

State Child Welfare Data Linkages Descriptive Study: Data User's Guide



STATE CHILD WELFARE DATA LINKAGES (SCW) DESCRIPTIVE STUDY

DATA USER'S GUIDE

September 2024

Eileen Gao, Joanne Lee, and M.C. Bradley

Submitted to:

Jenessa Malin and Christine Fortunato, Project Officers
Eleanor Eng, Project Specialist
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services
Contract Number: 47QRAA18D00BQ/75ACF121F80021

Submitted by:

Project Director: M.C. Bradley
Mathematica
1100 First Street, NE, 12th Floor
Washington, DC 20002-4221
Phone: (202) 484-9220
Fax: (202) 863-1763

This report is in the public domain. Permission to reproduce is not necessary. Suggested citation: Gao, Eileen, Joanne Lee, and M.C. Bradley (2024). *State Child Welfare Data Linkages Descriptive Study: Data user's guide*, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Disclaimer

The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research, and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

This report and other reports sponsored by the Office of Planning, Research and Evaluation are available at www.acf.hhs.gov/opre.

[Subscribe to OPRE News and Follow OPRE on Social Media](#)



Contents

Background.....	1
Archived Data	3
Accessing the data	3
Resources for using the data	3
Technical assistance.....	3
Acknowledgement of source.....	4
Publication submission requirement.....	4
Data Collection Procedures.....	5
Developing the instruments.....	5
Sample	5
Data collection.....	6
Initial Survey Data File	8
Data structure.....	9
Variables	9
Variable characteristics.....	9
Connected Data Survey Data File	10
Preparing the data	10
Data structure.....	10
Variables	11
Variable characteristics.....	12
Using the data	12
File formats.....	12
Analytic considerations	13
References.....	13
Glossary.....	13
Appendix A: Variables in the Initial Survey Data File.....	14
Appendix B: Variables in the Connected Data Survey Data File	18

Background

The Child Abuse Prevention and Treatment Act (CAPTA), first authorized in 1974 and reauthorized regularly since then, requires the examination of a wide range of topics related to the incidence of child abuse and neglect with the aim of informing efforts to better protect children from maltreatment and improve the well-being of victims of maltreatment. One promising approach to addressing these topics is to connect administrative records—such as those from child welfare, health, social services, education, public safety, and other agencies. Connecting data may help improve the quality, usefulness, interoperability, and availability of child maltreatment data.

The State Child Welfare Data Linkages (SCW) Descriptive Study aims to provide novel information regarding connected (linked or integrated) state data that may be leveraged to improve the ongoing and accurate surveillance of child maltreatment incidence and related risk. The study examined the extent to which child welfare agencies in 50 states, Puerto Rico, and Washington, DC (hereafter collectively referred to as states) connected administrative data on child maltreatment to other data sources and aimed to learn more about agency practices related to sharing and connecting data.

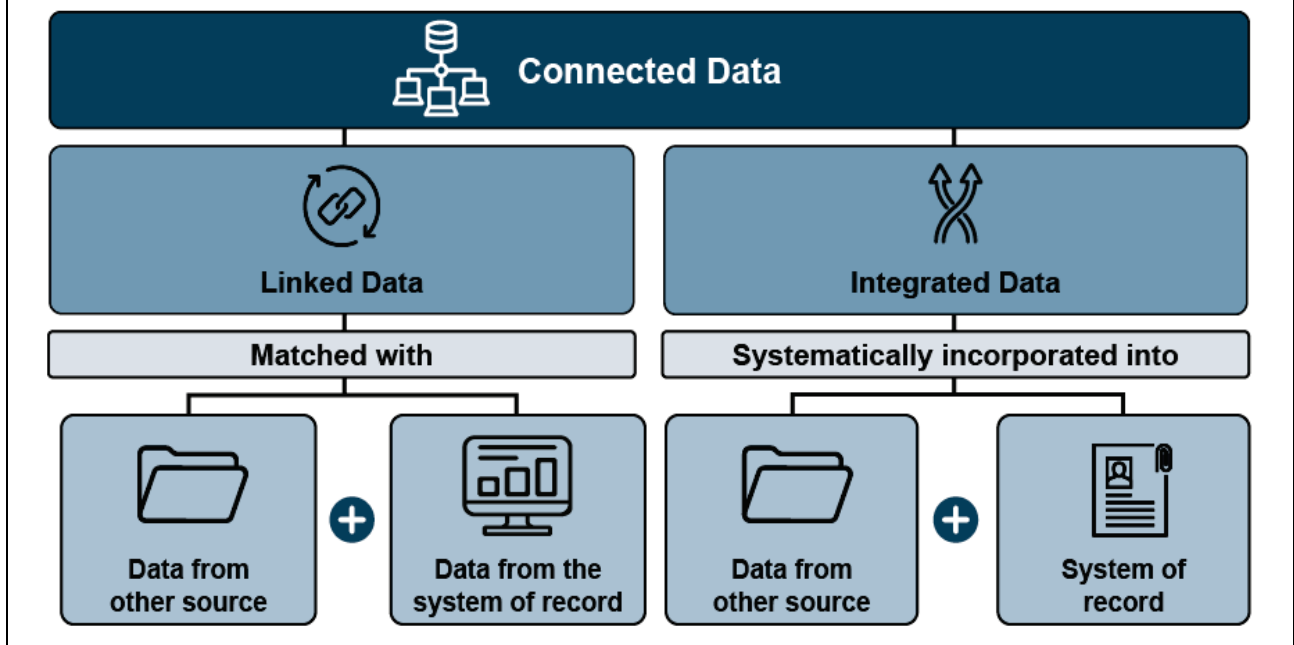
The SCW Descriptive Study explored the following research questions:

- What are the characteristics of states' data systems of record?
- What states have connected data and how are those data used?
- What data sources are being connected, and what are the characteristics of the connected data?
- How do states link, manage, and govern their connected data?
- What are states' plans for connected data and how are they building capacity?

Exhibit 1. The SCW Descriptive Study's definitions of linked data, integrated data, and connected data

The SCW Descriptive Study defined three terms for the purposes of the study:

- **Linked data:** A set of records that includes (1) data from the state system of record for child maltreatment reporting to the National Child Abuse and Neglect Data System and (2) data from other sources, based on a common identifier or other matching criteria
- **Integrated data:** Data that have been systematically incorporated (through direct entry by staff or a data exchange) into the state system of record for child maltreatment reporting to the National Child Abuse and Neglect Data System
- **Connected data:** Linked or integrated data



The SCW Descriptive Study team conducted the following three data collection activities:

1. A high-level web survey of state child welfare directors (or their designees).
2. A more in-depth web survey of a state agency staff person knowledgeable about connected data.
3. Semi-structured interviews with a subsample of state agency staff to further learn about efforts to develop linked data, the outcome of those attempts, methodologies used to link data, the policies and resources that support or hinder connected data efforts, and future plans.

This document provides information about the survey data that was archived from the study. The surveys were conducted from July 2022 through January 2023. Therefore, the data may not reflect the current status of states' connected data. Products from the SCW Descriptive Study, including briefs and technical reports, are posted at <https://www.acf.hhs.gov/opre/project/state-child-welfare-data-linkages-descriptive-study>.

