Navigating Data Systems When Integrating Home Visiting Data

Dale Epstein

Introduction

At state and local levels, home visiting programs may use one of several mechanisms to collect and store data from families and clients who participate in home visiting services. States that are interested in integrating home visiting data with other early childhood data[1] must often navigate the different ways in which home visiting data are stored and accessed and understand whether or how those systems are able to interact with other early childhood data systems.

The purpose of this resource is to 1) discuss common ways home visiting data are stored at the state and local levels and 2) provide recommendations for how data integration leaders can navigate these data storage systems when integrating home visiting data with other early childhood data.

The State-level Home Visiting Integration with Early Childhood Data Systems (SHINE) project aims to support states in integrating their home visiting data with other early childhood data. SHINE is a project of the Early Childhood Data Collaborative (ECDC), which focuses on the development and use of coordinated state early care and education data systems. This resource was funded by the Heising-Simons Foundation.

Systems for Storing Home Visiting Data

Home visiting data are collected and stored in a variety of systems. These may include proprietary data systems developed by a vendor, homegrown databases, basic electronic spreadsheets, or records stored on paper. Since home visiting data are stored in many different ways, state leaders interested in integrating these data across programs and with other early childhood data will need to consider how to navigate across these systems. For instance, leaders should understand how each method of storing data affects the capacity to integrate and use the data. We briefly describe typical ways that home visiting data are stored.

Proprietary data systems and software as a service

A common way of storing home visiting data is in a proprietary data system, which is typically developed and owned by a for-profit vendor for specific data collection, storage, and reporting purposes. For instance, Penelope and ChildPlus are both proprietary data systems, although they operate differently and have different terms and conditions.[2] Many national home visiting models require or suggest that states or local implementing agencies (LIAs) use certain proprietary data systems, which allows the national home visiting models to use the same process to access the breadth of data being collected, stored, and reported across multiple sites. Similar to the national

[1] Early childhood is the time of child development from prenatal through age 8, with most programs targeting children from birth to age 5. The early childhood system is a set of policies, approaches, and services that are delivered through existing systems, such as education (e.g., pre-K), health care (e.g., immunization), or social services (e.g., subsidies to offset the cost of child care).

[2] Examples of data vendors and data systems are classified based on our knowledge of these systems, but individual data vendors may describe themselves using different terminology.

models, states or LIAs that have multiple sites implementing the same home visiting model may require the use of a proprietary data system for the same reason.

Because these proprietary data systems are designed to serve specific purposes, such as complying with federal or model reporting requirements, the home visiting data that states can access from these systems are often limited to existing reports or data export features. Thus, requesting modifications to meet other needs may involve significant costs to states. Further, vendors of these systems sometimes have contracts that restrict access to individual-level data or limit the availability of identifying information, making it challenging for states to both access the data and integrate it with data from other early childhood programs.

Home visiting data are also often stored through software as a service (SaaS). This type of application can also be a proprietary data system. SaaS applications typically include software that is a ready-made product, but that can be modified to meet the needs of the user. (This type of software is typically a service hosted by a data vendor or third-party provider, and an agency or organization pays to access via the internet or a cloud-based platform on a monthly basis.) Many SaaS applications are built for the general public to be used in a variety of ways and can be adapted for specific uses. For instance, Efforts to Outcomes and Salesforce are examples of SaaS that are cloud-based services with customizable options.[3]

Homegrown data systems

Some states, LIAs, and/or local programs have developed their own systems to house home visiting data, often referred to as homegrown data systems. A homegrown data system is a customized database developed for a specific program, such as an individual home visiting program or for a set of programs that are funded similarly or use the same home visiting model. These systems are commonly developed to address a specific local, state, or federal reporting need or requirement. The complexity of homegrown data systems can range from having basic data collection and reporting capabilities to allowing data leaders and users to interface with the system.

State agencies, LIAs, or local programs sometimes develop homegrown data systems to collect and store their home visiting data as an alternative to a proprietary data system or SaaS. An agency may prefer to develop a homegrown system for several reasons. For instance, the agency may already have a data system in place that staff and data leaders know how to use, or the homegrown system may represent a more cost-efficient option than paying for a proprietary data system or SaaS. On the other hand, some of these homegrown data systems serve as a supplemental system—that is, home visiting programs enter data into both a proprietary and a homegrown system. Programs may do this because they have difficulty accessing data from the proprietary data system that a home visiting model requires them to use, so a second data system allows them to use data in a way that meets their needs. In addition, home visiting programs may actually want a separate system to track data elements not housed in the other system—for example, those related to state, local, or foundation activities such as program management—or to create custom reports not available in the other system. Depending on their level of sophistication and capabilities, homegrown data systems vary greatly in how readily their data can be accessed and integrated with other data systems. Quality and timeliness of data may also be issues, depending on how systems' data are cleaned and updated.

Electronic spreadsheets

While many states, LIAs, and local programs now have online, electronic data systems, some home visiting programs still collect their data using electronic spreadsheets (e.g., Excel spreadsheets or csv files) stored on a computer or shared drive within an organization or agency. While electronic spreadsheets can allow more flexibility in making changes to home visiting data entry and storage—especially compared to some data systems that might not have a way to add additional data elements

---

[3] Examples of data vendors and data systems are classified based on our knowledge of these systems, but individual data vendors may describe themselves using different terminology.

beyond the required data entry fields—electronic spreadsheets often do not allow for more complex forms of data analysis or reporting.

