDE were mapped.

Data Transfer, Transformation, Updates, and Retention

In order to minimize burden, data were submitted by cohorts in a format convenient to the cohort data team and in single or multiple files, although future data transfers are expected to be submitted in the format used for the initial transfer. A secure web-based file sharing system is used to submit

data. Each user has a password protected account and access limited to that user's cohort directory. Datasets are transformed using a commercial extract-transform-load (ETL) system. Each cohort has its own transformation which can be run independently, and is tested and validated to ensure data integrity. All CDE records are retained per regulatory requirements, but when new data are submitted, previous submissions are flagged as inactive. This allows a clear delineation of what data was available at any point in time, and, if needed, the data can be rolled back to a previous version.

In the most recent data transfer, data was expanded to include longitudinal data. Data mapping was performed by the data team of each individual cohorts based on mapping guide provided by the coordinating center. The cohort data team performs the mappings and submits the data to the coordinating center with documentation of the mappings.

The Database

Data are stored in a commercial relational database and, key fields are linked across tables. Database constraints are defined to enforce data integrity. It is expected that future data requests will expand the data included in the CDE set. The structure of the database is flexible so that new data elements can be easily incorporated. Participant data are stored in three normalized database tables. The core participant database table (cohortparticipantdataitem) holds demographic, socioeconomic, clinical, and sexual behavior data, e.g., sex at birth, age, income, weight, hemoglobin, number of male sex partners. There is one record per participant data item (multiple per participant), which allows for the inclusion of additional data items as the CDE is expanded with subsequent data transmissions without changing the original database structure. While substance use and biospecimen data could have been included in a single table, this would have required additional complexity not needed for both types of data. For both substance use and biospecimen data, it was determined that data were specialized to such a degree that it would be more efficient to maintain a separate database table for each. The substance use database table (cohortparticipantsubstanceuse) has fields for specific drug, reporting method, and administration route. Data on available biospecimens are stored in another database table (cohortparticipantbiospecimen) which likewise contains fields for number of aliquots, specimen type, additive, and derivative type.

VDR

Investigators can use the interactive platform to determine the number of participants and specimens that are available for their potential research study. Filters are utilized to allow the user to specify inclusion criteria for their study population. For example, the user can determine how many male HIV-positive people reporting use of heroin are in each C3PNO cohort and the number and type of specimens available for cohort participants meeting the inclusion criteria. Importantly, no individual participant-level data are displayed; only summary counts of the number of total participants and of each biospecimen type meeting the specified criteria are shown. In addition to counts, the results page displays the specified search criteria used in generating the tabulations for the benefit of the end user. More complex searches can be obtained by working with the C3PNO coordinating center.

Bioethics experts were consulted in the development of procedures to protect against inadvertent disclosure of personally identifiable information by allowing researchers to perform queries on the database on the interactive platform. For example, query results are not displayed if fewer than ten participants are included in a given category. If a search results in less than ten participants for a specific cohort, data are only displayed collapsed across cohorts and only when the minimum required numbers are displayed in the query result.

Results

Table 2 shows a sampling of key types of demographic and clinical data and Table 3 shows a sampling of substance use data available at the www.c3pno.org website. Data from 6,177 participants represent cohort participants who have been seen at the last full visit. Additional data for approximately 3,000 more participants in the Canadian cohorts, (ACCESS and V-DUS) will be available once participants are re-consented to allow their data to be shared with the consortium. Currently, a small fraction of participants from the V-DUS cohort have had an opportunity to provide their consent at a study visit.

Table 2: Number of participants with a recent full visit contributing specific data elements in C3PNO Virtual Data Repository (c3pno.org January 27, 2020)