Archived Data

Accessing the data

Survey data from the SCW Descriptive Study are archived for restricted use at the National Data Archive on Child Abuse and Neglect (NDACAN). Interview data gathered by the SCW Descriptive Study are not archived. Survey data files are available for researchers to use after following NDACAN procedures for requesting access to a restricted data set. Guidance for requesting access to a restricted data set can be found at <https://www.ndacan.acf.hhs.gov/datasets/request-restricted-data.cfm>.

The full data files available from NDACAN are (1) data from the Initial Survey, where each record is a state's general response about the state's child welfare data system and connected data, and (2) data from the Connected Data Survey, where each record is a state's response about integrated data and/or a single linked data set in the state, for up to four linked data sets.

Resources for using the data

The SCW Descriptive Study has made the following documentation available to support users of the data:

- This data user's guide has information about each data set, including an overview of data collection procedures and guidance on using the data, such as how to link the data sets to each other and with other data sources.
- A codebook for each data file is also available to support users. Each codebook provides information about each variable in the data set, including variable names, labels, definitions, variable type, and frequencies.
- [State Child Welfare Data Linkages Descriptive Study Technical Report: Study Background and Design](#) includes detailed information about methods and includes the survey instruments as appendices.
- [State Child Welfare Data Linkages Descriptive Study Technical Report: Study Findings](#) includes study findings from across the surveys and interviews.

This data user's guide, along with the two codebooks, can be found on the NDACAN website at <https://www.ndacan.acf.hhs.gov/datasets/dataset-details.cfm?ID=287>. Additional information about the study is available at <https://www.acf.hhs.gov/opre/project/state-child-welfare-data-linkages-descriptive-study>.

Technical assistance

Users can receive technical assistance by submitting a request to NDACANsupport@cornell.edu. NDACAN staff provide free user support for the data sets they distribute. They can address issues such as importing data to an analysis program, clarifying variable names or labels, or solving problems with the data as delivered. There are also support resources provided on the User Support page of NDACAN's website (<https://www.ndacan.acf.hhs.gov/user-support/user-support.cfm>).

Acknowledgement of source

Authors should acknowledge NDACAN and the original collector of the data when they publish manuscripts that use data provided by NDACAN. Users of these data are requested to include the following statement or an adaptation of it:

The data used in this publication were made available by NDACAN and have been used with permission. Data from the State Child Welfare Data Linkages (SCW) Descriptive Study were originally collected by: M.C. Bradley, Joanne Lee, Jeremy Page, Tara Strelevitz, and Rachel Sutton-Heisey at Mathematica. Funding for the project was provided by the Office of Planning, Research, and Evaluation in collaboration with the Children's Bureau in the Administration for Children and Families at the U.S. Department of Health and Human Services (Contract Number: 47QRAA18D00BQ/75ACF121F80021). The collector of the original data, the funder, NDACAN, Duke University, Cornell University, and their agents or employees bear no responsibility for the analyses or interpretations presented here. The information and opinions expressed reflect solely the opinions of the authors.

The bibliographic citation for this data collection is as follows:

Bradley, M.C., Lee, J., Page, J., Strelevitz, T., & Sutton-Heisey, R. (2024). *State Child Welfare Data Linkages (SCW) Descriptive Study Initial Survey and Connected Data Survey* [Data set]. National Data Archive on Child Abuse and Neglect. <https://doi.org/10.34681/ZYZJ-3X67>.

The citation for these data sets and other data sources at NDACAN can be found at the Child Abuse and Neglect Digital Library (canDL), a free database maintained by NDACAN, at <https://www.zotero.org/groups/421939/candl/library>. Users are strongly encouraged to review published works, based upon these data, before doing analyses.

Publication submission requirement

Users of this data are required, in accordance with the terms of the data license for this data set, to notify NDACAN of any published work or report based wholly or in part on these data. A copy of any completed manuscript, thesis abstract, or reprint should be emailed to NDACANSUPPORT@cornell.edu. Such copies will be used to provide NDACAN's funding agency with essential information about the use of NDACAN resources and to facilitate the exchange of information about research activities among data users and contributors.

Data Collection Procedures

This section provides a high-level overview of instrument development, sample, data sources, and data collection.

Developing the instruments

The study team, led by Mathematica, designed the various data collection instruments to build on one another and tailor data collection to different respondents. The survey items were created for the surveys and not sourced from other data collections or publications.

- The first survey of state child welfare directors (*Initial Survey*) set the foundation for future data collection by identifying the best respondents for each state and gathering background information on how the state's child welfare agencies are organized, what the system of record is, how data are managed, and any known data linkages.
- The second survey (*Connected Data Survey*) built on the Initial Survey by asking the identified contacts for connected data sets to expand on information provided in response to the first survey. This survey had two sections. The first focused on a specific linked data set and data sharing agreement identified in the Initial Survey. The study team sent this section on linked data sets to a contact identified in the Initial Survey as being knowledgeable about the specific linked data set and data sharing agreement. The second focused on data integrated into the system of record. Skip logic determined which section(s) a respondent would get based on whether the Initial Survey identified them as a contact for linked or integrated data. The study team sent the section on integrated data to the contact identified in the Initial Survey as being knowledgeable about non-child-welfare data integrated into the system of record.

The study team sought input from a range of experts—including current and former staff at state child welfare agencies, federal staff, and researchers—to help review and pre-test the data collection instruments.

Once the study team developed the draft instruments, the team asked volunteers from the Casey Family Programs Data Leaders group to pre-test them. The Data Leaders group is organized by Casey Family Programs and includes child welfare staff from state agencies with leadership roles related to research and evaluation, data analysis, federal reporting, quality assurance, and continuous quality improvement. Non-state partners also participate in the Data Leaders group, including individuals who work with state members on research topics and former state staff with data expertise. Pre-testers identified ways that their states' contexts could affect responses to questions about connecting state child welfare data with data from other agencies and organizations. Pre-testing was conducted in October 2021. Respondents provided feedback on question layouts, their comfort level when answering questions, clarity, and response options for multiple-choice questions.

The Office of Management and Budget (OMB) and the Health Media Lab Institutional Review Board approved the finalized instruments and outreach materials.

Sample

The samples for each survey are discussed separately in this section.

Initial Survey

The target population was comprised of 52 states. This survey focused on child welfare directors because they would generally know whether their agency partnered with other agencies and organizations to collect and connect data. The study team also expected them to be able to identify other state contacts who could answer specific questions about connected data in later data collection activities. The study team recognized that some child welfare directors would delegate part or all of the survey questions to other staff, such as their deputy or senior data or research staff. Thus, the Initial Survey was conducted by web. Each state received one web link to a survey, and each state could share the web link to its survey with any designees.

Connected Data Survey

This survey focused on staff who were knowledgeable about integrated data in their state or about linked data that fall under specific data sharing agreements. Staff in this target population were identified by child welfare directors (or their designees) through the Initial Survey if directors said their state had integrated data or data sharing agreements covering linked data sets. In states with integrated data, respondents were asked in the Initial Survey to name someone who could answer questions about integrated data in the Connected Data Survey. Staff who were identified as knowledgeable about integrated data were sent the section of the Connected Data Survey related to integrated data. In states with data sharing agreements covering linked data sets, respondents were asked in the Initial Survey to name someone who could answer additional questions about linked data sets from each data sharing agreement, for up to four agreements. Twenty-eight states reported having at least one data sharing agreement to support a linked data set; 17 states reported more than one data sharing agreement to support a linked data set; and 8 states reported having four or more data sharing agreements to support a linked data set. Staff who were identified as knowledgeable about those linked data sets were sent the section of the Connected Data Survey related to linked data, for up to four data sharing agreements. Staff who were identified in the Initial Survey as knowledgeable for integrated data and linked data sets from a data sharing agreement received one link to the Connected Data Survey that contained sections of questions on integrated data and linked data.

