

De-identification and Data Protection — For Your IRB or Privacy Office

Cannabis and Health Data Repository

De-identification and Data Protection

This one-page summary is written to be handed to a contributing institution's IRB or privacy office. It describes exactly what the Cannabis and Health Data Repository does and does not receive, and how contributed data is protected.

The Cannabis and Health Data Repository is a curated, de-identified research resource developed under a National Institutes of Health cooperative agreement (NIH/NIDA Award 1UM1DA059000), jointly administered by Johns Hopkins University and Realm of Caring.

Two-step, non-reversible de-identification

Step 1 — at your institution. Direct identifiers are removed and replaced with an opaque study identifier. The table mapping that identifier to your real participants is kept **at your institution** and is never transmitted.

Step 2 — at the Data Coordination Center. The Center derives the Repository's own subject identifier from your opaque ID using a **keyed one-way transform**, validates that the dataset meets the **HIPAA Safe Harbor** standard, and can assist with any remaining Safe Harbor de-identification where your IRB permits us to receive the data. The Repository's subject identifier cannot be reversed to your opaque ID, and the Repository never holds the map back to your real participants.

The single most important assurance

The Repository never receives your participants' names or original identifiers, and never receives the key that links research identifiers to real people. Direct identifiers are removed, dates and geographic detail are reduced to permissible granularity, and each participant carries an opaque, irreversible research identifier. The resource is therefore **structurally unable to re-identify individuals**, and recipients additionally agree by contract not to attempt re-identification.

Free text and indirect identifiers

Narrative free-text fields and any indirect identifiers that carry disclosure risk are handled separately from the openly released de-identified data. Where governance permits, they may be routed to a **restricted tier**: retained for research value but never released openly, and made available to a researcher only under additional controls and a separate restricted-use agreement. If applying de-identification would unduly harm the analytic value of your data, do not strip it silently: contact us and we will discuss a restricted-use arrangement. *(The handling of free-text and PHI-bearing narrative fields is pending governance review and not yet finalized.)*

Controlled access after de-identification

Contributed data is released **only** as de-identified data, **only** to verified researchers, and **only** under an executed Data Use Agreement that prohibits re-identification and specifies permissible uses and the term of use. A Steering Committee reviews each request on scientific merit, feasibility, and ethics. Data is transferred over encrypted, out-of-band channels, processed within a protected enclave, and subject to comprehensive access logging; downloads are time-limited and attributable. The identifiable data never reaches the public web application.

What the contributor retains

Contributors retain their own IRB approvals and participant consent. Before contributing, confirm that: (1) you hold IRB approval for the study that produced the data; (2) your consent permits sharing the data for secondary research; and (3) your consent describes how participant confidentiality is maintained. If your consent language is ambiguous on secondary sharing, contact us before submitting — this is easier to resolve up front than after data has been transferred.

For questions from an IRB or privacy office, use the contact page. Cannabis and Health Data Repository · NIH/NIDA Award 1UM1DA059000 · Johns Hopkins University and Realm of Caring.