

A PLAN FOR AN EARLY CHILDHOOD INTEGRATED DATA SYSTEM IN OKLAHOMA:

DATA INVENTORY, DATA INTEGRATION PLAN, AND DATA GOVERNANCE PLAN

January 31, 2020



The Oklahoma Partnership for School Readiness would like to acknowledge the *Oklahoma Early Childhood Integrated Data System (ECIDS) Project Oversight Committee* for advising and supporting development of this plan:

Steve Buck, Cabinet Secretary of Human Services and Early Childhood Initiatives

Jennifer Dalton, Oklahoma Department of Human Services

Erik Friend, Oklahoma State Department of Education

Becki Moore, Oklahoma State Department of Health

Funding for development of this plan was provided by the Preschool Development Grant Birth through Five (Grant Number 90TP0037), a grantmaking program of the U.S. Department of Health and Human Services, Administration for Children and Families, Office of Child Care.

Contents

Glossary.....	6
Image Credits	14
1. Executive Summary.....	15
1.1. Uses of an ECIDS	15
1.2. About this ECIDS Plan.....	16
1.2.1. Data Inventory	16
1.2.2. Data Integration	17
1.2.3. Data Governance	17
1.3. Next Steps	19
2. Introduction	22
2.1 History and Context	22
2.2 Vision, Use Cases, and Value Proposition for the ECIDS.....	23
2.2.1 Vision for the ECIDS	23
2.2.2 Priority Use Cases.....	24
2.2.3 Value Proposition for ECIDS.....	30
2.3 Challenges and Opportunities.....	35
2.3.1 Oklahoma’s Opportunity.....	38
2.4 Purpose of this document.....	43
2.4.1 Data Inventory	44
2.4.2 Data Integration Plan	44
2.4.3 Data Governance Plan.....	44
3. Data Inventory	46
3.1. Introduction and Background	46
3.1.1. Data Landscape Map.....	46
3.2. Methodology.....	48
3.3. Data Inventory Content and Findings	49
3.3.1. General Inventory of Agencies and Systems.....	49
3.3.2. Data Needed to Fulfill Use Cases	49
3.3.3. Feasibility of Use Cases	51
3.3.4. Summary of Use Case Feasibility	54
3.3.5. Risks and Mitigations	56
3.4. Future Work to Expand and Refine Data Inventory.....	57

4.	Data Integration Plan	59
4.1.	Introduction and Background	59
4.1.1.	Data Integration Preparation	59
4.2.	Data Integration Plan Options and Recommendation.....	60
4.2.1.	Choosing a Cloud provider	60
4.2.2.	Cloud Data Security.....	65
4.2.3.	Data Transportation	69
4.2.4.	Data Storage: The Data Lake	72
4.2.5.	Data Management and Processing	76
4.2.6.	Initial Analytics and Data Products	80
4.2.7.	Data Integration Roles and Responsibilities	87
4.3.	Key Considerations for Data Integration.....	89
5.	Data Governance Plan.....	90
5.1.	Introduction	90
5.2.	Data Governance Structure	90
5.2.1.	Data Governance Objectives.....	90
5.2.2.	Success Factors for the Governance Model.....	91
5.2.3.	Data Governance Framework	92
5.2.4.	Changes to Laws and Regulations.....	103
5.2.5.	Privacy Laws and Regulations	103
5.2.6.	Changes to Interagency Agreements	105
5.3.	Methodology for Developing Data Governance Framework.....	106
5.3.1.	Stakeholder Engagement.....	106
5.3.2.	Review of Data Governance Models.....	107
5.4.	Critical Issues for Data Governance	107
6.	Future Work Necessary to Implement Oklahoma’s ECIDS	108
6.1.	Introduction	108
6.2.	Draft Implementation Timeline	108
6.3.	Draft Hiring Plan.....	110
6.3.1.	Recruitment Challenges and Mitigations.....	111
6.4.	Estimated Five-year Implementation Costs	111
6.5.	Next Steps and Considerations	113
6.5.1.	Project funding and ownership.....	113

Appendices.....	115
Appendix A: OK Early Learning Stakeholders.....	115
Appendix B: System and Data Inventory Templates.....	116
Appendix C: Data Inventory	118
C.1 Agencies and Data Systems by Tier Status	118
C.2 System-Level Inventory of Data Systems.....	118
C.3 Program Eligibility Requirements	128
C.4 Availability of Data Needed to Fulfill Use Cases #1-3	134
C.5: Availability of Data Needed to Map to OKDH MPI	163
C.6: Distribution of Head Start Programs and System Utilization	168
Appendix D: Example of Data Elements and Sizes for Cost Exercise	172
Appendix E: More Information about Cloud Service Providers	173
Appendix F: Methodology for Determining Access Deserts	174
Appendix G: Draft Interagency Agreement	176
Appendix H: Survey of Agencies Regarding Work Groups.....	192
Appendix I: Results of Agency Work Group Survey	195
Appendix J: Center Administration Agreement Key Terms	197
Appendix K: Data Contributor Agreement.....	201
Appendix L: Governance Questions Shared with Agencies	216
Appendix M: Examples of Data Governance Models	218
M.1 Different Models of Data Governance	218
M.2 Deeper Descriptions of Executive-led data governance models.....	220
Appendix N: Center Roles and Responsibilities	223
Appendix O: Recruitment Challenges and Strategy.....	224
Appendix P: Preliminary Five-year Cost Estimate for Oklahoma ECIDS*.....	226
Appendix Q: Expense Categories and Descriptions for Oklahoma ECIDS.....	227

Glossary

AGILE METHODOLOGY

Agile methodology approaches the development of analytic and data products in iterative sprints (usually two weeks). With the Agile approach, development teams focus on quick deliverables that can run parallel to other workstreams. Agile strives to create a “draft” analytic or data product—called a Minimum Viable Product—as fast as possible so that users can try it out and provide feedback for the next iteration.

BATCH PROCESSING/UPLOADING

In the context of computing, batch processing refers to the processing of previously collected jobs in a single batch. Batch uploading refers to the uploading of previously collected jobs in a single batch.

CLOUD SERVICE PROVIDER

Cloud service providers (CSPs) are companies offering network services, infrastructure, or business applications in the cloud. Services are hosted in a data center that is accessible to the customer.

DATA LAKE

A data lake is a centralized repository that stores data as-is, without having to first structure or clean the data. A data lake is a vast pool of raw data, the purpose for which is not yet defined.

DATA SCIENTIST

A person employed to analyze and interpret complex digital data, such as the early childhood data on special education eligibility, especially in order to assist in decision-making.

DATA WAREHOUSE

In computing, a data warehouse is a central repository of structured, filtered data that has already been processed for a specific purpose from one or more disparate sources. This system can be used for reporting and analysis of data.

ENCRYPTION

Encryption is the process of encoding information in such a way that only authorized parties can access it and those who are not authorized cannot. Data encryption translates data into another form, or code, so that only people with access to a secret key or password can access it.

ENTERPRISE DATA WAREHOUSE

An enterprise data warehouse is a database or collection of databases, that centralizes structured data from multiple sources and makes it available for analytics and use across an enterprise.

