FORESIGHT LAW+POLICY

JUNE 2022

THE IMPORTANCE OF MODERNIZING TECHNOLOGY IN DEVELOPING EARLY CHILDHOOD INTEGRATED DATA SYSTEMS

BY ELLIOT REGENSTEIN

Foreword by Dean Folkers, the Council of Chief State School Officers

INTRODUCTION

For more than a decade, states around the country have been working to create early childhood integrated data systems (ECIDS). These systems bring together information from multiple agencies, helping policymakers to gain a comprehensive understanding of a fragmented early childhood ecosystem. For ECIDS to be successful requires technology capable of integrating data and supporting timely analysis – and then human systems for managing the data integration that are responsive to the needs of the early childhood community. Ultimately, the goal of this work is to improve outcomes for children and families. More effective data systems can play a meaningful role in supporting improved outcomes.

Until recently, the technology supporting ECIDS development was a major impediment. The processes needed to integrate data within on-premises servers were – and still are – cumbersome and time-consuming, making it very hard to produce timely reports based on integrated data. Now, however, cloud computing allows states to integrate and analyze data much more efficiently and at a lower cost – making it easier to meet the operational and analytical needs of early childhood stakeholders. But cloud technology is just one element of a modernization approach, which needs to include substantial attention to the human capacity of state government and the early childhood field. This paper will highlight the importance of modernizing state capacity in developing ECIDS. The paper is in four parts:

- ***A brief summary of why having an ECIDS is critical to early childhood systems;***
- ***An explanation of what the cloud is;***
- ***An analysis of why meeting the goals of ECIDS requires using the cloud; and***
- ***A discussion of some of the practical challenges states may face in building cloud-based ECIDS, and some strategies for addressing those.***

It is important that early childhood policymakers understand the importance of using modernized technology in developing an ECIDS. Newer technology can enable faster production of and access to information, offer more options for analysis, provide sophisticated privacy and security benefits, and can be cheaper and safer to maintain. As states design their data integration approaches, they should make every effort to leverage readily-available cloud technology, rather than investing (or reinvesting) in on-premises infrastructure.

This paper draws on the wisdom of numerous experts, who are listed in the appendix. In some cases their insights were collected through interviews; in other instances, they provided comments on a draft of this paper. Some experts contributed in both of those ways. In many instances this paper anonymizes comments from experts; experts were told that their comments would be anonymized, allowing them to speak more freely. In one subsection an expert is quoted directly, and that text was shared with that expert prior to publication to confirm accuracy.

Table of Contents

Table of Contents continues

Table of Contents

GLOSSARY

TERM	DEFINITION
Cloud computing	"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."[1]
Early Childhood Integrated Data System (ECIDS)	"An early childhood integrated data system (ECIDS) collects, integrates, maintains, stores, and reports information from early childhood programs across multiple agencies within a state that serve children and families from birth to age eight."[2]
Ed-Fi	"The Ed-Fi Data Standard is the set of rules for the collection, management, and organization of educational data that allows multiple systems to share their information in a seamless, actionable way."[3]
ELT (extract, load, transform)	"ELT is a process that extracts, loads, and transforms data from multiple sources to a data warehouse or other unified data repository."[4]
ETL (extract, transform, load)	"ETL is a process that extracts, transforms, and loads data from multiple sources to a data warehouse or other unified data repository."[5]
Data enclave	"A data enclave is a secure network through which confidential data, such as identifiable information from census data, can be stored and disseminated."[6]
Data engineer	"Data engineers are responsible for finding trends in data sets and developing algorithms to help make raw data more useful to the enterprise."[7]
Data integration	"Data integration is a … type of data sharing that involves record linkage, which refers to the joining or merging of data based on common data fields. These data fields include personal identifiers, such as name, birth date, social security number, or an encrypted 'unique ID' that is used to link or join records at the individual level."[8]
Data lake	"A data lake is an unstructured data repository that contains information available for analysis."[9]

Glossary continues

Data lakehouse	"[A] data lakehouse is built to house both structured and unstructured data."[10]
Data scientist	Data scientists "identify relevant questions, collect data from a multitude of different data sources, organize the information, translate results into solutions, and communicate their findings in a way that positively affects business decisions."[11]
Data warehouse	"[A] data warehouse pulls together data from many different sources into a single data repository for sophisticated analytics and decision support."[12]
On-premise server	"An on-premise server is a physical, on-site server that a company must manage and maintain individually."[13]
PaaS	"Platform as a service" describes a complete environment for analytics that includes both infrastructure and tools.[14]
Patch	"A patch is a software update comprised code inserted (or patched) into the code of an executable program. Typically, a patch is installed into an existing software program. Patches are often temporary fixes between full releases of a software package."[15]
Use case	"A programmatic use case describes the information needs in response to state early childhood priorities. This includes refining the specific key programmatic policy question(s) the analytic tool could answer, the purpose for integrating data, and to inform the planning of the integration of data elements from contributing programs."[16]

FOREWORD

BY DEAN FOLKERS
Education Data and Technology Director, Council of Chief State School Officers

For years, states have been working on the transition from having information to using information. The shift can be a particular challenge in the early learning field, where key data exist within different state agencies. States have wrestled with the complexity of bringing together raw data from multiple sources, and then using it effectively to improve child outcomes. But states are starting to think differently about the work, and this paper supplies insight and consideration to innovative approaches to a longstanding problem.

Early childhood education is incredibly important to the overall success of the education system. A condition for success to occur is better information about what specific services children receive and the impacts of those services. And to have that better information will require states to invest in modern technologies – ones that increase security, ensure privacy controls, and create better opportunities for sophisticated analysis. The historic approaches, often manual in nature, simply have not worked as well as they should, and different approaches are needed.

Technology alone will not lead to necessary changes. States must build a shared culture and develop relationships among various parts of the early childhood sector. That work will be most successful when states have a clear and compelling vision for their data use in early childhood and engage a wide range of stakeholders who can see themselves in that vision. The work of developing and using integrated data systems is not easy and there are obstacles. Those challenges each have solutions. Change can be difficult in any organization, but the cost of avoiding change is simply too high for states to continue paying. States can embrace a new era of information technology and find that they have better access to useful intelligence – and pay less for that insight.

The work starts by developing a "decision agenda." What are the important decisions to make with better information? This work requires state level leadership, but it is not just for state government and should involve strong partnerships with community-level and program leaders. While the state has an essential role, communities and providers do too, and their contributions are critical to the success of an integrated data system.

Once states have a start on their decision agenda, they can build the human systems needed to generate the knowledge needed to inform that agenda – and acquire the technology needed to support those human systems. Those human systems may look different than what states have done in the past, but that can be a good thing. States are often spending too much human capacity right now on maintaining legacy systems that are not meeting the community needs; breaking out of the cycle of technical debt is a key step toward the early childhood system states deserve.

Over time, states on this journey should partner with and learn from each other. These processes are not always linear, and there will be bumps in the road along the way. This paper intends to help states get started on the journey and supply insight into finding the smoothest path possible.

EXECUTIVE SUMMARY

State policymakers with responsibility for early childhood services have long faced a serious obstacle: the information they have about the early childhood system is limited and fragmented. New technology creates new opportunity for those state leaders, as it is now more feasible than ever to integrate data from multiple sources to inform decision-making. This report is meant to assist those policymakers in understanding what's possible, and in managing the process of developing a more effective Early Childhood Integrated Data System (ECIDS).

The report starts by explaining why integrating data are so important to early childhood policymakers, practitioners, and parents. Early childhood data is typically spread across multiple state agencies, and integration is needed to give states a coherent sense of what experiences young children are actually having. An ECIDS brings together information from multiple early childhood programs, allowing for analysis that supports policymakers, practitioners, and parents.

To date, state systems have not been well designed to produce useful data. A lack of resources has led to inadequate support for providers, outdated technology, mismatches between program needs and existing processes, and limited information. Moreover, the people responsible for collecting data often see very little benefit from the data.

Addressing these issues starts with a process to identify the data needs of the field. An ECIDS will only be successful if stakeholders see it as helping them to do their jobs, and those stakeholders need to be engaged in the system's design. Some states have already done meaningful work in this area, and started to build a stronger culture of data use.

Once states are ready to start building the infrastructure to integrate data, they will be well-advised to leverage cloud technology. Cloud computing can speed up the process of integrating data, allowing states to load data into a data repository – often referred to as a "data lake" – without having to first go through time-consuming transformation processes. In the cloud states can manage their data more effectively, creating access protocols to ensure security and conducting data transformations as needed to address the state's operational needs.

The cloud also makes possible the use of more sophisticated analytic tools, which states can leverage to maximize the impact of the data collected. The cloud is flexible, allowing states to pay only for the storage they need – generally at a very competitive cost.

But realizing the benefits of the cloud requires putting in place new kinds of oversight systems and human capacity, which can be a challenge in the state government context. Strong leadership will be needed to ensure that states have a shared agenda for data use, role clarity for key implementers, adequate resources to support the work on a sustainable basis, and a strong demand for the data produced. Some of the changes leaders may need to manage include:

- ***How the state structures its interagency data governance;***
- ***How the state acquires and manages technology infrastructure, including ensuring that sensitive data are held securely;***
- ***How early childhood programmatic staff oversee the programs for which they are responsible;***
- ***How agencies build trust and work together to integrate data;***
- ***How the state develops capacity for data science and analytics;***
- ***How the state manages access to sensitive data; and***
- ***How the state supports communities to use data effectively.***

Moreover, states should consider strengthening their analytic capacity, to make sense of the information produced. Key decision-makers may be overwhelmed by a cascade of new raw information without dedicated analysts to make sense of it.

All of these issues are navigable. This report provides guidance for state policy leaders who see the value of integrating early childhood data, and want to do so in a practical and cost-effective manner. If state leaders build an ECIDS that actually produces useful information, that information can be used to improve services for young children and their families – the ultimate goal of any ECIDS.

Why Early Childhood Integrated Data Systems Matter

A. The Fragmentation of the Early Childhood Field

The early childhood field has historically been a fragmented one. Unlike K-12 education, early childhood is not compulsory in either direction – there's no requirement that it be offered to parents, nor are children obligated to participate. In many states core early education and care programs—including state pre-k and child care—are administered by different agencies.[17] The federal-to-local Head Start program bypasses state government altogether, although the federal government funds a Head Start collaboration office in each state.[18]

A handful of states have now created standalone agencies focused on early childhood.[19] In states without a lead early childhood agency, though, there are often two or more agencies that play a key role in delivering early childhood services:

- ***Child care subsidies have been provided by the federal government for decades, with the Child Care and Development Fund dating back to 1990.[20] Many of the agencies administering those funds are social services or human services agencies, along with some education and workforce agencies.[21]***
- ***Pre-k is typically seen as an educational program, and in most states is overseen by the state education agency.[22]***
- ***Home visiting is generally administered by human services or family services agencies.[23]***
- ***Special education for preschoolers is generally administered by the state education agency, but special education for infants and toddlers is overseen by a mix of education, health, and human services agencies.[24]***

To understand early childhood outcomes and to design thoughtful early childhood policies can therefore require collaboration across multiple state agencies and key stakeholders. This is particularly true given the high interest in understanding the long-term impacts of early childhood supports; data about those impacts is often held by agencies that do not have direct engagement in early childhood.

It can be hard for states to develop a coherent approach to early childhood. In many states, there is no single senior official who is empowered to act on behalf of the system as a whole, with line authority over its programs. Each agency may also have very different cultures and hold different relationships. Even in states that have developed a single early childhood agency, it can be difficult to create a unified culture.[25] While states have attempted to develop clear and comprehensive visions for the system, those can be challenging to implement and sustain.

The ability to produce meaningful data is heavily impacted by administrative fragmentation. Each early childhood funding stream will typically have its own operating data system – also referred to as "management systems" – that providers use to track and report data for compliance purposes. Those data systems are often incomplete, and providers often collect additional data that they do not report to the state. And whatever data those management systems have has remained siloed in most states, as most states lack an overarching way to integrate data from multiple management systems. As noted above, an ECIDS "collects, integrates, maintains, stores, and reports information from early childhood programs across multiple agencies within a state that serve children and families from birth to age eight. "[26] The goal of ECIDS is to bring together data from multiple programs, giving states a comprehensive overview of their early childhood system.

