

Archiving Quantitative Child Maltreatment Data

National Data Archive on Child Abuse and Neglect (NDACAN)



Focus

- This video will provide an overview of the dataset archiving process by covering the following topics:
 - Introduction to NDACAN
 - Benefits of data sharing
 - Archiving is a collaborative process
 - Data sharing or data management plan creation
 - The process and steps to prepare and deposit documentation and data
 - Activities undertaken after a dataset is released

Terms

- Data archive - defined by the National Institutes of Health, as “A place where machine-readable data are acquired, manipulated, documented, and finally distributed to the scientific community for further analysis”
- Data contributor- Person or organization who engages in preparing and submitting data for archiving
- Dataset package- A collection of study documentation and data files which describe, or are the result of, a data collection effort.

Abbreviations

- ACF –Administration for Children and Families, a division of the Department of Health & Human Services
- ACYF- Administration on Children, Youth and Families, a division under ACF
- CB- Children’s Bureau
- CMRL- child-maltreatment-research-list serve
- HHS- United States Department of Health and Human Services
- NDACAN – National Data Archive on Child Abuse and Neglect

Introduction to NDACAN

- Trustworthy repository for data from quantitative research studies and administrative data systems on the topic of child maltreatment
- Located at Cornell University
- Founded in 1988
- Supported with a contract from the Children's Bureau (CB), ACF, DHHS
- PI: Chris Wildeman, PhD
- Our mission is to facilitate secondary analysis of research data relevant to the study of child abuse and neglect and provide an accessible and scientifically productive means for researchers to explore important issues in the child maltreatment field.

Benefits of data sharing

- Data sharing enables research transparency, through replication and validation of the original research findings, as well as, opportunities for collaboration between researchers and extension of the original research.
- Research funders have recognized the importance of data sharing and have incorporated data sharing requirements in their funding opportunities.
- Increased citation rate for investigators who share their data (Piwowar, Day, & Fridsma, 2007)

Archiving is a collaborative process

- NDACAN views data archiving as a collaborative process between us and the data contributor. The collaboration begins at the earliest stages of a research project and will extend for years beyond the release of the dataset to the child maltreatment research community.
- Although this video focuses on the duties of a data contributor, we would also like to point out that we are your partner and can provide assistance at almost any stage of your research.
- The following are examples of the ways in which NDACAN assists researchers:
 - We provide resources to help researchers with creating data sharing or data management plans for submission with their funding proposal.
 - Researchers have solicited input from NDACAN regarding commonly used, construct specific, measures and instruments, found in our other datasets. This helps to inform their decisions on which measures to use in their proposed primary data collection research project.
 - Data from the Archive has been used by researchers to calculate sample size targets and to construct weights that were applied to complex survey data which was designed to be nationally representative.

- Examples continued...

- Data from the Archive can be analyzed by researchers to provide supporting evidence for hypotheses appearing in their proposed primary data collection project.
- Once the project has been funded and data collection begins, the Archive is available to respond to questions on the topic of data management.
- NDACAN staff are available to respond to questions that arise while data contributors are preparing the dataset package. In the past, staff have provided guidance on topics such as, which variables to include in the archived data, how to recode problematic variables, and how best to structure (stacked/long vs wide) the data files.
- Data contributors are welcome to send inquiries to NDACANsupport@cornell.edu

Data sharing or data management plan creation

- Some funding agencies require applicants to submit a data sharing or data management plan with their funding proposal.
- The Contributor Data Management Guidelines are viewable at the following web address:
 - <https://www.ndacan.cornell.edu/contribute-data/contribute-contributor-data-management-plan-guidelines.cfm>
- From the Contributor Data Management Guidelines web page, data contributors can also access the Template for NDACAN Contributor Data Management Plan.
- NDACAN has a document entitled “Data Sharing Resources” which is a collection of resources that might be helpful when creating a data sharing or data management plan. Please visit the Contribute Data page of the NDACAN website at <https://www.ndacan.cornell.edu/contribute-data/contribute-data-general.cfm>
- Researchers interested in designating NDACAN as the recipient of their research data should contact NDACAN to discuss their proposed research. NDACAN can provide a letter of acknowledgement stating that we are aware of the research project and agree to archive the data resulting from the project.

Prepare and Submit a Dataset Package

This section of the video will provide an overview of the steps that must be undertaken by a data contributor to prepare the documentation and data files for submission to the Archive.

