Research Brief OPRE 2017-67
October 2017

Early Childhood Data Definitions: A Guide for Researchers Using Administrative Data

CCEEPRA
Child Care and Early Education Policy and Research Analysis

Early Childhood Data Definitions: A Guide for Researchers Using Administrative Data

OPRE Research Brief #2017-67
October 2017

Submitted by: Carlise King and Kelly Maxwell, Child Trends

Submitted to: Ivelisse Martinez-Beck, PhD, Project Officer
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Kathleen Dwyer, PhD, Project Officer
Office of Planning, Research and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Contract Number: HHSP23320095631WC

Project Director: Kelly Maxwell
Child Trends
7315 Wisconsin Avenue
Suite 1200W
Bethesda, MD 20814

Suggested citation: King, C. & Maxwell, K. (2017). Early Childhood Data Definitions: A Guide for Researchers Using Administrative Data. OPRE Research Brief # 2017-67. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Acknowledgments: The authors gratefully acknowledge the guidance from Kathleen Dwyer, senior social science analyst at the Office of Planning, Research and Evaluation. We appreciate the thoughtful review and feedback from experts Nicole Forry, Rena Hallam, Channa Hewawickrama, Naneida Lazarte-Alcalá, and Caitlin McLean. Finally, we appreciate the support of the Child Trends communications team for their assistance in the preparation of this product.

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

Early Childhood Data Definitions: A Guide for Researchers Using Administrative Data

Purpose of this guide

This guide is intended to increase researchers' awareness of existing resources that can help them define variables to support secondary analysis of administrative datasets. Administrative data are often difficult to analyze because the same or very similar variables are defined differently in different datasets. For example, something as seemingly simple as the age of a child could be defined using a child's exact date of birth, pre-defined age categories (i.e., infant, toddler, preschool-age, etc.), or number of years old at enrollment date (e.g., 3 years old, 4 years old, etc.). These variables may be defined differently among existing administrative datasets. In these instances, researchers must determine the best way to define variables (which may include recoding existing data) before using them. Although there are resources and tools available to help researchers make decisions to resolve these discrepancies, researchers may not be aware of them. This guide summarizes the resources specific to early childhood data and gives examples of how researchers may want to use them.

Who should use this guide

This guide is intended for researchers using administrative data collected by government and/or private entities engaged in early childhood-related services, supports, or initiatives. We use the term *researcher* to include any staff person working in a state agency, university, nonprofit, or other research organization conducting research using early childhood administrative data. We developed this resource to help researchers within and outside state agencies to:

- Identify common definitions for administrative data variables; and
- Identify strategies to refine operational definitions to align study variables when they do not align in administrative datasets.

How we define early childhood administrative data

Administrative data refers to information about individual children, families, and/or providers collected and maintained as a part of program operations, services, interventions, and policies. Early childhood administrative data may include information about:

- Child care subsidy, Head Start, state pre-Kindergarten, early intervention, home visiting program data
- Child, class/group, family, organization, program site, practitioner, and system-level data
- Demographics, enrollment, eligibility, measures of quality, workforce characteristics, and policy data

How to use this guide

The guide describes four data definition resources designed to support the collection and use of early childhood data.

1. Common Education Data Standards (CEDS): Data definitions resource for administrative data collected within and across state early childhood, elementary, high school, college, and workforce institutions
2. INQUIRE Data Toolkit (INQUIRE): Data definitions and guidance on how to answer research questions related to early care and education (ECE) access and quality initiatives
3. Core Data Elements for Early Childhood and School-Age Registries (Workforce-Data): Data definitions and best practices for collecting early childhood and school-age workforce data as part of state workforce registries
4. Data Dictionary for ChildCare.gov (ChildCare.gov): Data definition resource to align state/territory child care program data used to help parents locate and choose child care arrangements

Researchers using this guide will first want to determine the type of data they are interested in gathering to determine which resource best meets their needs. The CEDS resource provides the largest number of data elements and includes at least some data elements contained in the other three resources. The CEDS is also updated periodically. The INQUIRE Toolkit, Workforce-Data, and Data Dictionary for ChildCare.gov were developed for specific types of data collection. This means they may include data elements not included in CEDS because they were developed to address specific research questions or project goals. For example, the INQUIRE Data Toolkit includes research questions related to parents' child care decision-making and use of information about a program's quality, which are not reflected in the other resources. The Workforce-Data guide is specific to data collected for state workforce registries, and ChildCare.gov focuses on data collected for child care consumer websites. As you review the resources, note the purpose and type of data included to determine which resource would best support your own research and evaluation planning.