EXTRACT

Used in the context of data, an extract is a piece of data that has been pulled from a data source. Data extraction refers to the process of analyzing and crawling through a data source to retrieve relevant information in a specific pattern. Data can then be replicated to a destination, such as a data warehouse.

HEAD START

Head Start is a program of the United States Department of Health and Human Services that provides comprehensive services to low-income children and families in the areas of early childhood education, health, nutrition, and parent involvement.

KEY MANAGEMENT

Key management refers to the management of cryptographic keys in a cryptosystem, including the generation, exchange, storage, use, crypto-shredding (destruction), and replacement of keys. It includes cryptographic protocol design, key servers, user procedures, and other relevant protocols.

LICENSED CHILD CARE

This refers to programs and providers that have been evaluated by the state licensing agency, deemed to meet minimum health, safety, and educational standards, and have been issued a license.

LOAD

In the context of data, load is the process of taking transformed data from one or more sources and loading it into a destination system where the users can access it.

MASTER PERSON INDEX

A Master Person Index (MPI), also referred to as a Master Patient Index, is a database used to maintain accurate data on individuals across various departments and programs. Originally used in healthcare organizations to maintain patient data, MPIs are being used more widely to capture educational, health, behavioral, and other individual-level data. MPIs use a matching algorithm to establish unique identifiers for individuals with data in disparate data systems.

METADATA

Metadata is a set of data that describes and gives information about other data.

MINIMUM VIABLE PRODUCT

A minimum value product (MVP) is a product with just enough features to satisfy early users and provide feedback for future product development. In the context of the Early Childhood Integrated Data System (ECIDS), it is a “draft” analytic or data product that users can try out and provide feedback to improve the next iteration.

MIXED DELIVERY SYSTEM

A system of early childhood care and education services that are delivered through a combination of programs, providers, and settings, such as Head Start, licensed family and center-based child care programs, public schools, and other community-based organizations, that is supported by a combination of public and private funds. As a whole, Oklahoma’s ECCE mixed-delivery system includes eleven programs, policies, and funding streams, along with related services.

SEMI-STRUCTURED DATA

Data that is tagged with definitions but lacks structure and needs to be processed/parsed. Examples include XML and JSON data elements. Most web page data content is semi-structured.

SINGLE SOURCE OF TRUTH

Single source of truth is a concept in data management whereby everyone in an enterprise bases business decision on the same data. To put a single source of truth in place, an organization or enterprise must provide relevant personnel with one source that stores the data points they need.

SOONERCARE

SoonerCare (Oklahoma Medicaid) is a health coverage program jointly funded by the federal and state government. This program helps pay some or all medical bills for many people who can't afford them.

SOONERSTART

SoonerStart is Oklahoma's early intervention program. It is designed to meet the needs of families with infants or toddlers with developmental delays. In accordance with the Individuals with Disabilities Education Act (IDEA) the program builds upon and provides supports and resources to assist family members to enhance infant's or toddler's learning and development through everyday learning opportunities.¹

STRUCTURED DATA

Data that has been organized and processed and is confined to that structure. Examples include SQL database tables and Excel sheets.

SUBSIDIZED CHILD CARE

Subsidized child care benefits ensure high quality care for children while their parents or guardians are at work, in training, or receiving an education. Subsidized child care benefits may also be provided as part of a protective service plan to prevent abuse, neglect or exploitation. The subsidy is paid directly to the child care provider on the family's behalf. The family may have a co-payment for the child care based on their income, the number of family members and the number of family members needing services.

TRANSFORM

In the context of this report, the process of converting data from one format or structure into another format or structure.

UNIVERSAL PRE-K

Universal Pre-K (UPK) is a movement within the American education system to make access to preschool education available to all families, similar to the way kindergarten is available to all 5- and 6-year-olds. Oklahoma offers UPK in all but a few districts.²

UNSTRUCTURED DATA

Data that has no structure or organization. Data objects have no relationship to other data objects. Examples include media files, emails, database backup files.

¹ <https://sde.ok.gov/soonerstart>

² <http://nieer.org/2016/01/06/universal-pre-k-what-does-it-mean-and-who-provides-it>

USE CASE

A specific situation where the Early Childhood Integrated Data System (ECIDS) could be used.

USER CREDENTIAL MANAGEMENT

User Credential Management is a form of software used to issue and manage credentials as part of public key infrastructure (PKI).

Acronyms

ACE	Adverse Childhood Experiences
ACS	American Community Survey
AFS	Adult and Family Services
API	Application Program Interface
AWS	Amazon Web Services
BIE	Bureau of Indian Education
CC-MASS	Child Care, Monitoring, Administration, and Safety System
CECPD	Center for Early Childhood Professional Development (CECPD)
CPU	Central Processing Unit
CSP	Cloud Service Provider
DISCUSS	Deliver Interoperable Solution Components Utilizing Shared Services
DSS	Division Support Services
DRS	Department of Rehabilitation Services
EASI	Eligibility-Access-Service-Impact
ECCE	Early Childhood Care and Education
ECE	Early Childhood Education
ECIDS	Early Childhood Integrated Data System
eMPI	Enterprise Master Person Index
EPPIC	System used by Finance for Child Care Subsidy information and Medicaid payments for service
ETL	Extract-Transform-Load
FACS DB2	Field Automated Communication Systems Database 2
FDPIR	Food Distribution Program on Indian Reservations
FedRAMP	Federal Risk and Authorization Management Program
FERPA	Family Education Rights and Privacy Act

GCP	Google Cloud Platform
GICH	Governor's Interagency Council on Homelessness
GPU	Graphical Processing Unit
HIPAA	Health Insurance Portability and Accountability Act
HITECH	Health Information Technology for Economic and Clinical Health Act
HS	Head Start
IaaS	Infrastructure as a Service
IAM	Identify Access Management
IDEA	Individuals with Disabilities Education Act
IDT	Interactive Data Tool
JDBC	Java Database Connectivity
KPI	Key Performance Indicator
LEP	Local Expenditure Program (Electronic bookkeeping system for the county health departments)
LIHEAP	Low Income Home Energy Assistance Program
MDS	Mixed Delivery System
MMIS	Medicaid Management Information System
MPI	Master Person Index
MVP	Minimum Viable Product
OBDR	Oklahoma Birth Defects Registry
OCCRRA	Oklahoma Child Care Resource and Referral Agency
OCCY	Oklahoma Commission on Children and Youth
OCLPPP	Oklahoma Childhood Lead Poisoning Prevention Program
ODA	Official Depository Account
ODBC	Open Database Connectivity