B. Early Efforts to Develop an ECIDS

Pioneering work on the development of ECIDS was conducted in Pennsylvania Governor Ed Rendell's administration, from 2003 to 2011.[27] As part of a larger set of governance reforms, the state worked to bring together data from multiple sources.[28] This work meaningfully advanced the state's ability to produce useful reports that informed decision-making.[29]

The federal government then began to see integrating early childhood data as a priority. The bipartisan reauthorization of Head Start in 2007 created new state advisory councils that were required to "develop recommendations regarding the establishment of a unified data collection system for public early childhood education and development programs and services throughout the State."[30] In 2011 and 2013 federal investment through the Early Learning Challenge helped multiple states advance their work to integrate data.[31]

But when the federal money went away, efforts in many states stalled. The work done to date had not established ECIDS that could produce data rapidly enough to meet stakeholder needs – and so in many cases the work did not have enough of a constituency to spark continued investment. A 2018 survey by the Early Childhood Data Collaborative found that most states had a very limited ability to link data across programs.[32] That remains the case today.

C. How an ECIDS Works

Each early childhood service typically has its own statewide operating data system that regularly collects data from the field – and then produces reports based on the data. These statewide operating data systems are typically focused on the information needed for the state to ensure compliance with its oversight requirements, and in some instances providers collect additional data for their own purposes. An ECIDS takes the content from multiple statewide operating data systems and brings them together into an integrated space, helping states to build their understanding of the system as a whole.

Importantly, the ECIDS is not itself an operational one-stop shop system. States are working to build integrated enrollment systems that allow families to sign up for multiple programs through a single portal; those efforts are extremely valuable, but they are different.[33] An ECIDS is not focused on the transactional data tasks common in the management systems—such as determining eligibility, enrolling children, recording attendance, case management, or capturing staff credentials—but instead on warehousing and analyzing those and other data to support point-in-time and longitudinal analyses.[34] These analyses can inform decision-making by policymakers, practitioners, and parents.

One core aspect of an ECIDS is a matching process that allows states to match records across databases. In some cases states extend a single ID across multiple programs; in other cases they develop processes to match different IDs from different programs. These processes are not necessarily unique to the ECIDS, as states will have other initiatives to connect data across agencies – for example, tracking high school students through higher education and into the workforce. But matching processes are critical in early childhood, where understanding service patterns depends on knowing which children are participating in which programs. Because data collection can be inconsistent, inaccurate, and uncoordinated, there must be a process to determine which children are which.

For example, let's say that a girl named Emma Faulk was born September 9, 2018, and lives with her mother Hannah Bruce at 2316 Jones Way. She is enrolled in state-funded pre-k. If an identical name, birthdate, and address appears in the records of child care enrollment, we can be reasonably sure that it's the same child. And we can also be confident that she is not the same as a boy born the same day named Noah Jones who lives at 87 Dyson Boulevard. But what if there's a girl named Emma Bruce with the same birth date and address? Or a girl named Emma Faulk with a listed birth date of September 19, 2018, and an address of 2316 Jones Avenue—are those two different people, or are data entry errors obscuring the fact that it's the same girl?

Many states developed their matching processes years ago. Fortunately, those efforts can generally be leveraged as states adopt newer technology. Indeed, they provide a helpful starting point to ECIDS development. The technology of the process for integrating data from multiple sources is described below in "How the Cloud Works and Why it Matters."

For present purposes, the important thing to note is that the accuracy of the data produced by the ECIDS is directly dependent on the accuracy of the data in each underlying data management system. One benefit of establishing an ECIDS is that it can shed new light on the data in those underlying systems, creating a cycle of analysis and improvement that is beneficial to all of the underlying systems. This is important because in many early childhood data systems the underlying data are incomplete or inaccurate, for a number of reasons:

- ***Lack of support for providers.*** *Providers are given very limited support with the process of data entry. Many providers operate with very limited resources, and have not received sufficient training—or ongoing support—on the data entry process.*
- ***Hard-to-use and outdated data systems.*** *The state operating data systems into which they are entering their data are often outdated, and use technology that provides a poorly designed user experience. Design choices by the state can make it less likely that providers will be able to use the system effectively. In some cases this may be because the management system was originally designed for some other purpose, and then was pressed into service as a management system because it was easier for the state. (This problem is particularly acute for the many providers who utilize multiple funding streams, and may be burdened with separate data entry obligations for each of them.)*
- ***A mismatch between program needs and existing processes.*** *The state's operating processes are badly designed, and the management system reflects that poorly designed process. In some cases states blame the technology, when the reality is that there is no technology that could cover for the fact that the underlying processes and policies are not user-friendly. One informant noted that sometimes states design data systems based on badly-designed bureaucratic processes—and then because the data system is expensive and difficult to update, the data system becomes an obstacle to improving the underlying process.*

- ***Limited information.*** *The state's management systems may be collecting limited amounts of data, making it harder to match across systems.*
- ***The data doesn't help the people collecting it.*** *The data aren't ultimately useful to the providers themselves. Because the data are used for state compliance purposes and do not benefit the people entering it, they have little incentive to use their limited bandwidth to ensure its accuracy.*

Many state data sets are impacted by one or more of these interconnected problems. Importantly, many of these problems are caused by a lack of resources for information technology; states frequently lack the political will to budget for regular technology updates, or to pay competitive salaries for staff with the needed skills. And these are just the challenges in producing accurate raw numbers; there are additional challenges inherent in the process of making sense of those numbers, which we will turn to later.

D. Designing an ECIDS to Meet the Needs of the Early Childhood Field

For an ECIDS to have a positive impact requires designing it to meet the needs of the state's early childhood system. Ultimately, the ECIDS should be a "club" that people want to join; if it provides direct operational benefits to state agencies, providers, and others, those partners will see it as a good use of time, energy, and capital. To ensure that an ECIDS is meeting the needs of stakeholders, it should be developed using specific use cases and key questions that inform system design.[35] In prioritizing the questions state leaders should ask: What are the questions that, if we could answer them, would allow us to change our policies and practices in a manner that benefits children and families?

States should have their own process for identifying prioritized questions and use cases. There are national resources that can provide a useful starting point to that process.[36] But involving the early childhood stakeholder community in the process of prioritizing key questions is an important part of making sure that the ECIDS meets the state's needs – and that stakeholders are bought into the ECIDS process. Multiple informants talked about the value of mobilizing the early childhood community to inform the design of the ECIDS.

One reason to bring stakeholders together is to build understanding of how different participants in the system expect different benefits from data integration.[37] Conversations about system design need to be sensitive to that, and build understanding among stakeholders about the diverse range of needs and expectations within the system.

Another reason to be inclusive in the system design process is that many early childhood providers are eager to have better information. Head Start providers, child care providers, and schools all operate in a dynamic marketplace—and in the case of private providers, the ability to navigate that marketplace successfully is an existential imperative. Having those providers at the table can help to ensure that the ECIDS will support their work, and can also increase the sense of urgency about the importance of integrated data.

States generally realize that at the beginning of an ECIDS development process, they need to focus on basic enrollment data. States typically know how many children are enrolled in each individual program they offer, but do not know how many children are enrolled in multiple programs (and in what combinations).[38] States also do not have a sense of the overall supply and potential demand in different communities, including children who are eligible for services but not receiving them. Answering questions like these will often be a necessary first step toward answering any more sophisticated questions about the early childhood system.

For the most part, states that have started on this work have focused on some similar questions— including questions focused on eligibility, access, service quality, and impact:[39]

Eligibility

Which children and families are eligible for which services?

Access

Which children and families are actually accessing services? Where are there gaps?

Service quality

Are the services meeting the needs of families? Fulfilling the expectations of funders?

Impact

What are the child outcomes that follow from various combinations of services?

These categories can provide a helpful point of reference for state work, but should not be treated as a boundary or limitation. States also may not have data to answer all of these questions at the outset of an initiative to develop an ECIDS; in those instances, states can get started with the data they have, and think about how to build out the system over time.

The unfortunate reality is that for years, state early childhood ecosystems have operated without being able to answer these important questions (and many more).[40] Prior to COVID-19, states and providers got used to operating in a world without the information they really needed to make good decisions. During the pandemic, however, the early childhood field was pushed into chaos, and the ability of states and providers to respond to that chaos was severely compromised by the lack of information.[41] The impact of that inadequate response was felt by families around the country.

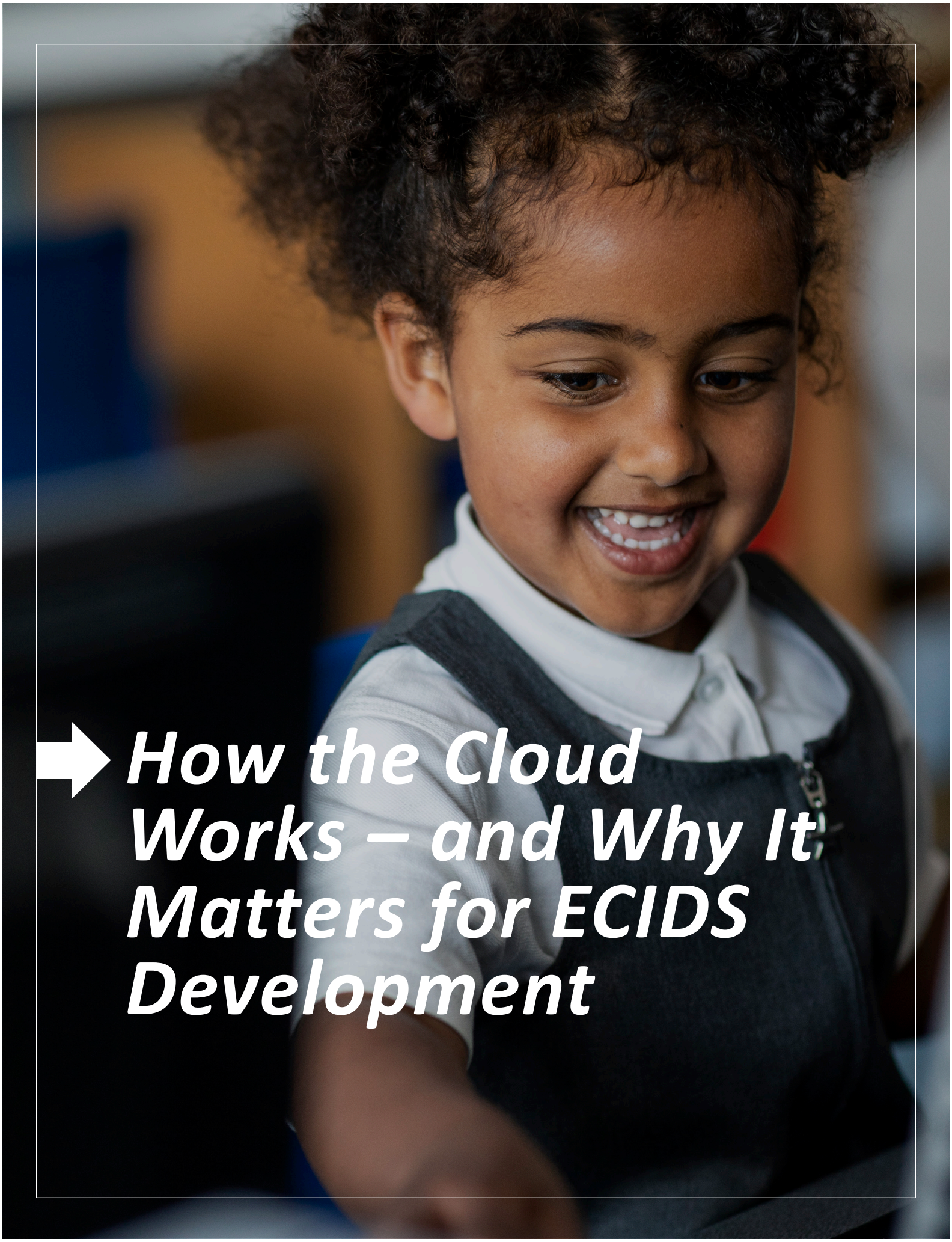
As one informant put it, states are providing billions of dollars in services while having only the most rudimentary understanding of what impact those expenditures are having – or even which children the state is serving. While ultimately there are many questions states should be able to answer with an ECIDS, states are well-advised to start with some of their most fundamental questions; prioritizing some critical questions to start with can be a key step in building a culture of data use. While it can be helpful to have a longer list of questions to inform longer-term system development, states increase the likelihood of success if they can be very clear on what outcomes matter most to them.

