



DATA DIALOGUE PROTOCOL FOR UNDERSTANDING RELATED ECE DATA SYSTEMS

sdp.cepr.harvard.edu
sdp@gse.harvard.edu

This tool is part of Preschool Supply & Demand, a collaboration between the Strategic Data Project and representatives from the California Department of Education, Denver Public Schools, Denver Preschool Program, South Carolina First Steps, Chicago Early Childhood Integrated Data System, National Head Start Association, Massachusetts Department of Early Education and Care, and Washington Education Research and Data Center. It is supported by the Gates Foundation.



Center for Education Policy Research
HARVARD UNIVERSITY

Data Dialogue Protocol for Understanding Related ECE Data Systems

Tool Description

These protocols and notetaking guides collect information on how different related ECE data systems function—how data is entered, processed, maintained, and updated. Having this information helps lead agencies hosting comprehensive early childhood data systems effectively address technical and operational integration across systems. Split across two or more data dialogue sessions, this collaborative tool builds relationships with relevant data stewards, creates an ongoing conversation about data quality and management, and documents dataset-specific information needed for analyzing supply and demand within a specific state or locality.

- The interview protocol for Data Dialogue #1 can be found starting on page 4.
- The interview protocol for Data Dialogues #2+ can be found starting on page 10.
- A sample table for recording programming setting, funding streams, and tracking can be found start on page 19.

Purpose

Structured, recurring engagement with data providers is necessary for building reliable, accurate data systems of ECE data. ECE systems and data are created in a fragmented landscape with operational and compliance requirements in mind. This means that even if systems look similar on the outside, these systems were built on various real-world processes and concerns that are specific to that system. Data engineers and analysts must understand these on-the-ground realities or run the risk of making incorrect—though seemingly reasonable—assumptions about what fields mean, when values change, and how records should be interpreted over time. These data dialogues help bridge the gap between engineers and analysts and the people who actually create, enter, manage, and report on ECE data. It is meant to be an ongoing process of information sharing and rapport building; emphasis on “ongoing,” as data systems are constantly shifting due to programmatic changes.

What a “Data Dialogue” Is:

Data Dialogues are recurring, structured engagements designed to surface and validate business rules before those rules are encoded into architecture or analysis. ***They are not audits, compliance checks, or one-time onboarding sessions.*** These conversations are essential inputs into mapping logic, deduplication strategies, documentation artifacts, and analyst guidance.



Interviewer Instructions

Your efforts to cultivate a positive, comfortable atmosphere can foster a productive and meaningful interview that is rich in detail and insight. The goal of a one-on-one interview is to create space for the participant to openly share their experiences, perspectives, and reflections in their own words. Participants may express uncertainty, nuance, disagreement, or evolving thoughts over the course of the conversation. Encourage honesty and depth rather than “right” answers. Your role is primarily to listen actively, ask open-ended and probing questions (see examples below), keep the conversation focused on the interview topics, and gently guide the discussion forward when needed.

Remember that more sessions can (and should) be scheduled. It is better to understand a topic fully and deeply rather than get through the entire checklist of materials. Importantly, the order of these questions can be changed if specific information is needed to move forward with the technical or analytical work related to these data systems.

Examples of probing questions include:

- I’m not sure I understand ____, can you say more?
- I want to make sure I understand. I am hearing you say ____, is that correct?
- Can you say more about that?
- You mentioned ____, can you elaborate?
- What did that experience look like for you?
- How did that affect your work/experience/decision-making?
- Can you give an example of that?
- What made that particularly challenging or successful?

Waiting time can feel uncomfortable, and that’s okay. Give participants time to think and respond by pausing for 10–15 seconds after asking a question or after a participant finishes speaking. If the participant does not respond, you may gently rephrase the question or ask whether any clarification would be helpful.

Below are some facilitation pointers you may find helpful.

Ways to foster a relaxed atmosphere	Ways to encourage open participation
• Be friendly and welcoming. • Smile and maintain appropriate eye contact. • Speak in a calm, pleasant tone of voice. • Use relaxed and attentive body language. • Incorporate humor where appropriate and natural. • Be patient and do not rush participants to respond. • Show genuine interest in participants' perspectives. • Acknowledge sensitive topics thoughtfully.	• Review expectations at the beginning of the conversation. • Use pauses and probing questions. • Ask open-ended, non-leading questions. • Reflect back key ideas to confirm understanding. • Encourage participants to expand on examples or experiences. • Allow participants to guide the discussion when appropriate. • Avoid interrupting unless necessary to redirect or clarify. • Do not judge responses.

Interview Protocol for Session #1

Agency:	
Program year:	
Date of interview:	
Interviewer(s):	
Attendee(s):	

Primary contact for follow-up:	
Role/title of primary contact:	
Preferred follow-up method:	

Introduction

We are reaching out as part of our work to better understand how each partner agency's data systems function—how data is entered, processed, maintained, and updated. This will help us more effectively address technical and operational questions and support smoother integration across systems. Ultimately, we are working toward a clearer, more connected picture of the early childhood education data landscape in our [state/city/county/locality]. We can only achieve this goal by learning from your experience and expertise with your data system.