For each resource, we will introduce you to the overall purpose, scope, and possible uses. Each resource description will include:

- An overview, background and purpose, description of how to use the resource, and examples of how you may use the resource to support your research; and
- A research scenario to illustrate how each resource may be helpful in your research.

The research scenario will start at the end of this section and continue throughout the guide, drawing examples from each of the data definition resources. While fictional, this scenario includes examples of situations you may face when using administrative data for research.

Our goal is to increase awareness and knowledge related to each resource. The examples described below represent only a sample of definitions available in each resource. We encourage you to explore each resource on your own. Links to each resource are provided throughout the guide when available.

We begin with the following scenario:

> *You are a university-based researcher who recently partnered with the Child Care and Development Fund (CCDF) Lead Agency to develop a proposal to study the characteristics of your state's subsidized child care program, and you just learned that you were awarded the grant. The CCDF Lead Agency has agreed to work with your team and share their data to answer questions about the quality of child care programs paid through its subsidy system, the qualifications of the workforce, and the characteristics of children served. Some of the questions you'd like to address are, "What is the quality of care received by children enrolled in the child care subsidy system? Do teachers in programs serving children receiving subsidies have the same level of training and experience as teachers in programs that do not serve children receiving child care subsidies?"*
>
> *As you consider the data needed to answer your study questions, you learn that you will need to gather data from several agencies and programs. After contacting a few programs, you notice that the definitions and methods for collecting the same information are inconsistent across programs. For example, the state's pre-Kindergarten program defines a child's income or age eligibility differently than the state's child care subsidy program. Or programs may describe staff positions differently, making it difficult for you to track qualifications by staff at similar levels.*

Researchers often encounter problems like these when analyzing administrative data. Are there common operational definitions you can use as a guide to determine how best to define variables for your research? Are there strategies for developing common operational definitions to align variables from different datasets? The answer is "yes." Review the description of each early childhood data definition resource to learn more.

Description of Early Childhood Data Definition Resources

To help you determine which data definition resource might be most helpful, we provide an overview of each resource and describe its purpose and how it may help you in conducting secondary analyses of administrative data. We also continue the previous scenario to demonstrate how you might use each resource to support your work (see blue boxes).

Common Education Data Standards

Overview: The Common Education Data Standards (CEDS) is a national collaborative effort, led by the National Center for Educational Statistics and established in 2009 to provide guidance to states on commonly collected data elements. CEDS includes standard data definitions, response options, and technical specifications to make it easier to share state administrative data across early learning, elementary, college, and workforce institutions.

Background and purpose: CEDS is intended to support the capacity of state data systems to exchange, compare, and understand data collected within and across early learning, K-12 education, postsecondary education, and workforce sectors. Researchers, administrators, practitioners, and other experts provided input on the early learning data elements regarding a range of early childhood services. For example, Head Start, state pre-Kindergarten, early intervention, home visiting, private child care programs, and other early childhood services are defined. Early learning definitions are organized by child, family, staff, class/group, and organization. At each level, methods for tracking demographics, enrollment, program characteristics, and workforce qualifications are described. CEDS could help you answer such questions as, "What are the definitions for commonly used early learning data elements?" or "What are recommended formats for documenting early childhood services to facilitate data sharing?"

Several resources were developed to promote standardization of educational data, including a list of data elements, definitions, formats, and online resources for aligning multiple data collection efforts. This resource allows researchers to examine how different data elements are collected and defined across education and workforce sectors. For example, a researcher or state administrator can search for education data elements by sector or upload an existing data dictionary to compare a state's data definitions to CEDS. The use of CEDS is voluntary and its information is updated based on stakeholder input. CEDS is currently on its seventh version.

How to use the CEDS resources: CEDS was developed to promote standard approaches for defining education data. This resource includes data definitions, response options, and technical specifications to support the collection and sharing of data across systems and can be used to compare data definitions and identify recommended response options and methods for recording data.