	COHORT:	ALIVE	Heart	HYM	JHHCC	MASH	mSTUDY	RADAR	V-DUS	Total
	Total Participants	**1328**	**580**	**400**	**948**	**1016**	**560**	**1030**	**315**	**6177**
Demographics	**Sex at Birth**	1328	580	400	948	1016	560	1030	315	**6177**
	Transgender Status	*	*	400	948	1016	*	1030	315	**3709**
	Race	1328	580	397	948	1016	547	1030	302	**6148**
	Ethnicity (Hispanic or Non-Hispanic)	1328	580	400	948	1016	547	1030	302	**6151**
	Homelessness	1328	*	400	*	49	10	19	314	**2120**
	Incarceration	*	*	398	*	1012	545	1030	315	**3300**
Health Care	**Health Insurance**	1324	*	400	*	975	525	913	*	**4137**
	Accessed Health Care Past 6 Month	1328	†	298	†	1015	546	666	315	**4168**
HIV-Related	**HIV Status**	1328	580	400	948	992	560	1030	315	**6153**
	Antiretroviral	372	406	66	942	480	*	209	*	**1995**
	CD4 Count	372	368	65	948	475	281	95	*	**2604**
	HIV-1 Viral Load	374	371	65	948	474	284	79	*	**2595**
Other Testing/ Diagnoses	**Hepatitis B Status**	1284	*	313	*	1015	81	*	315	**3008**
	Hepatitis C Status	1327	*	313	942	1016	*	*	313	**3911**
	Tuberculosis	*	*		948	*	*	*	*	**948**
	Chlamydia	*	*	386	*	*	544	1021	315	**2266**
	Gonorrhea	*	*	386	*	*	544	1023	315	**2268**
	Syphilis	*	*	384	*	*	143	*	315	**842**

Abbreviations: HIV, human immunodeficiency virus; cohort.

*No data of this specific type were reported by the cohort as available for use.

†Heart and JHHCC studies are clinical cohorts so all participants are in care.

Table 3: Number of participants with a recent full visit with data on drug usage, either by self-report within the last 6 months or by urinalysis (c3pno.org January 27, 2020)*

COHORT:	ALIVE	Heart	HYM	JHHCC	MASH	mSTUDY	RADAR	V-DUS	Total
Total Participants	**1328**	**580**	**400**	**948**	**1016**	**560**	**1030**	**315**	**6177**
Cocaine									
Self-report	1321	580	400	428	1014	545	1030	315	**5633**
Urinalysis	0	0	356	440	1012	559	1027	254	**3648**
Heroin									
Self-report	1321	580	400	63	1015	545	1030	315	**5269**
Urinalysis	0	0	356	47	1008	559	1027	0	**2997**
Methamphetamines									
Self-report	0	0	400	74	1016	545	1030	315	**2364**
Urinalysis	0	0	356	47	1011	559	1027	0	**3000**
Prescription Pain Killers									
Self-report	1321	0	399	0	1012	196	1030	315	**4273**
Urinalysis	0	0	0	404	0	0	0	254	**658**
Fentanyl									
Self-report	0	539	0	8	1014	349	0	315	**2225**
Urinalysis	0	0	0	56	997	398	0	254	**1705**
Cannabis									
Self-report	1321	580	400	486	1015	545	1030	315	**5692**
Urinalysis	0	0	356	439	1011	559	1027	254	**3646**
Alcohol (self-report)	0	553	400	942	1016	0	1030	315	**3736**
Nicotine (self-report)	1319	580	114	614	1014	0	0	315	**3956**
Speedball (self-report)	1321	577	0	0	0	0	0	289	**2187**
Hallucinogen (self-report)	1321	0	400	70	1014	0	1030	0	**3835**
Stimulants (self-report)	1321	0	400	22	1015	0	1030	293	**4081**

* When a cohort did not collect drug usage data by a method they are reported in the table as 0.

Data elements include race, sex, substance use, clinical history, and sexual behavior, and additional elements can be added with each update – refer to the c3pno.org website for data currently available. As some of the cohorts follow men who have sex with men (MSM) only, there are more participants who report male sex at birth than female overall (72% and 28%, respectively). At last visit, approximately 61% are Black/African American, 18% are Hispanic, and nearly half are HIV-positive.

All cohorts include assessments of self-reported recent substance use at each study visit, but the substances assessed varied across the cohorts (see Table 3). In the last data transfer, all cohorts assessed cocaine use; eight cohorts assessed heroin use; six cohorts assessed methamphetamine use; six cohorts assessed prescription pain medication use; and eight cohorts assessed cannabis use. Urinalysis results for toxicological screens for substance use are also available in some cohorts for some substances.

