



Data Direction 5: Linking data to understand children’s academic progress from early care and education through elementary school

Sarah Friese, Dale Epstein, and Kelly Maxwell, *Child Trends*

This is one of a set of five “Data Directions” that present issues ECE staff may encounter if they are interested in or attempting to share or link their data with other agencies. Each Data Direction presents a hypothetical scenario, and then outlines possible action steps programs could use to address the issues raised. This fifth Data Direction presents an example of a specific data linking project and possible steps to take in order to link ECE program data with public school data.

Ms. Booker is the director of an early care and education (ECE) program located in a neighborhood with a high percentage of both low-income and non-English speaking families. Her program serves children age three to five, and most children who graduate attend a public school in nearby school districts. She would like to track the progress of her children once they enter kindergarten and up through third grade to better understand students’ development in their early elementary school years and possibly identify areas for improvement in her program. Her staff complete assessments of children’s school readiness skills in the fall and the spring. She believes the school districts administer assessments two times each year starting at the beginning of kindergarten and continuing through the end of third grade. Ms. Booker knows that access to assessment scores from kindergarten to 3rd grade will help her and her students, but she is not sure what she needs to do to get the data from the school.

Accessing the K-3 assessment data, linking¹ it together with her ECE data, and then analyzing the linked dataset is a complex process that will require Ms. Booker to implement strategies that call on her own skills as a program leader, the skills of her staff who have expertise in data management, and possibly a community partner who can provide research and data analysis support and expertise.

Action Steps

Below are action steps that will help Ms. Booker link her program's data with data from the local school system(s) to address her question(s) of interest.

1. Specify the questions of interest that can be answered by linking ECE data to the K-3 data.

The first step in this process is for Ms. Booker to identify her questions that require K-3 data from school districts. To facilitate this process, she could write down a list of specific questions she most wants answered that would require combining her own data with data from the districts. If she is working with an external research partner, they can help her determine which questions she can analyze and answer herself and which ones require more advanced statistical expertise.² It is important for Ms. Booker to think about these questions early, before she requests the data, to ensure that she is obtaining the appropriate information from the district partners. Using these questions as a starting point will help her understand which data elements she needs to request from the district. As she develops the list of questions, it may also be useful to describe why these questions are important to her and how the information may also be useful to school district leaders. Possible questions she could answer using linked data include:

- ***Are some of the children better prepared for entry into kindergarten than others?*** Access to math and literacy assessments at the beginning of kindergarten will allow Ms. Booker to examine the school entry skills for specific sub-groups of children, such as those who attended her program for two years compared to those who attended for only one year, or Dual Language Learners compared to monolingual English speakers. She can use this information to help her identify how each of these different groups of children are doing in kindergarten, so that she can then consider ways her program can work with the school district partners to support each child during the transition to kindergarten.
- ***Are some of the children who participated in Ms. Booker's ECE program more likely to be reading at grade level by the end of third grade than others?*** Having the data to examine this relationship will help Ms. Booker understand how well her children meet an important school benchmark, and whether there are any differences in third grade reading scores for particular subgroups of children, such as those who attended Ms. Booker's program for two years versus one. Examining these data for each group of children that leaves her program and enters kindergarten, and comparing their data to children from similarly disadvantaged backgrounds, may provide some information about the relative skills of children who attended her program. Following cohorts of children allows Ms. Booker to see if there are changes from year to year in the percentage of children meeting benchmarks.

These are some examples of questions that Ms. Booker could answer by linking her data with data from the local school districts. The information from analyzing these data could offer possible ideas to explore further for program improvement,

¹ Data are considered linked when information from two or more separate data systems or databases are shared, connected, combined, or merged. These data systems or databases may be housed in the same program or in multiple programs or agencies. Linking can occur in various ways ranging from simple, (e.g., sharing a spreadsheet) to more complex (e.g., merging two databases into a single file), to very complex (e.g., fully integrating data across multiple agencies). However, all linked data have the potential to provide useful information to support ECE program improvement.

² For more information on this topic, see *Data Direction 3: Engaging an External Research Partner to Help Use Linked Data*, another resource in this series.

Action Steps

1. Specify questions of interest.
2. Develop a relationship with staff within the school districts who oversee the K-3 data.
3. Review policies and practices to ensure that the program protects the children and families' privacy.
4. Identify the specific data to be shared.
5. Develop and document a process for data linking.
6. Develop a long-term data management and reporting plan.
7. Merge the data.
8. Analyze the data.
9. Annually review the partnership, data sharing agreement, and data plan.

but it is important to note that the data would not allow Ms. Booker to know whether participation in her program *caused* children to be more successful in school. There could be other reasons for school success, like family characteristics, quality of the elementary school, characteristics of the communities families live in, transition supports, or other factors. If Ms. Booker was interested in understanding the impact of her program on children's outcomes in elementary school, she would need to work with a research partner to conduct a more sophisticated study, often known as an *impact study* or *impact evaluation*. Administrative data like the kind Ms. Booker collects and is linking with the districts can provide an excellent source of data for such studies, and a research partner can provide expertise about the quality of the data and how to use it appropriately in an impact study.

