Data Producer’s Guide to the Child & Family Data Archive

Child & Family Data Archive
ChildandFamilyDataArchive.org
Quick Reference Summary

The Child & Family Data Archive facilitates Early Care and Education research by:

- **Curating** research data, which means we ensure that the data and supporting documentation are accessible and usable for research for many years into the future;
- **Archiving** the files in a secure infrastructure that so that they are faithfully preserved and migrated to new technology when available; and
- **Sharing** research data in accessible and equitable ways to researchers and policy makers.

The Child & Family Data Archive specifically seeks to acquire and disseminate data from studies funded by the Office of Planning, Research, and Evaluation (OPRE). The archive is funded by a contract from OPRE and is executed by the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan.

**Technical assistance during the project planning phase** includes communication planning, establishment of a deposit workspace, and limited consultation on Institutional Review Board (IRB) documents, consent procedures, and other early-stage processes to ensure that these facilitate future archiving and sharing.

**Technical assistance during the data collection phase** includes consultation on data file design, help to ensure that best practices for variable-level metadata are utilized, and formulation of plans to remediate potential disclosure risks.

**Technical assistance during the data cleaning, processing, analysis, and reporting phase** includes consultation on documentation creation, implementation of disclosure risk remediation, and review of draft data files and documentation.

**Technical assistance during data depositing** includes training about the deposit system, consultation on which data and documentation files to include, and assessment of all files and metadata for completeness.
Table of Contents

Quick Reference Summary ............................................................................................................... 1
Table of Contents .............................................................................................................................. 2
Overview of the Child & Family Data Archive ................................................................................ 3
Introduction to Data Preparation and Technical Assistance ............................................................. 3
Identification of Data Producers/Projects between OPRE and ICPSR ............................................ 4
Three-way Communication .............................................................................................................. 4
Technical Assistance to Data Producers ........................................................................................... 6
Introductory Conference Call ........................................................................................................... 6
Technical Assistance During the Project Planning Phase ................................................................. 7
Technical Assistance During the Data Collection Phase ................................................................. 7
Technical Assistance During the Data Cleaning, Processing, Analysis, and Reporting Phase ........ 7
Technical Assistance During the Data Depositing Phase ................................................................. 8
Walk-through for the Deposit Workspace ........................................................................................ 8
File Management ............................................................................................................................ 10
Administrative Panel ...................................................................................................................... 10
Project Description ......................................................................................................................... 10
Related Publications ....................................................................................................................... 12
Project Communication Log ........................................................................................................... 13
Curation Process .............................................................................................................................. 13
Overview of Curation ....................................................................................................................... 13
Disclosure Risk Review and Remediation ..................................................................................... 13
Communication with Data Producers During Curation Workflow ................................................ 15
Other Considerations for Data Producers ....................................................................................... 15
Further Information ......................................................................................................................... 16
Appendix A: Suggestions for Facilitating Future Secondary Analysis ........................................... 17
Appendix B: Deposit Checklist ......................................................................................................... 18
Overview of the Child & Family Data Archive

The Child & Family Data Archive facilitates Early Care and Education research by:

- **Curating** research data, which means we ensure that the data and supporting documentation are accessible and usable for research for many years into the future;
- **Archiving** the files in a secure infrastructure so that they are faithfully preserved and migrated to new technology when available; and
- **Sharing** research data in accessible and equitable ways to researchers and policy makers.

The Child & Family Data Archive specifically seeks to acquire and disseminate data from studies funded by the Office of Planning, Research, and Evaluation (OPRE). The archive is funded by a contract from OPRE and is executed by the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan. The Child & Family Data Archive builds on ICPSR’s extensive experience in archiving, preserving, and disseminating data relevant to child care, Head Start, and other child and family issues. The Child & Family Data Archive also leverages ICPSR’s infrastructure that supports these activities. The overall goals for the project—relative to the child and family datasets and materials that projects produce—are to:

1. Establish and maintain a secure archive that ensures the confidentiality, integrity, and availability of these data;
2. Prepare these data for secondary analysis;
3. Create and maintain an interactive website to provide a range of data users easy and continuous access to these data;
4. Collaborate to promote the use of these data among diverse end-users through a variety of training, technical assistance, and dissemination activities; and
5. Remain up-to-date about advances and regulations in the field of data archiving and develop innovative strategies to improve access to these data.

