

METHODOLOGY AND ADVANCED ANALYTICS RESOURCE CENTER (MAARC)

INTRODUCTION

The Justice Community Overdose Innovation Network (JCOIN) 1.0 hubs, the Coordination and Translation Center (CTC), and the Methodology and Advanced Analytics Resource (MAARC) have all committed to share the data they have collected with the research community, consistent with the National Institutes of Health's 2023 Data Management and Sharing Policy and the data sharing requirements of the Helping to End Addiction Long-term® Initiative. This document outlines the approach.

QUANTITATIVE DATA

Quantitative data include the JCOIN core measures (both client and staff), and supplemental measures collected by the hubs during their clinical trials, stigma and jail surveys collected by the MAARC, and data from studies performed by the CTC. These data are being submitted to the MAARC in a standardized, open format (CSV) along with corresponding variable-level metadata (in JSON or YAML format). Data packages containing these data are being made available to JCOIN members via the JCOIN Data Commons (JDC), according to established procedures. Individuals authorized by a hub's principal investigator (PI) may access that hub's data at any time, without restriction. JCOIN members who wish to access data from other hubs must submit a project proposal to the JCOIN Steering Committee (SC), whose members may authorize the use of data from their hub as well as request to participate in the proposed project.

After a project has been approved, the MAARC will grant access to the specific data requested by those individuals (e.g., investigator[s] and analysts) explicitly identified in the proposal (if necessary, a package containing only those data will be created). Such data may be used solely for the purpose(s) described in the project proposal. Some exceptions to this procedure may be made to permit immediate access to data within the JDC by JCOIN members, subject to specific terms of use; examples include the Opioid Environment Policy Scan (OEPS) and the stigma surveys, both created by the MAARC.

Quantitative data submitted to the MAARC will be transferred from the MAARC directly to the National Addiction & HIV Data Archive Program (NAHDAP), a National Institute on Drug Abuse-funded data archive hosted by the Inter-university Consortium for Political and Social Research (ICPSR). NAHDAP will do the following:

1. Work directly with each hub to execute an agreement between the hub's institution and NAHDAP that authorizes NAHDAP to release that hub's data to researchers for scientific analyses.
2. Confirm that the data are consistent with the corresponding schema and translate the data and metadata into NAHDAP's internal format.
3. Perform data curation, including (but not necessarily limited to) the following:

(Cont)

METHODOLOGY AND ADVANCED ANALYTICS RESOURCE CENTER (MAARC)

VERSION 0.1

- a. Verify that data meet basic standards of completeness and usability (including adequate documentation).
 - b. Confirm that data are adequately deidentified, and if necessary, make recommendations to further minimize the risk of deductive disclosure (i.e., reidentification).
 - c. Assist in identifying one or more subsets of data that are adequate for specific analytic use cases and can be made available via registration-only access (see below).
4. Create a JCOIN project page on their data portal, together with a project page for each hub nested underneath. The former will provide an overview of JCOIN data and links (or information on requesting access) to data products containing combined data from two or more hubs (e.g., core measures aggregated over all hubs). The latter will provide links (or information on requesting access) to data products containing data from a single hub.
5. Assist in developing requirements for data access and process requests for data access according to those requirements. Requirements may include but are not necessarily limited to the following:
- a. Data to be used for scientific research only, including both research for publication and for internal use (e.g., preparation of grant proposals or study planning).
 - b. Approval by user's local IRB (typically just a letter confirming exempt status).
 - c. Requirements to ensure data security, including requiring analyses to be performed within ICPSR's data enclave or the JCOIN Data Commons.
 - d. User must acknowledge the data's source and funder in all publications, presentations, and written reports using a provided acknowledgment.
6. Provide basic user support (e.g., how to access and use the files). If permitted by the hub PI, NAHDAP will contact the PI in cases where a user asks a legitimate question about the data and/or study which NAHDAP staff are unable to answer.
7. Archive and maintain the data in perpetuity. Datasets archived at NAHDAP will continue to be available through the JDC via system interoperability.
- Note that a hub PI's responsibilities are limited to the following: (1) submit core and supplemental measures to the MAARC, using the procedures it has developed; (2) work with NAHDAP and their institution to execute the release to share data; and (3) review their hub's project page on NAHDAP's data portal, and provide any additional information about their study requested by NAHDAP; (4) inform NAHDAP whether they wish to be contacted in the event that a data user has a question about the study or data that NAHDAP cannot answer.

(Cont)

METHODOLOGY AND ADVANCED ANALYTICS RESOURCE CENTER (MAARC)

CONTROLLED VERSUS REGISTRATION-ONLY ACCESS

By default, data archived at NAHDAP will be controlled access, according to requirements like those described in Item (5) above. Approved users will be able to analyze the data within ICPSR's data enclave or the JCOIN Data Commons (now authorized at the FedRAMP moderate level). This provides one of the highest levels of data security and can be implemented immediately.

Despite the value of controlled access in ensuring data security, these restrictions present a substantial impediment to use (e.g., ICPSR's controlled-access datasets are used much less frequently than their less restricted datasets). Thus, whenever possible, researchers should determine whether they can create a companion dataset for which the risk of disclosure has been sufficiently mitigated that it may be released according to a registration-only paradigm. All that is required to use a registration-only dataset is to create an account on ICPSR's website (this permits them to track data use, as required by some depositors) and to agree to their terms of use (e.g., the user will not attempt to reidentify participants, the user will not redistribute the data, and the user will acknowledge the data and funder in publications, presentations, and written reports).

A standard way to mitigate disclosure risk is to exclude variables from the dataset. For example, a large dataset can be broken up into several smaller datasets, each with a subset of the original variables, and each with

a subset of the original variables, and each with a different participant ID so that the individual datasets cannot be relinked (e.g., NHANES uses this method to distribute their data via direct download from their website). As noted above in Item (3a), NAHDAP will work with the MAARC and JCOIN analysts to determine one or more limited data products that are analytically valuable and can be distributed via registration-only access.

QUALITATIVE DATA

Qualitative data collected by the hubs may be shared through the Syracuse Qualitative Data Repository (QDR). QDR's staff, expert in the curation and packaging of qualitative data, are ready to work one-on-one with each hub wishing to share qualitative data. Fees for data curation and archiving have been prepaid by the MAARC, and QDR's services may be accessed by contacting them directly and identifying yourself as a JCOIN hub.