ODMHSAS	Oklahoma Department of Mental Health and Substance Abuse Services
ODOC	Oklahoma Department of Commerce
OHCA	Oklahoma Health Care Authority
OJA	Oklahoma Office of Juvenile Affairs
OKACAA	Oklahoma Association of Community Action Agencies
OKDHS	Oklahoma Department of Human Services
OKFutures	Oklahoma's Future Begins with Children
OKHIN	Oklahoma Health Information Network
OKWORKS	Oklahoma Works
OMES	Office of Management and Enterprise Services
OPDR	Oklahoma Professional Development Registry
OPSR	Oklahoma Partnership for School Readiness
ORBS	Oklahoma Reconciliation and Billing System
OSDE	Oklahoma State Department of Education
OSDH	Oklahoma State Department of Health
OSIIS	Oklahoma State Immunization Information System
PaaS	Platform as a Service
PDG B-5	Preschool Development Grant Birth through Five
PHI	Protected Health Information
PHIDDO	Public Health Investigation and Disease Detection of Oklahoma
PHOCIS	Public Health Oklahoma Client Information System
PII	Personally Identifiable Information
SaaS	Software as a Service
SIF	Schools Interoperability Framework
SIS	Student Information Systems

SNAP	Supplemental Nutrition Assistance Program
SSI	Supplemental Security Income
SSRF	Server-Side Request Forgery
STRIDES	Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability
Systems_ETO	Case Management Client Information Systems
TANF	Temporary Assistance for Needy Families
VPN	Virtual Private Network
WIC	Women, Infants, and Children Supplemental Nutrition Program
WLS	Work Life System
YIS	Youth Information System
ZIS	Zone Integration Servers

Image Credits

“Hourglass” by CombineDesign from the Noun Project

“Despair” by Priyanka from the Noun Project

“Thumbs Down” by Ben Davis from the Noun Project

1. Executive Summary

Overwhelming evidence suggests that early childhood is the time when education has the greatest potential to impact a child's future development. A growing body of research indicates that early childhood education (ECE) programs reduce achievement gaps and improve school readiness.

Recognizing this potential, Oklahoma has long stood at the nation's forefront in ensuring access to early childhood education opportunities. Oklahoma has provided universal Pre-Kindergarten since 1998. The state legislature formed the Oklahoma Partnership for School Readiness (OPSR) in 2003 to ensure all Oklahoma children arrive at school with the knowledge, skills, and physical and emotional health to achieve success. And in 2018 Oklahoma launched the *Oklahoma's Future Begins with Children* (OKFutures) initiative to improve its mixed-delivery early childhood education system.

Critical to these efforts is access to timely, accurate data on early childhood programs and the children and families they serve. But accessing and using integrated early childhood data across programs is not a simple undertaking. Early childhood data resides in multiple agencies with different source systems; moreover, much of this information is sensitive, requiring rigorous data governance and management.

To maximize use and usefulness of early childhood data, therefore, Oklahoma must streamline the process of integrating and using early childhood data, while protecting the privacy and security of sensitive data. An effective way to do this is to build an early childhood integrated data system (ECIDS). An ECIDS transforms disparate data into valuable information that can be used to support mission-critical activities for Oklahoma's early childhood community and its constituents. Once implemented, an ECIDS will help Oklahoma enhance and accelerate decision-making, improve program performance, optimize public policies, assess outcomes, and enable multiple agencies and departments to work together more efficiently to improve child and family outcomes.

This document outlines a plan to develop and implement an ECIDS. The plan includes an overview of the system's potential uses, the data needed for the system, the capacity needed to store and integrate these data, and a data governance framework administered by a central governing body. OPSR led this planning with funding from the federal Preschool Development Grant Birth through Five (PDG B-5). This plan is the product of an intensive planning and stakeholder engagement effort over the past eight months. It builds on more than two decades of work to build partnerships, create early childhood systems, and engage state and tribal organizations, providers, parents, philanthropic and business leaders, and concerned community members.

1.1. Uses of an ECIDS

Stakeholder and executive input have been critical in developing this plan. In June and July 2019, stakeholders and executives identified a vision for the ECIDS:

- With linked early childhood data, Oklahoma government would be better equipped to assess, target, and improve specific early childhood program investments and services to meet the needs of Oklahoma children and families. Specifically, population and eligibility data would allow Oklahoma to identify and reach children and families who should be but are not receiving

services, and streamline and improve service delivery and effectiveness for those children and families who are receiving services.

- Linked data would allow stakeholders to evaluate long-term outcomes of early learning, such as school readiness and behavioral and health outcomes. For example, linked data would allow stakeholders to observe potential relationships between school readiness and interactions with the juvenile justice system and demonstrate the cost-benefit of early learning investments to lawmakers and voters. This, in turn, should catalyze and guide program and policy decisions.

Stakeholder and executive input also informed development of five priority use cases for the ECIDS. Use cases describe the specific questions the ECIDS could answer and articulate the data that would need to be integrated to answer these questions. The following priority use cases for the ECIDS reflect the input of stakeholders in the ECIDS and the emerging direction of Oklahoma’s needs assessment and strategic planning work currently underway:

Use Case #1: What and where are there gaps in early childhood education services and quality, and what options exist to close those gaps?

Use Case #2: What are the factors driving the significant gap between referrals for early intervention services for children with disabilities and developmental delays (SoonerStart) and children actually receiving these services? What options exist to close this gap?

Use Case #3: If a child or family is eligible for a social safety net program—such as Temporary Assistance for Needy Families (TANF) or subsidized early childhood education—what other social safety net programs are they eligible to receive? What options exist to better combine and deliver these services?

Use Case #4: What barriers do insured families face in accessing health services, especially in rural areas? What options exist to reduce these barriers?

Use Case #5: How do early childhood education programs affect longer term academic and behavioral outcomes?

These five priority use cases will focus Oklahoma’s early efforts to build an ECIDS, and there will be opportunities to refine, expand, and build upon them. Furthermore, as the ECIDS becomes more used and useful, we anticipate the Governor and participating agencies will propose additional use cases.

1.2. About this ECIDS Plan

This ECIDS Plan includes a Data Inventory, Data Integration Plan, and Data Governance Plan.

1.2.1. Data Inventory

An early step in establishing an ECIDS and addressing the priority use cases is to map the landscape of early childhood education data in Oklahoma. The data inventory describes the data elements needed to fulfill priority Use Cases #1, 2, and 3, and assesses their feasibility based on available data identified in the data inventory. Future iterations of the data inventory will address Use Cases #4 and 5, as well as additional use cases the Governor’s Office and participating agencies may identify.

Although there are data gaps that Oklahoma will need to address, all three of the initial use cases appear to be feasible. Incorporation of additional programs and data sources into the ECIDS will gradually improve the completeness of available data and lower the reliance on proxy data over time. When Oklahoma stakeholders are ready to develop new use cases, the data inventory will help identify available data assets and those that would need to be developed.

1.2.2. Data Integration

The data integration plan describes the recommended approach to combining data from several different sources into a unified format so they can be analyzed and used in decision-making. It leverages cloud technology to help drive down costs and increase flexibility and functionality. Specifically, this data integration plan describes:

1. The process of choosing a cloud service provider
2. Cloud data security
3. Data transportation
4. Data storage
5. Data management and processing
6. Initial analytics and data products
7. Data integration roles and responsibilities

To proceed with the data integration plan proposed in this report, Oklahoma stakeholders must first confirm a cloud service provider (CSP), because this decision will have implications for other data integration steps. 3Si/Foresight recommends that Oklahoma choose Microsoft Azure. Because Oklahoma already has infrastructure built on Microsoft products—including Azure—the 3Si/Foresight team’s analysis focused on determining whether there was a compelling reason to go with a CSP other than Azure. In our opinion, the advantage of being able to more easily transition to Azure outweighs any minor advantages offered by its major competitors, Amazon Web Services and Google Cloud Platform.