2. Develop a relationship with the staff within the school districts who oversee data.



Many large school districts have staff who collect, manage, and analyze data about their students. The lead person in this group is an essential ally to any ECE program hoping to access school district assessment data. The school district data leader will likely know the steps Ms. Booker needs to follow to get approval to access district data. This person may also know what steps are needed to actually link the data between ECE programs and elementary school or district data. Talking with this person is a key step in the data linking process.

The conversation should help Ms. Booker and the school district staff develop an understanding of how this collaboration could be useful to both the ECE program and school district. The school district, for instance, may be interested in this project because it would help them better understand the children they serve and strengthen collaborations with ECE programs in their community. More information on how to develop relationships with data owners in other organizations is available in another resource in this series, *Data Directions 4: Developing a Data Sharing Partnership*.

3. Review policies and practices to ensure that the program protects children and families' privacy.

Linking data will need to be done in a way that follows the data privacy and confidentiality requirements for the school district that are included in the Family Educational Rights and Privacy Act (FERPA),³ as well as any relevant data privacy or data governance policies for Ms. Booker's ECE program (e.g., the Head Start Program Performance Standards).⁴ The school district staff should be knowledgeable about FERPA rules regarding data privacy and can provide information to Ms. Booker about any requirements that must be met before linking the datasets. Ms. Booker may want to discuss the plan to link data with families in her program and seek their input. Depending on the privacy policies that are applicable to her program, she may also need to ask families to sign a consent form to share their children's data with the schools. This also may be a good time for Ms. Booker to have a larger discussion among staff and families about data privacy issues and the program's policies and procedures to protect privacy. If there is a data governance entity or a research partner, Ms. Booker may want to include them in these discussions.⁵

³ More information about FERPA is available on the U.S. Department of Education website: <http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html>

⁴ The Head Start program and school district may also want to review other existing privacy guidelines that might be applicable to their project, such as the Children's Online Privacy Protection Rule (COPPA) or the Children's Internet Protection Act (CIPA), which relate to children's data that are shared over the internet. More information about COPPA is available on the Federal Trade Commission website: <https://www.ftc.gov/tips-advice/business-center/guidance/complying-coppa-frequently-asked-questions>. More information about CIPA is available on the Federal Communications Commission website: <https://www.fcc.gov/consumers/guides/childrens-internet-protection-act>.

⁵ Two other resources in this series provide additional, relevant information on these topics: *Data Direction 1: Creating and Managing a Data Governance Entity* and *Data Direction 2: Discussing Data with Families*.

4. Identify the specific data to be shared.

Data elements are individual pieces of information that are collected on children, families, staff, and programs (such as child assessments or staff qualifications). Ms. Booker's staff and the school district staff will need to talk specifically about the exact data elements they each need. Not all of the schools' or program's data will need to be linked, so it is important to identify which data elements are necessary to answer the questions. For the example questions under step 1 above, some of the necessary data elements include total score on a kindergarten entry literacy assessment, score on a third grade reading assessment, which literacy curriculum children experienced in preschool, date of enrollment and attendance in the ECE program, and relevant background characteristics such as children's home language.

When identifying which data elements to link, both parties need to agree how children will be identified in each data file to ensure that each child in the ECE data file is matched correctly with their information in the school district data file. If Ms. Booker and the school district assign the same unique ID number to children, then this identifier could be used to match children across the data files. If a unique ID number cannot be assigned, they need to use some basic information about each child to appropriately match the information from the ECE program with the data from the school. Child name, date of birth, gender and address could be used to match child-level data across multiple datasets.

5. Develop and document a process for data linking.

All of the key steps in the data linking process should be described in a data sharing agreement that is signed by both parties. The data sharing agreement could include information like the names of specific files to be linked, the particular data elements in each file to include, the process for sharing and storing data that maintains the privacy of children and families, how the data will be linked, how often data will be linked, and who will have access to the data files. Ms. Booker and the districts can write a data sharing agreement once they have discussed and decided together how data will be shared in a way that appropriately accounts for these issues. The districts might have an existing data sharing template that can be modified to meet the needs of this project. For ideas about appropriate modifications, Ms. Booker can look at data sharing agreements that have been developed by other organizations, like those available through the Privacy Technical Assistance Center (PTAC).⁶



6. Develop a long-term data management and reporting plan.

When Ms. Booker and the school district staff talk about sharing data, they may want to develop a long-term plan for the data before they link and analyze their data. Where will the data be stored and how long will it be kept? Once they are done using the linked dataset, what will happen to the data? Ms. Booker and her district partners also may want to talk about any plans to produce a public report. If there will be a public report, they may need to discuss how best to report the data to protect the confidentiality of the children, particularly if they plan to report information on subgroups of children. They will also want to agree to both review any draft reports before they are made public.