One informant emphasized the importance of making cross-agency data integration truly longitudinal, with the ability to connect a child's experience over time. This informant also explained that child characteristics can vary over the course of time; for example, a child may go in and out of low-income status multiple times in the birth-to-five period. The ability to track that kind of data will make an integrated system truly longitudinal, rather than just a linked series of snapshots.

In some instances, states have already developed dashboards and data reports that are well accepted by the public; for those states, improving the back-end infrastructure will help to make those dashboards and reports more accurate and complete. One informant talked about the importance of developing new products that meet the needs of the field, including dashboards that are updated on a more regular basis. This informant said that every time the state had created a new data product, it was being used by the field – and leading the field to request additional data products.

Multiple informants emphasized the importance of the culture of data use. If key decision-makers are demanding data – and then actually using it to inform decisions – then an ECIDS can come to be seen as a core part of the state's early childhood system. If they aren't using it, then it won't be seen as beneficial enough to justify its cost. Building that culture will require leadership and vision.

Now that the urgency of providing good information about basic functions of the system is better understood, there may be a higher chance that states can find the political will to develop and use an ECIDS. And fortunately, developing an ECIDS should be easier than it was in years past. The arrival of cloud technology and new techniques in data integration opens up possibilities for states that simply did not exist a decade ago, when the last wave of major ECIDS progress occurred. The next section will focus on what cloud technology is and how it works – and why that matters for ECIDS development.



How the Cloud Works – and Why It Matters for ECIDS Development

Cloud technology offers many advantages for states considering an ECIDS. States seeking to reap the advantages of integrated data should consider those potential advantages in developing their system. This section looks at what some of those benefits are, and why they matter in the ECIDS context.

A. Overview of the Cloud

In September 2011, the National Institute of Standards and Technology (NIST) – part of the U.S. Department of Commerce – adopted a definition of cloud computing.[42] This definition provides the basics of what cloud computing is, and what it does.

According to NIST, "[c]loud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."[43] Essential characteristics of cloud computing include on-demand self service, broad network access, resource pooling, rapid elasticity, and measured service.[44] The cloud can be delivered through multiple service models and deployment models.[45] The term "Platform as a service" describes a complete environment for analytics that includes both infrastructure and tools.[46]

Cloud computing offers all of the analytical products and services that would be available on-premises, but eliminates the need to buy and maintain hardware – including the constant upgrading of software, hardware, and security protocols. All of those functions are performed by a cloud vendor. Cloud products are generally state of the art, with services offered evolving rapidly. Moreover, when new services become available they can be easily acquired.

Indeed, one advantage of the cloud is that it is "pay as you go." With on-premises servers, states are responsible for building and maintaining expensive infrastructure. In the cloud, however, there are no fixed infrastructure costs; instead, states pay only for the services they need, which can grow over time based on demand. States are thereby able to avoid the "sunk costs" that come with on-premises infrastructure.

B. Key Functions of the Cloud

There are two functions of cloud computing that are particularly important for ECIDS development: the streamlined process of integrating data from multiple sources, and integrated analytical tools that make analysis and product deployment simpler.

i. Integrating Data

As described above, by definition an ECIDS brings together data from multiple sources. This is one area in which the cloud is vastly superior to on-premises technology.

It is important here to distinguish between data linkage and data integration. Data linkage is about connecting data across two data sets; that is, establishing a link between A and B. Data integration, by contrast, takes A and B and integrates them to create a new set of data – an integrated dataset C that incorporates information from both A and B. This dataset C can contain information that answers questions that neither A nor B could answer on their own.

Data integration has been very difficult for states. With on-premises servers, data often undergo significant cleanup before they are loaded. That process takes time and money, and must be conducted regardless of whether the data will ever actually be used after being loaded. Indeed, this whole process has historically been a huge impediment to having a useful ECIDS – because by the time data have been loaded into the servers, they are already likely to be out of date. And that is before anyone has even initiated the potentially time-consuming process of trying to extract data. Moreover, resource constraints have sometimes led states to limit how much data they load to the server – in some cases by simply not including certain data sets in their data warehouse.

Because the cloud provides immense scale at extremely low cost, it is easier to integrate entire systems, using an integration process called ELT (extract, load, transform). This means that states no longer have to undergo the traditional ETL (extract, transform, load) process whereby they must pre-process their data to align to specific file specifications or storage limits before integrating it. By loading data first – without having to worry about cherry-picking data due to storage or cost issues – states can access more source data in more formats (structured, semi-structured, unstructured) with more processing power and analytical tools. This in turn can help make data transformation more efficient, because the raw data can be analyzed up front, which then informs the work of developing transformations.

States generally are loading data into the cloud on some regular cadence, which over time builds up a repository of data – sometimes referred to as a "data lake" – that can be used for multiple purposes. The state can then maintain the data lake in the cloud, and process it as needed for specific use cases. Given that, ELT can provide substantial advantages, because the transformation can be tailored to the state's needs at any given moment. While in some circumstances it will make sense to transform the data first, the state is better off having the option of ELT.

Multiple informants emphasized that for data integration, states should focus on building data sets that meet a variety of needs. Rather than just doing a series of one-off projects, states should take an approach that maximizes efficiency and impact. The ELT integration process creates flexibility in meeting diverse needs by providing quicker access to more data, while allowing for processing logic to emerge from the diverse use cases.

Multiple informants also emphasized that the ease with which states can move data into a central repository should not be confused for the ease with which that data can actually be integrated once it gets there. The cloud does have advantages for data integration, but that integration does not happen automatically; states will need dedicated capacity to conduct the integration work. (One informant expressed skepticism that state efforts to build a single repository would be successful, arguing that it was unlikely states would be able to successfully restructure the "pool of crap" that they would get from individual management systems.)

The cloud's ability to integrate data dramatically increases the odds that an ECIDS will be successful. Developing an ECIDS should be about more than producing new data reports; it should be about changing the culture of data use in early childhood policy, and creating an expectation that major decisions will be informed by timely data. Creating space for integrated data may be a new concept for some information technology (IT) leaders, who may be more used to cleaning up data in its original management systems. But by creating a storehouse of integrated data in the cloud, the state can work with the data much more efficiently. And it is essential to maintain that storehouse of integrated data, because those are the data that can be used to provide answers to the state's key questions.

ii. Analytic Tools

Cloud technology has the most modern tools for analytics, and those tools are evolving rapidly. That means that in the cloud, states can conduct analyses that would have been much more difficult previously. This gives states the opportunity to leverage their data for operational benefit in a way they could not have before.

To maximize the benefit of the tools, states will need to develop stable code and business rules that allow for the analysis they need. In the cloud, analytic tools are better integrated, and they work faster. Large-scale data science platforms can give states more analytic options, particularly with large data sets.

A benefit of the cloud is that it makes things easier for both data engineers and data scientists. While their roles can overlap, one simplified way to think about it is that data engineers are the people who prepare the data, and data scientists are the people who make meaning of the data. The cloud makes for simpler handoffs between these staff, with more transparency between the two – and both benefit from the increased computing power available in the cloud.

Some states have already made headway in utilizing analytic tools, and states could choose to share their progress in a way that maximizes efficiency – and allows states to learn from each others' experiences. Rather than writing custom code, states can just borrow from what is already available in the field in the open-source community. Moreover, new packages and libraries are deployed on an ongoing basis, and can generally be re-used or purchased for less than it would cost to develop them in-house.[47]

Another benefit of the analytic tools in the cloud is that they are easily adaptable. This means that the technology is more likely to meet the evolving needs of the state. One informant emphasized the importance of flexibility, which will allow the ECIDS to meet the ever-changing needs of the field.

Another key consideration in some states is using data to report on how different populations are using disaggregated data; some states are seeking to move to improved self-service products that allow users to explore the data, rather than answering a series of ad-hoc requests.

This approach to analysis is not just taking states' previous approach and doing it faster – this is potentially a whole new way for states to approach the work of analytics. That is particularly important in the early childhood space, where the state is engaged with a dynamic and complex market involving a range of providers and funding streams.

Common Education Data Standards in the Launch of an ECIDS

One centralized cost driver in the startup of traditional ECIDS has been the development of common definitions and common standards. The processes states have used to do this have been time-consuming and complicated. Cloud-based ECIDS make modern data integration techniques like ELT simpler, which means the transformation work is not necessary on the front end. It also allows the process of ECIDS development to launch faster, giving the project a greater chance of success.

Many K-12 leaders have invested in the use of Common Education Data Standards (CEDS),[48] and will want to see them used in the ECIDS. CEDS can and should be useful to the buildout of the ECIDS. The key is not to insist on using them to clean up data before it is put into the data lake – but to use CEDS as a reference for cleaning up the data already in the lake.

CEDS, like all standards, create limitations. Those limitations have value for bringing order out of chaos. But they also intentionally exclude certain information. In the K-12 context, the CEDS tradeoff – excluding some information to bring order to other information – has in many instances proven to be a net benefit. But the tradeoffs look very different in early childhood.

If the state attempts to get all early childhood programs in conformance with CEDS before integrating data, it will lose valuable time for potentially minimal gain. Given the current condition of data in early childhood, the process of mapping raw data to the CEDS before incorporating it in a data model is highly likely to result in mismatches. The process is much more likely to be successful if the data are incorporated into data models first. Over time, underlying management systems may be able to adjust to the CEDS, but the state should allow that to be an iterative process – not a precondition for launching the ECIDS.

Indeed, one benefit of using CEDS is that it may help the ECIDS connect to the state's longitudinal data system. One key use case for early childhood policymakers is the long-term impacts of early childhood services, and tracking those impacts is only possible if the ECIDS can connect upward to K-12 data – as well as higher education and workforce data. Using CEDS in that ongoing process can be helpful, in a way that using CEDS as a barrier to entry is not.

C. Scalability, Flexibility, and Speed

In addition to the improved ability to integrate data and conduct analytics, key benefits of using the cloud include cost savings, simplification, resilience, and elasticity. State Chief Information Officers say that scalability and flexibility are the biggest factors driving cloud adoption.[49] The fact that so much data can be stored so cheaply – and that it is easy to grow or shrink the state's capacity as needed – is a key advantage cloud computing has over on-premises servers. Informants talked about how the cloud allows states to integrate data much more efficiently, and then use data much more nimbly – with analytic tools available nowhere else.

Another major benefit is speed. States using the cloud have the opportunity to integrate data – and produce analytic reports – far faster than they do when using on-premises servers. This means that the officials responsible for overseeing state early childhood systems and programs can get much better information for ongoing decision-making, which should allow them to use funds more effectively. Integrating and analyzing data on a rapid cycle does require dedicated staff capacity, but the underlying technology allows those staff to have a greater impact. Ideally states can iterate and learn on an ongoing basis; state governments may not have operated that way in the past, but the cloud opens up the possibility of doing so.

The competitive commercial nature of cloud technology also means that states aren't trapped with a particular vendor. While many states establish ongoing relationships with a preferred cloud provider, the pay-as-you-go nature of the service means that they should never be trapped with that vendor; if that vendor's performance is not satisfactory, there are plenty of competitors waiting to take their place. And because there are no costs tied up in built infrastructure it is easier to move from one vendor to another – although any transition from one vendor to another will come with some costs and operational barriers.

Cloud-based ECIDS provide an affordable, scalable and centralized data integration point for other operational platforms, even decades-old management systems some states still maintain. If the state is able to update its older systems, the cloud infrastructure of the ECIDS will adapt. Developing an ECIDS does not necessarily require fixing the underlying management systems – although the weaknesses in those systems it exposes may put pressure on states to do so. Replacing management systems is generally far more expensive and labor-intensive than building an ECIDS, and raises completely different operational challenges.

Multiple informants said that another advantage of the cloud is improved security, an issue discussed further under "Addressing the Practical Reality."