Introduction to Data Preparation and Technical Assistance

This guide describes the procedures you will follow as a data producer to prepare and deposit data for inclusion in the Child & Family Data Archive as well as the assistance that ICPSR will provide you along the way. Data preparation and technical assistance (TA) for data producers must take into consideration the origins of a dataset, often referred to as the research lifecycle. That lifecycle can be divided into interdependent phases including:

- Project planning;
- Data collection;
- Data cleaning and processing;
- Data analysis and publication/reporting; and
- Data archiving.
ICPSR, the data producer, and OPRE all play roles throughout the research lifecycle, facilitating the research study data’s eventual archiving and curation in the Child & Family Data Archive. Table 1 provides a general overview and timeline from consultation to deposit to curation and beyond. For data producers working with the Child & Family Data Archive, ICPSR provides a high level of TA. For example, project managers with the Child & Family Data Archive will provide you with consultation during the Pre-Deposit Consultation and Data Deposit stages and actively assist you with resolving any issues during the Post-Deposit Consultation stage. Your Federal Project Officer (FPO) is included in all TA contacts with you, to the level that they prefer. Overall, the process can take from less than a couple of months up to over a year, depending on the complexity and number of datasets, remediation following the disclosure risk review, and any problems uncovered during ICPSR’s review of the data and quality control checks. On average, a typical OPRE complex/multi-source/multiple linked dataset with codebook(s) and user guide takes approximately 3-4 months. This is a rough estimate; ICPSR will provide a more refined estimate after your data are deposited and the Curation team completes its initial review.

Please note that this guide focuses mainly on quantitative data, but ICPSR will work with you to curate, archive, and share qualitative, video, or other types of data as well.