The data in the ECIDS will represent millions of stories of Oklahoma’s children, youth, and families, and telling those stories will require the collective efforts of trained data analysts, engineers, managers, and stewards working under the coordination and guidance of a sound governance framework. Therefore, another step Oklahoma will need to take in implementing this data integration plan is to assess what capacities it needs to build to implement and maintain an ECIDS. The Data Integration Plan (Section 4 of this document) discusses the roles and responsibilities needed for data integration, and the Data Governance Plan (Section 5 of this document) details the roles and responsibilities of those involved in governance of the ECIDS. Section 6 outlines future work to implement an ECIDS, including a hiring timeline and plan.

1.2.3. Data Governance

The data governance framework is designed to protect data privacy and security while making it faster and easier for stakeholder agencies to analyze and use data in decision-making. The data governance plan includes processes governing how data is collected, integrated, and reported; interagency decision-making; and adding new agencies in the future. It identifies updates to Oklahoma’s laws and regulations and modifications to existing and new interagency agreements necessary to support the governance framework. It also describes policies for managing and protecting personal information and privacy and

security protocols that meet federal and state requirements and ensure data is used for educational purposes only.

For Oklahoma's ECIDS to produce the desired impact, it needs a governance structure that can support interagency data use on an ongoing basis. The success of the ECIDS will require a continued focus on the collective benefit of data sharing and the value to the entire system of having broad agency participation. It will also require a design that supports collaboration among agencies, so each agency is advancing its own mission while contributing to the broader field. Leadership from the Governor's Office and individual agencies will be needed on an ongoing basis to sustain momentum.

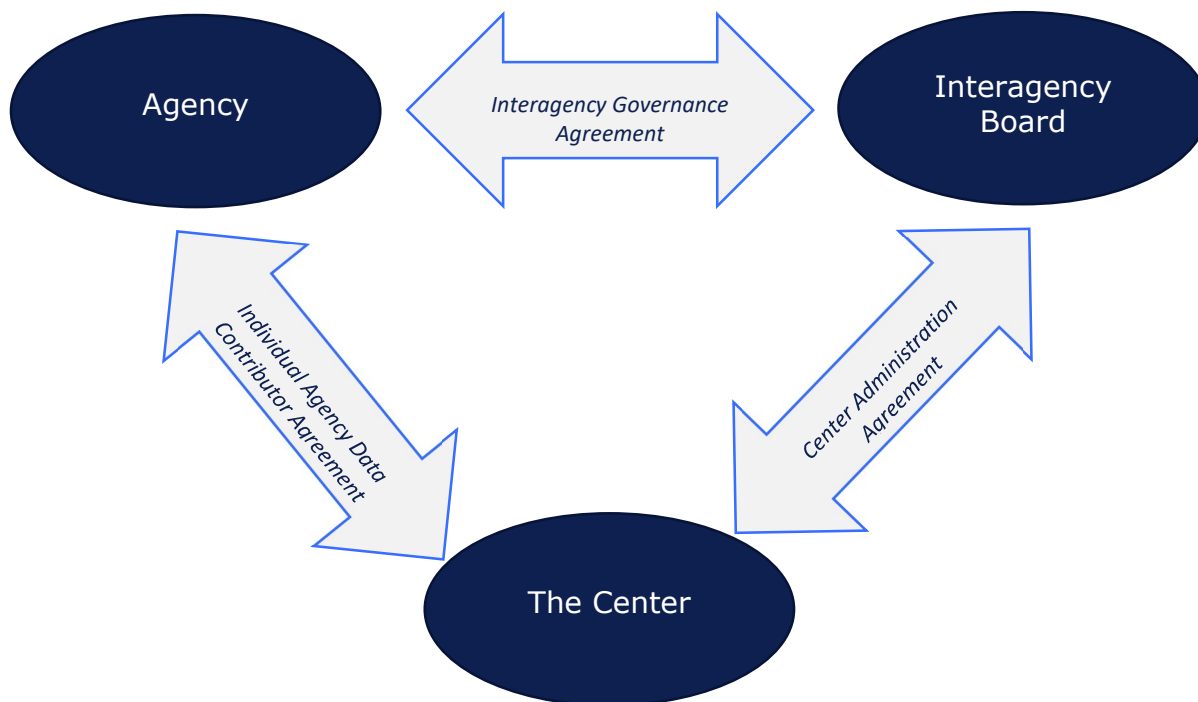
The 3Si/Foresight team developed six objectives for data governance, based on input from stakeholders and research on best practices. These six objectives informed development of the proposed data governance framework:

- *Focus on child and family outcomes:* The system will view all its activities through the lens of how they improve outcomes for children and families.
- *Usefulness:* The data system will be useful to end users—with an emphasis on program outputs and impact rather than program inputs and processes—while streamlining and reducing governance and time spent on preparing the data for analysis (to the extent possible and within appropriate legal and administrative parameters).
- *Action orientation:* The governance structure will orient toward meeting the operational needs of the Governor's Office and state agencies. The governance structure will focus on providing quick and efficient access to data for the Governor's Office and state agencies, so the data can be used to drive decision-making. It must also maximize the use of its existing capabilities at any given time, rather than waiting for additional capacity to be built.
- *Expandability:* The governance structure will be launched with a set of committed state agencies who see the value of partnership. Over time, that structure may be expanded to include other agencies. The system may also begin accepting data from partners outside of state government.
- *Inside-outside collaboration:* While the oversight of the system is a core state government function, many states have chosen to include external stakeholders in their governance structure. Oklahoma agency leaders expressed interest in this possibility. In particular, the system will seek to partner with Oklahoma's tribes and Head Start providers.
- *Inspiration:* Ideally, the new data system will allow agencies to fundamentally change the way they operate by providing them with information that is more comprehensive and useful than any information they had previously, delivered at a faster speed. This has significant implications for the way agencies operate and deliver services, both individually and collectively. Agencies will need to evaluate their capacity to use the new system on an ongoing basis and expand as needed to improve their operations.

To achieve these six objectives—along with the overarching goal of improving service to Oklahoma's children and families—we propose an interagency data governance framework for managing the ECIDS. The framework includes three critical components (shown in the diagram below):

1. A high-level interagency board of empowered agency representatives that provides leadership for the governance structure (referred to here as the Board).
2. A lean work group structure that integrates critical agency expertise to facilitate policy development.
3. A “Center” that provides centralized administrative capacity and technical infrastructure.

Figure 1-1 - Data governance framework: Relationship among agreements and entities



The Interagency Board will draw on the expertise of work groups and oversee the management capacity provided by the Center. The Board, work groups, and Center will have tightly defined responsibilities and be charged with maintaining an action orientation to meet the needs of participating agencies. Establishing the capacity needed to manage the ECIDS will require an interagency agreement that reshapes the relationship among child-serving agencies and accounts for the important role of the Office of Management and Enterprise Services (OMES).