Users can search CEDS by domain of interest or conduct a word search. Below are the domains collected as part of CEDS:

- Early Learning
- K-12
- Postsecondary
- Career and Technical Education
- Adult Education
- Workforce

- Assessments
- Learning Standards
- Learning Resources

Recommended definitions for early learning: Researchers can use the CEDS website to access recommended definitions and response options for each of the domains listed above. For example, if a researcher wants to track different types of early childhood services a child received over one year, this resource includes a recommended list of possible services a child may receive.

Examples:

Early Childhood Services Offered: "Types of services that may be identified for a child needing additional services as a result of an individual service plan or other referral intended to support the child's positive development" (Source: https://ceds.ed.gov/CEDSElementDetails.aspx?TermxTopicId=30840)

Recommended List of Services: "Assistive technology services, audiology services, family training/ counseling services, health services, medical services, nursing services, nutrition services, occupational therapy, physical therapy, psychological services, sign language and cued language services, service coordination, social work services, special instruction, speech-language pathology services, vision services, behavioral health, transportation" (Source: https://ceds.ed.gov/CEDSElementDetails.aspx?TermxTopicId=30840)

Recommended formats for early childhood data elements: For each data element, users are given a recommended format to document the information in a data system for record keeping and/or later analysis. For example, a researcher may want to record the exact dates that a child received some type of support services. CEDS includes examples of how individual data fields should be documented, and at what level. This information could assist the user in planning for how they will document data collected.

Example:

Family Income: Total family income from all sources. Income includes money, wages, or salary before deductions; net income from non-farm self-employment; net income from farm self-employment; regular payments from Social Security or railroad retirement; payments from unemployment compensation, strike benefits from union funds, workers' compensation, veterans benefits, public assistance (including Temporary Assistance for Needy Families, Supplemental Security Income, Emergency Assistance money payments, and non-Federally funded General Assistance or General Relief money payments); training stipends; alimony, child support, and military family allotments or other regular support from an absent family member or someone not living in the household; private pensions, government employee pensions (including military retirement pay), and regular insurance or annuity payments; college or university scholarships, grants, fellowships, and assistantships; dividends, interest, net rental income, net royalties, and periodic receipts from estates or trusts; and net gambling or lottery winnings.

Format: "Numeric - up to 2 digits after decimal place"

(Source: https://ceds.ed.gov/CEDSElementDetails.aspx?TermxTopicId=23674)

If you are interested in using the CEDS, there are tools to help you compare your data definitions with theirs. Align is an online tool you can use to upload your own data definitions to compare them with CEDS. After identifying any differences, you can decide if you want to refine your definitions to align with CEDS.

Scenario for researchers using the CEDS

Researchers frequently want to understand the experiences of children receiving child care subsidies to better serve this population (see the Appendix for a brief description of available child care subsidy data). To answer these questions, researchers need to understand how information about children, families, and programs is documented within an administrative dataset. For example, how should you determine whether a child is enrolled in a program or receiving other support? The following scenario illustrates how a researcher might use the CEDS resource to answer the question, "What are the characteristics of children who are enrolled in subsidized child care programs?"

For your study, you want to use administrative data to calculate the number of children served in subsidized programs during the last fiscal year, and to then gather some demographics on the population. You consider several methods for how to do this. One way is to use the number of children eligible based on state guidelines. The second is to use enrollment data based on the entry and exit dates for each child. A third option is to use the number of days in attendance to determine enrollment. For example, you may want to create a threshold for enrollment so that if a child is enrolled less than a certain number of days during the year, you would not include them in your analyses. These decisions will depend on the questions you are trying to address.

You consult CEDS on methods for defining program eligibility, enrollment, and attendance to help select the best method to determine the number of children enrolled in the subsidy program in the last year. After consulting CEDS, you decide to capture child eligibility status, enrollment start and exit dates, and the number of days attended to get a more accurate count of children in the subsidy program who participated in a child care center.

INQUIRE Data Toolkit

Overview: The INQUIRE Data Toolkit was developed in 2013 by the Quality Initiatives Research and Evaluation Consortium (INQUIRE) Data Work Group to help researchers answer questions related to state quality rating and improvement systems (QRIS) and other quality initiatives. This resource includes information on over 200 data definitions and guidance on how to analyze 20 research questions related to ECE access and quality initiatives.