Data collection

Data collection is discussed for each survey separately in this section.

Initial Survey

Before the start of the survey fielding period, the team worked with Children's Bureau regional program managers to confirm the accuracy of publicly available contact information for child welfare directors and to send the directors advance emails about the survey. The survey was open in July and August of 2022. Mathematica sent reminders by phone and email during the survey fielding period. A total of 44 states and their child welfare agencies completed the survey, for a final response rate of 84.6 percent.

Connected Data Survey

The Children's Bureau sent the intended respondents an advance email about the Connected Data Survey one week before sending them the link(s) to the web survey. In addition, the team contacted the Casey Family Programs' Data Leaders group, which had been contacted earlier to identify volunteers to conduct

pre-tests. The team shared reminders with this group about the study and the timing of its data collection activities after OMB approval and after the Initial Survey fielding period, anticipating that members of this group would respond to or help other state staff respond to the Connected Data Survey.

The survey opened on October 18, 2022, when the study team emailed the intended respondents a link to a survey including questions about their state's integrated data or a specific data sharing agreement (or linked data set). The survey included skip logic and autofill fields to tailor each survey to the states' integrated data or identified data sharing agreement. All states indicated they had integrated data and were thus eligible to respond to the integrated data section of the survey. Only the subset of states that indicated that they had linked data sets were eligible to respond to the linked data set section of the survey. Staff in some states received more than one survey if the state reported integrating or linking multiple data sets. Multiple surveys could be sent to staff within a state. Anyone who had the link was able to access that web survey, so the prospective respondents could easily share the survey with others in their agencies, if necessary. The team sent a total of 92 web surveys. The study team sent reminders by phone and email during the survey fielding period. The survey closed on January 13, 2023.

Exhibit 2. Respondents to the Initial Survey and Connected Data Survey

	Initial Survey	Connected Data Survey		
		Overall	Linked data set section	Integrated data section
Number of eligible states	52	44	28	44
Number of responding states	44	31	18	29
Percentage of responding states	84.6%	70.5%	64.3%	65.9%
Number of eligible surveys	52	92	67	44
Number of survey responses	44	57	37	29
Percentage of survey responses	84.6%	62.0%	56.7%	65.9%

Source: SCW Descriptive Study, Initial Survey and Connected Data Survey.

Of the 44 states receiving at least one Connected Data Survey, 31 (70 percent) responded to at least one survey (Exhibit 2). Among those 31 states, 92 surveys were sent and 57 were completed, for an overall response rate of 62 percent. The denominator for this response rate reflects all states that had integrated and/or linked data as indicated in the Initial Survey. Twenty-eight states received Connected Data Survey questions about their linked data sets, and 18 states (64 percent) responded. Across those 28 states, we sent 67 surveys that included questions about linked data sets, and we received responses for 37 of them (57 percent). Eight states gave information for only one linked data set, and 10 states responded to two to four surveys about separate linked data sets.

Although a state could indicate it had only linked data and no integrated data, and vice versa, every state that responded to the Initial Survey reported integrating some data into its system of record from the list of services provided in the survey (which included child welfare services). Consequently, all 44 states that completed the Initial Survey were asked questions about their integrated data system. Twenty-nine states (66 percent) responded to those questions. Because the Connected Data Survey could only ask about one integrated data system per state, the number of states and the number of surveys with questions about integrated data are equal.

Initial Survey Data File

This section describes how the Initial Survey data file was prepared, the structure of the data file, and key information about the variables in the data file. The Initial Survey instrument is included as an appendix to [State Child Welfare Data Linkages Descriptive Study Technical Report: Study Background and Design](#).

All of the survey items can be divided into three mutually exclusive groups: (1) multiple-choice questions, (2) items requesting contact information or for survey administration only, and (3) items requesting names of data-sharing agreements and partner agencies. During and after data collection, the team examined frequencies for all variables to confirm that survey skip patterns were accurate, the number of responses was correct, and open-text responses were plausible. The team cleaned each type of data as follows:

- **For multiple-choice questions**, the study team contacted the primary respondent for clarification if needed. Some respondents to the second survey (Connected Data Survey) wanted to correct information provided in the first survey, so the team did make those corrections. The study team did not ask all respondents to confirm information provided in the first survey, however, and the study team did not edit data provided by states unless specifically requested by the state.

Some multiple-choice questions included an "Other" response option which, if selected, prompted states to provide additional information in open text. The study team reviewed responses to open-ended "other" options to determine whether any response could be categorized under an existing response option or whether a new response option was warranted. Our preference was to categorize open-text responses under existing response options. For open-text responses that could not be categorized under an existing response option, we determined whether a revised category could cover open-text responses from at least two states. We required the revised category to include at least two states to limit information that could potentially identify the individual respondents. If there was only one open-text response, then we grouped it with the most similar existing response option and renamed this response option to cover the open-text response too. The team replaced the original open-text responses with the revised categories, so no variables in the final files include any open-text responses.

- **For contact information**,¹ the study team cross-checked the information provided in the survey with websites to verify name, affiliation, role, and contact information. In a few instances, the team corrected typos in contact information provided in the first survey by using publicly available information or by reaching out to the respondent for the first survey.
- **For names of data-sharing agreements and partner agencies**,² the study team cross-checked agencies' names provided in the survey with websites to verify their names. The study team corrected a few spelling errors because these fields fed into the other data collection activities.

¹ Contact information is mentioned here to describe the full data collection, but it is excluded from the distributed data files to protect confidentiality.

² Data sharing agreement names and partner agencies are mentioned here to describe the full data collection, but they are excluded from the distributed data files to protect confidentiality

Data structure

The data file has 44 records, with one record for each of the 44 states that responded.

Variables

The data file includes variables based on states' responses to multiple-choice questions in the survey instrument and other variables for research use. The only variables excluded from the archived file that were present in the full survey are those including personally identifiable information (PII) or related to survey administration, such as: names of individuals, organizations, and data sharing agreements (which were used to help respondents to the Connected Data Survey identify the focal data set); and other contact information that was gathered to field the surveys.

- **For multiple choice questions**, the team created categorical variables where the categories align to the choices in the survey item. As described above, the team combined responses for a few variables where open-text information from states was back-coded. In the data sets provided in the file formats native to Stata (.dta) and SAS, and in the codebook, the value ".s" denotes logical skips in the survey and the value ".m" identifies item non-response. In the files provided in SPSS and comma-delimited format, we use a numeric placeholder of -98 for logical skips and -99 for item non-response. In the Initial Survey codebook, the variable label includes the relevant question number in the Initial Survey instrument. No other masking or data manipulation have been done to the data.
- **Other variables for research use** include derived variables and state identifiers added by the study team. The team used derived variables to report publicly available findings in an infographic, briefs, and technical reports, which are available at <https://www.acf.hhs.gov/opre/project/state-child-welfare-data-linkages-descriptive-study>. Definitions of the derived variables are included in its variable label in the data file and in the codebook. The team added two string variables that can be used to identify states and merge this data file with other data sets.³ The variable **state** contains the state's two-letter postal abbreviation, and **fips** contains the state's numeric Federal Information Processing Standard (FIPS) code. The team also added information in the numeric variable **admin** on whether the child welfare system is state-administered system, using information from the Child Welfare Information Gateway (see [this 2018 fact sheet](#))⁴.