1.3. Next Steps

Initial ECIDS planning is complete and implementation is poised to begin as soon as the Governor or his cabinet approves the plan and OPSR obtains necessary support/funding. Section 6 of this document includes a draft timeline for implementing the ECIDS from 2020 through 2024. Figure 1-2 shows a proposed high-level timeline for implementing the ECIDS.

Figure 1-2 - Proposed high-level timeline for implementing ECIDS

2020		
BOARD	CENTER	OUTSOURCED SERVICES
<ul style="list-style-type: none"> Establish Board Convene work groups Contract with agency to serve as temporary Center until permanent Center can be established Hire Center Executive Director Serve as final authority over ECIDS and Center 	<ul style="list-style-type: none"> Hire outsourced services (contractor) Implement necessary legal frameworks and data agreements (in coordination with Board) Carry out initial steps of ECIDS implementation Begin developing Use Case #1 	<ul style="list-style-type: none"> Backfill temporary Center capacity until permanent staff can be hired Support technical aspects of ECIDS development Support development of Use Case #1
2021		
BOARD	CENTER	OUTSOURCED SERVICES
<ul style="list-style-type: none"> Continue serving as final authority over ECIDS and Center Work with Center to determine appropriate mix of in-house and outsourced resources 	<ul style="list-style-type: none"> Hire and train core staff Complete transition from temporary to permanent Center by the end of 2021 Develop data products for Use Cases #1, 2, and 3 	<ul style="list-style-type: none"> Continue supporting technical aspects of ECIDS development Support development of data products for Use Cases #1, 2, and 3
2022		
BOARD	CENTER	OUTSOURCED SERVICES
<ul style="list-style-type: none"> Continue serving as final authority over ECIDS and Center Work with Center to finalize sustainability plan to continue Center operations 	<ul style="list-style-type: none"> Develop and implement sustainability plan to continue Center operations Incorporate more data from partner agencies into ECIDS Continue developing data products for Use Cases #1, 2, and 3; add Use Cases #4 and #5, plus additional use cases 	<ul style="list-style-type: none"> Continue supporting technical aspects of ECIDS development Support development of data products
2023-2024		
BOARD	CENTER	OUTSOURCED SERVICES
<ul style="list-style-type: none"> Continue serving as final authority over ECIDS and Center Add new agencies as appropriate 	<ul style="list-style-type: none"> Continue implementing sustainability plan 	<ul style="list-style-type: none"> Continue supporting data transport, storage, and maintenance Support development of data products

Section 6 includes low and high five-year cost estimates for an ECIDS. Total estimated costs for 2020-2024 are \$17.2 million to \$22.9 million. These estimates take into account expenses for establishing a transitional Center and permanent Center; costs associated with setting up and maintaining the cloud; outsourced costs for data transport, storage and maintenance; and legal support. Section 6 discusses these estimates and Appendix P (Preliminary Five-year Cost Estimate for Oklahoma ECIDS) provides detailed projections.

Oklahoma is well-positioned to create an ECIDS. The state has already made significant advances in early childhood education, created data linkages and infrastructure that can be leveraged to build the ECIDS, and convened stakeholders to develop this plan. Furthermore, Oklahoma has secured high-level support for the ECIDS. These conditions create a unique opportunity for Oklahoma to develop an ECIDS and use better information to improve the lives of children and families.

2. Introduction

2.1 History and Context

In recognition of the abundant evidence that early childhood education benefits children, families, communities, and taxpayers, Oklahoma has set a goal of creating the most effective mixed-delivery early childhood education system in the United States. Critical to this effort is access to timely, accurate data on early childhood programs and the children and families they serve. The Governor's cabinet, the independently-elected Superintendent of Public Instruction, and numerous agency leaders have partnered to lead and support this critical work.

Accessing and using integrated early childhood data across programs is not a simple undertaking. Early childhood data resides in multiple agencies with different source systems, and these differences need to be addressed before these data can be integrated and used. Furthermore, much of this information—especially data on individual children—is sensitive, requiring rigorous data governance and management.

To maximize use and usefulness of early childhood data, therefore, Oklahoma must streamline the process of integrating and using early childhood data, while protecting the privacy and security of sensitive data. An effective way to do this is to build an Early Childhood Integrated Data System (ECIDS). The goal of an ECIDS is to transform disparate data into valuable information that can be used to support mission-critical activities for Oklahoma's early childhood community and its constituents. By efficiently collecting, processing, analyzing, packaging, and distributing early childhood education information to key stakeholders, an ECIDS can help Oklahoma enhance and accelerate decision-making, improve program performance, optimize public policies, assess outcomes, and enable multiple agencies and departments to work together more efficiently.

Oklahoma is well-positioned to create an ECIDS:

- The state has already made significant advances in early childhood education (see sidebar).

EARLY LEARNING IN OKLAHOMA

Oklahoma has a long history of commitment to early learning, having funded pre-Kindergarten in 1998. Oklahoma is only the third state to fund universal pre-Kindergarten. According to the National Institute for Early Education Research (NIEER), 75 percent of eligible four-years old children are enrolled in pre-Kindergarten in Oklahoma.

In 2003, the Oklahoma legislature formed the Oklahoma Partnership for School Readiness (OPSR), a public-private partnership that works to ensure all Oklahoma children arrive at school with the knowledge, skills, and physical and emotional health to achieve success. OPSR provides state-level coordination of policies, programs, and funding for early childhood education.

In December 2018, the U.S. Department of Health and Human Services awarded Oklahoma a grant to support the state's work to improve its mixed-delivery early childhood education system. OPSR is convening state agencies and stakeholders to realize this vision.

OPSR, which serves as the State Early Childhood Advisory Council in state law and under the Head Start Act of 2007, is authorized to develop recommendations for a unified data collection system.

- Oklahoma has already created or is working to create other data linkages and infrastructure that can be leveraged to build the ECIDS, including a Master Person Index (MPI) that is integrating roughly 20 data systems, a seven-agency data use agreement, and a data use agreement between the Departments of Health and Education to link home visiting and early intervention data.³
- OPSR has already taken steps toward developing an ECIDS: it has received technical assistance, convened stakeholders in Oklahoma’s early childhood education system to determine the work ahead, and secured funding through the federal Preschool Development Grant Birth through Five (PDG B-5) to plan for an ECIDS.⁴
- Because Oklahoma does not have a current ECIDS, it is not anchored to existing practices and has more flexibility to leverage new technologies and approaches.
- There is high-level support for the ECIDS. Oklahoma Governor Kevin Stitt has prioritized improving the use of data to monitor public investments. The Governor’s cabinet has actively supported an initiative to centralize and integrate multiple data systems in Oklahoma, including early childhood data.

In combination, these conditions create a unique opportunity for Oklahoma to develop an ECIDS and use better information to improve the lives of children and families.

2.2 Vision, Use Cases, and Value Proposition for the ECIDS

2.2.1 Vision for the ECIDS

To build on these favorable conditions and advance an effective plan for an ECIDS, stakeholders need a clear vision that articulates what Oklahoma wants to monitor and track with an ECIDS, why it matters, and what decisions or policies these data are intended to affect. (Appendix A includes a list of stakeholders in Oklahoma’s ECIDS.)