Contributor's Handbook

- In order to keep this video concise, only summary information about the archiving process is provided. For more detailed information about the archiving process, please consult the document entitled, “A Contributor’s Guide to Preparing and Archiving Quantitative Data” which can be found on the Contribute Data page of the NDACAN website:
 - <https://www.ndacan.cornell.edu/contribute-data/contribute-data-general.cfm>
- Web links to NDACAN resource documents discussed during this presentation can also be found in the summary description for the video

Overview of the steps for archiving a dataset

- **Step One:** Complete and submit Part I of the Study Submission Form and the Investigator Contact Cover Sheet.
- **Step Two:** NDACAN will set-up a call to discuss the dataset. The data contributor will have the opportunity to ask questions. NDACAN will decide whether the data are suitable for archiving at the data archive.
- **Step Three:** If NDACAN determines the data are suitable for archiving, the data contributor will prepare the remaining elements of the dataset package in accordance with the Contributor's Handbook and as summarized in this video.
- **Step Four:** Once the dataset package is assembled, create a compressed .zip folder which contains the entirety of the dataset package. Notify NDACANSupport@cornell.edu that the dataset is ready for submission.

Overview of the steps for archiving a dataset continued...

- **Step Five:** When NDACAN receives the request to submit the dataset package, staff will set-up a means for the files to be electronically transmitted.
- **Step Six:** Once NDACAN retrieves the file from the file transfer system, then they will conduct a quick review to be sure that the files received match what is required or was discussed in prior conversations. Processing the dataset package may not occur right away if other datasets were in the queue ahead of the dataset submitted.
- **Step Seven:** NDACAN will process the dataset in the order in which it was received in the queue of datasets waiting to be processed. This requires a study contact person to be available to respond to questions and review the final dataset package once it has been prepared.

What cannot be archived?

- NDACAN has established archiving exclusion criteria. If datasets meet any of the established criteria, they cannot be archived. The criteria can be found in a document entitled, “NDACAN Archiving Exclusion Criteria” located at the following link:

<https://www.ndacan.cornell.edu/contribute-data/contribute-application-process.cfm>

How to begin the archiving process?

- The first step in starting the archiving process is to complete and submit the Study Submission Form: Part I and the Investigator Contact Sheet found at the following NDACAN webpage:
 - <https://www.ndacan.cornell.edu/contribute-data/contribute-application-process.cfm>
- Once NDACAN receives the completed Study Submission Form: Part I and the Investigator Contact Sheet, NDACAN staff will contact study staff to set up a call/online meeting to discuss the data collection, unique attributes of the data, data disclosure issues and the next steps in the archiving process.
- This initial contact should be made as soon as possible after funding has been awarded in order to insure a smoothest archiving experience for the data contributor.

Study Submission Form: Part I

- The Study Submission Form: Part I collects the following information:
 - Study title
 - Abstract
 - List of investigators as they would appear in a publications
 - Keywords to describe the study
 - Sponsor/agency name
 - Award Number
 - Award Start and end dates
- Submit the completed form to NDACANsupport@cornell.edu

Study Submission Form: Investigator Contact Sheet

- Submit a completed Study Submission Form: Investigator Contact Sheet for each investigator involved in the study
- The Investigator Contact Sheet collects the following information:
 - Study title
 - Salutation
 - First, middle, and last name
 - Investigator Degree
 - Title- name of the position currently held at the institution
 - Institution or Organizational affiliation and address
 - Phone, fax, and email address
 - Investigator's role:
 - Principal Investigator
 - Contact Person for questions about the study

Prepare the study materials

- The next steps in the process are undertaken by study staff once the data collection effort has concluded.
- Study staff will prepare the following materials for archiving:
 - Codebook/Data Dictionary (unambiguous variable name, descriptive label, missing data codes, values and value labels for categorical variables, derivation logic for derived variables, and variable data type)
 - Electronic copies of data collection instruments/measures
 - Copies of Interim and final reports related to the data collection effort
 - List of bibliographic citations for published articles and reports
 - Institutional Review Board review/approval letter and informed consent template
 - Complete Study Submission Forms Parts II, III, and the Instrument Information Form which are located at the following web address:
<https://www.ndacan.cornell.edu/contribute-data/contribute-application-process.cfm>

Study Submission Form: Part II- Dataset Details

- Contact person
- Study title
- Types of data collected
- Date data collection started
- Date data collection ended
- Geographic area to which the data are relevant
- Unit of analysis
- Sample description
- Response rates
- Study design description
- Data collection procedures description
- APA formatted list of published works based off from the study
- Were the data being submitted collected by you or were the data obtained from another source?
- Are there any secondary identifiers which present challenges to preserving respondent confidentiality?
- Is this data submission part of a longitudinal study?
- Will the data in this contribution need to be updated?
- Is this a new edition or a special version of data already archived?
- A blank field is available for any additional information not covered by the form.

