



Linked and Integrated Data: Select Findings from the State Child Welfare Data Linkages Descriptive Study

Background

The [State Child Welfare Data Linkages \(SCW\) Descriptive Study](#) (1) examined the extent to which child welfare agencies in 50 states, Puerto Rico, and the District of Columbia (hereafter referred to as states) connected administrative data on child maltreatment to other data sources and (2) collected information about state practices related to sharing and connecting data. The SCW Descriptive Study aims to provide novel information regarding connected (linked or integrated) state data that may be leveraged to improve ongoing and accurate surveillance of child maltreatment incidence and related risk.

For more information about the SCW Descriptive Study and findings from other data collection activities, please visit the [project's web page](#). Data from both surveys are archived at the [National Data Archive on Child Abuse and Neglect](#).

This brief highlights information provided by states on their use of connected data, a term we use to refer to linked data and integrated data (Exhibit 1). This brief addresses the following questions:

1. How are linked and integrated data used?
2. Who can access and use linked and integrated data?
3. What data are linked and integrated?
4. How do states link and integrate data?

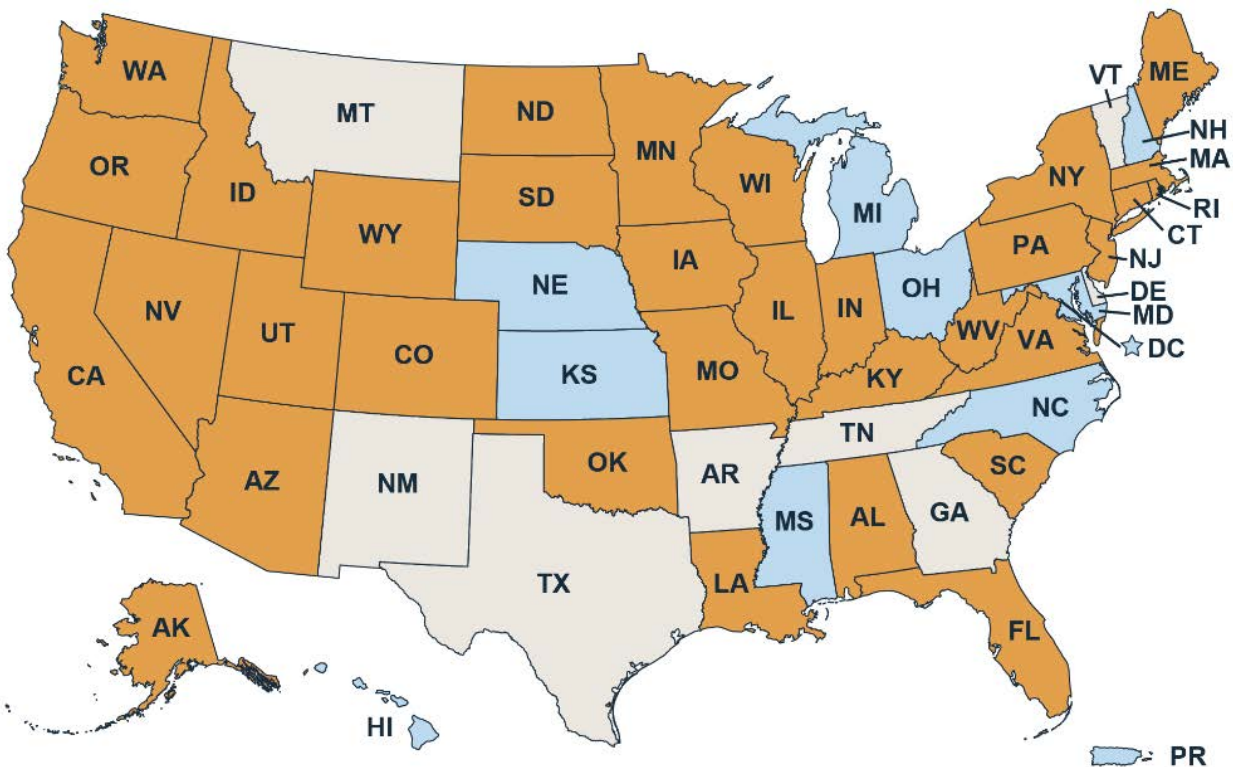
To answer these questions, we used data collected from two surveys that were conducted for the SCW Descriptive Study:

1. A high-level web survey of state child welfare directors or their designees (Initial Survey)
2. A more in-depth web survey of a state agency staff person knowledgeable about connected data (Connected Data Survey).

The Initial Survey gathered background information that fed into the Connected Data Survey, including the name of the state’s system of record for child maltreatment reporting to the National Child Abuse and Neglect Data System (“system of record”), whether the state had integrated data, and whether the state had data sharing agreements that supported linked data sets. This survey also collected contact information for staff who could share more about the linked and integrated data sets. The Connected Data Survey built on the Initial Survey by asking the identified contacts to provide more information about the state’s connected data. In this brief, we describe responses to items in the Connected Data Survey.

Exhibit 1 shows which states responded to both surveys, to the Initial Survey only, and to neither survey. This brief only includes information from the 33 states that responded to both surveys. These states are shaded in orange. [State Child Welfare Data Linkages Descriptive Study Technical Report: Study Background and Design](#) provides additional information on the survey samples and response rates.