For all variables in the file, we confirmed accuracy of the descriptions of numeric values (value labels) for variables as shown in the codebook, that variable labels were accurate, that the only string variables included in the file were state identifiers, and that missing value patterns and counts were plausible.

Appendix A shows how variables are ordered in the final data files and provides descriptions, related survey item numbers, and information on variable naming conventions.

Variable characteristics

The codebook shows the following information about each variable in the Initial Survey data file:

³ The [NDACAN User Support page](#) includes examples of using different statistical packages to merge data.

⁴ Children's Bureau. "State vs. County Administration of Child Welfare Services." Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services, 2018. <https://cwig-prod-prod-drupal-s3fs-us-east-1.s3.amazonaws.com/public/documents/services.pdf>.

- **Variable name:** Used to identify the data element in the data set that the variable corresponds to. Variables were named descriptively rather than in reference to a specific survey item.
- **Variable label:** A brief description of the variable. For variables that are not derived variables and are not state identifiers, the label also includes the specific relevant survey item(s).
- **Variable type:** Either numeric or string.
- **Universe:** The total number of cases possible for each variable in the data set. The universe for all variables is 44, which represents the 44 states who responded to the survey.
- **N:** The total number of cases with valid and non-missing values for the variable.
- **Frequencies:** The frequencies for each value are presented for all numeric variables.

Connected Data Survey Data File

This section describes how the Connected Data Survey data file was prepared, the structure of the data file, and key information about the variables in the data file. The Connected Data Survey instrument is included as an appendix to [State Child Welfare Data Linkages Descriptive Study Technical Report: Study Background and Design](#).

Preparing the data

States received surveys which could have included sections of questions about integrated data and linked data. The survey items in both sections included multiple-choice questions. During and after data collection, the team examined frequencies for all variables to confirm that survey skip patterns were accurate, the number of responses was correct, and open-text responses were plausible. The team contacted the primary respondent for clarification if needed.

Some multiple-choice questions included an "Other" response option which, if selected, prompted states to provide additional information in open text. The team reviewed responses to open-ended "other" options to determine whether any response could be categorized under an existing response option or whether a new response option was warranted. Our preference was to categorize open-text responses under existing response options. For open-text responses that could not be categorized under an existing response option, we determined whether a revised category could cover open-text responses from at least two states. We required the revised category to include at least two states to limit information that could potentially identify the individual respondents. If there was only one open-text response, then we grouped it with the most similar existing response option and renamed this response option to cover the open-text response too. The team replaced the original open-text responses with the revised categories, so no variables in the final files include any open-text responses.

The full survey also included contact information and names of data-sharing agreements and partner agencies. These items are not included in the distributed files.

Data structure

There are 57 records in the data file. Each record is a completed survey from a state, which could include information on integrated data and/or a particular linked data set. Information on states' responses to the Initial Survey was used to determine which section(s) each respondent would receive. Individuals

identified as knowledgeable about integrated data and linked data under a data sharing agreement received the full instrument. Nineteen records include responses to questions about both integrated data and linked data. Individuals who were identified in the Initial Survey as knowledgeable about integrated data were asked for their contact information and shown the section of questions about integrated data through their survey web link. Twenty-seven records are responses from individuals who only received questions about integrated data. Individuals who were identified in the Initial Survey as knowledgeable about linked data under a data sharing agreement were asked for their contact information and shown the section of questions about linked data through their survey web link. Eighteen records are responses from individuals who only received questions about linked data.

Variables

The data file includes variables based on states' responses to multiple-choice questions in the survey instrument and key state identifiers. The only variables excluded from the data file that were present in the full survey are those including PII or related to survey administration; names of individuals, organizations, and data sharing agreements; and other contact information that was gathered to field the surveys.

For multiple choice questions, the team created categorical variables where the categories align to the choices in the survey item. As described above, the team combined responses for a few variables where open-text information from states was back-coded. In the data sets provided in the file formats native to Stata (.dta) and SAS, and in the codebook, the value ".s" denotes logical skips in the survey and the value ".m" identifies item non-response. In the files provided in SPSS and comma-delimited format, we use a numeric placeholder of -98 for logical skips and -99 for item non-response. In the Connected Data Survey codebook, the variable description (contained in the variable label) includes the relevant question number in the Connected Data Survey instrument. No other masking or data manipulation has been done to the data.

Other variables for research use include binary indicators for whether the respondent received questions about integrated data (variable **svy_intdataquestions**) and linked data (variable **svy_linkeddataquestions**). These variables are based on responses from the state to the Initial Survey. In addition, the team has included derived variables and state identifiers. The team used derived variables to report publicly available findings in an infographic, briefs, and technical reports, which are available at <https://www.acf.hhs.gov/opre/project/state-child-welfare-data-linkages-descriptive-study>. The definition of the derived variable is included in its variable label and in the codebook. The team added two string variables that can be used to identify states and merge this data file with other data sets.⁵ The variable **state** contains the state's two-letter postal abbreviation, and **fips** contains the state's numeric FIPS code. The team also added information in the numeric variable **admin** on whether the child welfare system is a

⁵ The [NDACAN User Support page](#) includes examples of using different statistical packages to merge data.

state- or county-administered system, using information from the Child Welfare Information Gateway (see [this 2018 fact sheet](#))⁶.

Appendix B shows how variables are ordered in the final data files and provides descriptions, related survey item numbers, and information on variable naming conventions. For all variables in the file, we confirmed accuracy of value labels for variables as shown in the codebook, that variable labels were accurate, that the only string variables included in the file were state identifiers, and that missing value patterns and counts were plausible.

Variable characteristics

The codebook shows the following information about each variable in the Connected Data Survey data file:

- **Variable name:** Used to identify the data element in the data set that the variable corresponds to. Variables were named descriptively rather than in reference to a specific survey item.
- **Variable label:** A brief description of the variable. For variables that are not added for research use, the label also includes the specific relevant survey item(s).
- **Variable type:** Either numeric or string.
- **Universe:** The total number of cases possible for each variable in the data set. There are 57 records in the file.
- **N:** The total number of cases with valid values for the variable.
- **Frequencies:** The frequencies for each value are presented for all numeric variables.

Using the data

File formats

Each survey data set is provided in file formats native to SPSS (.sav), Stata (.dta), and SAS (.sas7bdat). There are also import program files for SAS (.sas), SPSS (.sps), and Stata (.do) to read in the text (.dat) data file, and comma-delimited (.csv) data file for use with spreadsheet programs. File formats native to Stata, SPSS, and SAS include variable labels, value labels, and formats.

The CSV or text data file may be imported into the user's preferred statistical package by using the import function for that package. Because labels are not included in CSV or text files, users should refer to the codebook. They can use the codebook to obtain the definition of each variable and the meaning of the numeric codes.

Guidance for using the import programs and instructions for importing the data into R can be found on the [NDACAN User Support](#) web page.

⁶ Children's Bureau. "State vs. County Administration of Child Welfare Services." Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services, 2018. <https://cwig-prod-prod-drupal-s3fs-us-east-1.s3.amazonaws.com/public/documents/services.pdf>.

Analytic considerations

The Initial Survey was sent to all states, and the Connected Data Survey was sent to individuals identified in the Initial Survey. However, not all states responded to the Initial Survey, and not all individuals who received the Connected Data Survey responded to it. Data gathered from the surveys are not nationally representative. The survey responses reflect information from participating states when the surveys were conducted (July 2022 through January 2023). Therefore, the data may not reflect the current status of states' connected data.