Study Submission Form: Instrument Information

- Complete an Instrument Information form for each measure or instrument used in the study and for which there are corresponding data in the data file.
 - Study nickname
 - Full measure name
 - Abbreviated or nickname name
 - Version
 - Bibliographic citation for the measure
 - If the measure is project derived, provide a general description
 - Description of how the measure was modified for use in the study

Prepare the data file(s)

- Study staff must **remove all direct identifiers** from the data files PRIOR to submission of the dataset package to the Archive
 - Direct identifiers include:
 - Names
 - Social Security Numbers
 - Phone Numbers
 - Medical Record Number
 - Insurance Card Number
 - Highly specific Geographic Variables (i.e., Street Addresses, Geo-coordinates, Census Block)
- Prospective Data Contributors can consult with NDACAN staff to determine how to identify and appropriately reduce the disclosure risk of problematic variables.
- NDACAN will also conduct a disclosure risk analysis upon receipt of the data.

Prepare the data file(s) continued...

- The data file should contain the following required data file elements:
 - An unambiguous variable name that matches the name appearing in the codebook.
 - A descriptive variable label – A textual description of the item, or a clear reference to its associated question in the data collection instrument.
 - A list of valid values and corresponding labels for categorical variables.
 - Missing/inapplicable data codes and their meanings.
 - Variable data type (i.e., numeric, character, date).
 - Column specifications for each variable.
 - Decimal settings should reflect the data contained in each variable.
 - Note: When there are multiple data files and Codebooks, include a document that maps the data file to its respective Codebook document.

Prepare the data file(s) continued...

- NDACAN accepts data in a variety of file formats. NDACAN currently can receive data files formatted for SPSS, Stata, and SAS and delimited text data files.
- If the data files are not in one of these file formats, please contact NDACAN to discuss what you have and to confirm that we will be able to directly open or convert the file to another format upon receipt.

Study Submission Form: Part III – Data File Characteristics

- Part III is a Microsoft Excel spreadsheet
- Enter the following information into a numbered row in the spreadsheet for each data file that will be a part of the submission:
 - File name
 - Number of records
 - Number of variables
 - Number of records per case
 - Format (i.e., tab delimited, SPSS, SAS, or Stata)
 - Indicate whether the following were performed:
 - consistency checks
 - checks for undocumented codes
 - Indicate whether the data contain variable and value labels

Collate the dataset package for submission

- Once prepared, the dataset package, including all documentation and data files, should be contained within a .zip folder prior to finalizing the upload arrangements with NDACAN.
- Data Contributors should check their dataset package against the NDACAN publication entitled, “NDACAN Archiving Checklist for Dataset Packages (PDF)” found at the following web address:
<https://www.ndacan.cornell.edu/contribute-data/contribute-application-process.cfm>
- Send an email to NDACANsupport@cornell.edu stating that the dataset package is ready for upload.
- The dataset package should be submitted no later than 8 months prior to study funding expiration.

Transmit the dataset package

When NDACAN receives the announcement that the dataset package is ready for submission, we will evaluate and choose one of the following file sharing methods for the Data Contributor to send the dataset package, based on what is known about the data at that time:

- Cornell Enterprise Box
- Cornell Dropbox which is also known as Cornell Secure File Transfer
- More information about the security specifications for these file transfer systems can be found on page 6 of the web document entitled, “NDACAN Archiving Process and Steps”
 - <https://www.ndacan.cornell.edu/contribute-data/contribute-application-process.cfm>
- These are the same file sharing methods used to send the dataset package to secondary analysts.

Processing the Deposited Data

This next section of video provides an overview of the steps that NDACAN takes, from the point when we receive the dataset package to processing and preparing it for release.

Receipt of the dataset package

- Once the data and documentation files are received by NDACAN, they will be saved to a secure file server, the dataset will be logged into the internal tracking system, and given a unique numeric identifier known as a dataset number.
- NDACAN will process the dataset in the order in which it was received in the queue of datasets waiting to be processed.
- It is important that study staff be available to respond to questions during processing and also to review the final NDACAN prepared dataset package prior to it being released to the child maltreatment research community.

Summary of dataset processing

- NDACAN will conduct a disclosure review of the data file(s) and work to reduce or eliminate disclosure risk by recoding or removing problematic variables in consultation with the Data Contributor.
- Create a Section 508 compliant User's Guide containing study level metadata about the original data collection effort which will accompany the data upon release. The information supplied by the Data Contributor in the Study Submission Forms is used to populate the contents of the Users Guide.
- Produce the final versions of the data file(s) in the following formats:
 - SPSS native (.sav)
 - SAS native (.sas7bdat)
 - Stata native (.dta)
 - Import program files for SPSS, SAS, Stata (.sps, .sas, .do & .dct)
 - Text data file (.dat)
 - Tab delimited data file (.tab)

Data contributor review

- Once NDACAN has prepared the dataset package it gets sent to the data contributor for review
 - Data contributors must have a signed Data Contributor's Agreement on file. NDACAN prepares the document and sends it to the data contributor for signing. The document can be signed at any time during the archiving process but must be on file prior to the dataset's release.
 - Data contributor supplied edits will be vetted by Archive staff and incorporated into the final version of the dataset package.