For these reasons, governments have been moving toward increased use of the cloud – and COVID-19 may have accelerated those efforts.[50] But wise leaders have sought to base the migration on operational need – which makes sense, given the history of state technology initiatives that spent a lot of money without fulfilling the state's mission. Being "cloud smart" means using the cloud when it meets the mission.[51]

D. Access Protocols

Another advantage of the cloud is that it allows for much more sophisticated identity and access management.[52] These tools give administrators the ability to control at a very granular level which users have access to which data. That includes the ability to define different roles for different kinds of data users, and to assign each of those roles appropriate levels of access. That makes it easier for the ECIDS administrator to ensure that personally identifiable information is not accessed improperly – which in turn should be reassuring to the agencies contributing data to the ECIDS.[53]

In a cloud environment, access control protocols are key to the state's compliance with privacy laws. Privacy laws generally govern who can properly access sensitive data, with strict limitations to protect individual privacy. Access controls are the method by which states ensure that only authorized personnel are able to see that sensitive data, while still allowing data to be used in permitted ways that help the state support improved services to children.

When data are stored in the cloud, properly credentialed analysts and researchers can explore the data – in some cases after it has been de-identified – to figure out how it might be used most effectively. Indeed, one advantage of the cloud is that the state does not need to export data sets to researchers; instead, it can use access protocols to allow those researchers to explore data within the secure space, in a manner that protects individual privacy. (In many instances sensitive data on the cloud is stored in a Secure Data Enclave, which offers protection against misuse.) This allows the state to shift away from "data sharing," in which the state actually exports its data to third parties; with the cloud, the state maintains control of the data at all times.

In traditional models, researchers have had to submit queries for data that may or may not be available. State staff then needed to spend time determining whether the data requested even existed. If they existed, that then generally triggered a time-consuming process to organize the data and respond to the request. Creating exploratory space in the cloud where researchers and analysts can see what is possible – without taking data out of the secure environment – has multiple advantages:

- ***It reduces the burden on state staff.*** *State staff will still need to manage the process of credentialing outside users, as that process still needs to be controlled; moreover, in many instances the data in the cloud may need to be "de-identified" before credentialed users can explore it. These tasks, however, can be far less time-consuming than the work previously needed to search through the data sets.*
- ***It allows for faster and more efficient use of state employee time.*** *Instead of stabbing in the dark, researchers can take a preliminary look at available data and then formulate requests that are far more precise – and realistic. State officials can then clean up only the data needed to fulfill the request, and provide information in a timely manner. If the data requestor has data access rights in the ECIDS – for example, as a researcher based at a state university – some or all of the cleanup burden can be shifted from state employees to the person making the data request.*

- ***It provides clear records of access and usage.*** *When users are granted access to the cloud environment, all of their behavior is logged and subject to the controls the state wishes to impose (such as preventing data download). This not only ensures that data will be confined to the environment, it also creates clear records of usage that can help improve offerings or investigate suspicious behavior. This stands in contrast to traditional models, where approved users are often given access to sensitive data and expected to secure it themselves. This self-security can be a major risk to data security, as approved users can still make simple data security mistakes like using insecure platforms (like email or public wifi) – or even losing entire laptops on which sensitive data is stored.*

In sum, the access protocols possible using the cloud can save the state money and allow researchers faster access to data, all while improving security.

E. Cost Containment

The cloud's ability to expand rapidly and cheaply is well suited to the development of an ECIDS. The development of an ECIDS is likely to be an iterative process, given the number of agencies involved; some may be early adopters, whereas others may join later. The flexibility and cost of the cloud benefit this process.

i. Startup Costs

Centralized data infrastructure can be the biggest expense states have in maintaining an ECIDS with on-premises technology. The cloud has much lower infrastructure costs, both with regard to physical infrastructure and staffing.[54] This allows states to get started on an ECIDS with much less up-front investment.

But the work of starting up an ECIDS will require a meaningful investment of time. Some of that time will have to come from state agency personnel; some amount of it also typically comes from outside vendors. Either way, that time has a real cost. Some of the startup activities states can expect to spend time on include:

- **Defining the key use cases that will inform system design;**
- **Putting in place a governance framework to manage integrated data space;**
- **Inventorying existing data held in program management systems, and determining the technical needs for importing the data into the shared space;**
- **Assessing, cleaning, and integrating the data to deliver on the prioritized use cases; and,**
- **Developing the analytic tools and reporting mechanisms needed to make the data useful.**

These costs will typically be borne largely by state government; some amount of them are in-kind (typically the agency personnel time), and others will require dedicated outlays of funds. In some instances philanthropies have helped to shoulder some of the one-time startup costs associated with ECIDS development.

ECIDS development does not take place on a blank slate, and the path toward establishing a cloud-based ECIDS will be shaped by the state's overall approach to technology.[55] If the state is already embracing the cloud, that should generally make the process easier; the state will already be familiar with the benefits of the cloud, and may have capacity focused on managing it. This may require the ECIDS to adapt to that larger context, but that should be feasible.

Indeed, for an ECIDS it may well be more efficient to partner with other data integration initiatives, such as work to integrate education data longitudinally or projects to integrate health data cross multiple agencies.[56] That may allow for the costs of maintaining data infrastructure to be shared across more agency partners. There will likely be some functions of the ECIDS that will require unique and dedicated capacity – particularly when it comes to analytics. But the ECIDS' needs for data storage and extraction will likely not be unique, opening the possibility of making the ECIDS part of a broader data integration effort. The risk to ECIDS of this approach is getting lost in the shuffle if other related initiatives are higher-profile or better resourced.

If the state is not already embracing the cloud, the ECIDS could be a worthwhile pilot project. That, however, brings with it all of the risks of early adoption. There are also other potential issues if the state has not already begun to move toward use of the cloud – including personnel, technology, and process issues. But if early childhood stakeholders are clamoring for better data, that may help key leaders appreciate the importance of developing the infrastructure for effective data access, analysis, and use.

The speed with which cloud projects can be launched can have political benefits. One informant noted that a benefit of the cloud is that with older technologies, an integrated data system simply could not be launched successfully within a three-year grant period – or even a four-year gubernatorial term. Now that the barriers to launch have been reduced, the development of an ECIDS is something that a governor can launch early in a term and then manage to completion before facing re-election; that improves the odds of consistent leadership and support throughout the project.

While the opportunities presented by a cloud-based ECIDS are real, the state's investment in legacy systems is potentially a political issue, as states have been known to chase their sunk costs.[57] If a state has already spent a lot of money on expensive infrastructure it may be reluctant to abandon that investment, even if the switch would produce meaningful savings on an ongoing basis. Moreover, if there are technology vendors attached to the existing infrastructures, those vendors will likely lobby the state to not make any changes.

A significant amount of the investment in state data systems has come from the federal government. There have been a number of initiatives that have included a specific focus on building capacity, such as the Statewide Longitudinal Data Systems program.[58] In other instances federal funds for specific services authorize states to use a portion of those funds for data system development. States can develop data systems as a permissible use of federal programmatic funds.[59]

ii. Carrying Costs

Not only does the cloud allow for an ECIDS to launch faster and with less expense, it reduces the carrying costs to the state over time.

One major way it does that is by not requiring a centralized staff to oversee physical infrastructure. Investing in a server puts the state on the hook for the ongoing maintenance of infrastructure whether it's being fully utilized or not; in the cloud, states can pay only for the services and computing power they are using. Because the cloud is pay-as-you-go, the state can better tailor its spending based on its operational needs. This flexibility can significantly reduce ongoing expenditures. It can also save states the effort of having to constantly purchase upgrades for hardware and software, and deal with patches in between updates.

Another way the cloud reduces costs is by reducing data sharing costs for individual agencies. One of the major burdens on agencies participating in an ECIDS is cleaning up data before sending it to the centralized repository. As explained above, this labor-intensive cleanup process is done in the speculative hope that at some point the data will be useful in the ECIDS. But the fact that the cleanup takes so long means the data may no longer be useful by the time it is even sent to the central repository. The cloud dramatically reduces this ongoing expense and delay; moving entire raw data sets into a data lake takes far less work, and accordingly far less money for the contributing agency.

This reduction of the cleanup burden is also potentially important in any effort to include Head Start providers. There will still be complex issues to navigate with regard to accessing Head Start data in the first place. But eliminating the need to clean up data before adding it to a central repository does lower at least one of the practical hurdles to integrating Head Start data into an ECIDS.

To be clear, a cloud-based ECIDS still has carrying costs – the state will still need a team of people to oversee its data integration and analysis space, including to manage access protocols and security. As described above, that oversight team need not be focused solely on early childhood; the state may see efficiencies from having ECIDS be part of a larger data integration infrastructure. If that is the case, there may still be a need for analytic staff focused solely on early childhood, and the early childhood agency or agencies may be responsible for sharing a portion of the overall system carrying costs.

Particularly if one-time grant funds are used to address startup costs, the state will need to identify sources of funds to cover ongoing expenses.[60] These costs will include not only cloud platform expenses, but the personnel needed to manage the system and conduct analytic work. In some instances integrated data systems charge user fees, or receive contracts for analytic services.[61]

One potential challenge for states is that the traditional model of funding for data projects has been a substantial up-front investment – often a federal grant – followed by a defined set of ongoing costs, some of which can still be paid for through that initial grant. For states that are used to this model, moving to a monthly pay-as-you-go approach may present some budgeting challenges. While overall the cloud approach should be less expensive at the outset and over time, the fact that it is such a different approach may require some engagement to educate the relevant fiscal staff – and legislative appropriators. States with centralized IT agencies may also have established funding mechanisms for that agency that would be impacted by a substantial change in technology approach, and states are regularly wrestling with how to structure funding for those agencies in a way that truly incentivizes them to deliver high-quality services in a cost-effective manner.

Technical Debt

A number of informants talked about the idea of "technical debt." With legacy systems, many states are paying staff to spend significant time each year to manually generate annual reports – rather than using technology on the front end to automate the process.

The ongoing expense in staff time can add up quickly, and some informants emphasized that automating those functions can allow staff to be used for more productive purposes. According to one informant, a state agency they used to work for estimated that it cost $1 million each time the state applied for a federal grant, given the staff time needed to navigate inefficient systems; by modernizing its data infrastructure, the state could realize significant savings on each future application. Other informants told similar stories.

The idea of technical debt goes beyond establishing a cloud-based ECIDS. Using outdated technology can be a false economy, with higher maintenance costs and greater technical risk. Informants emphasized that states would often be better off spending what it takes to keep their systems up to date, particularly in the context of an overall operation that could be made more efficient.

Importantly, shifting to the cloud does not eliminate the potential for technical debt. State choices relating to management practices and underlying codes can still create technical debt, even if the underlying technology offers the possibility for greater efficiency. The concept of technical debt is still an important one, though, and should be taken into account as the state plans and implements its ECIDS.

Addressing the Practical Reality

In many states, ECIDS development has proceeded in fits and starts. When federal money is available — through State Advisory Council startup grants, Early Learning Challenge grants,[62] or the Preschool Development Grant-Birth to Five[63] — many states have been able to build some ECIDS infrastructure. Some states have also leveraged federal Statewide Longitudinal Data System grants. But when federal money hasn't been available and states have had to rely on their own funds, progress has generally been slower. And even in many of the states that have spent significant time and energy in developing an ECIDS, the operational impact to date has been limited.

In many instances, states have built up a set of human capacities based on processes that rest on older technologies. There are some systems and practices that are needed regardless of what technology a state is using; accordingly, improved technology gives states the opportunity to develop a useful ECIDS – but actually implementing an ECIDS using that technology will require changes in how the people in state government go about their work. That is difficult, and the challenges should not be underestimated. Implementing an ECIDS will require actively engaged leaders at the state level – and will create or change the roles of technology staff, analytic staff, program staff, and community leaders.

A. Leadership

i. Getting Started

Ideally leaders will see a move to cloud technology – and the improved processes that can be built around cloud technology – as a change that benefits their work, and the work of people around them. A consistent theme in conversations with informants about ECIDS development is that leadership matters a great deal. Several informants emphasized that ECIDS is just a means to an end, with the end being improved outcomes for children and families through improved early childhood services. If the state is truly committed to improving outcomes and services it will likely need an ECIDS to do so, but if it lacks that commitment then an ECIDS is not truly necessary. So if senior officials are investing political capital in the development of an ECIDS as part of an overall effort to improve the early childhood system, that gives the work a chance to succeed; if they are not investing that political capital, the project is far less likely to make headway. Governors, legislators, and agency heads are among the leaders who can play this role.