Table 1: Overview of Deposit Consultation to Post-Release Timeline

<table>
<thead>
<tr>
<th>Step</th>
<th>Stage</th>
<th>Timeframe</th>
<th>Description</th>
<th>Persons Involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introductory Conference Call</td>
<td>Initial discussion with data producers about archiving their data with ICPSR. Discuss the level that the FPO wishes to be involved on subsequent archiving calls.</td>
<td>Data producers, study director, study assistants, federal project officers (FPOs), and Child and Family Data Archive project manager (PM). If changes are needed in the standard agreement for confidential data, the University of Michigan lawyers will be consulted.</td>
<td></td>
</tr>
<tr>
<td>Step</td>
<td>Stage</td>
<td>Timeframe</td>
<td>Description</td>
<td>Persons Involved</td>
</tr>
<tr>
<td>------</td>
<td>-----------------------------------------------------</td>
<td>-------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2</td>
<td>Technical Assistance Pre-Deposit</td>
<td>2-24 weeks, depending on study completion and presence of confidential data subject to special restrictions</td>
<td>Provided during the research planning, data collection, data cleaning and analyses, and reporting phases (page 8). Ongoing pre-deposit consultation includes explaining the process and what can be expected, what is needed in addition to the data (e.g., codebook, questionnaires, etc.), and how to prepare the data for submission. Additionally, if the data contain confidential information, dissemination options and restricted data use agreements are discussed.</td>
<td>Data producers, study director, study assistants, and Child and Family Data Archive PM. FPOs included to the extent agreed upon.</td>
</tr>
<tr>
<td>2</td>
<td>Technical Assistance During Data Deposit and Submission</td>
<td>2-12 weeks, depending on delivery of complete data/documentation</td>
<td>Assist the data producer with the deposit, review the deposit to prepare for curation, assist with submission, and track and coordinate with the curation team.</td>
<td>Data depositor, Child and Family Data Archive PM, and curation supervisor. FPOs included to the extent agreed upon.</td>
</tr>
<tr>
<td>3</td>
<td>Curation Initial Review</td>
<td>1-4 weeks, depending on existing queue of studies and priority</td>
<td>Disclosure Risk Review (page 16) and creation of the Data Processing Plan. The Review identifies any disclosure concerns in the data. The Plan records changes the curation team expects to make to files in order to release. After the curation team completes this review, they provide an estimate of how long it will take until the data can be released.</td>
<td>Curator and curation supervisor</td>
</tr>
<tr>
<td>Step</td>
<td>Stage</td>
<td>Timeframe</td>
<td>Description</td>
<td>Persons Involved</td>
</tr>
<tr>
<td>------</td>
<td>-------</td>
<td>-----------</td>
<td>-------------</td>
<td>------------------</td>
</tr>
<tr>
<td>4</td>
<td>Communication During Curation Workflow</td>
<td>0-12 weeks, depending on disclosure review findings, documentation needs, and depositor response time</td>
<td>Discuss disclosure review findings with the data producer, obtain any missing documentation or files, and resolve missing items (e.g., value labels).</td>
<td>Data depositor, Child and Family Data Archive PM, curator, and curation supervisor. FPOs included to the extent agreed upon.</td>
</tr>
<tr>
<td>5</td>
<td>Curation Process</td>
<td>2-16 weeks, depending on level of curation, state of the data, existing queue, and the study’s priority</td>
<td>Develop the metadata record, process the data, and perform quality control review on the product</td>
<td>Curator and curation supervisor</td>
</tr>
<tr>
<td>6</td>
<td>Data Release</td>
<td>1 day</td>
<td>Release the study and its documentation, perform post-release checks, and announce the release.</td>
<td>Curator, curation supervisor, and Child and Family Data Archive PM</td>
</tr>
<tr>
<td>7</td>
<td>Post-Release Study Support</td>
<td>Ongoing, as needed</td>
<td>Provide data users with technical support to access data, supply restricted data application support, process and monitor restricted data use applications, perform study updates, and process data resupplies as needed.</td>
<td>User support staff, Child and Family Data Archive PM and project assistant, and curation staff</td>
</tr>
</tbody>
</table>

### Identification of Data Producers/Projects between OPRE and ICPSR

Your project was selected for inclusion in the Child & Family Data Archive. This selection most likely happened in one of three ways:

1. Your Federal Project Officer (FPO) contacted the Child & Family Data Archive’s Project Officers at OPRE;
2. OPRE and ICPSR identified your project during our annual review of OPRE’s grants and contracts for the year; or
3. You reached out to ICPSR or OPRE about archiving your project’s data with the Child & Family Data Archive. Additionally, your project’s FPO and the Child & Family Data Archive’s Project Officers agreed to include the data in the archive.
Three-way communication

Soon after your study was identified, Child & Family Data Archive’s Project Officers at OPRE initiated communication between you, ICPSR, and your FPO (i.e., three-way communication). While most of this communication will be TA between you and ICPSR, your project’s FPO is included in this communication to ensure that all three parties have the same information and share the same expectations. At the time that communication was initiated, you received a copy of this guide.

Throughout each study’s lifecycle, OPRE and ICPSR will continuously assess whether the study’s data should be included in the Child & Family Data Archive. Changing information about the study, its data, project timelines, resources, and priorities might impact the decision to exclude a study from further consideration. Key points at which inclusion assessments take place are during the initial identification of the study, before deposit of the data and documentation, before the ICPSR curation team begins their work, and after the curation team review that may uncover concerns with the study files for release. If the decision is made to no longer consider your study for the archive, you and your FPO will be informed of the that decision. We will also provide you with information on alternative ways to archive the data with ICPSR.