Dataset Access Procedures and Release

In this section, the two pathways that secondary analysts will have to follow to access the data are discussed and the process of announcing the availability of the dataset to the child maltreatment research community, also known as the “release.”

Dataset access procedures

- Archive staff will assess the final version of the dataset to determine which data ordering access procedures are appropriate. The order procedures outlined below will determine the steps and eligibility criteria requirements that must be met before a prospective data user can receive the data.
 - **Terms of Use Agreement** – This data ordering access pathway is the least restrictive and provides access to a wide audience of researchers. The Terms of Use Agreement document can be found at the following web address:
 - https://www.ndacan.cornell.edu/datasets/order_forms/TermsOfUseAgreement.pdf
 - **Restricted Access Data Licensing** – More information about the restricted access data license process and forms can be found at the following web address:
 - <https://www.ndacan.cornell.edu/datasets/request-restricted-data.cfm>

Release of the dataset

- After the final version of the dataset package is reviewed by all stakeholders, the dataset will be released to the child maltreatment research community.
 - The dataset title, abstract, and User's Guide are posted to the Datasets page of the NDACAN website with instructions for how to order the data.
 - The dataset's availability is announced to the child-maltreatment-research-list serve (CMRL) and to the NDACAN Twitter account

Post Dataset Release: Ongoing Activities

After a dataset is made available for use by the child maltreatment research community, NDACAN's work is not done. We promote the use of datasets, provide technical support, and actively track publications produced using datasets in our holdings.

Promote use of the dataset

- The dataset will appear in the quarterly electronic newsletter entitled, “The NDACAN Updata”
 - The newsletter is sent to over 6,000 NDACAN mailing list subscribers
- NDACAN hosts an annual Summer Research Institute.
- NDACAN hosts periodic dataset specific webinars and trainings which are recorded and added to the website and NDACAN YouTube channel

Technical support

- NDACAN provides robust technical support for datasets contained within its holdings.
- There is a dedicated support email which is connected to a technical support ticket tracking system.
 - The support email is NDACANsupport@cornell.edu
- Archive technical support serves as the first point of contact for dataset technical assistance requests
 - Data Users are directed to send support requests to this email address in study documentation
 - If there is a question that Archive staff cannot answer using the supplied data documentation, then Archive staff will reach out to the study contact to request assistance in responding to the question.
 - For the most popular datasets at the Archive, the number of times Archive staff need assistance from study staff is rather low, around 5 times a year, with the highest volume in the first year the dataset is released and then it trails off each year after.
- Staff also develop user support documents to assist secondary analysts working with the datasets

Track publications

- NDACAN maintains an online searchable digital citations management database called the child abuse and neglect Digital Library (canDL)
- This database is where we store bibliographic citations for publications relating to our archived datasets
 - Users of our datasets are required to submit their bibliographic citations for published works to NDACAN
 - Reminder calls ,for data users to submit citations, are published in the quarterly Updata newsletter
 - NDACAN periodically reviews the published literature to identify citations for published works not previously captured in the canDL
 - The canDL is accessible to data users and contributors on the Publications page of our website at the following web address:
<https://www.ndacan.cornell.edu/publications/publications.cfm>

Conclusion

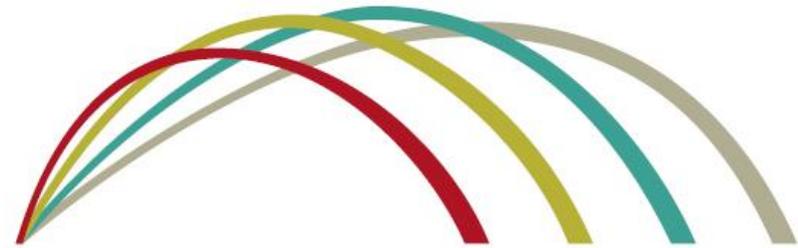
This concludes the substantive content portion of the video presentation entitled, “Archiving Quantitative Child Maltreatment Data.”

Please submit questions about archiving data at NDACAN to NDACANsupport@cornell.edu

Bibliography

- National Institutes of Health (2003, March 5). NIH Data Sharing Policy and Implementation Guidance. Retrieved May 10, 2019, from https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#archive
- Piwowar H.A., Day R.S., & Fridsma D.B. (2007). Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE* 2(3): e308. <http://doi.org/10.1371/journal.pone.0000308>

The National Data Archive on Child Abuse and Neglect is a project of the Bronfenbrenner Center for Translational Research at Cornell University.



**Bronfenbrenner Center
for Translational Research**

Funding for NDACAN is provided by:



Children's Bureau

An Office of the Administration for Children & Families