There are a series of actions leaders can take to expedite the process:

- ***Asking questions that model the need for data use, and emphasize the importance of data in the decision-making process – including the need for continuous improvement and higher quality;***
- ***Pushing for the development of a shared agenda for data use that informs the ECIDS design, while acknowledging that different sectors will have different data needs;***
- ***Establishing role clarity for all of the participants in the ECIDS process;***
- ***Ensuring that resources are available to support the ongoing work;***
- ***Clearing out bureaucratic hurdles where possible; and***
- ***Using the data that are actually produced, and becoming a voracious consumer and disseminator of the ECIDS' output.***

Having leaders who understand the importance of ECIDS and fulfill some or all of those functions has been a significant asset for states in the development process. Moreover, one informant pointed out that it helps to surround strong leaders inside state government with a supportive community outside state government – including advocates, providers, and philanthropy.

Role clarity is important in the development and maintenance of an ECIDS. In many states there is a centralized information technology agency – but in some states authority for information technology is dispersed.[64] Regardless of the state structure, a successful ECIDS will involve both programmatic staff and IT leaders.[65] IT leaders must ensure that infrastructure decisions are coherent and aligned with larger IT strategy, while programmatic staff must be included as the "user voice" that can advocate for particular abilities and workflows.

Because these projects involve so many agencies, building trust will be key. Agencies may not have a culture of trusting each other with sensitive data, nor can it be assumed that they trust the centralized IT agency. It will take strong leadership and content knowledge to develop project plans that build trust among the core players while continuing to respect their individual prerogatives. Ultimately an ECIDS will only succeed if each participant sees it as being in their own institutional best interest.

Some informants said that legislatures can play a valuable role in the process by demanding data, and by passing laws requiring agencies to partner on data integration. While some states have preferred to work through voluntary cooperation, other states have found that having the legislature impose requirements on the executive branch is the best way to move forward. Legislatures also play an important role in appropriating funds for ongoing data use, which is another reason to get them involved.

ii. Ongoing Management

The cloud allows states to adopt entirely new processes and tools for managing and using data. But many state governments are not particularly adept at adopting new processes for managing and using data. That will require capacity, and an intentionally focused strategy.

For the ECIDS to be successful on an ongoing basis will require it to have an interagency governance structure. There is already substantial literature about data governance, which states should take advantage of.[66] These governance efforts will require some amount of capacity, and states will need to build that capacity into their sustainability plan. The governing body will play a role in addressing many of the issues described in this paper.

More than that, though, the state's policy leaders – including agency heads, their top deputies, and the managers running early childhood programs – will need to adapt new ways of management. For years, they have gotten used to compliance-oriented data systems focused on individual segments of a complex market. An ECIDS will produce an entirely new kind of data, and even with analytic support it's not obvious that state managers will be well positioned to use that data in decision-making. As described further below, support for that change management process will be essential to the success of the ECIDS.

States will need to establish a governance and management framework capable of taking advantage of the cloud's power and speed. In some cases, there may be active opposition to that change. In other instances – potentially more instances – there will not be active opposition, but there will be inertia and a lack of pressure to change. Political and policy leaders with a sense of urgency about the need for data will have to use their power and influence to drive a thoughtful process of change; otherwise, the status quo is likely to persist.

B. Technology

i. Technology Staff

The state's current technology workforce may not have the skill sets and expertise needed to work with the cloud.[67] That's not only a practical challenge, it's a political one, as IT employees may see the use of cloud technology as a threat to their livelihoods. Indeed, in some states the financing model of the state's IT agency will create a strong incentive for that agency to favor the technological status quo.

Informants identified numerous personnel roles that states will need to fill. States will need staff who understand technology, who understand business processes, and understand how the data will ultimately be used. One informant referred to "data product management skills" – including the ability to understand data user requirements, build solutions, manage ongoing improvements, and in time manage the transition to new solutions when the need demands.

In addition to staff capacity, the state will need leaders who know how to manage the technology staff to "prevent them from running amok," in the words of one informant. This informant said that technology staff do not think about problems the same way as programmatic staff, and that managers who are sensitive to both worldviews are extremely valuable.

State work rules may be relevant here. In some cases there may be no incentive for current staff to change, as staff may be protected in their job regardless of whether they develop the skills to use cloud technology. If that is the case, the state will have to consider whether those staff can be deployed to other projects, and what that means for the cost of the ECIDS. One informant suggested splitting the staff between developing the future state and managing the legacy systems, and being clear with staff that there is a timeline for retiring the legacy systems – after which their roles will need to be transitioned.

States should also document their practices clearly to provide for continuity when staff turnover. Some informants talked about a struggle with institutional memory; in many cases there is valuable knowledge that lives only in the heads of individual employees, and if those employees leave there may be substantial challenges in transitioning. The difficult part for states is that the more idiosyncratic the knowledge base of the employee, the more likely they are to remain at the state.

State salary practices are a key challenge in the technology sector. Because state government salaries are often not competitive with those of private firms, it can be hard for the state to attract talent with the most up-to-date knowledge. When the state provides training for IT staff to bring their knowledge up to date, that increases the odds that those staff will leave. States are sometimes able to attract highly skilled employees who are committed to public service and therefore willing to accept a lower salary, but those individuals then can get overloaded with work – leading to burnout and turnover. Indeed, in general informants relayed that states frequently underpay their IT staff, which can lead to turnover, which then makes it very difficult to provide continuity on major initiatives. In many instances it might be more efficient for states to pay competitive salaries to keep a talented IT staff that can use technology efficiently to produce actionable data.

One possibility for building state capacity in some states may be to develop public-private partnerships. Private and philanthropic partners may be willing to support fellowships or secondments that add staff capacity to the public sector.[68]

ii. Procurement and Vendor Management

Procurement processes and vendor management have historically been a challenge for state government.[69] Overall, state procurement and vendor management rules and procedures are widely seen as a challenge to developing a successful ECIDS. One informant described hearing "countless horror stories" from state colleagues dealing with procurement issues. For this reason, the informant recommended that to the extent possible states use technology that is as flexible as possible, so as to avoid being stuck with a particular vendor.

A different informant told one of those horror stories. This informant described an RFP that was written to acquire technology services that did not move forward until six years after it was written. By that point, the state's needs had changed – as had the available technology. But amending the RFP would have meant that the procurement could not move forward, so the state went ahead and issued the outdated RFP. One government official emphasized that they tried to build on existing infrastructure in a strategic and iterative manner, for fear that a large data project would inevitably go off track.

States have often been trapped with vendors who develop products that state staff cannot work with independently, leading to an ongoing relationship where the vendor continues to extract significant payments from the state. Some informants did describe positive relationships with reliable vendors, but even those informants acknowledged that the history of failed IT projects in state government provides a sobering context for the development of an ECIDS.[70]

As described above, there are benefits to making ECIDS part of a larger data integration effort – but there are also challenges. Informants noted that some states have centralized information technology offices that view themselves as the key provider of services – and these agencies may have business practices and funding models that cannot be adapted easily to a cloud-based environment. While some central IT agencies have embraced the idea of change, others have not, and they can be an obstacle to the development of an ECIDS. Where that is the case strong leadership will be needed to overcome the resistance.

C. Privacy and Security

i. Establishing Access Protocols

As described above, the cloud allows for more efficient processes for data use and access – but those processes might be unfamiliar to agency legal staff, requiring extra effort to establish.[71] The ECIDS should allow for the creation of entire new data sets that can be used in creative ways. This may be a challenge for legal departments to understand; they may be more familiar with a process based on itemized queries, with legal department signoff for each one. Informants report that legal departments have proven capable of making this paradigm shift, but that requires support from leadership – and a substantial investment of time in educating the staff.

Informants emphasized that policy leaders can set the tone for the legal review process. Informants generally reported that with the right support lawyers can find their way to a place where they understand and sign off on new kinds of data integration systems. One informant emphasized that if lawyers perceive that leaders are determined to get to yes and really want to move forward, they will respond to that, and find the best ways to implement data integration in full compliance with existing privacy laws (which informants universally believe to be possible). But if lawyers perceive that leaders are looking for excuses not to move forward, they will find those excuses.

One informant said that establishing different levels of access to data can help lawyers understand how sensitive data will be protected, and also how data will be useful to stakeholders – giving them a sense of the benefits of the system. Having different logins and permissions can allow for the maintenance of a system with multiple levels of access. Only a select few people ever need to see personally identifiable information, and many users can use data effectively after it has been de-identified.

A different informant said that interagency data sharing can be a struggle regardless of the underlying technology, and expressed frustration with the idea that different legal and contract offices within the same state government could spend months arguing back and forth over relatively small language changes. While most of the issues that arise in these processes could be resolved through some creative problem-solving, not every state legal or contracts office has an adequate supply of creative problem-solvers. This informant said that while relationships among leaders can sometimes help to expedite resolution, having the legislature force action may be a more effective strategy.

ii. Security

Cloud security is generally considered extremely robust, but it does have to be managed on an ongoing basis.[72] There are certain security features that should be built into any interagency data sharing infrastructure.[73] Multiple informants emphasized the importance of having staff with specific skill sets and training employed to manage security.[74]

Jim Reavis, CEO of the Cloud Security Alliance, explained that security is not strictly a technology problem – it's often a problem of processes. He said that the nature of technology use has shifted dramatically, so the thinking about security has to shift as well. Instead of redundancy and hardware, technology use focuses on software – which puts a premium on managing access, and also on having an IT strategy well suited to the technology.

Informants generally agreed that cloud providers are well-positioned to provide the level of security states will appropriately demand. Because cloud vendors are competing for market share in a robust commercial environment, they face significant pressure to have state-of-the-art security; any failure of security would be economically devastating to the cloud provider. Major cloud providers invest significant resources in ensuring that their clients' data is secure – and because of the competitive nature of the market are updating cloud security on a near-constant basis. Cloud providers also typically offer a "government cloud" product that meets federal data security standards.[75]

States will still have to provide dedicated staff to manage security, but those staff will have better security tools at their disposal in the cloud. In many instances on-premises servers use outdated technology that vendors struggle to support – and if states themselves are responsible for updating the security, that can be a challenge given the political realities of information technology funding. Reavis explained that the number of tools available for protecting hardware is dwindling, and that soon states will be unable to get support for older technology infrastructure; he described these older systems as a "ticking time bomb" – adding that maintaining security for older systems will only get more expensive, and the likelihood and consequences of cyberattacks or breaches will only get worse. He said that while states may be inclined to hang on to existing technology as long as they can, they can reduce their risk by updating technology sooner rather than later.

D. Making Data Useful

i. Analytic Staff

While cloud technology makes improved analytics possible, states will need staff capacity to take advantage of the available tools. While many states have data analysts already on staff, many of them are focused on compliance and operational reporting. Very few state agencies have the personnel needed to do the kind of sophisticated analysis the cloud makes possible – and which often requires real expertise in the early childhood field. In addition to having analytic staff in programmatic agencies, some states have created centralized units focused on analysis; multiple informants cited the Kentucky Center for Statistics as a model they would like to replicate in their state.[76]

In particular the state will likely benefit from new data engineers and data scientists. Think of data engineers and data scientists as the chefs, who take the raw ingredients of data and combine them together to make the data digestible to broader audiences. This kind of business intelligence is key, because to date policy leaders have struggled to digest the buffet of raw data with which they are often presented. For that reason, those data chefs will play a key role in the early childhood ecosystem.

Moreover, those chefs do not have to develop all of the recipes themselves; as described above, they can share analytic tools with data chefs from other states.

There are already emerging efforts to leverage early childhood data to improve policy. For example, in Miami an initiative called the Miami-Dade Integrating Data for Effectiveness Across Systems (IDEAS) Consortium for Children uses shared data to analyze child experiences – and environmental factors impacting those experiences.[77] The Iowa Integrated Data System for Decision-Making has looked at systemic issues in early childhood, including child care.[78] Virginia's Link B5 system is meant to help policymakers and providers understand the early childhood care and education landscape, including the quality of experiences children are having.[79] While these partnerships involve major universities, other kinds of partnerships to leverage interagency data are also possible. And improving the underlying technology for data integration would open up new possibilities.