Technical Assistance to Data Producers

ICPSR will provide TA to you throughout the research lifecycle. Assuming that your study was identified early in the lifecycle, TA will begin with the project planning phase. ICPSR will create a deposit workspace for your project and share it with you and, if desired, the FPO. The workspace will be a place:

- To share working files throughout the research lifecycle, allowing ICPSR to review draft data and documentation;
- To deposit final versions of data and documentation; and
- To support three-way communication through the deposit system communication log tool (see Project Communication Log section, below)

The deposit workspace is the preferred means for project tracking and file sharing between you, ICPSR, and OPRE. Within the deposit workspace, the communication log records messages and allows file sharing between all parties. In the rare instance that you cannot use the deposit workspace, we will collaborate to identify other procedures or tools for sharing your study’s resources for deposit.
Introductory conference call

Also soon after the three-way communication is initiated, you, your FPO, and ICPSR will have an introductory conference call. ICPSR will provide all parties with an agenda in advance of the call as well as a brief summary after the call, documenting any decisions made. The call will be used to:

- Discuss depositor TA and the data preparation process;
- Provide clarification on this guide;
- Discuss what types of early-stage consultations are needed (Institutional Review Board (IRB) documents, consent procedures, etc.);
- Establish a schedule of communication and determine the best means of communication to be used throughout the lifecycle of the study (though this can be adjusted as needed); and
- Introduce the deposit workspace which was created for the study.

If your study is identified after the project planning stage of the research lifecycle, this introductory call will also discuss the status of the data collection and what data preparation and other consultation TA that ICPSR should provide.

Technical Assistance During the Project Planning Phase

Beyond initial communications, ICPSR may provide some TA in the early stages of your project. We have several ideas about ways you can plan now for easier data sharing later and can consult with you as you develop your project. ICPSR will provide limited consultation on IRB documents and consent procedures to ensure that these facilitate future archiving and sharing. ICPSR has existing documentation to specifically support IRB and consent development.

Technical Assistance During the Data Collection Phase

As your project transitions from planning to data collection, ICPSR will continue supporting you. Communication between you, your FPO, and ICPSR will continue according to the schedule and means identified in the introductory call (see above). ICPSR will review considerations from the project planning phase, provide consultation on data file design, and help to ensure that best practices for variable-level metadata are utilized. You, your FPO, and ICPSR project staff will discuss disclosure risk, identify data elements to be collected that increase risk, and begin formulating plans to remediate those risks.

Technical Assistance During the Data Cleaning, Processing, Analysis, and Reporting Phase

Once data are collected and your team begins to clean and analyze the data, ICPSR project staff will shift to TA appropriate for this stage. As before, communication between you, your FPO, and ICPSR will continue according to the schedule and means identified in the introductory call (see above). ICPSR will provide consultation on documentation creation, ensuring that decisions made in the processing of data are not forgotten (e.g., what variables were recoded and why, how were index variables calculated, and how were data altered to reduce disclosure risk) and that metadata
are well documented. When possible, any documentation you archive must be 508 Compliant. You, your FPO, and ICPSR will revisit disclosure risks and implement plans to remediate those risks. We will request that any disclosure risk review or remediation work you complete be briefly documented and shared with ICPSR. Appendix A provides several suggestions for how to make files user-friendly and facilitate secondary analyses. We will also revisit timelines for deposit and release of data, identify priorities (e.g., the official report will be released before the data), and ensure that schedules for the release of reports and the data are aligned.

**Technical Assistance During the Data Depositing Phase**

Once data are cleaned and prepared, ICPSR will begin providing deposit-related TA. Communication between you, your FPO, and ICPSR will continue according to the previously established schedule and means. The communication plan may change if your team would benefit from more frequent interaction. This phase includes finalizing data and documentation files, adding study-level metadata, and submitting the deposit to ICPSR.

ICPSR will provide consultation to:

- determine which data and documentation files to curate for release,
- identify the disclosure risk remediation process, and
- assess all files and variable metadata to maximize utility for secondary analysis.