This first “data dialogue” is meant to be a broad overview of your current system, with future sessions diving deeper into the nuts and bolts of your data. We may not get through everything, but this is the start of our ongoing communication. This is meant to be a collaborative process, so we also hope to address any pain points you have experienced in delivering your data.

With all that said, do you have any questions?

Protocol and Notetaking Template

Question	Notes
Section 1: Getting to Know the Interviewee	
Can you tell me a bit about your role at your agency and how long you've been here?	
How does data show up in your day-to-day work?	
In terms of data, how do you see your agency's use of data and what would you like to be moving towards?	
Section 2: Confirm Agency Programs for the Current Program Year	
Confirm which programs are running (e.g., center-based, home-based, home visiting)	



What funding sources do you receive? • Who is the grantee of these sources? Are they passed to partners, delegates or sub-grantees? • For each, approximately how many of your served children receive them? <i>Create a table that shows which grants are there, where they live, and then ask about how many slots are in each (see end of document for example).</i>	
Who, if any, are your sub-delegates or community partners [for Head Start or other funding sources]? How do you define these relationships?	
What is the start and end of the program year? • Is there a summer program?	
Section 3: Overview of the Agency's Data System	
Can you tell me about how data is managed in your organization?	



Who is responsible for data management and oversight at your agency and what does the data team look like? • Are there multiple people involved in data preparation, quality checks, or final submission? • Can you tell us about their roles?	
How often is data documentation updated?	
Can you share this documentation with us?	
What data system(s) do you use for managing child, staff, and program data (e.g., ChildPlus, Salesforce, Excel)? • Is all staff, program, child enrollment, attendance, child health, child assessment data housed in the same system? • Do these include all programs you fund/service? • How long has the agency been using this system? What's your overall experience with the system?	



Agencies that receive funding from different sources sometimes must enter a child's record into multiple systems. Is this the case here? If so, what does that look like?	
Where are the legacy systems? • What year did it sunset? • What moved or stayed? • Were there any issues with the transition of data? What prompted the decision to move to the new system?	
Section 4: Submission Workflow <i>NOTE: This section specifically deals with how data from the provider's system is submitted to an integrated data system (ECIDS); it might not be relevant for all users of this tool.</i>	
<i>Provide your current understanding of how the agency has historically provided data (e.g., frequency, method of transfer).</i> • Does this ring true? What would you add or clarify to this understanding?	
Is this an automated connection for pulling data from your agency? • Are there any additional considerations we should have about when we pull data?	



What challenges have you encountered with data submission to our system? • Are there any parts of the current process that feel unclear, burdensome, or manual?	
What is the best way for us to communicate data cleaning/processing questions to you?	

Next Steps

The next step is to schedule follow up data dialogues focused on child- and family-level data as well as site- and classroom-level data. Between now and the next session we'll review what you've shared to identify gaps in our knowledge and surface any questions we may have in the interim.

Session 2 date/time:	
Session 3 date/time:	

Interview Protocol for Sessions #2+

Agency:	
Program year:	
Date of interview:	
Interviewer(s):	
Attendee(s):	

Primary contact for follow-up:	
Role/title of primary contact:	
Preferred follow-up method:	

Introduction

Before we dive into specific child- and family-level data, it's helpful for us to understand the everyday processes that generate that data. Our goal is to better interpret what we see in the data you share with us to ensure our system design aligns with how your operations work. A key challenge we face is making sure we can accurately identify individual children—both within a single agency and across multiple agencies.

To do this well, we need to understand how certain child-level variables are stored in your system and what their limitations might be when used for matching purposes. We also need to understand how children are grouped within classrooms and sites; these relationships help us organize and analyze the data we receive.

With all that said, do you have any questions?



Protocol and Notetaking Template

Question	Notes
Section 1: Child Information	
What unique identifiers are used for children (e.g., internal ID)? • Do these IDs stay the same across program years, as long as the child is enrolled? • If they leave and re-enroll, are they issued another ID?	
How is date of birth stored? • Is it ever approximated or replaced with placeholder data? • For prenatal enrollments (expectant families), what is entered for the child's date of birth? Are parents' dates of birth used, or something else?	
How is age group (infant, toddler, pre-K) categorized (if at all) by your system? • Is this automatically calculated based on date of birth or manually assigned? • Do children ever move age groups during the program year?	



How often is family information (address, income, etc.) updated?	
Section 2: Enrollment and Status Tracking	
Do families apply directly through your program or do they use another portal to apply?	
Can you import those applications directly into your native data system, or do they need to be reentered manually? Is this done by an enrollment specialist, case worker, etc.?	
How do you identify currently enrolled children in your system (e.g., enrollment status, enrollment start and end dates)?	
What are the stages of enrollment? What statuses are marked in the system?	
During recruitment season, how are next year's enrollments captured in your system (e.g., once a family accepts, is there a projected enrollment start date)?	