In June and July 2019, OPSR and the 3Si/Foresight team convened stakeholders to identify a vision for the ECIDS:

- With linked early childhood data, Oklahoma government would be better equipped to assess, target, and improve specific early childhood program investments and services to meet the needs of Oklahoma children and families. Specifically, population and eligibility data would allow Oklahoma to identify and reach children and families who should be but are not receiving

³ A Master Person Index (MPI), also referred to as a Master Patient Index, is a database used to maintain accurate data on individuals across various departments and programs. Originally employed by healthcare organizations to maintain patient data, MPIs are being used more widely to capture educational, behavioral, and other individual data. MPIs apply a matching algorithm to establish unique identifiers for individuals with data in disparate data systems. OSDH’s MPI uses personally identifiable information (PII) common across data systems (e.g., Social Security Number, name, etc.) to determine if records from disparate systems are referring to the same individual.

⁴ PDG B-5 has been funded by the U.S. Department of Health and Human Services, Administration for Children and Families, Office of Child Care, Grant No. 90TP0037.

services, and streamline and improve service delivery and effectiveness for those children and families who are receiving services.

- Linked data would allow stakeholders to evaluate long-term outcomes of early learning, such as school readiness and behavioral and health outcomes. For example, linked data would allow stakeholders to observe potential relationships between school readiness and interactions with the juvenile justice system and demonstrate the cost-benefit of early learning investments to lawmakers and voters. This, in turn, should catalyze and guide program and policy decisions.

Governor Stitt, who came into office at the beginning of 2019, has made data-driven decision-making a priority for his administration and explicitly endorsed centralizing data—including early childhood data—to improve its use and usefulness. The Governor’s cabinet has agreed to take a key leadership role in supporting and accelerating development of an ECIDS. This leadership is essential to the success of the ECIDS, as the Governor is uniquely positioned to motivate stakeholder agencies to participate in contributing data to the ECIDS.

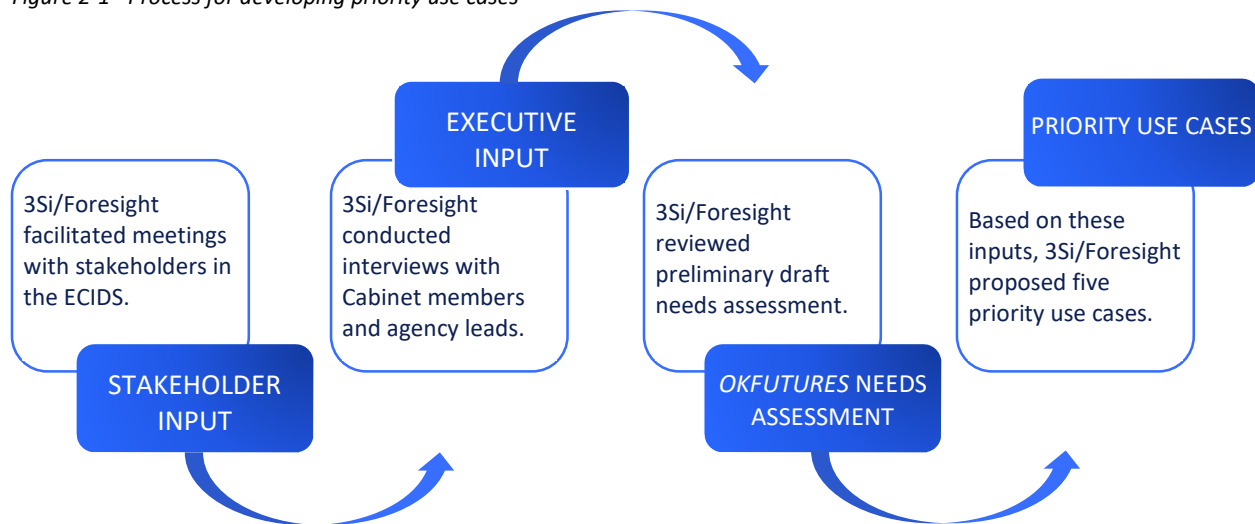
That said, for the ECIDS to succeed and grow in the long term, it will have to provide benefits to participating agencies that they believe demonstrate value worthy of the investment. Therefore, it is important to articulate the key questions that agencies will be able to answer with linked data, and how answering these key questions will help agencies fulfill their mandates and better serve Oklahomans.

2.2.2 Priority Use Cases

Identifying priority use cases is an important next step in the process of developing an ECIDS. Use cases describe the specific questions the ECIDS could answer and articulate the data that would need to be integrated to answer these questions. Consequently, use cases provide a rationale for key holders of these data to come to the table as part of a shared governance structure.

Use cases for the ECIDS reflect the input of stakeholders in the ECIDS and the emerging direction of Oklahoma’s needs assessment strategic planning work that is currently underway. 3Si/Foresight developed five priority use cases with input from stakeholders and executives and informed by outcomes from the *OKFutures Needs Assessment—Preliminary Draft for Review*. Figure 2-1 (on the following page) shows the process for developing these use cases:

Figure 2-1 - Process for developing priority use cases



STAKEHOLDER INPUT

In July 2019, 3Si/Foresight worked with stakeholders to brainstorm some initial questions the ECIDS could answer. OPSR hosted and the 3Si/Foresight team facilitated a large cross-sector meeting, which generated a set of initial questions.

Which children are in which programs?

Stakeholders expressed interest in capturing an unduplicated count of children being served by early childhood programs for youth ages birth to five. Stakeholders also expressed interest in capturing the count of children not being served by any of these programs.

What assessments are being done on children? How are assessment results being used to help children?

Stakeholders expressed concern that children are being assessed many times, often in duplicative ways and the information gleaned from those assessments is not always following the child as they transition to other programs. Stakeholders were interested in understanding how and when assessment is conducted, how the results are used, and how the assessment system could be more efficient to reduce duplication of effort for providers and families.

What are the long-term outcomes of children who had access to early childhood programs in different combinations compared to those who did not?

Stakeholders want to measure and compare the educational, behavioral, and life outcomes for children served by different combinations of early childhood education programs, or at different durations (for example, stakeholders expressed interested in comparing outcomes for children who only attended pre-kindergarten to children who attended multiple years of early childhood education). They also want to compare those results to outcomes for children who are not served by any early childhood education programs.

What is the impact on the Oklahoma workforce of early childhood education?

Stakeholders expressed interest in understanding the extent to which access to child care programs for children ages birth to five affected the ability of parents and guardians to work, and conversely, how inadequate supply of child care impacts working parents. There were also interested in learning more about the impediments to maintaining enough child care to meet the needs of working parents.

What mental health services are being used by children, families, and professionals?

Stakeholders believe that mental health services play a critical role in supporting children, families, and early childhood professionals. They perceived that there is a significant gap in available services and there are not enough trained professionals to meet the need. Stakeholders wanted to quantify the need for mental health services for children, families, and early childhood professionals and identify gaps in services.

Are children making smooth transitions among early childhood programs? Are they making smooth transitions from these programs into kindergarten?