Throughout the earlier stages of the lifecycle, please feel free to provide ICPSR with draft versions of data and documentation for feedback. However, when the time comes to submit your deposit, you must submit final versions of the data and documentation. Changes in the data and documentation generally result in resubmission and therefore redundant work for the archive, leading to delays in the release of the data.

ICPSR will provide guidance on adding study-level metadata to the deposit form (most will display on the study homepage) and will review the draft metadata prior to submission. Before submitting a finalized deposit, please review the Deposit Checklist provided in Appendix B. Most studies (including restricted-use studies) are deposited using ICPSR’s standard deposit terms. If the standard terms do not satisfy your IRB, other institutional requirements, or study-specific requirements, ICPSR will work with you to negotiate a restricted-use data deposit and dissemination agreement.

As deposits are nearing submission, ICPSR will provide you with further training about the deposit system and discuss how to submit final deposits. Next, we provide extended guidance on the system.

---

1 Some examples of studies with very complete metadata include:
- National Survey of Early Care and Education;
- Child Care and Development Fund (CCDF) Policies Database;
- Head Start Impact Study; and
- Head Start Family and Child Experiences Survey (FACES).
Walk-through for the Deposit Workspace

The deposit workspace facilitates collaboration, communication, and file sharing throughout your project’s research lifecycle. Figure 1 shows what you can expect to see the first time you log into your deposit workspace. There are five main functional areas: File Management, Administrative Panel, Project Description (metadata), Related Publications, and Project Communication Log.

Figure 1: Screenshot of ICPSR Deposit Workspace
**File Management**

The File Management area is located in the upper left area of the deposit workspace (see Figure 1).

**Create Folder.** You can organize files by creating folders and sub-folders in this area by clicking the “Create Folder” button. Navigate back to the main screen by using the breadcrumbs that appear at the top of the screen. You will not be able to modify study-level metadata or submit the deposit while inside a folder. The “more” button in the file management area allows you to download, delete, or copy the entire project.

**Upload Files.** To upload files to the workspace, click on the “Upload Files” just below the study title. This will open a dialog box into which you can drag and drop your data, documentation, or other files. You can also import files from a zip file (including the folder structure of a complex zip package) by clicking on “Import from Zip” and dragging files into the dialog box that appears.

**Administrative Panel**

The Administrative Panel is located in the upper right corner of the deposit workspace (see Figure 2). This area includes buttons for submitting your completed deposit to ICPSR (see more information in the Deposit Submission section below), sharing the workspace with a colleague, changing ownership of the workspace, viewing the log of actions made on the deposit form, or reporting a technical problem to ICPSR.

**Project Description**

The Project Description area is located below the File Management area (see Figure 1). Information in the Project Description (i.e., study-level metadata) will improve the ability for data users to discover your study’s data. By providing detailed metadata, you allow future users to understand your study and determine if it will be useful in answering their research questions. You can click on the question mark next to any field for more information on that field and can always contact ICPSR for more support around metadata entry or any other aspect of the deposit process.

**Project Description** (see Figure 2). This section that includes fields for Principal Investigator(s), Summary, Funding Sources, Grant Manager(s), and Distributor(s). Please be sure to enter the name(s) of the Principal Investigator(s) exactly how you would like it/them to appear in the data citation. The names are not necessarily limited to the Principal Investigator(s) on your grant/contract from OPRE—the names listed here are the people who should be cited for creating this study and its data. Similar to the authors of a publication these should only be the people who did substantive work, and ICPSR will use the names you provide in the order you provide them. If you would like to acknowledge other members of the research team that information can be entered in the collection notes under Scope of Project (see Figure 4).
**Scope of Project** (see Figure 3). This section includes fields for Subject Terms, Geographic Coverage, Time Period(s), Collection Date(s), Universe, Data Type(s), and Collection Notes.

Figure 2: Project Description Section of ICPSR's Deposit Workspace

---

Figure 3: Scope of Project Section of ICPSR's Deposit Workspace
**Methodology** (see Figure 4). This section of the metadata editor includes fields for Response Rate, Sampling, Collection Mode(s), Scales, Weights, Unit(s) of Observation, and Geographic Unit.

