NYSPI National Data Archive
Data Sharing Standards

Background:

National Institute of Mental Health (NIMH)-funded grants mandate submission of de-identified data to the public National Data Archive repository (NDAR) (https://data-archive.nimh.nih.gov).

The NDA policies, in conjunction with NYSPI's federal, state and contractual obligations, necessitate investigators take proactive steps to protect both subject privacy and the fidelity of their grants, while supporting the NDA's public benefits.

This document guides investigator efforts to maximize appropriate and efficient data sharing.

Summary

- NDAR submissions must use pseudo-GUIDs and may not use GUIDs.
- Submissions to NDAR may not include PHI.
- Studies must include informed consent inclusive of NDAR submission.
- Investigators should review NDAR data dictionaries prior to designing data models.
- Initial or significantly altered submissions should be reviewed by the director of psyIT

Considerations:

NYSPI, as a HIPAA covered entity, must comply with numerous state and federal privacy regulations, including HIPAA regulations, when sharing data. These regulations include prescription for de-identification of data.

The NDA data linkage principles richen data sets by connecting otherwise disconnected studies involving common research subjects. As described in HIPAA regulations, even in the absence of explicit identifiers, there is a potential for data linkage and re-identification through the use of combinations of variable values. Sharing of discrete values, whether non-health identifiers (such as IQ or years of education), or health identifiers (such as clinical test scores), increases risk of re-identification of subject data. Researchers who produce or share anonymous public use data files need to consider whether the data they are using or releasing could be used in combination with other publicly available data to infer individual identities. For instance, those with access to identifiers in either study potentially become aware of identifiable results from both studies. Researchers must make best effort to limit that potential for reidentification.

In some instances, formatting and data definitions submitted to the NDA repository may be specific to the individual project. In other instances, such as multi-site initiatives, data must be collected and/or submitted in keeping with a common data dictionary and standardized format.

Informed consent must clearly describe plans for data sharing through the NDA.

Procedure:

For purposes of data sharing, PIs must generate and assign pseudo-GUIDs, through the NDA processes, for all subjects. Actual GUIDs, requiring the use of identifying information, will not be used without explicit approval by the NYSPI IRB and OMH HIPAA Privacy Officer.

PIs must structure their data to support alignment with existing NDA data dictionaries, in instances where these have been established for a particular project. Refer to the NDAR Data Dictionary (https://ndar.nih.gov/data_dictionary.html?type=All&source=All&category=All) for the complete listing of the NDA's defined data structures PIs and their data management teams should review these structures prior to development of data collection / analysis tools and processes to ensure consistent data definitions, or should configure their data in a way that facilitates "cross-walk" of data to the required format.

Where NYSPI has developed and published data dictionaries for standardized measures (e.g. data definitions for subject demographics) prior to study execution, Institute standardized data definitions should be used when appropriate for a given study. However, specifications contained in NDA dictionary definitions, when required, will supersede any existing NYSPI data dictionary definitions.

To maintain compliance with HIPAA regulations, all 18 identifiers listed under the HIPAA regulations are to be removed. Dates that identify individuals should be substituted for dates whenever possible. For example, rather than providing date of birth, date of enrollment, date of follow-up evaluation, provide age (in years) at enrollment and age at follow-up evaluation. HIPAA allows age in years up to age with IRB approval.

The informed consent process must include language making it clear that the participant's data will be shared for research purposes through the NDA. The text of the consent form should describe both the benefits and risks of data sharing, indicating that while every effort will be made to ensure the confidentiality of this data, there is a potential risk of identification. Consent to data sharing can be a requirement for participation in the study; an alternative is to allow for participation, but to allow the subject to opt out of data sharing. If a subject opts out, there must be a flag in the database that can be programmed to exclude the subject's data from submission to the NDA.

Questions regarding submissions should be directed to the Director of psyIT, who will also review and approve the plan for submissions.

References:

National Data Archive Policy:
https://data-archive.nimh.nih.gov/s/sharedcontent/about/policy
HHS.gov. (2018). *Methods for De-identification of PHI*. [online] Available at: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html [Accessed 30 Jun. 2018].

National Data Archive Homepage:
https://data-archive.nimh.nih.gov/
Data-archive.nimh.nih.gov. (2018). *NDA*. [online] Available at: https://data-archive.nimh.nih.gov/ [Accessed 30 Jun. 2018].

HHS De-Identification Requirements:
https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html
HHS.gov. (2018). *Methods for De-identification of PHI*. [online] Available at: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html [Accessed 30 Jun. 2018].