In addition, the Connected Data Survey focused on specific linked data sets from a subset of data sharing agreements, which may not represent all linked data sets or data sharing agreements of participating states. Respondents were not required to complete all items in the survey, and some items have more missing data than others. Users of the data files should tabulate variables to check sample sizes in their analyses and distinguish between data that are missing due to logical skips and data that are missing due to item non-response.

Lastly, the team did not ask states to confirm the accuracy of responses for either survey. Individual respondents could have answered survey items without confirming their accuracy or forwarding the survey to appropriate designee(s). It is possible that individuals within an agency could have provided different responses based on their own knowledge, given the wide range of questions.

References

Child Welfare Information Gateway. "State vs. County Administration of Child Welfare Services." U.S. Department of Health and Human Services, Administration for Children and Families, Children's Bureau, 2018.

Glossary

- ACF: Administration for Children and Families
- Connected data: Linked or integrated data
- Linked data: A set of records that includes data from the state system of record for child maltreatment reporting to the National Child Abuse and Neglect Data System and data from other sources, based on a common identifier or other matching criteria
- Integrated data: Data that have been systematically incorporated (through direct entry by staff or a data exchange) into the state system of record for child maltreatment reporting to the National Child Abuse and Neglect Data System
- NDACAN: National Data Archive on Child Abuse and Neglect
- SCW Descriptive Study: State Child Welfare Data Linkages Descriptive Study
- System of record: The state system of record for child maltreatment reporting to the National Child Abuse and Neglect Data System

Appendix A: Variables in the Initial Survey Data File

Variables, grouped by topic and ordered as in the data file	Description, survey item, and variable naming convention
state, fips, admin	State identifiers including state abbreviation, FIPS code, and child welfare program administration type
cw_standalone	How the child welfare agency is situated in the state (survey item 2)
standalonecwag_svc_data_invest – contract_svc_data_other	<p>Responses to survey grid about entities that provide services and collect related data (Survey items 4a_1-4s_5)</p> <p>Variable names are structured with a prefix, “_svc_data_”, and a suffix. The prefix identifies the type of agency that provides the service:</p> <ul style="list-style-type: none"> • standalonecwag = child welfare standalone agency • largerag = larger agency • other = other agency • localcwag = local or county child welfare agency • contract = contracted service provider <p>The suffix identifies the type of service:</p> <ul style="list-style-type: none"> • invest = Child abuse and neglect investigations • place = Foster care placements • casem = Child welfare case management after investigations • socsvc = Social Service assistance • ss = Social Security benefits • vital = Issuance of vital records (birth and death) • cs = Child support services • dd = Intellectual or developmental disability services • prev = Prevention services • mhlth = Mental health services • sub = Substance misuse or substance use services • hous = Housing assistance • jj = Juvenile justice services • cc = Child care services • dv = Domestic violence/intimate partner violence services • ref = Refugee assistance services • emp = Employment/labor services • ffpas = Services defined in the state’s Prevention Plan for the Family First Prevention Services Act • other = Other services
cwsystem_legacy – cwsystem_other	<p>System of record (survey item 6)</p> <p>Variable names are structured with the prefix “cw_system” and a suffix to identify the type of child welfare system of record:</p> <ul style="list-style-type: none"> • legacy = suffix for legacy systems of record • ccwis = suffix for Comprehensive Child Welfare Information System (CCWIS) • legacytoccwis = suffix for migrating the legacy system to CCWIS • othertonccwis = suffix for other non-CCWIS statewide system • county = suffix for only county-based systems

Variables, grouped by topic and ordered as in the data file	Description, survey item, and variable naming convention
devstatus_cwsystem	<ul style="list-style-type: none"> • other = other system of record <p>Status of work developing or replacing the child welfare system of record (survey item 7)</p> <p>Variable name is structured with the prefix "devstatus" and suffix "cwsystem"</p>
directentristaff_cwfrontline – directentristaff_other	<p>Staff who enter data into the child welfare system of record (survey item 8)</p> <p>Variable names are structured with the prefix "directentristaff" and a suffix to identify the type of staff who enter data:</p> <ul style="list-style-type: none"> • cwfrontline = frontline staff • cwnotfront = state child welfare staff who are not frontline • cwca = staff of child welfare contributing agencies (CWCA) • otherstaff = staff of other state or county agencies • noncwcapprov = staff of non-CWCA service providers • other = other staff
analyticdata_ncandsfile – analyticdata_nofile	<p>Use of NCANDS or other analytic data sets (survey item 9)</p> <p>Variable names are structured with the prefix "analyticdata" and a suffix. The suffix identifies the data sets used for analytic purposes:</p> <ul style="list-style-type: none"> • ncdandsfile = the state uses the NCANDS file for analytic purposes • createsfile = the state creates another analytic data set for analytic purposes • nofile = the state does not use NCANDS or another file for analytic purposes
int_samsysrec_invest – nint_other	<p>Responses to survey grid about administrative data connected with the system of record (survey items 10a_1-10z_3)</p> <p>Variable names are structured according to the type of integration and the type of service in question.</p> <ul style="list-style-type: none"> • int = integrated into system of record • nint = not integrated into system of record • samsysrec = same system of record • exch = integrated via a data exchange • wh = integrated via data warehouse <p>The suffix identifies the type of service:</p> <ul style="list-style-type: none"> • invest = Child abuse and neglect investigations • place = Foster care placements • casem = Child welfare case management after investigations • socsvc = Social Service assistance • ss = Social Security benefits • vital = Issuance of vital records (birth and death) • cs = Child support services • dd = Intellectual or developmental disability services • prev = Prevention services • mhlth = Mental health services • sub = Substance misuse or substance use services • hous = Housing assistance • jj = Juvenile justice services • cc = Child care services

Variables, grouped by topic and ordered as in the data file	Description, survey item, and variable naming convention
	<ul style="list-style-type: none"> • dv = Domestic violence/intimate partner violence services • ref = Refugee assistance services • emp = Employment/labor services • ffpsa = Services defined in the state's Prevention Plan for the Family First Prevention Services Act • other = Other services
num_dsa_practice – num_dsa_linkeddata	<p>How many data sharing agreements the state has for a particular purpose (survey items 11a-11c)</p> <p>Variable names are structured with a prefix "num_dsa" and a suffix. The suffix identifies the use of data sets:</p> <ul style="list-style-type: none"> • practice = data sharing agreements to support caseworker practice • extres = data sharing agreements to support external research studies • linkeddata = data sharing agreements to support a linked analytic data set
dsa_1_statewide – dsa_10_endyear	<p>Responses to survey grid about data sharing agreements that support linked data sets (survey items 11c.d-11c.e.)</p> <p>Variable names are structured with a prefix "dsa" and the base 1-10 to identify the data sharing agreement, and a suffix to identify characteristics of the data sharing agreement:</p> <ul style="list-style-type: none"> • statewide = data sharing agreement include children or families statewide • end = data sharing agreement has an end year, is ongoing, or status is unknown • endyear = end year of the data sharing agreement
Plan_newdsa	<p>Plans for new data sharing agreements (survey item 12)</p> <p>Variable name is structured with a prefix "plan" and a suffix "newdsa"</p>
Statereq_datafromotheragency – statereq_expungemaldata	<p>State requirements for linking and expunging child welfare data (survey items 13-14)</p> <p>Variable names are structured with a prefix "statereq" and suffix. The suffix identifies the statutory requirements for data linkages involving child welfare records:</p> <ul style="list-style-type: none"> • datafromotheragency = state requires another agency to submit their data to the child welfare agency • cwdatatoootheragency = state requires child welfare agency to submit their data to another agency • crossagencydatastd = state requires data linkages involving child welfare to meet privacy standards for multiple programs • cwspecificdatastd = state requires data linkages involving child welfare to meet privacy standards for child welfare • none = state has no requirements for data linkages involving child welfare records • unknown = respondent does not know of statutory requirements for data linkages involving child welfare records • expungemaldata = state requires child welfare agency to expunge historical child maltreatment data (survey item 14)
expunge_scrnout_noinv – expunge_unsub_groups	<p>State timeline to expunge child welfare data (survey items 15-16)</p>