Figure 4: Methodology Section of ICPSR's Deposit Workspace

**Related Publications**

The Related Publications area (shown in Figure 5) is where you and your team will add citations to documents which your team or any other researchers have published using the deposited data. This would include reports submitted to OPRE if those are available publicly. We encourage you to enter these citations because they will improve ability for the user to find your study’s data as well as relevant publications.

Figure 5: Related Publications Section of ICPSR's Deposit Workspace
**Project Communication Log**

**Project Communication Log** (see Figure 6). This is the area that you, ICPSR, and (if desired) OPRE will use to send each other messages about the project and share files other than the project’s main data and documentation, which are uploaded as described above in the File Management section.

Figure 6: Project Communication Log Section of ICPSR's Deposit Workspace

**Deposit Submission**

The deposit workspace is an effective place to share temporary, in-progress files for review and possible revisions. Once the deposit is complete with all finalized files, the Submit to ICPSR button is used to send the deposited files and metadata entered to ICPSR staff to begin curation.

Although several members of the research team can contribute to the deposit, the person who submits the deposit will be presented with terms and conditions and fields to complete. The person who has the capacity to sign for the research team is determined by the Principal Investigator(s) and the research team. Questions about this can be discussed during a three-way communication call or email.

**Terms and Conditions** (see Figure 7). This area is where the deposit signer responds to prompts accessing disclosure risk (i.e., Is there confidential information in the data?) and copyright (i.e., Is there copyrighted material in the deposit?) regarding the files in the deposit. The Additional Information box allows the signer to explain more about these two items, if needed. By signing the ICPSR Data Deposit Agreement electronically, the signer attests that they have the right to make the files in the deposit available through ICPSR; that ICPSR can redistribute, promote, catalog, and store and preserve the files; and ICPSR can change the files to protect respondent confidentiality and to improve usability.
Figure 7: Terms and Conditions Section of ICPSR's Deposit Workspace

Terms and Conditions

Assessing Disclosure Risk

Is there confidential information in the data?
- Yes
- No (required)

Copyright

Is there any copyrighted material in the deposit that requires special consideration regarding whether to disseminate it?
- Yes
- No (required)

Additional Information

Is there any additional information you would like ICPSR to know about this deposit?

Deposit Agreement

For ICPSR to archive and distribute this data collection, we require an electronic signature attesting to the following statements:

1. I have all rights needed to make this Data Collection publicly available through ICPSR.
2. I give my permission for the Data Collection to be used by ICPSR for the following purposes, without limitation:
   - To disperse copies of the Data Collection in a variety of media formats
   - To promote and advertise the Data Collection in any publicity (in any form) for ICPSR
   - To describe, catalog, validate and document the Data Collection
   - To store, translate, copy or re-format the Data Collection in any way to ensure its future preservation and accessibility
   - To incorporate metadata or documentation in the Data Collection into public access catalogues
3. I give my permission to ICPSR to enhance, transform and/or rearrange the Data Collection, including the data and metadata, for any of the following purposes:
   - Protect respondent confidentiality
   - Enhance usability

(required)
- I attest that the information provided in this deposit form is accurate and I agree to the terms stated above.
- I do not have the authority to agree to these terms.
- I have verified that the files included in this deposit are in the public domain and can be distributed by ICPSR.
- The following document grants me the right to deposit on behalf of the owner.

Submit Data close

Curation Process

Overview of Curation

Once submitted, ICPSR will curate your data. Curation of data is one of the main pillars of ICPSR’s work. Curation is the process of assessing, cleaning, documenting, and transforming your deposited data and documentation into complete, well-described, and accessible resources.
for secondary analysts. The end products of curation (data, metadata, and documentation) are released through a study’s homepage that is searchable and accessible from both the ICPSR and Child & Family Data Archive websites. The study will be assigned a Digital Object Identifier (DOI) when curation is complete and data are released; DOIs are persistent, unique identifiers which facilitate data citation. Data are released in a full suite of statistical software package file formats including SAS, SPSS, Stata, and R as well as ASCII data files in fixed-column (with setup files) and tab-separated values (TSV) formats. Documentation may include codebooks, user guides, questionnaires or other data collection instruments, or other documents which will enable end users to successfully conduct secondary analyses of the data. ICPSR’s curation team executes, manages, and oversees all curation work, maximizing standardization of work and flexibility. After the curation team’s initial review and during the curation process, the Child & Family Data Archive project managers will inform you and your FPO about any problems uncovered and keep you updated on the status of the data release.