Exhibit 1. States that participated in the Initial Survey and the Connected Data Survey



- States that responded to the Initial Survey and the Connected Data Survey
- States that responded to only the Initial Survey
- States that did not respond to either survey

This brief draws on findings that are presented in fuller detail in [State Child Welfare Data Linkages Descriptive Study Technical Report: Study Findings](#). This brief highlights findings related to integrated data first, then findings related to linked data. The survey items from the Connected Data Survey used in the analyses are listed at the end of the brief.

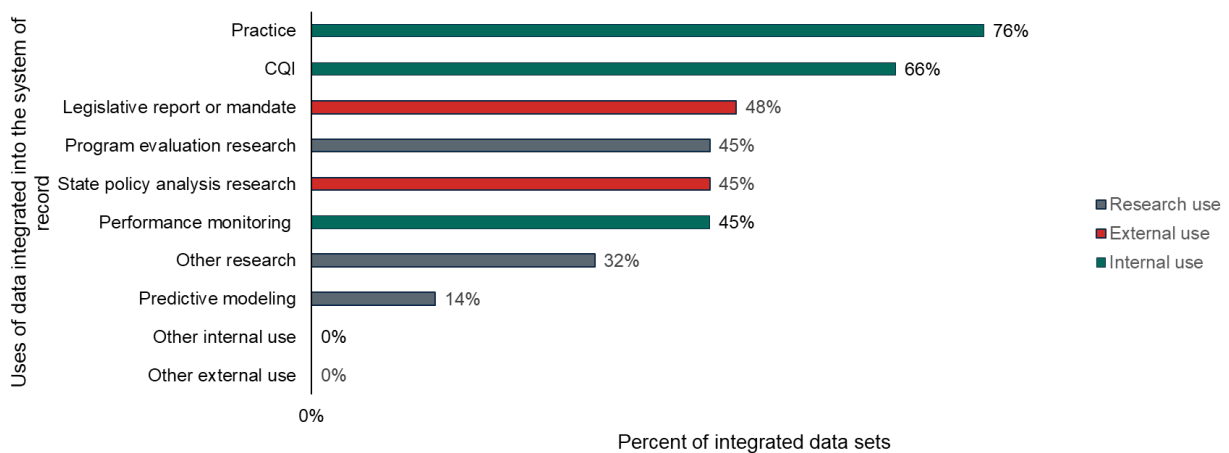
Findings related to integrated data

How are integrated data used?

Integrated data is defined by the project as data that have been systematically incorporated (through direct entry by staff or a data exchange) into the state system of record. Findings pertaining to integrated data are reported at the state level as counts or percentages.

The Connected Data Survey asked states about their uses of data integrated into their system of record. States could select more than one use (Exhibit 2). The most common use was for practice, such as casework, case management, or service delivery (76 percent of states), followed by use for continuous quality improvement (CQI; 66 percent of states).

Exhibit 2. Uses of integrated data



Source: SCW Descriptive Study Connected Data Survey, Question 32.

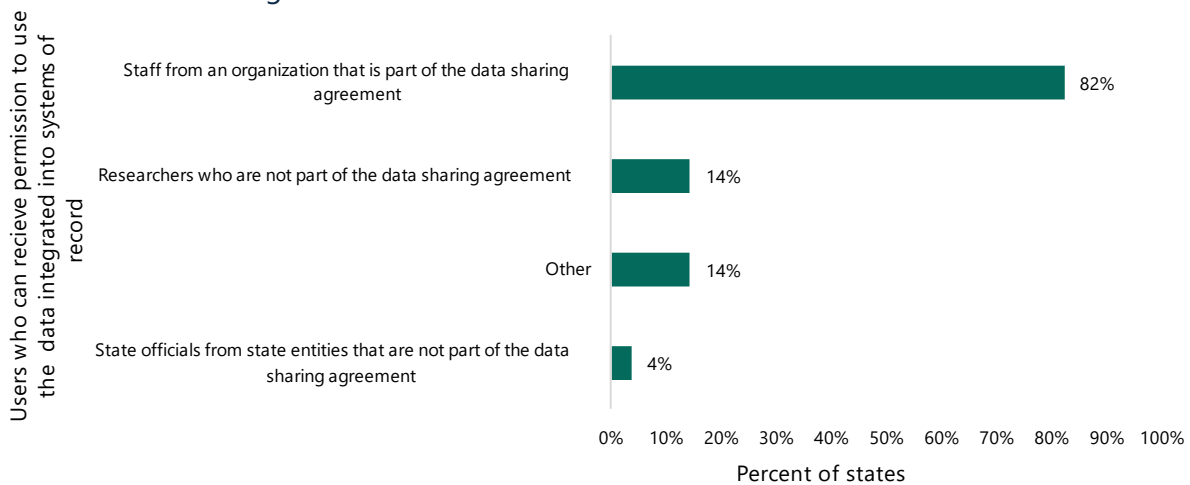
Note: This table reports percentages of 29 integrated data sets, as described by 29 participating states. Percentages sum to more than 100 percent because respondents could say they used data integrated into their system of record in more than one way.

CQI = Continuous Quality Improvement.