This is also an area where public-private partnerships may be most possible. Philanthropic leaders could fund "data fellows" who work with the state to analyze data. Large firms that hire data scientists could provide some amount of pro bono services, or secondments where staff embed in state government for a period of time. Data scientists are in high demand.[80] Accordingly, state government might benefit from some assistance in acquiring data science capacity.

One informant emphasized that to date state IT operations have not been designed to engage in the kind of "storytelling" needed to make data useful for broader audiences.[81] Moreover, the data systems states use typically collect basic data that leaves out important context; the state's data may give some sense of what happened, but it may say very little about why it happened. Focusing on the story data tell – and going beyond describing what happened to understanding why it happened – is an important but difficult leap forward for states to take.

One informant argued that the analytics infrastructure will be most useful if the state has a clear idea of its "decision agenda" – starting with the initial use cases and building outward from there. It will also be important for states to distinguish between questions that can be answered through analytics, and questions that require a full-fledged research design. Capacity will be needed to answer either kind of question.

Informants noted that state analytic capacity will have to be right-sized for the exercise. Historically, the burdensome nature of state data processes meant that it might take an inordinate amount of time to produce relatively basic results. But with new processes for integrating data, the volume of material that could be produced increases substantially. Having the capacity to manage that process will be essential to maximizing the impact of the state's data investment.

One informant questioned the value of state analytic capacity, noting that providers may be concerned that states will focus on using data to make providers look bad. This informant explained that states have a history of playing "gotcha" with providers, and where that is the case providers may not want the state to increase its analytic capacity. This informant said that in their state the focus has been on making data publicly available to the broader public, and on having a decentralized concept of expertise that focuses on the wisdom of the field. Other informants indicated that the problem of states playing "gotcha" is more about the state's mindset and approach than about its analytic capacity, and that ideally states will take a supportive approach – while building the analytic capacity needed to inform that supportive approach.

Regardless of how much analytic capacity the state has, informants identified independent benefits to making aggregated data sets available to the public. Multiple informants talked about efforts to put data on websites that would allow members of the public to "slice and dice" it their own way to answer their own questions. Having public engagement with the data can help build support for the enterprise of data use, and also can create pressure to improve data quality.

ii. Program Staff

One consistent theme from informants was that the programmatic staff working in early childhood programs are a key constituency in this work, and their resistance can make it difficult for data integration to have its intended impact. From the perspective of these programmatic staff, data sharing can often feel like an added burden for somebody else's benefit. These staff may have a mindset of protecting their data against outside use.

States should make a focused effort to help line staff understand how their work fits into the larger context of an ECIDS – and more specifically, how that might benefit them. Informants said this this work may need to be extremely granular, so that staff have a very specific understanding of how data can be used. They will likely also need training in using data in new ways. Some of these staff may also be concerned that the availability of new data will cast their division's work in a negative light, and may act to prevent expanded data use as a rational form of self-preservation.

Indeed, technological innovation can have all kinds of impacts on line staff. While managers may see value in automating certain functions and redeploying time to more important issues, for line staff those changes can be extremely disruptive. Informants related that if staff are used to doing their job a certain way, they may resist change and be reluctant to cede control. Some informants did identify pockets of success in data use among programmatic staff.

Informants also emphasized the value of having programmatic staff and technology staff working together.[82] One informant pointed out that the two groups "speak different languages" and have different cultures, and that it can take time to build relationships. Another informant explained that technology staff often do not know how to ask for guidance in a way that program staff understand, and the program staff often do not know how to describe what they want in a way that the tech staff can implement. Multiple informants talked about the need to spend time building those relationships in order to ensure the success of data integration projects.

iii. Community Leaders

Data use should not be limited to state government. Multiple informants emphasized the importance of making the analytic capacity accessible to communities and providers, and providing them with data products.[83] One informant suggested having staff who can help facilitate stakeholder exploration of de-identified data, rather than just staff who produce reports on their own. One informant emphasized the importance of focused outreach to the success of any community-based data initiative – with community leaders involved from the very beginning, not just as an add-on at the end.

One informant pointed out that simply providing data to communities does not always mean that those communities use data in decision-making, and that a meaningful amount of process support may be necessary. Communities might ultimately need facilitators and analysts to work at the local level, with support from statewide coaches. If the state does not have the resources to help communities with that support, philanthropic partners may be able to play a constructive role. Indeed, community engagement was described by informants as an ongoing process that will likely benefit from the active involvement of philanthropic partners.

As noted above, the ECIDS should be a club that communities and providers want to join. One approach would be to set up the ECIDS in a way that allows communities and providers to add their data, with defined parameters set for their participation. Communities and providers could be enticed to join by clearly articulated benefits, including the ability to integrate their data with that of the state and other providers (in a manner that protects individual privacy and complies with all relevant legal requirements). This approach can help build a critical mass of useful data from multiple sources, and make the enterprise all the more appealing to communities and providers that are not early adopters.

RECOMMENDATIONS

An ECIDS will only thrive if it is actually being used. Many states have had too many experiences where technology projects ate up substantial resources without producing meaningful information. The technology of an ECIDS is important, but only because it can support changes in decision-making that improves the odds for young children and their families. That commitment to improved service must be a core driver of any ECIDS.

For states that choose to implement an ECIDS as part of their commitment to young children, there are key action steps they need to take:

Engage

the early childhood field to identify prioritized data use cases.

Choose

a cloud-based technology platform that supports the creation of a data repository while enabling improvements in transformation processes.

Adapt

budgeting approaches to account for the fact that the cloud is pay-as-you-go, with a fundamentally different structure for both startup costs and ongoing carrying costs.

Work

with information technology staff to implement new management approaches, including access protocols that protect individual privacy while allowing for appropriate data use.

Develop

analytic capacity that leverages the analytic tools available in the cloud.

Work

with programmatic staff to adopt new processes that leverage updated technology.

Support

community leaders to take advantage of state technology resources.

All of that work will require strong leadership and sufficient resources. But for a state that is committed to improving outcomes for children and families, the investment in an ECIDS can fundamentally change the way it manages its early childhood system.

Appendix

The author is grateful for the many people who made themselves available to be interviewed for this project – as well as the people who assisted in other ways, including by making connections to additional sources and by sharing useful background information. In addition, several people reviewed early drafts of this report, and the final version is better for their efforts. The author's work leading to the publication of this paper was supported by the Bill & Melinda Gates Foundation. The author is solely responsible for the views expressed in the final report, which do not necessarily represent the opinion of any particular interview subject or reviewer, nor do they necessarily represent the views of the Bill & Melinda Gates Foundation.

As noted previously, all interviewees and expert reviewers were told that their comments would be anonymized. In one subsection an expert is quoted directly, and that text was shared with that expert prior to publication to confirm accuracy.

Interview Subjects

NAME	ROLE	ORGANIZATION
Lupita Alcala	Director, Education Policy & Outcomes	WestEd
Benjamin Boer	Data Coordinator	Office of Governor JB Pritzker (State of Illinois)
Kathy Booth	Project Director, Educational Data and Policy	WestEd
Jocelyn Bowne	Director of Research and Preschool Expansion Grant Administration	Massachusetts Department of Early Education and Care
Jason Breslin	Director of Early Learning	New York State Education Department
Jenna Conway	Deputy Superintendent, Division of Early Childhood Care and Education	Virginia Department of Education
Darren Fleischer	Policy Analyst	District of Columbia State Board of Education
Dean Folkers	Education Data and Technology Director	Council of Chief State School Officers
Carolyn Gosse	Director of Link B5	Center for the Advanced Study of Teaching and Learning, University of Virginia
Erin Kenny	Managing Director	Alvarez & Marsal
Christina Krasov	President	CK Impact Strategies

Interview Subjects continues

NAME	ROLE	ORGANIZATION
Jannelle Kubinec	Chief Administrative Officer	WestEd
John-Paul Hayworth	Executive Director	District of Columbia State Board of Education
Sara Mead	Assistant Superintendent of Early Learning	District of Columbia Office of the State Superintendent of Education
Misty Moody	Assistant Commissioner of Early Learning	Tennessee Department of Education
Lilla Pivnick	Strategic Data Project Fellow	Tennessee Department of Education, Office of Early Learning
Tim Norris		Education Research and Data Center, Office of Financial Management, State of Washington
Jim Reavis	Co-founder and Chief Executive Officer	Cloud Security Alliance
Andrew Rice	Chief Executive Officer	Education Analytics
Paul Tearnen	Managing Director	Alvarez & Marsal
Jessica Whittaker	Director of Birth to Eight Initiatives	Center for the Advanced Study of Teaching and Learning, University of Virginia
Vickie Ybarra	Director, Office of Innovation, Alignment, and Accountability	Washington Department of Children, Youth, and Families
John Yeoh	Global Vice President of Research	Cloud Security Alliance

Further assistance was provided by Rachel Anderson, Julia Bloom-Weltman, Tanya Forsheit, Winona Hao, Carlise King, Winnie Li, Abby Potts, Andrew Rice, Chris Strausz-Clark, and Sarah Weber. Earlier drafts were reviewed by Julia Bloom-Weltman, Ben Boer, Missy Coffey, Amy Hawn Nelson, Erin Kenny, Carlise King, Snow Li, Winnie Li, Dave Moyer, Chris Strausz-Clark, Paul Tearnen, and Sarah Weber. The author is also grateful to his colleagues at Foresight Law + Policy for their substantial support at every stage of the project, including reviewing drafts.

Endnotes

[1] Mell, P. and Grance, T. (September 2011). *The NIST Definition of Cloud Computing*. National Institute of Standard and Technology. Retrieved from: https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf.

[2] Coffey, M., Chatis, C., Irvine, S., Sellers, J, & Duarte, S. (2014). *An early childhood integrated data system: What is an ECIDS?* U.S. Department of Education. Washington, DC: National Center for Education Statistics.Retrieved from: https://slds.ed.gov/services/PDCService.svc/GetPDCDocumentFile?fileId=33126.

[3] Ed-Fi Alliance. *Ed-Fi is a Data Standard*. Retrieved from: https://www.ed-fi.org/what-is-ed-fi/ed-fi-data-standard/.

[4] IBM Cloud Education (August 3, 2021). ELT (*Extract, Load, Transform*). Retrieved from: https://www.ibm.com/cloud/learn/elt.

[5] IBM Cloud Education. (April 28, 2020). ETL (*Extract, Transform, Load*). Retrieved from: https://www.ibm.com/cloud/learn/etl.

[6] National Library of Medicine. *Data Enclave*. Retrieved from: https://nnlm.gov/guides/data-thesaurus/data-enclave.

[7] White, S.K. (July 26, 2018). *What is a data engineer? An analytics role in high demand*. CIO. Retrieved from: https://www.cio.com/article/222004/what-is-a-data-engineer.html.

[8] Hawn Nelson, A., Jenkins, D., Zanti, S., Katz, M., Burnett, T., Culhane, D., Barghaus, K., et al. (2020). *Introduction to Data Sharing and Integration*. Actionable Intelligence for Social Policy, University of Pennsylvania. Retrieved from: https://www.aisp.upenn.edu/aisp-intro/.

[9] Transforming Data With Intelligence. *What is a data lake?* Retrieved from: https://tdwi.org/portals/what-is-a-data-lake-definition.aspx.

[10] Marr, B. (January 8, 2022). *What Is A Data Lakehouse?* A Super-Simple Explanation for Anyone. Forbes. Retrieved from: https://www.forbes.com/sites/bernardmarr/2022/01/18/what-is-a-data-lakehouse-a-super-simple-explanation-for-anyone/?sh=25531c216088.

[11] Berkeley School of Information. *What is Data Science?* Retrieved from: https://ischoolonline.berkeley.edu/data-science/what-is-data-science/.