**Disclosure Risk Review and Remediation**

Disclosure risk review and remediation is a crucial part of ICPSR’s curation work. The Federal Committee on Statistical Methodology states, “Statistical disclosure occurs when released statistical data (either tabular or individual records) reveal confidential information about an individual respondent.” The disclosure risk review and remediation process requires balancing research participant confidentiality with secondary analytic utility. The process includes reviewing:

- Variables for direct identifiers;
- Variables for indirect identifiers;
- Variables for data which may be linked to external data sources which have identifiers;
- Cross-tabulations of variables to determine if deductive identifications can be made; and
- Text variables for direct or indirect identifiers.

The curation team at ICPSR will also review any disclosure risk reports you submit as part of our internal review process.

The team establishes a disclosure risk remediation plan to minimize the disclosure risks identified in the review. The remediation plan may include:

- Restricting access levels,
- Deleting or masking entire variables,
- Deleting or masking individual disclosive data elements, and
- Recoding data (changing continuous data to categorical data, collapsing categories of categorical data, etc.).

When possible, ICPSR seeks to create a public-use dataset during curation. However, when confidentiality concerns exist, data will be released as restricted-use. Restricted data at ICPSR are disseminated via three main methods:

---

**Secure Download:** Upon approval, researchers will receive via e-mail the link to an encrypted file stored in ICPSR’s secure transfer box that they must download to the secure location specified in the data security plan of the application. ICPSR sends the temporary password to decrypt the file in a separate email once the researcher confirms that the encrypted file has been placed in the secure location.

**Virtual Data Enclave (VDE):** Rather than receiving a downloadable copy of the data, researchers are approved to access the data, stored on ICPSR’s secure server that is the VDE, via a remote desktop connection. Researchers are unable to move files from the VDE to their own local desktop or to the Internet. To receive output from the VDE, researchers must request that ICPSR conduct a disclosure review on the desired files. Output that “clears” disclosure review is then sent to the researcher.

**Physical Data Enclave (PDE):** The PDE is located onsite at ICPSR’s offices in Ann Arbor, MI. Data in the PDE typically contain highly sensitive personal information or data from protected populations. To analyze data located in the PDE, researchers must travel to ICPSR. To receive output, researchers must request that ICPSR conduct a disclosure review on the desired files. This review may happen after the researcher leaves Ann Arbor. Output that “clears” disclosure review is then sent to the researcher.

The level of restriction will be determined based on the sensitivity and identifiability of the data, requirements of consent and IRB, and discussions between you, your FPO, and ICPSR. If a public version of the restricted data can be made in addition to the restricted version (within the bounds of resources and data utility), ICPSR will seek to create one.

**Communication with Data Producers During Curation Workflow**

Throughout the curation process, ICPSR will maintain communication with you and your FPO based on previously determined schedules and means, adjusting as necessary (see above). ICPSR will contact you and your FPO immediately when major issues arise such as obstacles to timelines, disclosure concerns, or other issues which would result in the delayed release of your study’s data. ICPSR will collaborate with you on solutions. ICPSR will also provide you and your FPO with updates as milestones are reached.

