Appendix H

Guidelines and Precedures for Preparing a Data File for the Prevalence Study

Feasibility Study for a National Registry of Child Maltreatment Perpetrators

Guidelines & Procedures
For Preparing a Data File for the Prevalence Study
April 2010

Table of Contents

1.0 Background

The Assistant Secretary for Planning and Evaluation (ASPE), US Department of Health and Human Services is conducting a study to assess the feasibility of developing and maintaining a National Registry of Child Maltreatment Perpetrators as mandated under the Adam Walsh Child Protection and Safety Act of 2006. Walter R. McDonald & Associates Inc (WRMA) has been contracted to conduct the study. The study has two parts: a Prevalence Study and a Key Informants Survey.

The purpose of the Prevalence Study is to estimate how frequently child maltreatment perpetrators have substantiated investigations in multiple states. In order to make these estimates possible, States are asked to provide date of birth and encoded (and therefore not identifiable) names for all substantiated perpetrators reported for the last five years of NCANDS submissions. It is estimated that providing the requested data will require approximately 30 hours of staff time on the part of each participating state. Participation is voluntary. We are hoping that your state will participate so that its experience will be represented in this study.

This document provides instructions for preparing the data that will be used in the prevalence study.

2.0 Data Submissions

To efficiently and accurately conduct the prevalence study a database containing information about perpetrators will be created. States participating in the study are requested to submit the date of birth and encoded name information for perpetrators previously reported in the NCANDS Child Files for the last five years. *Exhibit 1-1, Perpetrator Data Collection Process,* graphically depicts the activities that comprise the data collection processes.

States will *receive* from the study contractor, WRMA, the **Perpetrator Extract File (PErp File)** which contains all unduplicated perpetrator IDs (associated with substantiated maltreatments) for the State, for each of the last five years. The data is extracted from the NCANDS Child Files containing case level data. A single record will consist of the State abbreviation, year, perpetrator ID, report date and the county of report (refer specification in *Exhibit 3.1.1 Data File Specification for the PErp File*). The perpetrator IDs are unduplicated only for each submission year, and not unduplicated for the entire 5 years, to account for inconsistencies, if any, in the encryption methods employed each year.
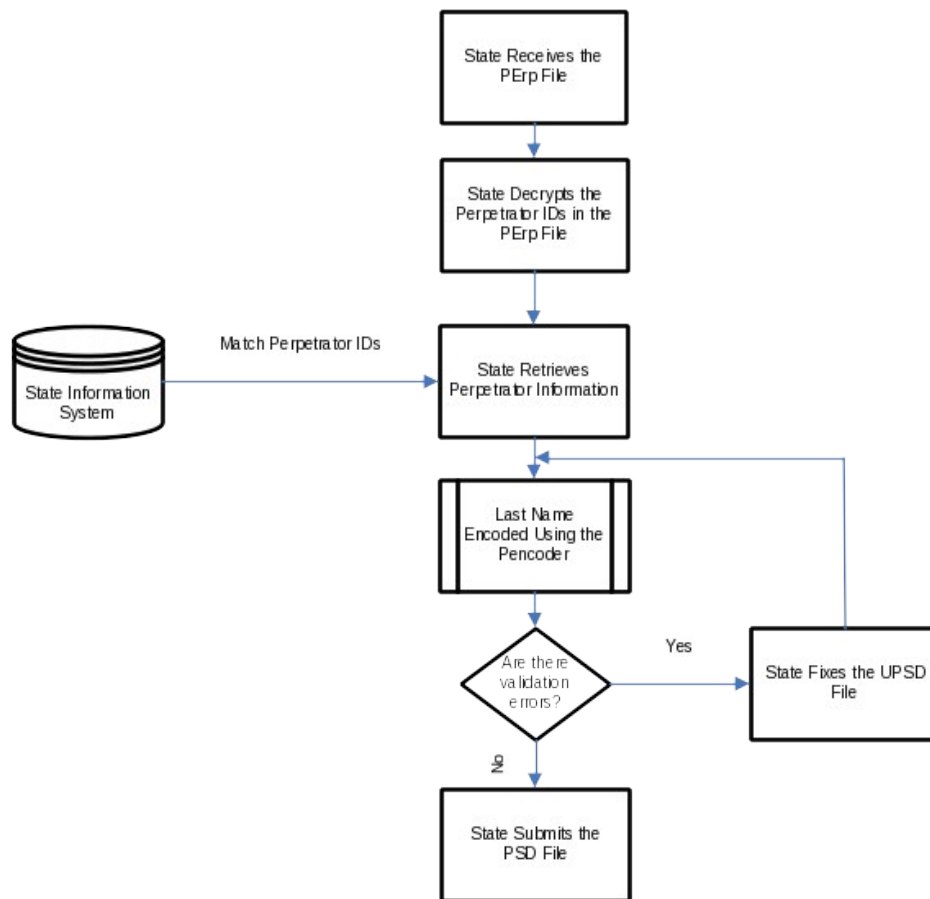
The State decrypts the perpetrator IDs in the received PErp file and matches each with the corresponding perpetrator in the State information system. Once the perpetrator is identified, a data extract file with additional information is created. This file is called the Unencoded Perpetrator State Dataset (UPSD) file