Stakeholders want to know if children are able to access early childhood programs and services—including kindergarten—without experiencing gaps in needed services, encountering problems with information sharing, or other potential issues. Specifically, they expressed concern that when children transition among programs, important information about them—such as assessments—does not follow them. Participants were also interested in what supports could be provided to parents and families who are navigating transitions, and what barriers exist that could be removed.

Are the services we are implementing cost-effective?

Stakeholders want to know if the benefits of early childhood education services align with or outweigh the costs.

EXECUTIVE INPUT

3Si/Foresight shared the stakeholder input with a smaller group of Cabinet members and agency leaders. Executives discussed the stakeholder input and identified their top two priorities among them:

1. Which children are served by which programs?
2. What assessments are being done on children? How are assessment results being used to help children?

Agency executives shared some additional insights and priorities that informed development of the priority use cases:

- Many of the executives interviewed wanted to learn more about outcomes for children and families who receive services. Specifically, they wanted to learn how different early childhood programs—as well as duration and quality of those programs—affect outcomes for children once they enter the K-12 education system.
- Executives wanted to understand how transitions among programs and services work for children, particularly children receiving educational and early intervention services under the Individuals with Disabilities Act (IDEA). They expressed interest in learning if some children are being dropped from the system in the transition, or if they need to get reassessed when they transition.

- The Oklahoma Department of Human Services (OKDHS) expressed interest in improving eligibility determinations for social safety net programs and enhancing coordination between programs. Similarly, the Oklahoma State Department of Health (OSDH) expressed interest in identifying and informing families of their eligibility for the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) and Supplemental Nutrition Assistance Program (SNAP), based on their Medicaid eligibility.
- OSDH also expressed interest in integrating tribal data, as tribes run social safety net and early childhood programs that serve some of the same children and families.

OKFUTURES PRELIMINARY DRAFT NEEDS ASSESSMENT

Next, 3Si/Foresight reviewed a preliminary draft needs assessment prepared by Urban Institute.⁵ Building on stakeholder and executive input and incorporating outcomes of the preliminary draft needs assessment, the 3Si/Foresight team fleshed out five potential priority use cases.

Use Case #1: What and where are there gaps in early childhood education services and quality, and what options exist to close those gaps?

The *OKFutures Needs Assessment—Preliminary Draft for Review* estimates that approximately 124,000 eligible children are not being served by one of the three primary early childhood education programs (licensed child care, pre-kindergarten, or Head Start/Early Head Start). This statistic suggests that at least some children who are eligible for and could benefit from these services are not accessing them. Furthermore, while Oklahoma closely monitors and tracks the quality of its licensed child care programs, there are no disaggregated data by race, ethnicity, and income on the quality of programs young children attend.⁶ Identifying and reaching children and families who should be but are not receiving services is part of the vision for the ECIDS and a key reason why stakeholders and executives prioritized understanding which children are served by which programs.

Program administrators, policymakers, and advocates need to know the entire birth-to-five population in Oklahoma by child and family demographics, unduplicated counts of how many of these children are currently being served by which combinations of early childhood programs, and at which levels of quality. They also need to know which children are currently unserved or underserved, by child and family demographics. This information would help Oklahoma identify and reach children and families who should be but are not receiving services, address potential inequities in quality, and isolate

⁵ From *OKFutures Needs Assessment—Preliminary Draft for Review*, Erica Greenberg, Natalie Spievack, Victoria Rosenboom, Michael Katz, Grace Luetmer, Mary Bogle, and Catherine Kuhns at the Urban Institute, July 2019: “The OKFutures needs assessment is the first of five activities to be completed under the OKFutures grant. It provides the rationale for a future five-year strategic plan and will inform new efforts to maximize parental choice; share best practices to increase program quality, collaboration, and efficiency; and improve overall quality across the early childhood education mixed delivery system. The needs assessment also serves as a baseline against which to measure future change. The OKFutures needs assessment is both a cumulative assessment of the current early childhood education mixed delivery system and a roadmap for its path to excellence.” Page 2.

⁶ Greenberg, Spievack, Rosenboom, Katz, Luetmer, Bogle, and Kuhns. *OKFutures Needs Assessment—Preliminary Draft for Review*, pages 19, 25.

geographic areas that do not have enough supply of quality early childhood education services to meet demand.

Use Case #2: What are the factors driving the significant gap between referrals for early intervention services for children with disabilities and developmental delays (SoonerStart) and children actually receiving these services? What options exist to close this gap?

The *OKFutures Needs Assessment—Preliminary Draft for Review* indicates that Oklahoma’s early childhood education system struggles to meet the needs of children with disabilities. As evidence, it cites that 71 percent of children who are referred to SoonerStart (Oklahoma’s early intervention program for children with disabilities and developmental delays) are deemed ineligible or opt out of service. It further cites that the rate of children receiving services under the federal Individuals with Disabilities Education Act (IDEA) is well below the national average for children birth to three (0.71 percent in Oklahoma versus 1.24 percent nationally) and for three- to five-year-olds (1.65 percent in Oklahoma versus 3.12 percent nationally). The Needs Assessment postulates that two factors drive this trend: First, state budget constraints have led to reductions in service providers and resource coordinators. Second, Oklahoma’s eligibility requirements for services under IDEA are more restrictive than some states, which may explain why many children who are referred to SoonerStart are deemed ineligible for services.⁷

As with the previous use case, this use case was developed in response to stakeholders’ and executives’ priority of identifying and reaching children and families who should be but are not receiving services. It also responds to interest at the Governor’s office in understanding if and how transitions for children receiving services under IDEA were affecting their access to needed services.

To address these questions, SoonerStart administrators from the Oklahoma Department of Education and policymakers would need to know:

- The extent to which limited service availability affects whether children with disabilities are able to access educational and early intervention services.
- The extent to which restrictive eligibility requirements or issues with eligibility assessments following children as they transition between programs impacts their access to educational and early intervention services.
- How these potential barriers differ across geographic regions and/or populations.
- What options exist to address these potential barriers.

Use Case #3: If a child or family is eligible for a social safety net program—such as Temporary Assistance for Needy Families (TANF) or subsidized early childhood education—what other social safety net programs are they eligible to receive? What options exist to better combine and deliver these services?

According to the Needs Assessment, early childhood education programs play a key role in connecting eligible children and families with social safety net services, often providing referrals and assistance with completing applications. The Needs Assessment examines the extent to which early childhood education

⁷ Greenberg, Spievack, Rosenboom, Katz, Luetmer, Bogle, and Kuhns. *OKFutures Needs Assessment—Preliminary Draft for Review*, page 29.

settings help vulnerable and underserved children and families access food assistance, housing support, and economic assistance. It reports that only a portion of the eligible population that could benefit from these services receives them, and barriers to access persist for low-income families, communities of color, and families in rural areas.⁸

As with Use Cases #1 and #2, this use case supports the vision for the ECIDS that was articulated by stakeholders and executives: identifying and reaching children and families who should be but are not receiving services. This use case also reflects two priorities for OKDHS: streamline eligibility determinations for social safety net programs to reduce the administrative burden on state agencies, providers, and families, and enhance coordination between services.⁹

Program administrators and policymakers need linked data across social safety net programs to streamline the process for determining eligibility for social safety net programs based on eligibility for other programs.¹⁰ Especially for children who are not enrolled in early childhood education programs and therefore not benefitting from referrals and assistance with applications for programs, linked information would provide program administrators with other ways to refer vulnerable children and families to food, housing, and economic support.