Who can access and use integrated data?

We asked whether anyone outside the agency had permission to access the data integrated into their system of record or extracts from that data, and if so, who (Exhibit 3). States could choose more than one type of user. Most states said staff from an organization that is part of the data sharing agreement could access the integrated data. Four or fewer states identified researchers or officials from state entities that were not part of the data sharing agreement as having permission to access integrated data or extracts from the data. The survey did not distinguish between researchers who were or were not affiliated with the state.

Exhibit 3. Users of integrated data



Source: SCW Descriptive Study Connected Data Survey, Question 31.

Note: This exhibit reports percentages of states responding to this question (n = 28). Percentages do not sum to 100 percent because respondents could choose multiple types of users who could receive permission to access or use the integrated data. Responses in the Other category included system users, contractors, none, and individuals with appropriate agreements.

What data are integrated?

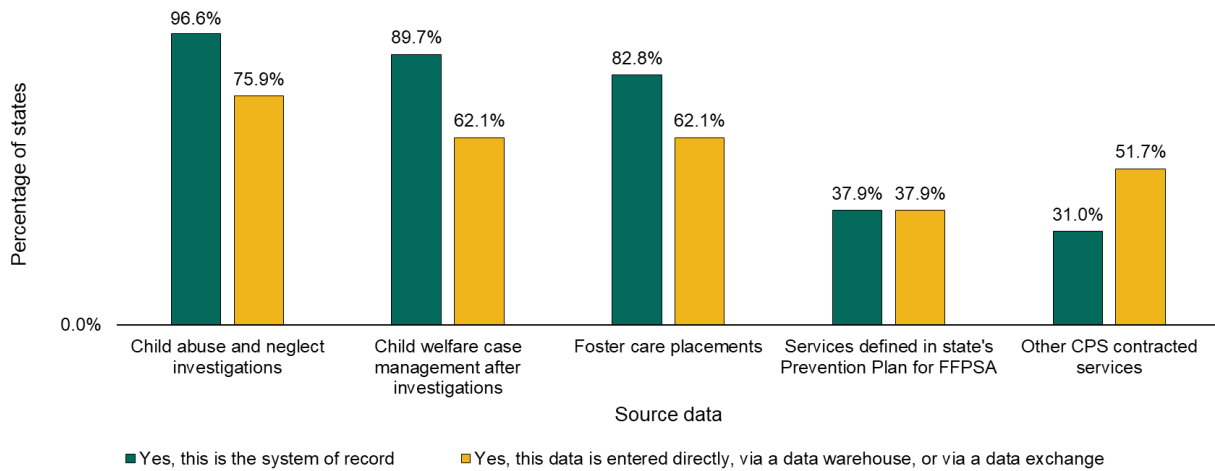
During the survey, states were given a list of services and asked whether and how data from each service were connected to the system of record. [A Snapshot of States' Child Welfare Data Systems of Record](#) includes the full list of services and information on how data may be connected to the system of record. In this brief, we summarize the most and least common data sources integrated into systems of record.

Overall, the services that were most likely to have data integrated into the system of record were those related directly to child maltreatment: child abuse and neglect investigations, child welfare case management after investigations, foster care placements, services defined in states' Prevention Plans for the Family First Prevention Services Act, and other services contracted by Child Protective Services (CPS) (Exhibit 4).

Gaps in integrated data

Data sources that were least likely to be integrated into the system of record (that is, five or fewer states included these data sources) were Supplemental Nutrition Assistance Program (SNAP) records, birth and death records, housing assistance, refugee assistance services, and labor and/or employment services.

Exhibit 4. Most common data sources integrated into the system of record



Source: SCW Descriptive Study Connected Data Survey, Questions 25.1.a–25.99.z.

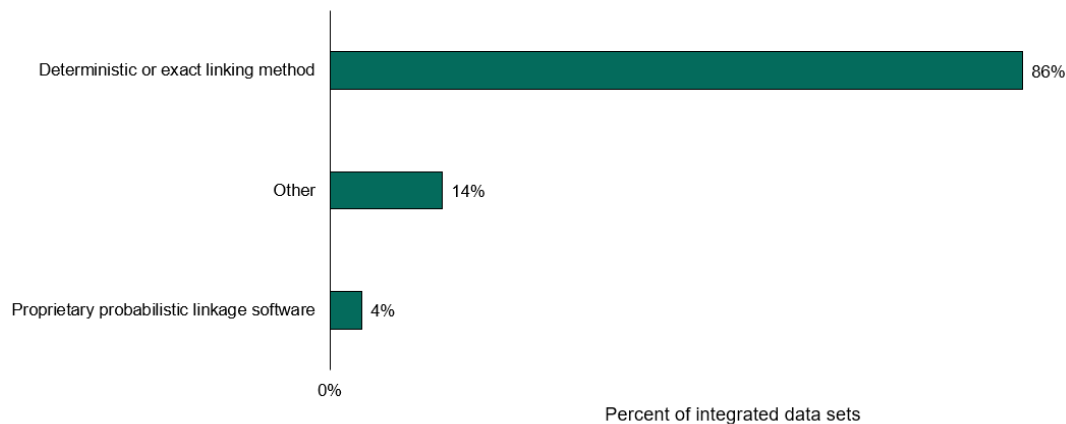
Note: This exhibit reports counts and percentages of participating states (N = 29). Total percentages do not sum to 100 percent because data from services can have multiple statuses and states can integrate multiple data sources into their systems.