Background and purpose*:* INQUIRE Data Work Group (INQUIRE) is a project supported by the Office of Planning, Research and Evaluation in the Administration for Children and Families in the U.S. Department of Health and Human Services. INQUIRE is a learning community of early childhood researchers with expertise related to research and evaluation of quality improvement initiatives such as QRIS.

The INQUIRE Toolkit was developed through input from the INQUIRE Data Work Group. Members helped identify common definitions used by states and developed guidance on how to answer research questions related to state quality initiatives. The toolkit is intended to support researchers conducting research on QRIS and other quality improvement efforts.

How to use the INQUIRE Toolkit resources*:* The toolkit is available as a written document and an online search tool. Users can select policy-related questions and review data definitions by category. The written document is separated into two sections: 1) *Linkages Guide* and 2) *Dictionary of Common Data Elements*.

The INQUIRE Toolkit could be used to help answer such questions as:

- What data definitions could you use to answer questions about state quality improvement initiatives?
- What data elements are needed to help evaluate programs and policies?
- How do definitions align with other data collection efforts?

The *Linkages Guide* provides analysis recommendations for a series of questions. Each analysis section contains hyperlinks to guide readers back to specific data fields as needed for the analysis. The online look-up tool can be accessed at http://inquiredatatoolkit.org/.

Recommended policy questions about state quality improvement initiatives: The *Linkages Guide* section includes 20 questions that can be used for state monitoring, reporting, and evaluation efforts. Each of these questions is related to state quality improvement initiatives.

> **Example:** A user could select the following policy-relevant question: "What percentage of at-risk children was enrolled in high-quality care?" They could then use this toolkit to identify the multiple data elements needed to answer this question. The user would also receive information on the level (i.e., child, staff, and classroom) for which each data element would need to be collected. For example, to track the quality of care a child receives, information about the quality of the classroom or program in which a child is enrolled would need to be analyzed, with demographic data specifying whether the child meets the researchers' definition of at-risk.

Recommendations for combining data elements for analysis

> **Example:** "The first step in this analysis is to determine the total number of at-risk children in the state or territory. With the goal of aligning definitions across early childhood data collection and reporting efforts, we recommend using the criteria outlined in the Race to the Top – Early Learning Challenge application in determining how many children fall into each of seven high-needs categories. These categories (and the corresponding data elements) are: low-income (Family Income, Number of People in Family); developmental disability (Primary Type of Disability); English learner (Language Code); Indian Lands (Child Resides on Indian Lands); migrant (Child: Migrant Status); homeless (Child: Homelessness Status); and, foster care status (Child: Foster Care Status). Children are considered at-risk/high needs if they meet one of these criteria" (INQUIRE Toolkit, 2013, p.20).

The *Dictionary of Common Data Elements* section contains a list of data definitions and recommendations for data collection. It is organized by the level of data collection: child, class/group, family, practitioner, program site, organization, and system. For example, a system-level data element includes definitions for the state's QRIS system, such as QRIS levels or eligibility requirements. Each data element includes a definition, methods for documenting data collected, and its alignment with federal reporting requirements (e.g., the Quality Performance Report) or parallel data collection efforts (e.g., Common Education Data Standards).

Recommended definitions to measure quality improvement: The Dictionary of Common Data Elements is organized by the level of data collection. For example, data that would be collected about a classroom—such as child-staff ratio or classroom quality observations—would be listed under the class/group section of the document. Similarly, information about early childhood staff would be available under the practitioner section. There are 7 levels of data collection included: 1) child, 2) family, 3) practitioner, 4) class/group, 5) program site, 6) organization, and 7) system.

> **Example:**
>
> *Recommended definition:* "Accrediting organization (program site level): Site is accredited by one of the following organizations" (INQUIRE Toolkit, 2013, p.115).
>
> *Recommended list of organizations:* "National Association for the Education of Young Children, National Early Childhood Program Accreditation, National Accreditation Commission, Council on Accreditation, National Association for Family Child Care, Southern Association of Colleges and Schools, or other accrediting organization" (INQUIRE Toolkit, 2013, p.115-116).

Recommended variable type: The data variable should be saved as a numeric field instead of a text field. This means the user would assign a numeric value to the organizations listed above. For example, National Association of the Education of Young Children = 1, National Early Childhood Program Accreditation = 2, National Accreditation Commission = 3, etc.

Users can review the *Dictionary of Data Elements* and *Linkages Guide* sections in the written document or select categories and questions using the online search tool.