[12] IBM Cloud Education (March 5, 2020). *Data Warehouse.* Retrieved from: https://www.ibm.com/cloud/learn/data-warehouse. A data warehouse includes structured data – in contrast to data lakes, which can include unstructured data. Marr, B. *What Is A Data Lakehouse?*

[13] Munk, D. (March 22, 2019). *Cloud-Based Vs. On-Premise Servers.* Forbes. Retrieved from: https://www.forbes.com/sites/forbestechcouncil/2019/03/22/cloud-based-vs-on-premise-servers/?sh=5d9381ae79e2.

[14] Carey, S. (July 6, 2021). *What is PaaS? A simpler way to build software applications.* InfoWorld. Retrieved from: https://www.infoworld.com/article/3223434/what-is-paas-a-simpler-way-to-build-software-applications.html.

[15] Techopedia. (February 15, 2017). *What does patch mean?* Retrieved from: https://www.techopedia.com/definition/24537/patch.

[16] Morrison, H., Coffey, M., & Sirinides, P. (2022). *ECDataWorks Programmatic Use Case for the Development of an Analytic Tool.* Philadelphia, PA: University of Pennsylvania. Retrieved from https://repository.upenn.edu/cgi/viewcontent.cgi?article=1007&context=ecdataworks.

[17] Fischer, A., Rafa, A., Atchison, B., Regenstein, E., Multon, M., Weyer, M., McCann, M., and Smillie, S. (November 10, 2020). *50-State Comparison: Early Care and Education Governance.* Education Commission of the States. Retrieved from: https://www.ecs.org/50-state-comparison-early-care-and-education-governance/; Education Commission of the States (November 2020). *Early Care and Education Governance Chart.* Retrieved from: https://c0arw235.caspio.com/dp/b7f93000ac9474f0c2524b34b062.

[18] Head Start Early Childhood Learning & Knowledge Center. *Head Start Collaboration Offices*. U.S. Department of Health & Human Services. Retrieved from https://eclkc.ohs.acf.hhs.gov/state-collaboration/article/head-start-collaboration-offices.

[19] Education Commission of the States. *Early Care and Education Governance Chart*.

[20] Office of Child Care (June 23, 2021). *OCC Fact Sheet.* U.S. Department of Health & Human Services. Retrieved from: https://www.acf.hhs.gov/occ/fact-sheet.

[21] Office of Child Care (May 25, 2021). *State and Territory Child Care and Development Fund Administrators.* U.S. Department of Health & Human Services. Retrieved from: https://www.acf.hhs.gov/occ/contact-information/state-and-territory-child-care-and-development-fund-administrators.

[22] National Institute for Early Education Research (2021). *The State of Preschool 2020*, p. 203. Retrieved from: https://nieer.org/wp-content/uploads/2021/08/YB2020_Full_Report_080521.pdf.

[23] Fischer, A., Rafa, A., Atchison, B., Regenstein, E., Multon, M., Weyer, M., McCann, M., Smillie, S. (November 10, 2020). *Early Care and Education Governance: State Profiles.* Education Commission of the States. Retrieved from: https://www.ecs.org/early-care-and-education-governance-state-profiles/.

[24] Early Childhood Technical Assistance Center (September 14, 2020). *Part C Lead Agencies*. Retrieved from: https://ectacenter.org/partc/ptclead.asp.

[25] Regenstein, E. (2020), *Early Childhood Governance: Getting There from Here*, Foresight Law + Policy, p. 65. Retrieved from: https://www.flpadvisors.com/uploads/4/2/4/2/42429949/flp_gettingtherefromhere_061120.pdf.

[26] Coffey, Chatis, Irvine, Sellers, & Duarte, *An early childhood integrated data system: What is an ECIDS?* U.S. Department of Education.

[27] Pennsylvania Historical & Museum Commission. *Pennsylvania Governors.* Retrieved from: http://www.phmc.state.pa.us/portal/communities/governors/.

[28] Sirinides, P. M. (2013). *Pennsylvania's Early Childhood Data Systems: History, Uses & Opportunities.* CPRE Working Papers. Retrieved from https://repository.upenn.edu/cpre_workingpapers/16; Stedron, J. M. (2010). *A Look at Pennsylvania's Early Childhood Data System.* National Conference of State Legislatures. Retrieved from: https://www.ncsl.org/portals/1/documents/educ/paearlychild-stedron.pdf.

[29] Stedron, *A Look at Pennsylvania's Early Childhood Data System.*

[30] 42 U.S.C. §9837b (b)(1)(d)(4).

[31] Jordan, E. and King, C. (2015). *Stacking the Blocks: A Look at Integrated Data Strategies in Rising to the Challenge: Building Effective Systems for Young Children and Families, a BUILD E-book.* Build Initiative, Chapter 7. Retrieved from: https://www.childtrends.org/wp-content/uploads/2015/08/2015-35BuildChap7.pdf.

[32] King, C., Perkins, V., Nugent, C., and Jordan, E. (September 2018). *2018 State of State Early Childhood Data Systems*. The Early Childhood Data Collaborative. Retrieved from: https://www.childtrends.org/wp-content/uploads/2018/09/ECDC-50-state-survey-9.25.pdf.

[33] These data systems also would benefit from using updated technology. Stoney, L. (January 2022). *Bridging the Data Gap: Diverse Delivery Requires 21st Century Technology.* Opportunities Exchange. Retrieved from: https://static1.squarespace.com/static/5f4d7a7ef6c82325c5ec80c0/t/61e73be1a3ef005db06fd478/1642544097841/OppEx_2022_IssueBrf_BridgingDataGap.pdf. For a broader description of state efforts in this area, see Gebhard, B. (2022). *Strengthening Connections: State Approaches to Connecting Families to Services.* Zero to Three. Retrieved from: https://www.zerotothree.org/resources/4309-strengthening-connections-state-approaches-to-connecting-families-to-services?utm_medium=email&utm_source=email_link&utm_content=whats_new_02.28.2022&utm_campaign=Q2_2022_Policy+Center_Resources.

[34] For more on the relationship between intended use and data integration design, see Hawn Nelson, Jenkins, Zanti, Katz, Burnett, Culhane, Barghaus, et al., *Introduction to Data Sharing and Integration*, p. 6.

[35] Cochenour, M., Porowski, S., and Regenstein, E. (October 2013). *Answering Key Questions with an Early Childhood Data System.* SLDS Issue Brief. Retrieved from: https://nces.ed.gov/programs/slds/pdf/IssueBrief_Answering_key_questions_with_an_early_childhood_data_system.pdf; Regenstein, E. (August 22, 2017). *An Unofficial Guide to the Why and How of State Early Childhood Data Systems.* Policy Conversations, No. 7, V 1.0. Start Early, formerly known as the Ounce of Prevention Fund. Retrieved from: https://startearly.org/app/uploads/2020/09/PUBLICATION_An-Unofficial-Guide-to-the-Why-and-How-of-State-Early-Childhood-Data-Systems.pdf; Data Quality Campaign and Foresight Law + Policy (June 2020). *Early Childhood Data Systems: Responding to COVID-19 and Building for the Future*, pp. 1-2. Retrieved from: https://dataqualitycampaign.org/wp-content/uploads/2020/06/DQC-FPL_Early-Childhood-Data-Systems_Responding-to-COVID-and-Building-for-the-Future.pdf.

[36] Regenstein, *An Unofficial Guide to the Why and How of State Early Childhood Data Systems* (Appendix); Cochenour, Porowski, and Regenstein, Answering Key Questions with an Early Childhood Data System, pp. 3-4.

[37] Hawn Nelson, Jenkins, Zanti, Katz, Burnett, Culhane, Barghaus, et al., *Introduction to Data Sharing and Integration*.

[38] Regenstein, *An Unofficial Guide to the Why and How of State Early Childhood Data Systems.*

[39] Sirinides, P., Coffey, M., Matherly, S. (2019). *Community Assessment: A new approach to using community-level early childhood service data.* Philadelphia, PA: University of Pennsylvania. Retrieved from: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1002&context=ecdataworks.

[40] E.g. Ryberg, R., Wiggins, L., Moore, K.A., Daily, S., Pina, G., and Klin, A. (December 12, 2021). *Measuring state-level infant and toddler well-being in the United States: Gaps in data lead to gaps in understanding.* Child Indicators Research. Retrieved from: https://link.springer.com/content/pdf/10.1007/s12187-021-09902-4.pdf; Falgout, MK, and Gibbs, H. (March 23, 2022). *5 Strategies for Equitable Implement of Public Investments in Child Care.* Center for American Progress. Retrieved from: https://www.americanprogress.org/article/5-strategies-for-equitable-implementation-of-public-investments-in-child-care/; Fullerton, J. (October 2021). *Bridging the Gaps in Education Data.* Retrieved from: https://www.aei.org/wp-content/uploads/2021/11/Bridging-the-gaps-in-education-data.pdf?x91208.

[41] Data Quality Campaign and Foresight Law + Policy. *Early Childhood Data Systems*, pp. 1-2.

[42] Mell and Grance, *The NIST Definition of Cloud Computing*; see also *From Cloud First to Cloud Smart*. Office of the Federal Chief Information Officer. Retrieved from: https://cloud.cio.gov/strategy/.

[43] Mell and Grance, *The NIST Definition of Cloud Computing*, p. 2.

[44] Mell and Grance, *The NIST Definition of Cloud Computing*, p. 2.

[45] Mell and Grance, *The NIST Definition of Cloud Computing*, pp. 2-3.

[46] Carey, S. *What is PaaS? A simpler way to build software applications.* InfoWorld.

[47] When developing and deploying code that makes predictions or inferences, states should ensure that they are ethical and unbiased. Edward Dieterle, Beth Holland, and Chris Dede, "The Cyclical Effects of Ethical Decisions Involving Big Data and Digital Learning Platforms," in *The Ethical Use of Data in Education: Promoting Responsible Policies and Practices*, eds. Ellen B. Mandinach and Edith S. Gummer (New York: Teacher's College Press, 2021), p. 206.

[48] Common Education Data Standards. https://ceds.ed.gov/.

[49] National Association of State Chief Information Officers, Grant Thornton Public Sector LLC, and CompTIA (October 2021). *Driving Digital Acceleration: The 2021 State CIO Survey*, p. 19. Retrieved from: https://www.nascio.org/wp-content/uploads/2021/10/2021-State-CIO-Survey.pdf.

[50] Stone, A. (September 1, 2020). *2020 Puts Cloud Computing in Government to the Test*. Government Technology. Retrieved from: https://www.govtech.com/computing/2020-puts-cloud-computing-in-government-to-the-test.html; Government Technology (June 2, 2021). *The State of Cloud in State and Local Governments*. Retrieved from: https://www.govtech.com/cloud-different/the-state-of-cloud-in-state-and-local-governments; Perepa, S. (September 2013). *Why the U.S. Government is Moving to Cloud Computing*. Wired. Retrieved from: https://www.wired.com/insights/2013/09/why-the-u-s-government-is-moving-to-cloud-computing/.

[51] Goldstein, P. (November 30, 2020). *Moving from Cloud First to Cloud Smart for State Governments.* StateTech. Retrieved from: https://statetechmagazine.com/article/2020/11/moving-cloud-first-cloud-smart-state-governments-perfcon; Goldstein, P. (September 25, 2018). White House Unveils New 'Cloud Smart' Strategy. FedTech. Retrieved from: https://fedtechmagazine.com/article/2018/09/white-house-unveils-new-cloud-smart-strategy; Office of the Federal Chief Information Officer, From Cloud First to Cloud Smart.

[52] Strom, D. (April 8, 2021). *What is IAM? Identity and Access Management Explained*. CSO. Retrieved from: https://www.csoonline.com/article/2120384/what-is-iam-identity-and-access-management-explained.html.

[53] The major cloud vendors all provide materials describing their Identity and Access Management tools. E.g. Amazon Web Services, *AWS Identity and Access Management User Guide*. Retrieved from: https://docs.aws.amazon.com/IAM/latest/UserGuide/iam-ug.pdf; Google Cloud, Identity and Access Management. Retrieved from: https://cloud.google.com/iam; Microsoft Azure, *Identity and access management (IAM)*. Retrieved from: https://azure.microsoft.com/en-us/product-categories/identity/.