How do states integrate data?

States described the methods they used to populate the system of record with integrated data. States could say they used any or all of the following methods: (1) deterministic or exact linking, (2) proprietary probabilistic linkage software, (3) open-source probabilistic linkage software, and (4) other.

Eighty-six percent of states used deterministic or exact linking method, meaning that records across data sources had the same values as identifiers. Only one state used more than one method. States that selected “other” referenced matching processes and code without referencing specific linkage software.

Exhibit 5. Data linkage methods for integrating data into the system of record



Source: State Child Welfare Descriptive Study Connected Data Survey, Question 27.

Note: This exhibit reports percentages of integrated data sets described by states. Percentages sum to more than 100 percent because respondents could select multiple data linking methods.

Findings related to linked data

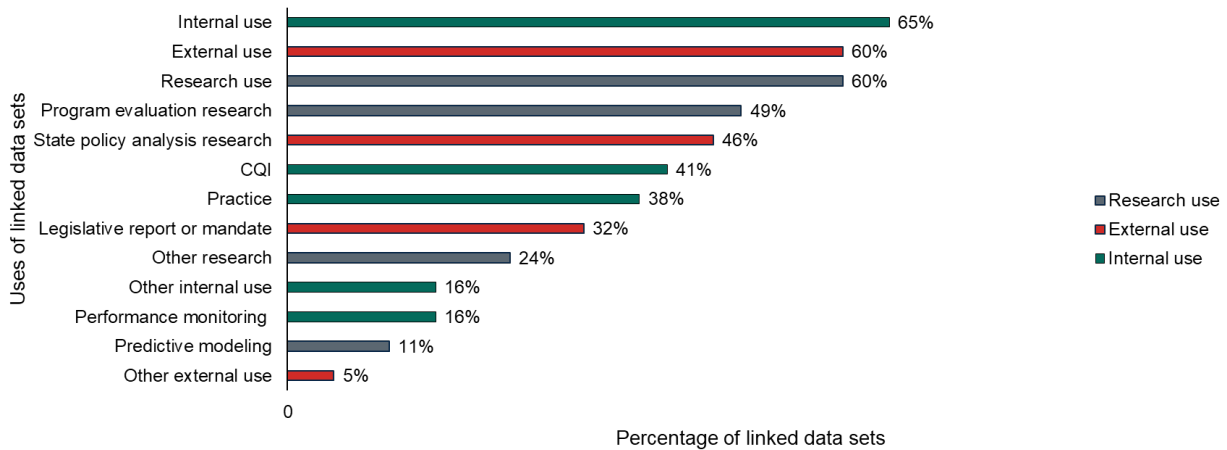
How are linked data used?

For the purposes of the study, we defined linked data as a set of records that includes data from the system of record and another source, matched based on a common identifier or other matching criteria. In the Connected Data Survey, states were asked to select the uses for their linked data sets—including, practice, continuous quality improvement (CQI), program evaluation research, state policy analysis research, legislative report or mandate, other research, performance monitoring, predictive modeling, other internal uses, and other external uses. Among linked data sets, the most common uses were program evaluation research, state policy analysis research, and CQI (Exhibit 6).

States described select linked data sets

The Connected Data Survey asked states about one linked data set per data sharing agreement, as identified in the Initial Survey, for up to four data sharing agreements. For each data sharing agreement, they were asked to focus on the most utilized linked data set that was relevant to child maltreatment incidence. If none of the linked data sets under the data sharing agreement were relevant to child maltreatment incidence, then states were asked to select the data set that was most utilized. Eighteen states reported on 37 linked data sets. In this section, we report counts or percentages of these 37 linked data sets.

Exhibit 6. Uses of linked data sets



Source: SCW Descriptive Study Connected Data Survey, Question 19.

Note: This exhibit reports percentages of uses for 37 linked data sets, as described by 18 participating states. Individual states may have reported on up to 4 linked data sets. Percentages sum to more than 100 percent because respondents could say they used each linked data set in more than one way. Other internal uses included sharing demographics between data systems and producing metrics for a public dashboard. Other external uses included use for a specific program or making foster parents aware of children’s health needs.

CQI = Continuous Quality Improvement.