Scenario for researchers using the Inquire Toolkit

Scenario for researchers using the INQUIRE Toolkit: The implementation of QRIS has increased the availability of administrative data about program quality (see the Appendix for a description of available QRIS data). To conduct research on the relationship between these initiatives and increased program quality, researchers need clear definitions of data such as which programs participate in QRIS, the types of support provided to improve quality, and tools used for determining the rating. Our scenario continues with an example of how a researcher might use the INQUIRE toolkit to answer this question, "Are programs serving children who receive child care subsidies more likely to participate in the QRIS compared to programs that do not serve children who receive subsidies?"

After your team established the number of children served in subsidized programs, you now want to understand the quality of those programs. You review the INQUIRE toolkit to identify which variables about a state's QRIS are important to collect about the programs serving children receiving child care subsidies. As you review the INQUIRE resource, you find additional variables for tracking QRIS participation history, date of rating, and expiration date of rating. Although you originally planned to request only the program's current rating, you now decide to expand your data request to include more historical data regarding each program's quality rating. You request data from the QRIS inception date in 2010. This gives you more information about how long the program was part of the QRIS, how many times it was rated over the years, and how the quality of care changed over time.

Core Data Elements for Early Childhood and School-Age Registries

Overview: The Core Data Elements for Early Childhood and School-Age Registries (Workforce-Data) resource was developed in 2013 by the National Registry Alliance (TNRA) to provide definitions and best practice standards for collecting early childhood and school-age workforce data. This tool will provide guidance on data you may want to collect at the individual, organizational, and event levels related to practitioners and trainers. It provides details on tracking professional development activities, education, and credential information.

Background and purpose: TNRA is a private, nonprofit, voluntary organization of state early childhood and school-age workforce registry and professional development leaders committed to enhancing, strengthening, and supporting the work of registries. In 2013, TNRA established a recommended list of core data elements and suggestions for best practices to guide state-level workforce data collection. The guidelines are intended to promote standard data collection practices for state-level workforce data systems. They also serve as requirements for states to participate in TNRA's national workforce dataset. Workforce data elements are aligned with CEDS where possible.

How to use the Workforce-Data resource: Researchers interested in collecting workforce data will find this resource helpful to identify core data elements they may want to collect. The guidance is organized into

three sections: personnel, organization, and event information. Standards for best practices are included for select data elements. These core data elements could help you answer questions such as, "What data should you track in state-level workforce data systems?" or "What are best practices for defining information about the early childhood workforce?" Below is a description of the data elements included in each of the sections:

1. **Person information**: This section includes recommendations for collecting data about individual staff, trainers, or mentors working in early childhood settings. Data related to persons include demographics, roles, employment, education, certification, and training.
2. **Organization information:** This section includes recommendations for collecting data about places of employment or community technical assistance organizations. Data related to organizations include contact information, work policies, and quality measures.
3. **Event information:** This section includes recommendations for collecting data about professional development and technical assistance provided to persons employed by early childhood organizations. Data related to events include training content, target audience, duration, and evaluation of activities.

The data elements included in this resource align with definitions provided as part of CEDS and the INQUIRE Data Toolkit. This resource differs from others in that it prioritizes the collection of specific workforce data elements and shares best practices on how to collect these types of data. Below are examples of definitions and recommended practices for data collection.

Recommended definition to document the early childhood workforce: An important part of tracking workforce data is to first have a process for identifying individual persons working across organizations in a way that does not personally identify them. Once you have a way to track individual teachers, directors, and trainers, you will be able to analyze work patterns and career development over time and across sectors (e.g., center-based, home-based, school-based programs). The example below defines a workforce identification number used in workforce registries to track information about persons working in early childhood programs.

Example: Workforce Identification Number

Definition: "The unique (single, non-duplicated) identification number for an individual assigned by the registry data system. This number is used to track all related data for the registry member over time. The registry workforce identifier may/may not be used to link registry data to other state-level early childhood/school-age data systems" (The National Registry Alliance, 2013, p. 5).

Recommended practices for collecting early childhood workforce data elements: This tool also provides guidance on how to collect workforce data. Experts experienced in collecting workforce data informed the development of recommended data collection practices for some data fields. These recommended practices advise users on how to enter, update, and restrict the information used to define each field. Below is an example of the standard practice recommended for developing a workforce identification number.