Table 4 highlights the numbers of participants with a recent full visit who have biospecimens available by each cohort by type. Plasma is available on 68%, serum on 47%, and PBMCs on 37% of participants. For some cohorts, additional biospecimens are available, including whole blood, oral rinse, passive drool, rectal swabs and sponges, nail, hair, buffy coat, and pellet specimens. Investigators can propose research using these biospecimens for consideration by the C3PNO Steering Committee.

Table 4: Number of participants with a recent full visit for whom biospecimens are available (c3pno.org January 27, 2020).

COHORT	ALIVE	Heart	HYM	JHHCC	MASH	mSTUDY	RADAR	V-DUS	Total
Total Participa	**1328**	**580**	**400**	**948**	**1016**	**560**	**1030**	**315**	**6177**
PBMC	386	*	*	690	1016	184	17	*	**2293**
Plasma	925	566	*	730	1016	184	755	*	**4176**
Serum	695	*	*	678	1016	490	*	*	**2879**
Buffy	*	*	*	568	*	*	*	*	**568**
Whole	*	*	*	*	1016	*	*	*	**1016**
Hair	*	*	*	*	*	187	*	*	**187**
Nail	*	*	*	*	*	412	*	*	**412**
Oral	*	*	*	*	*	452	*	*	**452**
Passive	*	*	*	*	*	490	*	*	**490**
Pellets	*	*	*	*	*	183	*	*	**183**
Rectal	*	*	*	*	*	123	*	*	**123**
Rectal	*	*	*	*	*	489	778	*	**1267**

Abbreviations: PBMC, peripheral blood mononuclear cells.
* No biospecimens of the specific type were reported by cohort as available for use.

Subject to minimum threshold requirements (to guard against any potential unintentional disclosure of identifying information), the number of biospecimens from individuals meeting specific criteria can be obtained by adding filters. For example, the VDR can provide a tally of the number of plasma specimens from HIV-positive persons who have CD4+ cell counts less than 200 cells/mm [5].

We previewed the C3PNO VDR at a preconference for the 22nd International AIDS Conference (AIDS 2018), and have disseminated availability of data and specimens through NIDA, International Workshop on HIV and Hepatitis Observational Databases (IWHOD), 25th Scientific Conference of the Society on NeuroImmune Pharmacology (SNIP), and additionally through links on a growing number of websites.

Limitations

Questionnaires were designed by each cohort team to best suit their population and study objectives. This results in heterogeneity in measures available for the cohorts. For example, substance use recall periods and assessment of frequency of use are not uniform across cohorts. The C3PNO consortium is currently conducting projects to allow data linking of drug use and other key data measures to facilitate cross-cohort analyses.

Conclusions

The benefits of data sharing are readily acknowledged by researchers, including those participating in the C3PNO consortium. Data linking work to address the use of different data instruments and to address differing definitions across cohorts for analytic purposes is ongoing. Cross-cohort analyses are in various stages of planning and execution.

A number of other NIH-funded consortia have also created VDRs to improve access to their data and specimens. The VDRs differ in the types of data and populations that are housed. The NIH is taking a leadership role in funding and requiring participation in VDR efforts. In the context of HIV treatment and prevention, the ACTG (AIDS Clinical Trials Group), IMPAACT (International Maternal Pediatric Adolescent AIDS Clinical Trials Network), and HVTN (HIV Vaccine Trials Network) have developed a combined VDR that allows an investigator to perform an interactive search to learn about the specimens available (http://www.specimenrepository.org). The VDR allows filtering on the types of study and participant characteristics. Concept proposals are sent to specific network for their review. While conceptually the HIV VDR is similar to the C3PNO VDR, the populations and key data elements differ.

The i2b2 (Informatics for Integrating Biology & the Bedside) now n2c2 (National NLP Clinical Challenges) clinical research platform for precision medicine is a NIH-funded research platform that makes clinical data in electronic health records into analyzable data by using natural language processing to make unstructured text into data sets (i2b2.org). Software is available to run queries and tranSMART tools are available for use in data exploration, display and analysis (https://i2b2transmart.org).

These and other VDRs share the benefit of reducing the effort and time required by a researcher to a minimum for determining the feasibility of many concepts. The VDRs differ in the provenance of and type of data available which are tailored to the specific populations that are of interest. They may also have associated specimens that can be used for running assays.