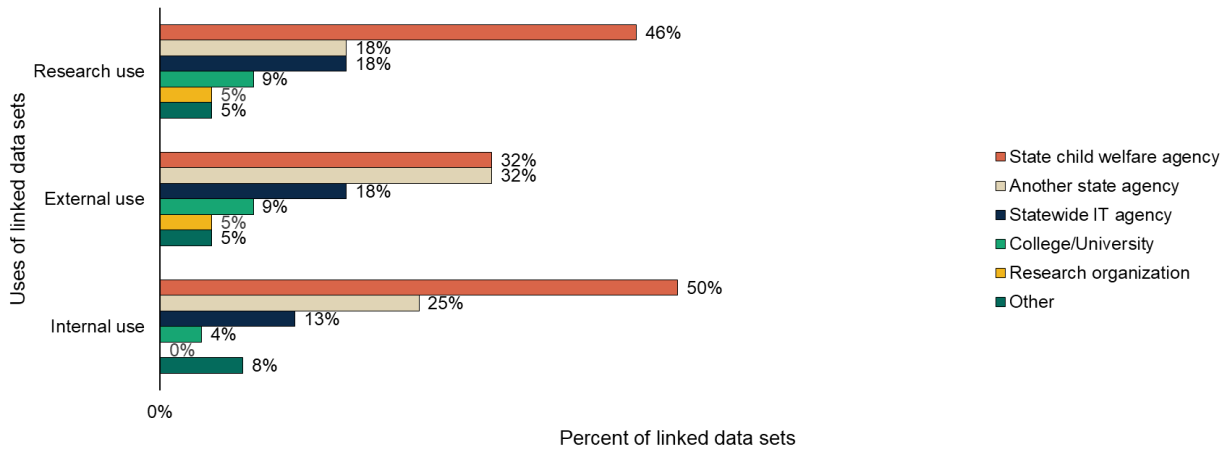
The study team combined the 10 uses in Exhibit 5 into the following three main groups for further analysis: (1) research, including program evaluation research, predictive modeling, and other research; (2) internal use, including CQI, practice, performance monitoring, and other internal use; and (3) external use, including state policy analysis research, legislative report or mandate, and another external use. We allowed linked data sets to appear in multiple groups but counted them only once within each group.¹

¹ Linked data sets often had multiple uses. For example, 11 linked data sets were reported as having research, internal, and external use. Five additional linked data sets were reported as having research and external use.

Who can access and use linked data?

We asked states about the organization that manages the data sharing agreement for the linked data set, active users other than staff at the managing organization, and people who could receive permission to use the linked data set or extracts from it other than staff at the managing organization. States indicated whether the organization that manages the data sharing agreement was the state child welfare agency, another state agency, a statewide IT agency, a college/University, a research organization, or other. Across all linked data sets, forty-six percent were managed by the state child welfare organization. Exhibit 7 presents managing organizations grouped into research, external, and internal uses.

Exhibit 7. Managing organizations of linked data sets, by use

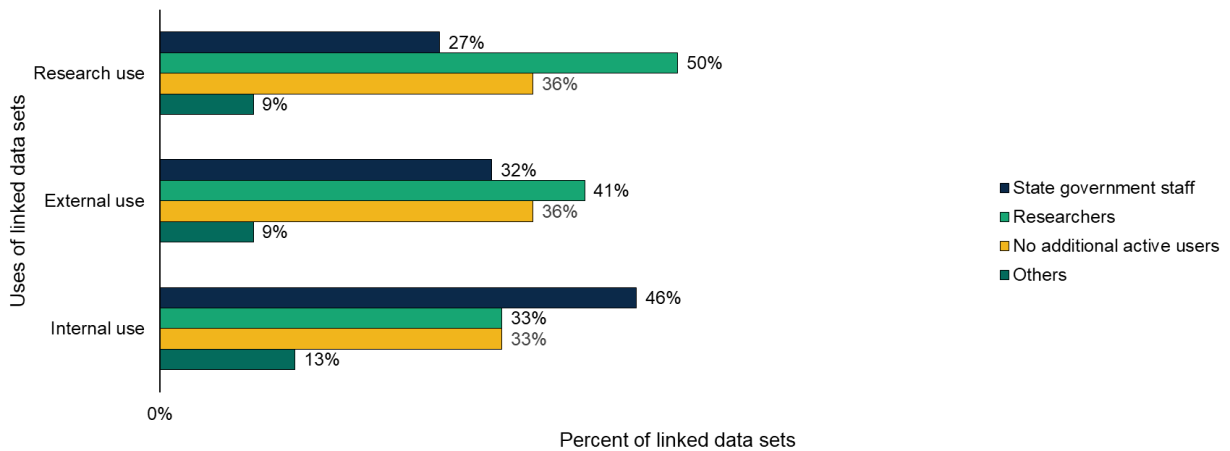


Source: SCW Descriptive Study Connected Data Survey, Question 19.

Note: This exhibit reports percentages of linked data sets, as described by participating states.

States identified active users and potential users outside the managing organization. We did not give states definitions or examples of active use, and the survey did not distinguish between researchers with and without state affiliation. Exhibits 8 and 9 present users by linked data set use.