Variables, grouped by topic and ordered as in the data file	Description, survey item, and variable naming convention
	<p>Variable names are structured with a prefix "expunge" and suffix. The suffix identifies the requirements for expunging data involving child welfare records:</p> <ul style="list-style-type: none"> • scrnout_noinv = when the child welfare agency expunges data on child maltreatment reports that are screened out or not investigated (survey item 15) <ul style="list-style-type: none"> – groups = categorized time periods when the state begins expunging data • unsub = when the child welfare agency expunges data on child maltreatment reports that are unsubstantiated (survey item 16) <ul style="list-style-type: none"> – groups = categorized time periods when the state begins expunging data
plan_changemaldef	<p>State plans to change the child maltreatment definition (survey item 17) Variable name is structured with a prefix "plan" and suffix "changemaldef"</p>
datainf_funding_suff	<p>State use of state and federal funding to support data infrastructure and funding sufficiency (survey item 18) Variable name is structured with a prefix "datainf", the base "funding", and suffix "suff"</p>

Appendix B: Variables in the Connected Data Survey Data File

Variables, grouped by topic and ordered as in the data file	Description, survey item, and variable naming convention
state, fips, admin	State identifiers including state abbreviation, FIPS code, and child welfare program administration type
svy_intdataquestions, svy_linkeddataquestions	<p>These variables identify whether the survey respondent received questions on integrated data or linked data, based on the state's response to questions 19a and 19b in the Initial Survey. Each row represents one survey.</p> <p>Variable names are structured according to the types of questions that the respondent received in the survey:</p> <ul style="list-style-type: none"> • intdataquestions = the respondent's survey included questions about integrated data. • linkeddataquestions = the respondent's survey included questions about linked data.
linked_managingorg	<p>The type of organization that manages the linked data set (survey item 3b)</p> <p>Variable name is structured with the prefix "linked" to show that these answers pertain to linked data and the suffix "managingorg".</p>
linked_agreement_dsa – linked_agreement_other	<p>The type of agreement that governs the sharing and use of data (survey item 4)</p> <p>Variable names contain the prefix "linked" to show that these answers pertain to linked data, the base "agreement, and a suffix to define the type of agreement:</p> <ul style="list-style-type: none"> • dsa = data sharing agreement • baa = business associates agreement • other = other agreement
linked_sametimeperiod – linked_sameupdatesched	<p>The timing of updating data sources in the linked data set (survey items 5a and 5b)</p> <p>Variable names contain the prefix "linked" to show that these answers pertain to linked data and a suffix to define the timing:</p> <ul style="list-style-type: none"> • sametimeperiod = whether all data sources in the linked data set cover the same time period • sameupdatesched = whether all data sources in the linked data set are updated on the same schedule
linked_invest – linked_other	<p>Responses to survey grid about sources of data included in linked data sets (survey items 6a1-6a25)</p> <p>Variable names contain the prefix "linked" to show that these answers pertain to linked data and a suffix to identify the type of service:</p> <ul style="list-style-type: none"> • invest = Child abuse and neglect investigations • place = Foster care placements • casem = Child welfare case management after investigations • socsvc = Social Service assistance • ss = Social Security benefits • vital = Issuance of vital records (birth and death) • cs = Child support services • dd = Intellectual or developmental disability services • prev = Prevention services • mhlth = Mental health services

Variables, grouped by topic and ordered as in the data file	Description, survey item, and variable naming convention
	<ul style="list-style-type: none"> • sub = Substance misuse or substance use services • hous = Housing assistance • jj = Juvenile justice services • cc = Child care services • dv = Domestic violence/intimate partner violence services • ref = Refugee assistance services • emp = Employment/labor services • ffpsa = Services defined in the state's Prevention Plan for the Family First Prevention Services Act • other = Other services
<p>linked_invest_startyr – linked_other_incldem</p>	<p>Responses to survey grid about sources and characteristics of data included in linked data sets (survey items 6b_1-6f_25)</p> <p>Variable names are structured with the prefix "linked" to show that these answers pertain to linked data, the base to identify the type of service, and the suffix to define the grid questions.</p> <p>The base identifies the type of service:</p> <ul style="list-style-type: none"> • invest = Child abuse and neglect investigations • place = Foster care placements • casem = Child welfare case management after investigations • socsvc = Social Service assistance • ss = Social Security benefits • vital = Issuance of vital records (birth and death) • cs = Child support services • dd = Intellectual or developmental disability services • prev = Prevention services • mhlth = Mental health services • sub = Substance misuse or substance use services • hous = Housing assistance • jj = Juvenile justice services • cc = Child care services • dv = Domestic violence/intimate partner violence services • ref = Refugee assistance services • emp = Employment/labor services • ffpsa = Services defined in the state's Prevention Plan for the Family First Prevention Services Act • other = Other services <p>The suffix identifies the characteristics of data sources:</p> <ul style="list-style-type: none"> • startyr = First year this data source is represented in the linked data • lastyr = Whether the data source is recurring or has an end year in the linked data • yr = Most recent year this data source is represented in the linked data • consec = Whether the data source covers all consecutive years • incldem = Whether the data source contributes demographic information to the linked data

Variables, grouped by topic and ordered as in the data file	Description, survey item, and variable naming convention
linked_levelind – linked_leveloth	<p>Linkage level used to link data sources (survey item 7)</p> <p>Variable names contain the prefix “linked” to show that these answers pertain to linked data and a suffix to indicate the linkage level:</p> <ul style="list-style-type: none"> • levelind = data linked at individual level • levelfam = data linked at family level • levelhh = data linked at household level • levelgeo = data linked at geographic location • levelorg = data linked at organization level • leveloth = data linked at other level
linked_mthd_det – linked_mthd_other	<p>Linkage method used to link data sources (survey item 8)</p> <p>Variable names contain the prefix “linked” to show that these answers pertain to linked data, the base “mthd” and a suffix to specify the data linkage method:</p> <ul style="list-style-type: none"> • det = deterministic or exact linking method • open = open source probabilistic linking software • prop = proprietary probabilistic linking software • proj = project specific method • unknown = unknown method • other = other linking method
linked_mthd_det_ssid – linked_mthd_det_otherid	<p>Common identifier used to link data sources with a deterministic linking method (survey item 9a)</p> <p>Variable names contain the prefix “linked” to show that these answers pertain to linked data, the base “mthd_det” and a suffix to identify the common identifier:</p> <ul style="list-style-type: none"> • ssid = Social Security number • sid = state ID • medid = Medicaid ID • sysid = System ID • dem = demographic data • loc = location data • otherid = other identifier
linked_mthd_ndet_childname – linked_mthd_ndet_raceeth	<p>Data element used to link data sources with a non-deterministic linking method (survey item 9b)</p> <p>Variable names contain the prefix “linked” to show that these answers pertain to linked data, the base “mthd_ndet” and a suffix to identify the data element:</p> <ul style="list-style-type: none"> • childname = child/individual name • mother = mother’s name • father = father’s name • dob = date of birth • addr = address • sex = sex • raceeth = race and/or ethnicity