It is necessary for external researchers to be aware that such cohorts exist in order to enhance utilization of the data. Furthermore, external researchers and those across the C3PNO consortium need access to the characteristics of the study population, data collection methods, and

storage/types of specimen in order to develop research proposals and plan analyses. For some research questions, data from multiple cohorts must be combined to have sufficient power. A VDR can facilitate the process. The advantages of a VDR in the public domain are numerous, perhaps the greatest being the enhanced ability to promote collaborative science through curated data sharing. Data sharing and transformation is a critical step in ensuring that existing data can be used in additional research studies. The barriers to data sharing are being overcome by transforming information collected into a set of CDEs for the purpose of VDR display. The C3PNO VDR is available to inform consortium and external investigators interested in submitting concept sheets for research proposing to use consortium data for consideration by C3PNO. A system is in place to streamline the submission of concept sheets, and to track the review, data use agreement, data and/or specimen transmission, and publication process. These data can be further utilized by other investigators for scientific inquiry.

Acknowledgements

First and foremost, we are most grateful for the involvement of our cohort participants for making this research possible. It is their time, commitment and involvement that allows us to collect the necessary data to draw meaningful scientific conclusions and advance our collective research via the consortium. We thank the C3PNO Cohort Principal Investigators (PI) and Data Managers (DM) for their leadership, expertise, and technical contributions to the consortium. In alphabetical order by Cohort: **ACCESS**: PI, M-J Milloy, DMs, Wing Yin (Janet) Mok and Ekaterina Nosova; **ALIVE**: PIs, Greg Kirk and Shruti Mehta, DMs, Jacquie Astemborski and Todd Noletto; **HYM**: PI, Michele Kipke, DMs, Julia Moore, Ji Hoon Ryoo, and Su Wu; **Heart Study**: PI, Shenghan Lai, DM, Shaoguang Chen; **JHHCC**: PI, Richard Moore, DMs, Jeanne Keruly, Steven Xu, Li Ming Zhou, and Charles Collins; **MASH**: PI, Marianna Baum, DM, Qingyun Liu; **mSTUDY**: PIs, Pamina Gorbach and Steven Shoptaw; DMs, India Richter, Fiona Whelan, Shahrzad Divsalar, Alexander Moran, Allison Rosen, and Stone Shih; **RADAR**: PI, Brian Mustanski, DMs, Antonia Clifford, Daniel Ryan and Kitty Buehler; and **V-DUS**: PIs, Kora DeBeck and Kanna Hayashi, DMs, Wing Yin (Janet) Mok, and Ekaterina Nosova. Lastly, we thank the additional members of the Frontier Science technical team; David Goss, Astrid Fuentes, Lynn Strusa and Kris Ricusso.

This project is supported by the National Institute of Drug Abuse (NIDA) of the National Institutes of Health under award numbers: U24DA044554, U01DA0251525, U01DA036297, U01DA036926, U01DA040325, U01DA036935, U01DA040381, U01DA036267, U01DA036939, 2U01DA038886. Additional information about each C3PNO cohort can be found at www.c3pno.org including objectives, current research, contact information, and links to cohort specific websites.

Financial Disclosure

No Financial Disclosures.

Competing Interests

No Competing Interests.

References

1. Lee DJ, Stvilia B. 2017. Practices of research data curation in institutional repositories: A qualitative view from repository staff. *PLoS One*. 12, e0173987. PubMed https://doi.org/10.1371/journal.pone.0173987

2. Final NIH Statement on Sharing Research Data. 2003. (Accessed August 27, 2019, at https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html.)

3. RFA-DA-17-019 Coordinating Center for the HIV/AIDS and Substance Use Cohorts Program (U24). at https://grants.nih.gov/grants/guide/rfa-files/RFA-DA-17-019.html.)

4. Gorbach P, Siminski S, Ragsdale A, Javanbakht M, Kim S, Chandler R. Cohort Consortium Profile: The Collaborating Consortium of Cohorts Producing NIDA Opportunities (C3PNO). 2019.

5. NIH Announces Draft Statement on Sharing Research Data. 2002. (Accessed August 27, 2019, at https://grants.nih.gov/grants/guide/notice-files/not-od-02-035.html.)