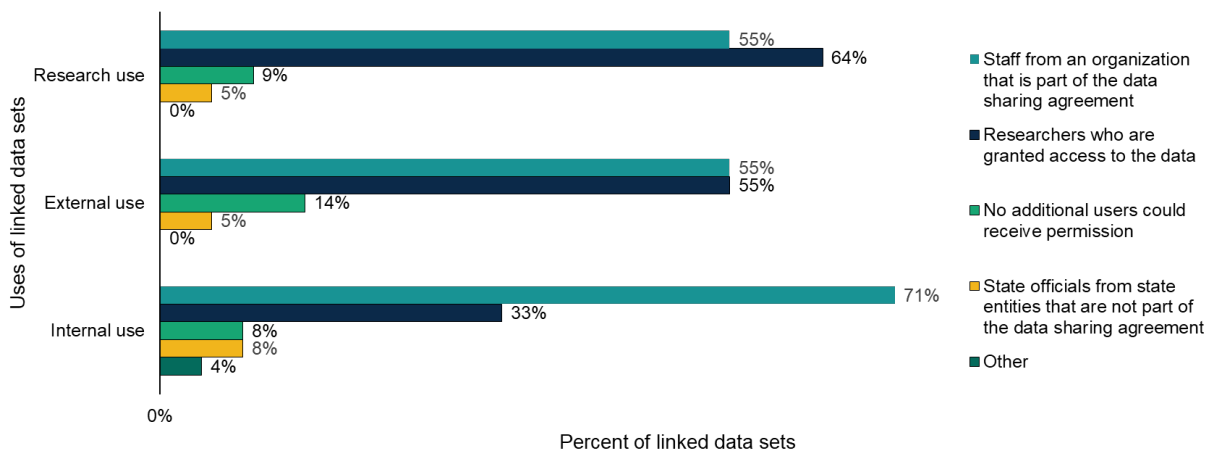
Exhibit 8. Active users of linked data sets, by use



Source: SCW Descriptive Study Connected Data Survey, Question 20.

Note: This exhibit reports percentages for linked data sets, as described by participating states. Percentages sum to more than 100 because respondents could answer that they had more than one type of active user. "Others" included foster parents and IT developers.

Exhibit 9. People who could receive permission to use the linked data set, by use



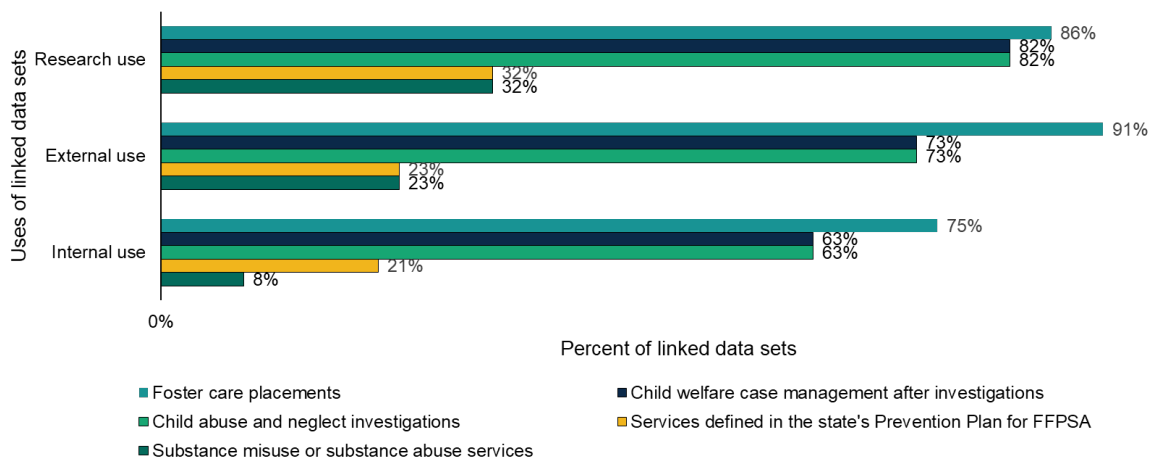
Source: SCW Descriptive Study Connected Data Survey, Question 21

Note: This exhibit reports percentages for linked data sets, as described by participating states. Percentages sum to more than 100 because respondents could answer that they had more than one type of potential user. "Others" included foster parents, researchers meeting certain criteria, and that no one can receive permission.

What data are linked?

States were asked which data sources were included in their linked data sets. Foster care placements, child welfare case management after investigations, and child abuse and neglect investigations were most likely to be included in linked data sets (Exhibit 10).

Exhibit 10. Most common data sources for linked data sets by use



Source: SCW Descriptive Study Connected Data Survey, Questions 6a1–6a25.

Note: This exhibit reports percentages for the number of linked data sets by use that includes each data source.

FFPSA = Family First Prevention Services Act.

Gaps in linked data

We examined which services were least likely to be present in the linked data sets that states reported on in the Connected Data Survey.

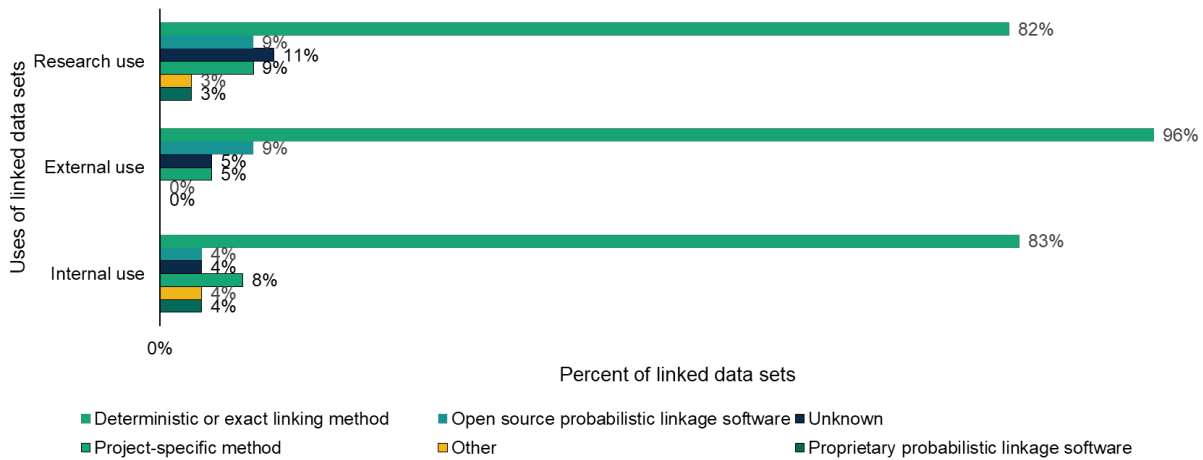
Only four linked data sets included information from intellectual or developmental disability services. Only two or three linked data sets included information from any of the following data sources: Social Security benefits, child support systems, employment and labor services, and domestic violence or intimate partner violence services. No linked data sets included information on refugee assistance services.