Example: Workforce Identification Number

Standard Practice: "The workforce identifier is not the SSN or any other number or combination of numbers and letters containing personally identifying information" (The National Registry Alliance, 2013, p. 5).

Scenario for researchers using the Workforce-Data resource

A strong and stable early childhood workforce is an important part of program quality. As more states develop workforce registries, the amount of data on early childhood educators is increasing (see the Appendix for a description of available workforce data). Researchers will need to develop consistent definitions for workforce elements such as education, training, and turnover to effectively conduct research that can inform workforce policies across early childhood settings. We continue our scenario by showing how the Workforce-Data resource could help you examine the question, "What is the education and training of staff working in subsidized child care programs?"

> *After assessing the programs' quality, the team decides they want to address questions about staff experience and training. There are multiple ways to assess this, and the team decides to analyze data on early childhood degree status, degree or certificate type, and early childhood credential held by a person. After reviewing the Core Data Elements for Early Childhood and School-Age Registries, you find recommendations for collecting training and technical assistance variables. You had not thought to include community-level training in your documentation so you revise your data request to include community training hours, type, and content.*

Data Dictionary for ChildCare.gov

Overview: The Data Dictionary for ChildCare.gov (version 1.0) was released in 2017 as part of the ChildCare.gov Project funded by the Administration for Children and Families. This resource provides guidance on data elements for child care provider searches as required in CCDF rules for state and territory child care consumer websites. Data elements include information about available child care options needed to help families identify high-quality early learning programs that meet their needs. The data dictionary defines data you may want to collect related to tuition rates, licensed capacity, licensing history, quality ratings, hours of operation, schedules offered, and other program characteristics (i.e., curriculum, transportation, accreditation, etc.).

Background and purpose: The Childcare.gov Project was funded to help states meet new requirements in the 2014 reauthorization of the Child Care and Development Block Grant (CCDBG). The law requires that ChildCare.gov, in addition to state and territory websites, must be easy to navigate for families and include specific information to help families as they search for high-quality child care arrangements.

The CCDBG reauthorization requires ChildCare.gov to include a list of all licensed and license-exempt child care programs from each state and territory. Information on each licensed program must include a program's health and safety record and quality rating for families to use as they search child care services in their communities. The law requires the national website to link with existing state/territory websites or other sources like child care resource and referral agencies that collect information about child care providers. Because states collect this information using multiple databases and different data definitions, a data dictionary was created to help child care administrators assess their current data definitions and make changes, as needed, to meet the requirements for ChildCare.gov.

The data dictionary was informed by the National Data Set for Early Learning and School Age Programs, the Common Education Data Standards, assessments of existing state/territory websites, and child care resource and referral networks and ChildCare.gov data collection required by CCDBG.

How to use the Data Dictionary for ChildCare.gov resource: This data dictionary resource includes a list of data elements from state/territory child care databases that are required for the ChildCare.gov website. Each data element includes a definition and specific guidelines on the data values and format. For example, monthly tuition for infant care is defined as the highest amount charged before the application of any discounts or subsidies. All tuition data elements should be exported in a dollar format with two

decimal place (e.g., $1,000.00). Data elements with specific data values are listed in a separate table. For example, the values defined for the state and federal program type data element include Head Start, state pre-Kindergarten, state vouchers, state contracts, and private scholarships. Child care administrators use the data dictionary definitions and instructions to assess how well their definitions and formats align with ChildCare.gov requirements.

The data dictionary is organized into five tables:

- Table 1. General Program and Contact Information
- Table 2. Information on Quality Ratings
- Table 3. Information on Program Features
- Table 4. Information on Operating Schedules
- Table 5. Information on Tuition and Fees

For each data element, users are given a recommended format for documenting the value of the element. For example, licensed capacity information must be reported as a whole number. The tables include each of the data elements that have been incorporated into the design for ChildCare.gov, along with a high-level definition and the priority level of each element. Data elements labeled as "required" are the minimum data elements required for a successful connection with ChildCare.gov. Data elements labeled as "needed" are elements that states and territories are strongly encouraged to make available, to improve the information available to families. Data elements labeled as "desired" are elements that states may want to consider collecting and sharing in the future. See examples of the data element definitions and data element type value tables below:

Data Element Definitions

Sample Data Definitions:

License type: "General category that identifies how the program is regulated by a state agency or partner agency. Licensed programs are regulated or certified by a state licensing agency. Registered programs are required to register with a state agency or partner agency. Exempt programs can legally operate without being licensed or registered" (Data Dictionary for ChildCare.gov (version 1.0), 2017, p. 5).