Do you store enrollment by program or by classroom? Can children be enrolled in multiple programs or classrooms simultaneously?	
How do you track part-time vs. full-time enrollment? How is that recorded (a flag, hours there, by program type, etc.)?	
What does your year-end process look like? How does this vary by the program and layers of subcontractors?	
Section 3: Eligibility and Funding	
How do you determine or record a child's eligibility for funding sources (e.g., Head Start)? How is this marked in the system? Is it all in one system?	
Are funding sources tied to children in your system? • If so, is this by site, classroom, or otherwise?	



Is family income/eligibility tracked? What codes are used? Do you enter parental employment in your databases?	
How do you define a “slot” in your system—both a “program” slot and a “funded” slot? What attributes are associated with each funded slot? (e.g., funding source, age group, full/part-day, duration, program setting)	
Are funding sources tracked separately or together?	
Is funding information all held in the same database?	
Is slot data tied to specific classrooms or individuals? Can children or classrooms have more than one funding source? How is this denoted (multiple fields, a single field with combinations)?	
Is the actual utilization of slots tracked and denoted in the system?	



How often is funded slot information updated, and by whom?	
Section 4: Attendance Tracking	
How is daily attendance recorded in your system? Is it captured per child, per classroom, or both?	
Is attendance stored as daily records or aggregates?	
What statuses are recorded (e.g., present, absent, excused, holiday)?	
How is attendance matched against scheduled times? Are scheduled times tracked separately (e.g., program dates)?	
Section 5: Staff & Classroom Assignments and Sites & Locations	
<i>Next, we would like to discuss how children are grouped in your system, mainly by classroom and site location. Often agencies will use these to denote groupings beyond physical space (e.g., having a classroom represent a home visiting group). We are trying to fully understand how you use and organize these types of data.</i>	



How do you define and track classrooms in your data system (e.g., by space, by teacher, etc.)? Do classrooms have consistent names or IDs over time? How are these updated? • If children move between rooms, how is that handled?	
What are the age ranges for each room? • How is age assignment done? • What are the maximum capacities for each room? Are they determined solely by the age of the child? Do these differ for home visiting caseloads? • Are there any “floating” or mixed-age rooms? How are they handled?	
What staff roles are tracked in your data system? (e.g., lead teacher, assistant, floater) • What data do you store for each staff member (e.g., name, credentials, training)? • Can staff be associated with multiple classrooms or programs?	
How do you define a “site” in your organization? Is it at the physical location, by program or a mix?	



How are site IDs determined? • How do they relate to other IDs like licensing numbers? • Do IDs align with other sources? When are these tied to other data in your system (funding, child membership, etc.)?	
Is program information stored at the site level? Are start/end dates used?	



Follow-up Instructions

Future sessions should be conducted in order to cover any questions that were not covered in the first few sessions.

Moving forward, additional data dialogue sessions should be conducted regularly to review any changes to data systems. This is especially likely to happen between program years as funding availability changes, systems go through updates, and reporting requirements may change how data are handled.

Below is a sample lifecycle for data dialogues that follow a typical program year timing (Fall–Spring). It can be adapted or modified as needed.

	Fall	Winter	Spring/Summer
Topic	Program-Year Confirmation Dialogue	Mid-Year Check-in	End of Year Checks
Timing	September–October	January–February	July–July
Purpose	• Check current program year changes and continuations. • Confirm open site lists and agency relationships.	• Identify any gaps in understanding. • Check any unexpected analytical findings • Confirm enrollment and other counts.	• Lock in current program year facts and implementations. • Prepare for any known changes for next program year. • Validate end-of-year counts.



Sample Table for Recording Programming Setting, Funding Streams, and Tracking

<i>Program Setting</i>	<i>Funding Stream</i>							
	Present?	EHS	HS	CCAP	MIECHV	EI	Funding Program A	Funding Program B
Directly Operated Sites								
Center Based								
Home Visiting								
After School								
FCCH								
Partners / Delegates								
Center Based								
Home Visiting								
EHS/HS Sub-Recipients								
Center Based								
Home Visiting								

Legend

Unknown	Yes + Tracked in SIS	Yes + Not Tracked in SIS	No Funding	Maybe (not tracked in SIS)	N/A
Please confirm whether the funding stream is used.	Funding stream is used and shared with organization.	Funding stream is used, but not in the data shared with organization.	Funding stream is not used.	Funding stream may be present but is not tracked by the data governor.	Program does not exist and/or not supported by funding stream.

NOTE: Funding Programs A and B are placeholders and can be replaced with other funding programs present in the system or removed. Similarly, not all providers have partners/delegates or EHS/HS sub-recipients. This table should be customized to the local context.