How do states link data?

Data linkage methods

States were asked how they linked the data sources for their linked data sets. More than three-quarters of linked data sets were constructed by using a deterministic or exact linking method, meaning that records across data sources had the same values as identifiers.

Exhibit 11. Data linkage methods for linked data sets, by use

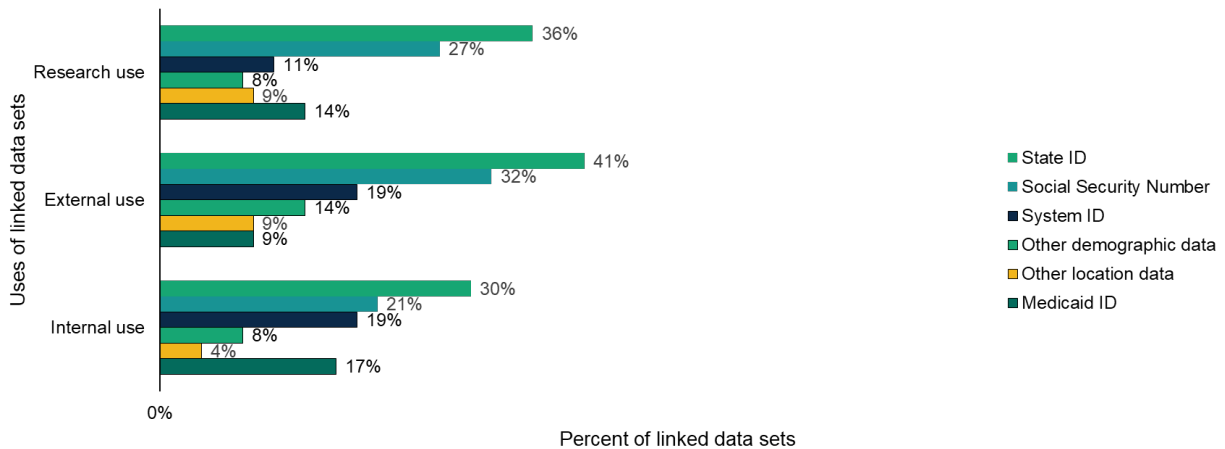


Source: State Child Welfare Descriptive Study Connected Data Survey, Question 8.

Note: This exhibit reports percentages of linked data sets described by states. Percentages sum to more than 100 percent because respondents could select multiple data linking methods.

Respondents then described what identifiers were used for deterministic linking. Thirty-five percent of linked data sets overall used a state ID, either alone or in combination with a social security number, Medicaid or system ID, or other data, including demographics and location data. Data sets used for internal or external uses were more likely to use a system ID, such as a case number to link data, than those used for research purposes. Data sets used for external uses were also more likely to use other demographic data, such as names, dates of birth, and other data, in combination with other IDs for the deterministic match.

Exhibit 12. Data linkage identifiers used for deterministic or exact linking by use



Source: State Child Welfare Descriptive Study Connected Data Survey, Question 9a.

Note: This exhibit reports percentages of linked data sets described by states. Percentages sum to more than 100 percent because respondents could select multiple identifiers used to link data. No combinations of identifiers used to link the data were selected by respondents for more than five linked data sets.

Conclusion

This brief provided a snapshot of state child welfare agencies' connected (integrated and linked) data. The study was based on data gathered in 2022–2023.

[State Child Welfare Data Linkages Descriptive Study Technical Report: Study Findings](#) provides additional detail, including context for these findings from interviews with 33 states. For researchers interested in further exploring the data highlighted here, survey data are archived with the National Data Archive for Child Abuse and Neglect.

References

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

Child Welfare Outcomes Data. "In Foster Care on the First Day of FY; Number of Children in Care & Median Length of Stay." Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Children's Bureau, n.d.

<https://cwoutcomes.acf.hhs.gov/cwodatasite/inCareOctoberOne/index>.

Acknowledgments

We thank states for participating in the study and sharing information with the SCW Descriptive Study team. We appreciate the feedback that Tammy Cordova, Dennis Culhane, Michelle Johnson-Motoyama, Cara Kelly, Zsuzsanna Kocsis, Colleen McGroarty Seifer, Fred North, Emily Putnam-Hornstein, Terry Shaw, Dori Sneddon, and Christopher Wildeman provided on topics of interest for this study. In addition, we acknowledge Yvonne Marki for developing the graphics shown here, Effie Metropoulos for editing, and Jill Miller for production support.