[54] Consortium for School Networking. *Define On-Cloud Premise Service Strategy.* Retrieved from: https://sites.google.com/site/cosnsend/steps-to-smart-network-design/step-2-define-data-storage-approach; State Educational Technology Directors Association (SETDA). *Networks for the Future.* Retrieved from: https://www.setda.org/priorities/equity-of-access/broadband-imperative/broadband-imperative-iii/networks-for-the-future/; Consortium for School Networking (2018). *Saving Money in the Cloud.* Retrieved from: https://www.cosn.org/wp-content/uploads/2021/09/Saving-Money-in-the-Cloud.pdf.

[55] Regenstein, *An Unofficial Guide to the Why and How of State Early Childhood Data Systems*, p. 21; see also Statewide Longitudinal Data Systems State Support Team (May 2021). *Early Childhood Integrated Data Systems Toolkit: Self-Assessment Tool*, pp. 7, 28. U.S. Department of Education. Retrieved from: https://slds.ed.gov/services/PDCService.svc/GetPDCDocumentFile?fileId=41273.

[56] Zanti, S., Jenkins, D., Berkowitz, E., Hawn Nelson, A., Burnett, T., & Culhane, D. (May 2021). *Building and Sustaining State Data Integration Efforts: Legislation, Funding, and Strategies*, p. 10. Actionable Intelligence for Social Policy. University of Pennsylvania. Retrieved from: https://aisp.upenn.edu/wp-content/uploads/2021/05/AISP-Policy-Report-2021.pdf; King, C. and Perkins, V. (February 2020). *Strategies for Financing the Integration of Home Visiting and Early Childhood Data Systems*, pp. 3-4. The Early Childhood Data Collaborative. Retrieved from: https://www.childtrends.org/wp-content/uploads/2020/02/SHINEbrief4_ChildTrends_Feb2020.pdf.

[57] Government Technology, *The State of Cloud in State and Local Governments*; see also Richard H. Thaler, *Misbehaving: The Making of Behavioral Economics* (New York: W.W. Norton & Co., 2016), p. 65.

[58] National Center for Education Statistics. *Statewide Longitudinal Data Systems Grant Program*. Institute of Education Sciences. Retrieved from: https://nces.ed.gov/programs/slds/.

[59] King and Perkins, *Strategies for Financing the Integration of Home Visiting and Early Childhood Data Systems*, pp. 6-7; Data Quality Campaign (May 2021). State and Local Governments Can Use Federal Funding for Education Data. Retrieved from: https://dataqualitycampaign.org/wp-content/uploads/2021/05/DQC-Federal-Funding-for-Ed-Data-Summary-Infographic_updated-May-2021.pdf; Zanti, Jenkins, Berkowitz, Hawn Nelson, Burnett, & Culhane. *Building and Sustaining State Data Integration Efforts*, p. 9.

[60] King and Perkins, *Strategies for Financing the Integration of Home Visiting and Early Childhood Data Systems*, pp. 5-6.

[61] Zanti, Jenkins, Berkowitz, Hawn Nelson, Burnett, & Culhane. *Building and Sustaining State Data Integration Efforts*, pp. 8-9.

[62] For more on what the Early Learning Challenge accomplished, see Jordan and King. *Stacking the Blocks*.

[63] Sirinides, P. & Coffey, M. (2021). *Using data to support systems building activities: PDG B-5 Application Review and Impact of COVID-19 on Grant Activities*. SRI International. Retrieved from: https://childcareta.acf.hhs.gov/sites/default/files/public/pdgb5ta_datasystem_applicationreview_acc.pdf.

[64] Miller, B. (January 8, 2019). *State IT Structure Landscape Changes Dramatically*. Government Technology. Retrieved from: https://www.govtech.com/data/state-it-structure-landscape-changes-dramatically.html.

[65] Nguyen, J. & Coffey, M. (2020). *Lessons learned about ECIDS planning teams: State infrastructure and composition*. SRI International. Retrieved from: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1004&context=ecdataworks.

[66] National Center for Education Statistics (February 2017). *Best Practices Brief: P-20W+ Data Governance – Tips from the States*. Institute of Education Sciences, Brief 4. Retrieved from: https://slds.ed.gov/services/PDCService.svc/GetPDCDocumentFile?fileId=25962; Cochenour, M., Chatis, C., and Irvine, S. *Early Childhood Data Governance in Action! An Introduction*. Institute of Education Sciences. Retrieved from: https://nces.ed.gov/programs/slds/pdf/EC_DataGovernance.pdf; Zanti, Jenkins, Berkowitz, Hawn Nelson, Burnett, & Culhane. *Building and Sustaining State Data Integration Efforts*, pp. 2-5; Petrila, J., Cohn, B., Pritchett, W., Stiles, P., Stodden, V., Vagle, J., Humowiecki, M., Rozario, N. (March 2017). *Legal Issues for IDS Use: Finding a Way Forward*. Actionable Intelligence for Social Policy. Retrieved from: https://aisp.upenn.edu/wp-content/uploads/2016/07/Legal-Issues.pdf; Gibbs, L., Hawn Nelson, A., Dalton, E., Cantor, J., Shipp, S., and Jenkins, D. (March 2017). *IDS Governance: Setting Up for Ethical and Effective Use*. Actionable Intelligence for Social Policy. Retrieved from: https://aisp.upenn.edu/wp-content/uploads/2016/07/Governance.pdf; Regenstein, *An Unofficial Guide to the Why and How of State Early Childhood Data Systems*; Cochenour, M. and Hebbeler, K. "Early Childhood Data Governance: A Prerequisite for Answering Important Policy Questions," in Kagan, S.L. and Gomez, R. (eds.), *Early Childhood Governance: Choices and Consequences* (2015), New York: Teachers College Press, pp. 112-120.

[67] Government Technology, *The State of Cloud in State and Local Governments*; conversations with state officials, Office of the Federal Chief Information Officer, From Cloud First to Cloud Smart, (federal); Fantuzzo, J., Henderson, C., Coe, K., and Culhane, D. (September 2017). *The Integrated Data System Approach: A Vehicle to More Effective and Efficient Data-Driven Solutions in Government*, pp. 25-27. Actionable Intelligence for Social Policy. Retrieved from: https://www.aisp.upenn.edu/wp-content/uploads/2017/09/The-IDS-Approach_Fantuzzo-et-al.-2017_Final.pdf; Wyld, D. C., (2009). *Moving to the Cloud: An Introduction to Cloud Computing in Government*, pp. 45-47. IBM Center for The Business of Government. Retrieved from: https://www.businessofgovernment.org/sites/default/files/CloudComputingReport.pdf.

[68] For example, "[t]he Tech Talent Project is nonpartisan, nonprofit project dedicated to increasing the ability of the U.S. government to recruit modern technical leaders in order to achieve critical economic, policy, and human outcomes." Retrieved from: https://techtalentproject.org/about-us/.

[69] Weatherford, M. (May 4, 2021). *'The Easy Button' for Taking Government to the Cloud*. Governing. Retrieved from: https://www.governing.com/security/the-easy-button-for-taking-government-to-the-cloud; *From Cloud First to Cloud Smart*. Office of the Federal Chief Information Officer, (procurement); Regenstein, *An Unofficial Guide to the Why and How of State Early Childhood Data Systems*, p. 19; Wyld, *Moving to the Cloud*, pp. 40-42.

[70] For a resource that may assist states, see Actionable Intelligence for Social Policy and Asemio, *Narrowing Technology Solutions for Integrated Data System Initiatives: A Provider Framework Toolkit* (2022). Retrieved from https://docs.google.com/document/d/1CnIIh3qY57vBf0yRJZ2vInr0wNOmuYoYXuWzrxqD9sU/edit.

[71] Regenstein, *An Unofficial Guide to the Why and How of State Early Childhood Data Systems*, pp. 14-17; Fantuzzo, Henderson, Coe, and Culhane, *The Integrated Data System Approach*, pp. 20-23; Wyld, D. C. (2009). *Moving to the Cloud*, pp. 42-44. See also Hawn Nelson, Jenkins, Zanti, Katz, Burnett, Culhane, Barghaus, et al., *Introduction to Data Sharing and Integration*, pp. 12-16 (providing a framework for the legal considerations involved in data sharing and integration).

[72] *From Cloud First to Cloud Smart*. Office of the Federal Chief Information Officer, (security); Regenstein, *An Unofficial Guide to the Why and How of State Early Childhood Data Systems*, pp. 30-33; Fantuzzo, Henderson, Coe, and Culhane, *The Integrated Data System Approach*, pp. 23-25.

[73] E.g. Patterson, D., Brennan, N., Haeberlen, A., Schroeder, A., Smith, A., and Steif, K. (March 2017). *Towards State-of-the-Art IDS Technology and Data Security Solutions*, pp. 14-18. Actionable Intelligence for Social Policy, Expert Panel Report. Retrieved from: https://aisp.upenn.edu/wp-content/uploads/2016/07/Technology-Data-Security.pdf.

[74] For an example of a state report that address privacy and security issues, see California Data System (June 2021), *Cradle to Career Data System: Final Report to the Legislature*. Retrieved from https://cadatasystem.wested.org/system/resources/W1siZiIsIjIwMjEvMDYvMjkvMjMvMDYvNTYvYzlkMThiYjItMTdkZC00NzM4LWI0YTMtZmNlYTk3ZjVhNjRkL0NyYWRsZS10by1DYXJlZXIgRGF0YSBTeXN0ZW0gSnVuZSAyMDIxIExlZ2lzbGF0aXZlIFJlcG9ydC5wZGYiXV0/Cradle-to-Career%20Data%20System%20June%202021%20Legislative%20Report.pdf?sha=d1fb5c8b2870a0dd

[75] Third Sector Intelligence and Foresight Law + Policy (January 31, 2020), *A Plan for an Early Childhood Integrated Data System in Oklahoma: Data Inventory, Data Integration Plan, and Data Governance Plan*, pp. 65-68. Retrieved from https://www.flpadvisors.com/uploads/4/2/4/2/42429949/flp_3si_oklahoma_school_readinees_a_plan_for_ecids_in_oklahoma_final.pdf.

[76] Frequently referred to as "KYStats." See https://kystats.ky.gov/.

[77] *Miami-Dade Ideas Consortium for Children*. University of Miami. Retrieved from: https://ideas.psy.miami.edu/#:~:text=The%20Miami%2DDade%20IDEAS%20Consortium,Head%20Start%2FEarly%20Head%20Start.

[78] Iowa's Integrated Data System for Decision-Making. *What is I2D2?* Retrieved from: https://i2d2.iastate.edu/publications/.

[79] *Link B5: Linking Virginia's Birth to 5 Early Childhood Care and Education Communities*. Center for the Advanced Study of Teaching and Learning, University of Virginia. Retrieved from: https://vecf.org/wp-content/uploads/2021/11/2021_LinkB5_2_pg_FINAL.pdf.

[80] Carew, Joseph M. (May 7, 2021). *Data scientist job outlook in post-pandemic world*. TechTarget.com. Retrieved from: https://www.techtarget.com/searchbusinessanalytics/feature/The-data-scientist-job-outlook-positive-post-pandemic.

[81] ECDataWorks. (2019). *Data Story: State and Local Leaders Need Analytic Tools to Tell their Data Stories*. CPRE Working Papers. Retrieved from: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1019&context=cpre_workingpapers.

[82] E.g. "The ECIDS Program Lead and the ECIDS Technical Lead play different but critical roles when creating a plan for an operational ECIDS." Nguyen, & Coffey, *Lessons learned about ECIDS planning teams*, SRI International, p. 6.

[83] See also Regenstein, *An Unofficial Guide to the Why and How of State Early Childhood Data Systems*, pp. 27-29; Fantuzzo, Henderson, Coe, and Culhane, *The Integrated Data System Approach*, pp. 27-28; Actionable Intelligence Through Social Policy *A Toolkit for Centering Racial Equity Throughout Data Integration*. University of Pennsylvania. Retrieved from: https://aisp.upenn.edu/wp-content/uploads/2020/08/AISP-Toolkit_5.27.20.pdf.

FORESIGHT
LAW+POLICY

www.flpadvisors.com
800 Maine Avenue, SW, Suite 200
(p) 202.499.6996 (f) 202.644.9779
Washington, D.C. 20024
Foresight@FLPAdvisors.com

Author: Elliot Regenstein