(refer specifications in *Exhibit 3.2.1 Data File Specification for the UPSD File*).
The last name information in the UPSD file is not encoded. The UPSD file is the
input for the encoding software.

The State will also receive an encoding software called the Pencoder. The UPSD
file is the input to the Pencoder. The Pencoder will encode the last names of the
perpetrators in the UPSD file using the New York State Identification and
Intelligence System (NYIIS) algorithm. The Pencoder will also validate the input
file to make sure that all fields confirm to the specifications.

The output file of the Pencoder software is called the Perpetrator State Dataset
(PSD) file (refer specifications in *Exhibit 3.4.1 Data File Specification for the PSD
File*). This file should be submitted to the study team through the secure web site
described in section 5.0. The perpetrator information in the PSD file will be used
to link to the NCANDS Child Files, to create the database for the prevalence
study.

Exhibit 1-1, Perpetrator Data Collection Process

3.0 Specification for Creating the Data Files

This section explains how the State creates the perpetrator dataset file. This includes describing the structure of the file, the data records in the file, the data elements in the records, and the procedures used for constructing the data file. Each State receives the PErp file and submits the perpetrator dataset file.

3.1 Perpetrator Extract File (PErp File)

This file contains all unduplicated perpetrator IDs (associated with substantiated maltreatments) for a State, for each of the last five years. The data is extracted from the NCANDS Child Files containing case level data. A single record consists of the State abbreviation, year and the perpetrator ID. This file is submitted to the State in text format (TXT).

Exhibit 3.1.1 Data File Specification for the PErp File

FIELD # (POSITION)	LONG NAME (SHORT NAME)	FIELD TYPE & CODES (Example)	FIELD LENGTH
1 (1-2)	STATE/TERRITORY (STATEAB)	ALPHABETIC (CT, VA)	2
2 (3-6)	SUBMISSION YEAR (SUBYR)	NUMERIC (2007, 2008)	4
3 (7-18)	NCANDS PERPETRATOR ID (NPERPID)	ALPHANUMERIC (00004356ABDF)	12
4 (19-26)	REPORT DATE (RPTDT)	NUMERIC [mmddyyyy] (12052008)	8
5 (27-29)	REPORT COUNTY (RPTCNTY)	NUMERIC	3

Example Record:

CT200700004356ABDF12202007040

3.2 Unencoded Perpetrator State Dataset (UPSD) file

The State decrypts the perpetrator IDs in the PErp file and matches it with the corresponding perpetrator in the State information system. Once the perpetrator is identified, a data extract file with additional information is created. This file is called the Unencoded Perpetrator State Dataset (UPSD) file. The last name information in the UPSD file is not encoded. The UPSD file is the input for the encoding software. This file is in text format (TXT).