**Other Considerations for Data Producers**

**Initial Data Submission.** After the initial release of the data, some datasets may require changes to the released data and documentation. However, please make every effort to deposit final versions of the data and documentation at time of initial submission. We also request that you notify your FPO and ICPSR as soon as possible if your study plans to release more data in the future (follow-up surveys, etc.), and we will work together to formulate a plan for releasing the additional data.
Data Trainings. Data trainings are an excellent way to introduce a study to prospective secondary users of data. If you think that you should provide a webinar or other training on the data you are depositing, please inform ICPSR of this plan as early in the planning process as possible. We also recommend including your involvement in data trainings as part of future grant/contract proposal budgets and timelines.

Collaboration. OPRE and ICPSR want researchers to be intentional about the archiving and curation of the data from their studies. As we work together to identify, prepare, deposit, and curate studies’ data and documentation, it is important to plan for FPO and producer involvement in these processes. As you work on future grant/contract proposals, we suggest that you include your involvement in archiving process as part of the proposed budget and timelines.

Further Information

To learn more about the Child & Family Data Archive or about data preparation and curation, please refer to the following:

- Child & Family Data Archive website, Deposit Data section under Support for Researchers
- ICPSR Guide to Social Science Data Preparation and Archiving
- Recommended Informed Consent Language for Data Sharing

Please contact Child & Family Data Archive project staff at ICPSR-help@umich.edu with any questions or requests for support.
Appendix A: Suggestions for Facilitating Future Secondary Analysis

The following will help future users better utilize your study’s data. If you would like support or suggestions from ICPSR on any of the following, please feel free to contact us.

☐ Be consistent with coding and labeling of values across variables (try not to use a range of 0-4 in some places and 1-5 in others).

☐ Label everything—even the things you think are obvious (example: a binary dummy variable coded 0 and 1).

☐ Avoid abbreviations, especially acronyms/shorthand/jargon used internally by your team or institutions.

☐ Ensure that your data and metadata only use ASCII characters. Examples of common non-ASCII characters include tabs, apostrophes or quotation marks that curve in a particular direction (’, ‘, “, and ” are not ASCII, while ‘ and ” are ASCII), em dashes (—) and en dashes (–) (hyphens are ASCII though: -), and characters with accents, umlauts, or other diacritical marks.

☐ Consider recoding special missing values (.A, .M, etc.) to numeric missing values (-9, -99, etc.) because not all statistical software packages can read special missing values.

☐ Document methods for creating constructed/recoded/index/derived variables. What rules, cut-points, or formulas were used to create those variables? What original variables were used to create those derived variables?

☐ Include the technical variables in the data file that are needed in order for statistical inference to be valid, such as weights, non-response adjustment, survey design variables, case disposition indicators, and other related variables.

☐ Consider creating a document, spreadsheet, or other crosswalk to show the association between variable names and original question text from data collection instruments. This is especially helpful in creating variable-level metadata and can be simple to create, depending on what types of software were used in planning and implementing the survey. We prefer to have the variable name in a separate column from the question text. This makes it easier to add any additional information (such as variable label) or adjust formatting before populating ICPSR’s question text file. For example:

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1a1</td>
<td>How many years have you lived there?</td>
</tr>
</tbody>
</table>
Appendix B: Deposit Checklist

Before submitting a deposit to ICPSR, please ensure that the following have been completed:

- Final version of the data file(s) have been created
  - Variable labels and value labels included (embedded or included in syntax/code/formats catalog)
  - Missing values designated (embedded or included in syntax/code/formats catalog)
  - Disclosure risk review complete and remediation taken
  - ICPSR feedback incorporated

- Final version of the data file(s) have been uploaded
- Final version of the data collection instrument(s) uploaded (if applicable)
- Codebooks uploaded (if variable labels, value labels, and missing values are not embedded in the data file(s) or accompanying syntax)
- Final version of the user guide has been uploaded (if applicable)
  - User guide is 508 Compliant
- Final versions of any other necessary documentation has been uploaded
  - Documentation is 508 Compliant

- Study-level metadata has been reviewed and finalized
  - Metadata has been reviewed by ICPSR

- Citations to related publications have been added

- The person submitting the deposit has the capacity to electronically sign the terms of deposit on behalf of the research team