Variables, grouped by topic and ordered as in the data file	Description, survey item, and variable naming convention
linked_pop_level	<p>Population level contained in the linked analytic data set (survey item 10)</p> <p>Variable name contains the prefix "linked" to show that these answers pertain to linked data, the base "pop" and the suffix "level" to indicate the population level covered by the linked data set</p>
linked_data_pop_childinrep – linked_data_pop_othercw	<p>Population characteristics pertaining to child welfare involvement contained in the linked analytic data set (survey item 11)</p> <p>Variable name contains the prefix "linked" to show that these answers pertain to linked data, the base "data_pop" and a suffix to identify each population characteristic:</p> <ul style="list-style-type: none"> • childinrep = children reported for maltreatment • childininv = children investigated for maltreatment • parinrep = parent(s) named in child maltreatment report • sibirrep = siblings of children reported for maltreatment • childinfc = children in foster care • famincasemgt = in-home child protection case management services • faminprev = child maltreatment prevention programs • othercw = other populations with child welfare involvement
linked_managingorg_startyr	<p>Year that the managing organization began using the linked data set (survey item 12)</p> <p>Variable name contains the prefix "linked" to show that these answers pertain to linked data, the base "managingorg" and the suffix "startyr" to refer to the year that the managing organization began using the linked data set</p>
linked_updatefreq	<p>Frequency that the data in the linked data set is updated (survey item 13)</p> <p>Variable name contains the prefix "linked" to show that these answers pertain to linked data and the suffix "updatefreq" to indicate the frequency that data in the linked data set is updated</p>
linked_governance_docs – linked_governance_str	<p>Data governance documentation structure (survey items 14-15)</p> <p>Variable name contains the prefix "linked" to show that these answers pertain to linked data, the base "governance" and a suffix:</p> <ul style="list-style-type: none"> • "docs" = whether there are documented data governance procedures (survey item 14) • "str" = data governance structure (survey item 15)
linked_methods_docs	<p>Whether data linkage methods and procedures are documented (survey item 16)</p> <p>Variable name contains the prefix "linked" to show that these answers pertain to linked data, the base "methods" and the suffix "docs"</p>
linked_protected_unauth	<p>Whether data are protected from unauthorized access (survey item 17)</p> <p>Variable name contains the prefix "linked" to show that these answers pertain to linked data, the base "protected" and the suffix "unauth"</p>
linked_user_access	<p>Whether there are processes for individuals to gain access to use the data (survey item 18)</p> <p>Variable name contains the prefix "linked" to show that these answers pertain to linked data, the base "user" and the suffix "access"</p>

Variables, grouped by topic and ordered as in the data file	Description, survey item, and variable naming convention
linked_use_practice – linked_use_othexternal	<p>Uses of linked data (survey item 19)</p> <p>Variable name contains the prefix “linked” to show that these answers pertain to linked data, the base “use”, and a suffix to identify uses:</p> <ul style="list-style-type: none"> • practice = linked data used for practice • cqi = linked data used for continuous quality improvement • progeval = linked data used for program evaluation research • statepolres = linked data used for state policy research • legmandate = linked data used for legislative report or mandate • othres = linked data used for other research • perfmonitor = linked data used for performance monitoring • predmodel = linked data used for predictive modeling • othinternal = linked data used for other internal purpose • othexternal = linked data used for other external purpose
linked_usedby_mngorg_othstate – linked_usedby_mngorgonly	<p>Active users of the linked data set outside of the managing organization (survey item 20)</p> <p>Variable name contains the prefix “linked” to show that these answers pertain to linked data, the base “usedby_mngorg”, and a suffix to identify active users:</p> <ul style="list-style-type: none"> • othstate = state government staff actively use the linked data set • othres = researchers actively use the linked data set • other = others actively use the linked data set <p>The variable name contains the prefix “linked” to show that these answers pertain to linked data and the base “usedby_mngorgonly” when the linked data set is only actively used by the managing organization</p>
linked_canuse_mngorg_none – linked_canuse_mngorg_other	<p>Potential users outside of the managing organizations that can request access to the linked data set (survey item 21)</p> <p>Variable name contains the prefix “linked” to show that these answers pertain to linked data, the base “canuse_mngorg”, and a suffix to identify potential users:</p> <ul style="list-style-type: none"> • none = not permitted for any staff outside of the managing organization • dsaorgs = staff from an organization that is part of the data sharing agreement • othstate = staff of other state entities that are not part of the data sharing agreement • res = researchers who are not part of the data sharing agreement • other = others
linked_access_paperapp – linked_access_other	<p>Methods to access the linked data set (survey item 22)</p> <p>Variable name contains the prefix “linked” to show that these answers pertain to linked data, the base “access”, and a suffix to identify methods:</p> <ul style="list-style-type: none"> • paperapp = paper application • online req = online request • direct req = direct request • dsa = establish new data sharing agreement • none = there are no methods to access the linked data set • other = other unlisted method

Variables, grouped by topic and ordered as in the data file	Description, survey item, and variable naming convention
int_agreement_dsa – int_agreement_other	<p>The type of agreement that governs the sharing and use of data (survey item 23)</p> <p>Variable names contain the prefix “int” to show that these answers pertain to integrated data, the base “agreement”, and a suffix to define the type of agreement:</p> <ul style="list-style-type: none"> • dsa = data sharing agreement • mou = memorandum of understanding • baa = business associates’ agreement • loa = letter of agreement • other = other agreement
int_sametimeperiod	<p>Whether all data sources integrated into the system of record cover the same time period (survey item 24)</p> <p>Variable names contain the prefix “int” to show that these answers pertain to integrated data and the suffix “sametimeperiod” to indicate whether data sources integrated into the system of record cover the same time period</p>
int_invest_sysrec – int_other_other	<p>Responses to survey grid about external sources of data that are systematically connected with the system of record (survey items 25a_1_a-25a_99_y)</p> <p>Variable names are structured with the prefix “int” to show that these answers pertain to integrated data, the base to identify the type of service, and the suffix to define if the data are connected to the system of record.</p> <p>The base identifies the type of service:</p> <ul style="list-style-type: none"> • invest = Child abuse and neglect investigations • place = Foster care placements • casem = Child welfare case management after investigations • socsvs = Social Service assistance • ss = Social Security benefits • vital = Issuance of vital records (birth and death) • cs = Child support services • dd = Intellectual or developmental disability services • prev = Prevention services • mhlth = Mental health services • sub = Substance misuse or substance use services • hous = Housing assistance • jj = Juvenile justice services • cc = Child care services • dv = Domestic violence/intimate partner violence services • ref = Refugee assistance services • emp = Employment/labor services • ffpsa = Services defined in the state’s Prevention Plan for the Family First Prevention Services Act • other = Other services <p>The suffix identifies whether the data source is connected with the system of record</p> <ul style="list-style-type: none"> • sysrec = this service data already exists in the system of record • directentry = this service data is entered directly into the system of record