Exhibit 3.2.1 Data File Specification for the UPSD File

FIELD # (POSITION)	LONG NAME (SHORT NAME)	FIELD TYPE & CODES (Example)	FIELD LENGTH
1 (1-2)	STATE/TERRITORY (STATEAB)	ALPHABETIC (CT, VA)	2
2 (3-6)	SUBMISSION YEAR (SUBYR)	NUMERIC (2007, 2006)	4
3 (7-18)	NCANDS PERPETRATOR ID (NPERPID)	ALPHANUMERIC (1234C45D40KL)	12
4 (19-26)	REPORT DATE (RPTDT)	NUMERIC [mmddyyyy] (12052008)	8
5 (27-29)	REPORT COUNTY (RPTCNTY)	NUMERIC (040, 001)	3
6 (30-41)	STATE ENCRYPTED PEPETRATOR ID (STPERPID)	ALPHANUMERIC (345DES987QKP)	12
7 (42)	UNENCODED FIRST INITIAL (FIRSTINI)	ALPHABETIC (D, S)	1
8 (43-92)	UNENCODED LAST NAME (LASTNM)	ALPHABETIC (SMITH, Johnson)	50
9 (93-100)	PERPETRATOR DATE OF BIRTH (PERPDOB)	NUMERIC [mmddyyyy] (12052008)	8

Example Record:

CT200700004356ABDF12202007040345DES987QKPDSMITH   12052008

Special Instructions:

1. The State encrypted perpetrator ID should be left-filled with zeroes, as needed to the 12 character length. For example, a perpetrator ID of "7856SDFG" is invalid. It should be reported as "00007856SDFG".

2. The perpetrator date of birth must be in month-day-year (mmddyyyy) format.

3. The unencoded first initial should be in caps. Ex: D for David.

4. The unencoded last name should be right-filled with spaces, as needed, to the 50 character length.

5. If data for a field is unavailable, the filed should be filled with blank spaces in accordance with the field length.

3.3 Perpetrator Encoding (Pencoder) Software

The last name information in the UPSD file is encoded by the Pencoder software. The output file from the Pencoder will be similar in format as the input file. However, the last name information will be encoded using the NYSIIS algorithm. The Pencoder software is provided to the State along with the PErp file.

The Pencoder will also validate the input file to make sure that all fields confirm to the specifications. The validation rules enforced in the Pencoder are as follows:

1. A valid State code should be entered in the STATEAB field. If invalid data is found, the entire record is removed.
2. The Submission year should be between 2004 and 2008. If invalid data is found, the entire record is removed.
3. The NPERPID and STPERPID fields should be 12 characters in length. If invalid data is found, the entire record is removed.
4. The PERPDOB filed should have valid month, day and year values. If invalid data is found, the PERPDOB filed is blanked.

The errors, if any, found during the validation and encoding process are reported in Error.txt file. This file is an output of the Pencoder. States are requested to review this document, fix the errors in the UPSD file, and run it through the Pencoder again. This process should be repeated until no errors are found.

More information about using the Pencoder application is under section *4.0 Using the Pencoder Application.*

3.4 Perpetrator State Dataset (PSD) file

This file is the output of the Pencoder software. This file should be submitted to the registry program. The perpetrator information in the PSD file will be used to link to the NCANDS Child Files, to create the prevalence database. This file is in text format (TXT).

Exhibit 3.4.1 Data File Specification for the PSD File

FIELD # (POSITION)	LONG NAME (SHORT NAME)	FIELD TYPE & CODES (Example)	FIELD LENGTH
1 (1-2)	STATE/TERRITORY (STATEAB)	ALPHABETIC (CT, VA)	2
2 (3-6)	SUBMISSION YEAR (SUBYR)	NUMERIC (2007, 2006)	4
3 (7-18)	NCANDS PERPETRATOR ID (NPERPID)	ALPHANUMERIC (1234C45D40KL)	12
4 (19-26)	REPORT DATE (RPTDT)	NUMERIC [mmddyyyy] (12052008)	8

5 (27-29)	REPORT COUNTY (RPTCNTY)	NUMERIC (040, 001)	3
6 (30-41)	STATE ENCRYPTED PEPETRATOR ID (STPERPID)	ALPHANUMERIC (345DES987QKP)	12
7 (42)	UNENCODED FIRST INITIAL (FIRSTINI)	ALPHABETIC (D, S)	1
8 (43-92)	ENCODED LAST NAME (LASTNM)	ALPHABETIC (SRRIA, JFDGTFD)	50
9 (93-100)	PERPETRATOR DATE OF BIRTH (PERPDOB)	NUMERIC [mmddyyyy] (12052008)	8

Example Record:

CT200700004356ABDF12202007040345DES987QKPDSRRIA   12052008

4.0 Using the Pencoder Application

The State will receive an encoding software called the Pencoder. The UPSD file is the input to the Pencoder. The Pencoder will encode the last names of the perpetrators in the UPSD file using the New York State Identification and Intelligence System (NYIIS) algorithm. The Pencoder will also validate the input file to make sure that all fields confirm to the specifications. The Pencoder can be downloaded by the States from the Collaborator, a secure internet data storage site.