7. Merge the data.

Once Ms. Booker and the school district staff have identified the data elements to be linked and developed a data sharing agreement and data management plan for linking and managing the data, the data from both organizations will need to be merged into one unique dataset. Both parties should agree upon who will do the merging when they develop and document the process for linking data. For example, the technical work of merging the data may be done by Ms. Booker's staff with restricted access to school district data files or the district partners may require that Ms. Booker send her data for

⁶ The Privacy Technical Assistance Center (PTAC) in the U.S. Department of Education has more information about data sharing agreements. The data sharing checklist available on their website is a helpful resource in developing a new agreement. That checklist as well as other data privacy resources can be found on the PTAC website here: <http://ptac.ed.gov/early-childhood-data-privacy>

their own staff to merge and de-identify before sending back. If a third party research partner is engaged, that partner may do the merging and provide both parties with a de-identified merged dataset. Typically, the merge will include statistical matching of two or more datasets based on either a unique ID that is assigned to each child, or other information about individual children such as first name, last name, date of birth, and mother's last name.

8. Analyze the data.

Once the data are merged, Ms. Booker can analyze the merged data file to answer her questions of interest, including whether certain subgroups of children are better prepared at kindergarten entry than others and/or more likely to be reading at grade level by the end of third grade. The type of data analysis needed to answer Ms. Booker's questions of interest may include some simple statistics like averages as well as some more complex analyses. Although it is possible that Ms. Booker may have a staff member who can help her with these analyses, she may need to partner with an outside expert or research organization to do this project. Support from a research partner could help Ms. Booker answer more sophisticated questions like, "How are factors like length of enrollment in the ECE program (e.g., one year vs. two), ECE program attendance, and family characteristics (e.g., parent marital status; parent employment status; family income level) related to kindergarten and third grade achievement?" A research partner will also be able to help Ms. Booker identify the limits of her data, what types of analyses she *cannot* do, and what additional data she would need to answer more complex questions.⁷

9. Annually review the partnership, data sharing agreement, and data management plan.

If Ms. Booker and the school district work together for a few years to link their data, then they may find it useful to annually review their collaboration and data linking effort to determine whether any changes are needed to their written agreements and plans. These revisions could include making changes to the type of files exchanged, the data that are shared (i.e., sharing more data elements or fewer), or the frequency with which the data are linked. If any data security or confidentiality concerns arise, they will need to work together to resolve the issues, and this should be considered under steps #3, 4, and 5 above.

Summary

These action steps highlight the type of work needed to link ECE program data with school district data. ECE program directors may find it useful to review these actions steps before discussing a possible data linking project with school district personnel. The steps included in this resource are by no means inclusive of everything needed to develop, implement, and maintain a data linking project. ECE program directors are encouraged to consider the unique needs of their program as well as their school district's openness to data linking when adapting these action steps.

⁷ More information about developing research partnerships is available in another resource in this series, *Data Direction 3: Engaging an External Research Partner to Help Use Linked Data*.

Acknowledgments

Child Trends is grateful for the financial support of the *Building Capacity to Use Linked Early Childhood Administrative Data* project, provided by the Office of the Assistant Secretary for Planning and Evaluation, in partnership with the Office of Planning, Research and Evaluation, in the U.S. Department of Health and Human Services. The goal of the project is to help early care and education programs build their capacity to link and use linked administrative data for program improvement and research purposes.

We would like to thank the federal staff who provided importance guidance throughout this project: Lindsey Hutchison and Kimberly Burgess, Office of the Assistant Secretary for Planning and Evaluation; Christine Fortunato, Office of Planning, Research and Evaluation; and Theresa Rowley, Office of Head Start.

We also appreciate the contributions of the project's expert panel, whose members dedicated time and provided advice to strengthen the resources developed as part of this project. Expert panel members include: Leigh Bolick, South Carolina Department of Social Services; Craig Detweiler, Georgia Department of Early Care and Learning; Emmalie Dropkin, National Head Start Association; John Fantuzzo, Penn Child Research Center, University of Pennsylvania; Vanessa Rich, Chicago Department of Family and Support Services; and Amanda Schwartz, National Center on Health, Amanda Schwartz Consulting.

We extend a special thank you to the early care and education program staff who provided examples of using linked data, explained the opportunities and challenges in linking data, described how they linked data and how it helped inform continuous quality improvement, and talked with us about their programs' data practices.

Suggested citation: Friese, S., Epstein, D., & Maxwell, K. L. (2016). *Data direction 5: Linking data to understand children's academic progress from early care and education through elementary school*. Report #2016-35. Bethesda, MD: Child Trends.

This report is in the public domain. Permission to reproduce this document is not necessary.

This and other reports sponsored by the Office of the Assistant Secretary for Planning and Evaluation are available at <https://aspe.hhs.gov/reports> and other products from the *Building Capacity to Use Linked Data* project are available at <https://aspe.hhs.gov/building-capacity-use-linked-data-program-improvement-and-research-initiatives>.

© Copyright 2016 by Child Trends, Inc.