Data integration leaders will need to explore several issues when considering how best to integrate home visiting data currently stored on individual spreadsheets with other more complex data systems. For instance, data stored on electronic spreadsheets may need to be modified so they can be exported through secure transfer processes to a statewide or LIA data system. This process includes modifying variable names so that they are consistent with other data sources or systems. In addition, data leaders will need to understand how the electronic spreadsheet data are collected, tracked, and stored. For example, it is important to know whether data are stored over previous data so that the history cannot be tracked and whether there is a record of all changes and updates made to the data, including fixes to data entry errors. It is also important to know how data cleaning is done; this includes knowing how data entry errors are fixed (i.e., manually or with an automated process).

If a state or LIA is interested in integrating data and currently stores data on electronic spreadsheets, they should consider one of two options: 1) Ensure the data are cleaned and tracked in a way that is compatible with the other data system; or 2) consider moving to a secure online data platform with the capabilities to communicate with other early childhood data systems.

Paper forms

Finally, some individual home visiting programs or LIAs use nonelectronic methods to collect data; these data are then recorded and stored in locked filing cabinets or storage rooms. Often these programs lack the resources or data-savvy staff needed to collect and store their home visiting data in electronic systems. Storing data in paper format can make it difficult to ensure the confidentiality and security of data and add challenges in reporting these data or integrating them.

To integrate local programs' home visiting data stored in paper form within state-level systems, state leaders must determine a way for these data to first be entered into an electronic data system. State agency staff must then work with these local agencies and organizations to determine how to link these data with other data systems, including how to modify variable names to match the naming conventions of the other data systems. Doing so may add several additional steps to the data integration process, creating both additional burdens for local and state staff and an opportunity for new data entry errors. However, not all home visiting data stored in paper form must be entered into an electronic system. One way that state data integration leaders can reduce the burden of entering data from paper into electronic format is to only enter those data that are most important and useful— in terms of integration—into electronic systems and leave the remainder in paper form.

Recommendations for Navigating Data Integration Challenges

Once data integration leaders understand what type(s) of home visiting data they want to integrate, as well as the types of systems in which these data are stored, it becomes important to understand more about the opportunities or constraints for linking data across various data systems. We recommend the following considerations for state leaders beginning to integrate home visiting data across different data systems.

- **Identify where data are stored for home visiting programs of interest**. Data integration leaders will need to first determine which home visiting data they want to integrate--and, in turn, identify which types of data systems currently house these data. Once they understand what types of systems house the home visiting data of interest, leaders will be able to plan their integration of data across different types of data systems.[4]

---

[4] For more information about how to develop an inventory of home visiting data to integrate with other early childhood data, see *Identifying Home Visiting Data to Integrate with Other Early Childhood Data available* at https://www.childtrends.org/publications/identifying-home-visiting-data-to-integrate-with-other-early-childhood-data.

- **Partner with data owners.** Data integration leaders should also work closely with the owners or data managers of the source data systems early in the data integration process. Data system owners and vendors can explain how their system works, guide data integration leaders on what information they need to know and help them access or export the data. Integration leaders should inform data owners or managers of the goals and purposes of data integration.

- **Be flexible**. Prior to integration, data integration leaders may want to develop several options for integrating data across data systems. Sometimes, a preferred approach for data integration may not be feasible once more information is gathered about how and where data are stored; thus, it is important for data integration leaders to remain flexible with their approach. Leaders should ask questions to determine the most feasible integration strategy: Can two or more systems directly connect with one another? Does the data system allow for sharing only exported data? Who may be available to assist with sharing data securely?

- **Consider new ways to collect and store data if needed to support integration.** If current data systems do not allow for data integration, leaders should consider how data can be moved to other systems. For example, if data are stored in spreadsheets or in paper form, leaders could move to an online data storage system to ease the integration process.

- **Review contract language.** When switching to a new data system, data integration leaders should examine the contract between the data vendor and agency to understand what is permissible related to data access and integration. For instance, if the contract language specifies that the vendor owns the data, data integration leaders may request that the originating organization or agency become the data owners, or an authorized user. Changing data ownership or allowing the origin agency to be an authorized user allows more flexibility to use the data for additional purposes. Leaders may also consider the length of the contract. If there is a financial penalty for closing out a contract with a current vendor, leaders may decide to wait until the contract has ended before transitioning data to a new system.

- **Determine the costs.** When planning for data integration, leaders should understand the financial cost of linking home visiting data across different data systems. Costs can range greatly, depending on how and where the data are stored, since they include costs for individual home visiting programs to aggregate, prepare, and share their data with a state system or LIA. For instance, data integration leaders should consider the following:

    - What would it cost the state, LIA(s), or local program(s) to have the data entered into a new system if they are currently stored in an electronic spreadsheet? What if the data are currently only on paper files?

    - Would the state incur any costs associated with modifications to current data systems (either proprietary data systems or homegrown systems) to facilitate data integration or extraction?

    - Once costs are determined, how will the integration process be funded at the state level?[5]

- **Data integration leaders should work with peers from other agencies in their state.** As they plan data integration efforts, leaders should determine whether other organizations or agencies are working on or interested in early childhood data integration efforts. By connecting with others involved in similar efforts, leaders can leverage existing resources, hardware, or software. At the very least, sharing knowledge and collaborating with others engaging in similar work is beneficial.

---

[5] For more information about identifying funding sources for data integration, see *Strategies for Financing the Integration of Home Visiting and Early Childhood Data Systems* available at https://www.childtrends.org/publications/strategies-for-financing-the-integration-of-home-visiting-and-early-childhood-data-systems.

Conclusion

States, LIAs, and home visiting models and programs collect, store, and report their home visiting data in a variety of ways, making it challenging to integrate home visiting data with data from other early childhood programs. States and organizations interested in integrating these data should first identify how and where home visiting data are stored and then develop a plan for integrating home visiting data across systems. The recommendations in this resource can provide a starting point for data integration leaders to work with vendors, data owners, and managers to determine the best ways to plan and implement their efforts to integrate home visiting and early childhood data.