4.1 System Requirements

The Pencoder is implemented as a relational database application in Microsoft Access 2003. Users must have Access 2003 or later installed on their computer. It operates in both Microsoft Windows XP/Vista/7 environments. A single Pencoder Access database file contains all of the system data, programming modules, tables, queries, forms, and reports necessary for the operation of the application. The input to the Pencoder is the State Unencoded State Dataset (UPSD) File and the output from Pencoder is the PSD File along with a report with results from the validation and encoding processing. All intermediate data sets created during the processing are contained within the Pencoder Access database. The Pencoder is distributed as a compiled Access file (MDE) along with essential documentation, including this Guidelines document. The distributed size of the Access file is about 8mb. The distribution of Pencoder to the States is accomplished via the Collaborator.

The hardware needed to run Pencoder includes: processor speed of 2.0 GHz, 1-2 GB of RAM and sufficient hard drive space to dedicate a gigabyte to the Pencoder database file (smaller States would need less hard drive space).

4.2 Downloading Pencoder from the Collaborator

The Pencoder is packaged in a single zip file and contains the following components:

- The Pencoder Access database file (.mde);
- The Bench Mark (BM) State UPSD File with sample records.

The State user can download the Pencoder zip file package from the Collaborator by:

1. Go to www.wrma.com
2. Click on "Extranet" on the top menu. A new page for the Collaborator will open.
3. Enter the login information provided to you.
4. Click on Registry_<State Name>
5. Click on "Documents"
6. Click on "Registry"
7. The "Registry Pencoder <Date>.zip" is the entire Pencoder application. Select "Download Document" from the dropdown list on the right.
8. The Pencoder distribution zip file should be downloaded into the **C:** folder on the State computer.

4.3 Installing the Pencoder on a Local Computer

The Pencoder is installed by unzipping and extracting the single distribution zip file:

C:\ Registry Pencoder <Date>.zip

into the user's C:\ folder. When the extraction occurs, a file structure will be created under the C:\Registry\ folder with the following folder structure and file contents being populated:

- C:\Registry\Pencoder\Input\BMRegistry.txt (BM test file);
- C:\Registry\Pencoder\Output\Log.txt (Test file);

4.4 Launching the Pencoder

The Pencoder is launched by double clicking on the Access file located at:

C:\Registry\Pencoder\Registry Pencoder.mde.

4.5 Location of Input and Output Files

The UPSD input file is typically stored in the C:\Registry\Pencoder\Input\ folder. The Pencoder places the encoded File in the C:\Registry\Pencoder\Output\ folder. The filename of the output PSD File is of the format: <State>_yyyymmdd_hhmmss.txt, where the date and time in the file name is set to the time the file was opened for writing. The results report created by Pencoder is placed in the same folder with the filename format of Results<State>_yyyymmdd_hhmmss.txt.

5.0 Submitting Files to the Registry Study

A State submits the Perpetrator State Dataset (PSD) file. Typically, this file will have been generated from the Pencoder application. The data files are submitted to and from the States on the Collaborator, a secure internet data storage site. The primary State contact for the study will receive a State username and password to access their State folder on Collaborator.

The State user can upload the PSD file to the Collaborator by:

1. Go to www.wrma.com
2. Click on "Extranet" on the top menu. A new page for the Collaborator will open.
3. Enter the login information provided to you.
4. Click on Registry_<State Name>.
5. Click on "Documents".
6. Click on "Registry".
7. Click on "Add Document" link on the top of the page.
8. On the "Add Document" page, scroll to the bottom of the page and click on the "Add" button.
9. On the "Open Document" dialog screen, browse and select the PSD file. Click on "Open".
10. Click on the "Submit" button to submit your file.