This brief was written by Tara Strelevitz, Eileen Gao, Joanne Lee, and M.C. Bradley of Mathematica, 1100 1st Street NE, Washington, DC, 20002, under contract with the Office of Planning, Research, and Evaluation (OPRE) in the Administration for Children and Families at the U.S. Department of Health and Human Services (#47QRAA18D00BQ). OPRE Project Officers: Jenessa Malin and Christine Fortunato. Mathematica Project Director: M.C. Bradley.

This brief is in the public domain. Permission to reproduce is not necessary. Suggested citation: Strelevitz, Tara, Eileen Gao, Joanne Lee, and M.C. Bradley. "Linked and Integrated Data: Select Findings from the State Child Welfare Data Linkages Descriptive Study" OPRE Report #2024-133. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2024.

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

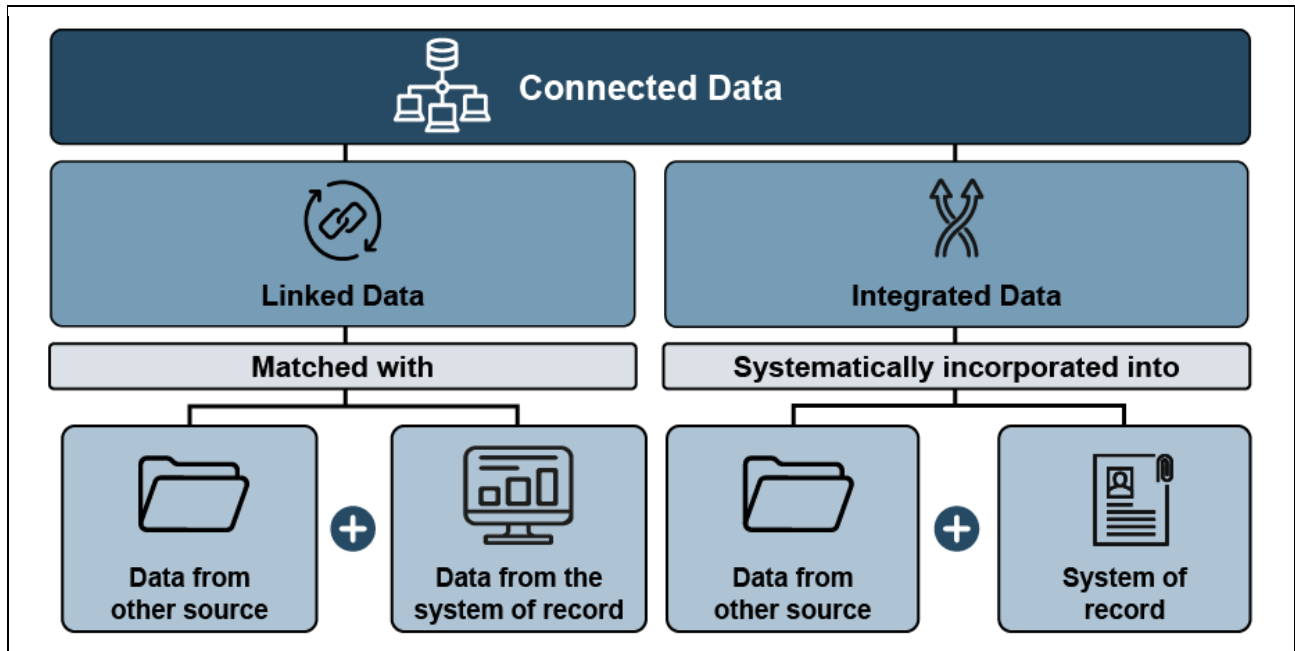


Appendix

Definitions

In Exhibit A.1, we summarize definitions of linked data, integrated data, and connected data that were developed for the purposes of the study.

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



Glossary of terms

Connected data: Linked or integrated data.

Data exchange: A separate agency systematically shares administrative records that are linked to records in the system of record.

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.

Investigated: Reports that were screened in are investigated to determine whether the alleged maltreatment occurred.

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 joined with data from other sources, based on a common identifier or other matching criteria.

Screening: Process by which reports of child maltreatment are evaluated before an investigation to determine whether they meet the definition of child maltreatment.

System of record: Child welfare information system used to report to the National Child Abuse and Neglect Data System.

Items from the Connected Data Survey used in this brief

Q6a. What are the sources of data included in the linked data?

Q8. What linking method(s) are used to merge the different sources of data?

Q9a. What common identifiers are used to link the data?

Q9b. What data elements are used to link the data?

Q11. Please identify which, if any, of these population characteristics pertaining to child welfare involvement are contained in the linked analytic dataset:

Q19. What are the linked data used for?

Q20. Who, in addition to staff from the managing organization, actively use the linked dataset?

Q21. Who, in addition to staff from the managing organization, may receive permission to use the linked dataset or extracts from it?

Q25a. What are the external sources of data that are systematically connected with the system of record?

Q31. Who, in addition to staff in the child welfare agency, may receive permission to access/use the integrated data?

Q32. What are the integrated data used for?