Variables, grouped by topic and ordered as in the data file	Description, survey item, and variable naming convention
	<ul style="list-style-type: none"> • dataexch = this service data is connected with the system of record via a data exchange • datawh = this service data is connected with the system of record via a data warehouse • planning = this service data is not connected to the system of record, but there is a plan to connect it • excl = this service data is not connected with the system of record • other = this service data has some other status
<p>int_invest_mostrecentyr – int_other_incldem</p>	<p>Responses to survey grid about external sources of data that are systematically connected with the system of record (survey items 25b_1_a-25f_99_y)</p> <p>Variable names are structured with the prefix "int" to show that these answers pertain to integrated data, the base to identify the type of service, and the suffix to define characteristics of the data.</p> <p>The base identifies the type of service:</p> <ul style="list-style-type: none"> • invest = Child abuse and neglect investigations • place = Foster care placements • casem = Child welfare case management after investigations • socsvc = Social Service assistance • ss = Social Security benefits • vital = Issuance of vital records (birth and death) • cs = Child support services • dd = Intellectual or developmental disability services • prev = Prevention services • mhlth = Mental health services • sub = Substance misuse or substance use services • hous = Housing assistance • jj = Juvenile justice services • cc = Child care services • dv = Domestic violence/intimate partner violence services • ref = Refugee assistance services • emp = Employment/labor services • ffpsa = Services defined in the state's Prevention Plan for the Family First Prevention Services Act • other = Other services <p>The suffix identifies the characteristics of data sources:</p> <ul style="list-style-type: none"> • mostrecentyr = Whether the data source is recurring or has an end year in the integrated data • lastyr = Most recent year this data source is represented in the integrated data • consec = Whether the data source covers all consecutive years • incldem = Whether the data source contributes demographic information to the integrated data
<p>int_sysreconnect_person – int_sysreconnect_other</p>	<p>Linkage level to connect data integrated into the child welfare system of record (survey item 26)</p>

Variables, grouped by topic and ordered as in the data file	Description, survey item, and variable naming convention
	<p>Variable names contain the prefix "int" to show that these answers pertain to integrated data, the base "sysreconnect", and a suffix to indicate the level of the linkage:</p> <ul style="list-style-type: none"> • person = the integrated data is connected with the child welfare system at the individual level • family = the integrated data is connected with the child welfare system at the family level • hh = the integrated data is connected with the child welfare system at the household level • geo = the integrated data is connected with the child welfare system at the geographic location level • org = the integrated data is connected with the child welfare system at the organization level • other = the integrated data is connected with the child welfare system at another level
int_mthd_det – int_mthd_other	<p>Linkage method used to connect data integrated into the child welfare system of record (survey item 27)</p> <p>Variable names contain the prefix "int" to show that these answers pertain to integrated data, the base "mthd", and a suffix to indicate the method of the linkage:</p> <ul style="list-style-type: none"> • det = deterministic or exact linking method • osproblink = open source probabilistic linking software • propproblink = proprietary probabilistic linking software • other = other linking method
int_mthd_det_ssid – int_mthd_det_otherid	<p>Common identifier used to connect integrated data with a deterministic linking method (survey item 28a)</p> <p>Variable names contain the prefix "int" to show that these answers pertain to integrated data, the base "mthd_det", and a suffix to indicate the common identifier:</p> <ul style="list-style-type: none"> • ssid = Social Security number • sid = state ID • medid = Medicaid ID • sysid = System ID • dem = demographic data • otherid = other identifier
int_mthd_ndet_childname – int_mthd_ndet_raceeth	<p>Data element used to connect integrated data with a non-deterministic linking method (survey item 28b)</p> <p>Variable names contain the prefix "int" to show that these answers pertain to integrated data, the base "mthd_ndet", and a suffix to indicate the data element:</p> <ul style="list-style-type: none"> • childname = child/individual name • mother = mother's name • father = father's name • dob = date of birth • addr = address • sex = sex

Variables, grouped by topic and ordered as in the data file	Description, survey item, and variable naming convention
	<ul style="list-style-type: none"> • raceeth = race and/or ethnicity
int_data_population	<p>Population level covered by the integrated data (survey item 29)</p> <p>Variable name contains the prefix "int" to show that these answers pertain to integrated data, the base "data", and the suffix "population".</p>
int_updatefreq_annual – int_updatefreq_other	<p>Frequency that integrated data are updated (survey item 30)</p> <p>Variable names contain the prefix "int" to show that these answers pertain to integrated data, the base "updatefreq", and a suffix:</p> <ul style="list-style-type: none"> • annual = data is updated annually • semiannual = data is updated semiannually • quarterly = data is updated quarterly • monthly = data is updated monthly • daily = data is updated daily • realtime = data is updated in real time • varies = data is updated at varying times • other = data is updated at some other frequency
int_canuse_mngorg_dsaorgs – int_canuse_mngorg_other	<p>Users in addition to the managing organization who can receive permission to access the integrated data (survey item 31)</p> <p>Variable names contain the prefix "int" to show that these answers pertain to integrated data, the base "canuse_mngorg", and a suffix to identify potential users:</p> <ul style="list-style-type: none"> • dsaorgs = staff from an organization that is part of the data sharing agreement • othstate = staff of other state entities that are not part of the data sharing agreement • res = researchers who are not part of the data sharing agreement • other = others
int_use_practice – int_use_othexternal	<p>Uses of integrated data (survey item 32)</p> <p>Variable names contain the prefix "int" to show that these answers pertain to integrated data, the base "use", and a suffix to identify use:</p> <ul style="list-style-type: none"> • practice = integrated data used for practice • cqi = integrated data used for continuous quality improvement • progeval = integrated data used for program evaluation research • statepolres = integrated data used for state policy research • legmandate = integrated data used for legislative report or mandate • othres = integrated data used for other research • perfmonitor = integrated data used for performance monitoring • predmodel = integrated data used for predictive modeling • othinternal = integrated data used for other internal purpose
research_link – external_link	<p>Groups of uses for linked data sets (survey item 19)</p> <p>Variable names contain a prefix to identify groups of uses and a suffix "link" to show that these answers pertain to linked data sets:</p> <ul style="list-style-type: none"> • research = linked data set used for research, including program evaluation research, predictive modeling, and other research. • internal = linked data set used for internal uses, including practice, continuous quality improvement, performance monitoring, and other internal uses.

Variables, grouped by topic and ordered as in the data file	Description, survey item, and variable naming convention
	<ul style="list-style-type: none"> • internal_s1 = linked data set used for improving service delivery, including continuous quality improvement and performance monitoring. • internal_s2 = linked data set used for other internal uses, including practice and other internal use. • external = linked data set used for external uses, including state policy research, legislative mandates, or other external uses.
linked_managingorg_r – linked_access_other_i2	Linked data set variables for each group of uses (survey items 3b, 4, 5a, 5b, 6a_1-6f_25, 7, 8, 9a, 9b, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22) Variable names contain the same structure as the original variables described above for linked_managingorg-linked_access_other, with a suffix to identify the group of uses: <ul style="list-style-type: none"> • r = linked data set used for research, including program evaluation research, predictive modeling, and other research. • i = linked data set used for internal uses, including practice, continuous quality improvement, performance monitoring, and other internal uses. • i1 = linked data set used for improving service delivery, including continuous quality improvement and performance monitoring. • i2 = linked data set used for other internal uses, including practice and other internal use. • e = linked data set used for external uses, including state policy research, legislative mandates, or other external uses.