Training and experience to support special needs: "Type of training and experience that program staff have regarding accommodating children with special needs" (Data Dictionary for ChildCare.gov (version 1.0), 2017, p. 9).

Data Element Type Values

Sample Data Values:

Table 1. Data Element Number and Name. #20 License Type:

- 1=Licensed
- 2=Registered
- 3=License exempt

(Data Dictionary for ChildCare.gov (version 1.0, 2017, p. 5)

Table 3. Data Element Values. #37 Training and Experience to Support Special Needs:

- 1=Developmental delay
- 2=Emotional/behavioral
- 3=Hearing impairment
- 4=Visual impairment

- 5=Physical challenges
- 6=Special health needs
- 7=Speech impairment

(Data Dictionary for ChildCare.gov (version 1.0, 2017, p. 9)

To help families access specific licensing information, states are asked to include a link to the state's licensing website. For example, the ChildCare.gov definition for the licensing complaints data element is "URL link to state licensing complaint reports on the child care program. One of three types of URL links can be used: (1) direct link to program-specific report, (2) link to program profile that includes direct link to program-specific report, or (3) link to child care licensing search tool." (March 2017 Data Dictionary for ChildCare.gov (version 1.0), 2017, pf.5).

Scenario for researchers using the Data Dictionary for ChildCare.gov

The type of child care arrangement a family may choose is often dependent on the location and price of options in their area. Other factors such as a program's health and safety record, hours of operation, staff experience, and type of curriculum offered may influence the selection of care. Languages spoken by staff is another important factor for families who want their child's native language supported and who want to communicate easily with staff about their child's needs. To understand reasons a family may select a certain type of care, researchers need to understand what type of care options are available to families. We end our scenario by answering the following question: "How do the characteristics of programs selected by a family using a subsidy compare with the characteristics of other programs in a geographic area?"

You have identified your data elements for the number of children receiving a subsidy, staff qualifications, and quality ratings for programs selected. You also want to understand how the types of care selected by families using a subsidy compare to other programs in a geographic area. You use the Data Dictionary for ChildCare.gov to get a list of data elements and definitions for information about programs families may consider when selecting care. You select information about program tuition to determine which programs cost more than the state's maximum subsidy payment amount. Next, you collect information about the hours of operation to see how many programs offer flexible schedules that would accommodate families with different work schedules. Last, you look at the languages spoken by staff to compare how well they align with languages spoken by families receiving a subsidy. You can use these data to understand how programs selected by families compared to other programs in the geographic area not selected. Now that you have developed your research questions and refined your data definitions, you are ready to conduct your study.

Appendix: Sources of Early Childhood Administrative Data

Although the resources we describe in this appendix were not specifically developed to provide guidance on common data definitions, they do include some data definitions. We believe that they may be useful to researchers interested in analyzing administrative data because they also describe some of the administrative data collected by states.

Child Care Subsidy Data: States are required to develop monthly reports on children, families, and child care settings that receive child care subsidies funded through the Child Care and Development Fund. The ACF-801 data reporting guidelines include a data dictionary that can help researchers understand what data are collected about families using subsidized care. Examples of data elements collected include type of care, child care subsidy payment, hours of care per month, and reasons for needing care.

Quality Rating and Improvement Data: Quality Rating and Improvement System (QRIS) initiatives provide a system for collecting data on the quality of early care and education programs. These ratings are based on criteria tied to specified state-defined quality standards. Quality indicators typically include staff qualifications, the physical learning environment, family engagement, and program management. The QRIS compendium includes information reported from state QRIS leaders about a range of variables such as the quality standards included in the QRIS, program participation, and technical assistance provided.

Early Childhood Workforce Data: According to the 2016 Early Childhood Workforce Index, 47 states collect some data about early childhood educators working in center- and home-based settings. Most states (42) use a professional workforce registry to house demographic, education, certification, training, and employment information on early childhood staff. Researchers can click here to find out what workforce data are collected and defined in their state's professional workforce registry. These data can be used to address research questions related to the early childhood workforce.