Use Case #4: What barriers do insured families face in accessing health services, especially in rural areas? What options exist to reduce these barriers?

According to the Needs Assessment, even families who have health care coverage have difficulty accessing basic health services.¹¹ Like social safety net programs, issues of health and health care overlap with education. Indeed, Oklahoma Head Start directors identify health as a number one priority, citing lack of access to screenings and medical services across the state and particularly in rural areas as a key concern.¹²

State health policymakers and program administrators need information on which children ages birth to five are receiving which health services, based on geographic location, demographics, family income, and health insurance coverage. They also need information on what barriers might inhibit access to those services, such as lack of available providers or providers who will accept their insurance, quality, or affordability. While there is no agreement on how best to close the gaps in healthcare access, this information will help clarify where there are gaps and who is affected. This information, in turn, would inform decisions about how to address these gaps.

⁸ Greenberg, Spievack, Rosenboom, Katz, Luetmer, Bogle, and Kuhns. *OKFutures Needs Assessment—Preliminary Draft for Review*, page 44.

⁹ 3Si/Foresight interview with Steve Buck, Secretary of Human Services (July 31, 2019).

¹⁰ Note that applications for enrollment in TANF, SNAP, child care subsidy, and Medicaid are currently possible with a single point of contact, so some level of this integration already exists in Oklahoma. Table 2-1 details other services relevant to these use cases.

¹¹ Greenberg, Spievack, Rosenboom, Katz, Luetmer, Bogle, and Kuhns. *OKFutures Needs Assessment—Preliminary Draft for Review*, page 41.

¹² Ibid.

Use Case #5: How do early childhood education programs affect longer term academic and behavioral outcomes?

Based on the vision for the ECIDS developed by stakeholders and executives, linked data would allow stakeholders to evaluate long-term outcomes of early learning, such as school readiness and behavioral and health outcomes. According to the Needs Assessment, data on children’s outcomes from early childhood through elementary school would provide an opportunity to study how variations in program features—including quality—translate into children’s success in school and beyond.¹³

Policymakers and program administrators need information on how early childhood education services tie to educational and behavioral outcomes, by location and child/family characteristics, to better understand the impact of those services. This information, in turn, would help inform investments in early childhood education.

2.2.3 Value Proposition for ECIDS

Table 2-1 describes the five priority use cases, the agencies that have data to contribute to each use case, and the value propositions for participating agencies (i.e., the benefits that participating agencies receive in return for their investment in the ECIDS). These five priority use cases are under development and there will be opportunities to refine, expand, and build upon them. Furthermore, as the ECIDS becomes more used and useful, we anticipate the Governor and stakeholder agencies will propose additional use cases.

¹³ Greenberg, Spievack, Rosenboom, Katz, Luetmer, Bogle, and Kuhns. *OKFutures Needs Assessment—Preliminary Draft for Review*, page 37.

Table 2-1 - Priority Use Cases, Agencies Involved, and Value Propositions

Use Case	Agencies Involved	Value Proposition
USE CASE #1: GAP ANALYSIS FOR EARLY CHILDHOOD EDUCATION		
<p>To develop an effective mixed-delivery early childhood education system, program administrators, policy-makers, and advocates need to identify service gaps, potential inequities in the level or quality of service, and geographic locations where there is not enough or adequate services to serve eligible children. Specifically, they need data on:</p> <ul style="list-style-type: none"> • The entire population of Oklahoma children, age birth to five, by child and family demographics. • Unduplicated counts of how many of these children are being served by which combinations of early childhood programs, and at which levels of quality. • The number of children who are currently unserved or underserved, by child and family demographics. 	<p>Head Start programs</p> <p>Oklahoma Child Care Resource and Referral Agency (OCCRRA)</p> <p>Oklahoma Department of Human Services (OKDHS)</p> <p>Oklahoma Department of Mental Health and Substance Abuse Services (ODMHSAS)</p> <p>Oklahoma Health Care Authority (OHCA)</p> <p>Oklahoma State Department of Education (OSDE)</p> <p>Oklahoma State Department of Health (OSDH)</p>	<p>Use Case #1 would help fulfill the vision for the ECIDS developed by stakeholders and executives: to identify and reach children and families who should be but are not receiving services. Specifically, the Governor’s Cabinet, participating agencies, and other key stakeholders would be able to:</p> <ul style="list-style-type: none"> • Understand which children are being served by which programs. • Understand how their programs are being used in combination with other government services. • Gain a clearer picture of the range of services available or unavailable within a specific community and allocate resources accordingly. • Develop options to close these gaps to deliver service(s) to unserved and underserved children.

Use Case	Agencies Involved	Value Proposition
USE CASE #2: GAP BETWEEN REFERRALS AND RECIPIENTS OF EARLY INTERVENTION SERVICES FOR CHILDREN WITH DISABILITIES		
<p>To improve Oklahoma’s ability to serve children with disabilities or developmental delays and to understand what happens to children when they are not referred for further screening/program services, administrators, public health officials, and policy advocates need to know:</p> <ul style="list-style-type: none"> • The extent to which lack of available or adequate services limit access to educational and early intervention programs. • The extent to which restrictive eligibility requirements or issues with eligibility assessments following children as they transition between programs limit access to educational and early intervention programs. • How these potential barriers to access differ across regions and/or populations. • What options exist to address these barriers. 	<p>Oklahoma Department of Human Services (OKDHS)</p> <p>Oklahoma State Department of Health (OSDH)</p> <p>Oklahoma Health Care Authority (OHCA)</p> <p>Oklahoma State Department of Education (OSDE)</p>	<p>Use Case #2 would help fulfill the vision for the ECIDS developed by stakeholders and executives: to identify and reach children and families who should be but are not receiving services. Participating agencies would be able to:</p> <ul style="list-style-type: none"> • Determine if eligibility requirements need to be adjusted. • Identify and potentially address geographic areas where there are gaps in services. • Resolve other barriers preventing children in need of early intervention services from accessing them.

Use Case	Agencies Involved	Value Proposition
USE CASE #3: ELIGIBILITY DETERMINATION FOR SOCIAL SAFETY NET PROGRAMS		
<p>OKDHS is interested in linking its data systems to streamline eligibility determinations and improve coordination among services. Especially for children ages birth to five who are not enrolled in early childhood education programs, linked information would provide program administrators with other ways to refer eligible children and families to food, housing, and economic support.</p>	<p>Oklahoma Department of Human Services (OKDHS)</p> <p>Oklahoma State Department of Health (OSDH)</p> <p>Oklahoma Health Care Authority (OHCA)</p> <p>Oklahoma State Department of Education (OSDE)</p>	<p>Use Case #3 would help fulfill part of the vision for the ECIDS that was developed by stakeholders and executives: to identify and reach children and families who should be but are not receiving services and streamline and improve service delivery and effectiveness for those children and families who are receiving services. Participating agencies would be able to:</p> <ul style="list-style-type: none"> • Streamline eligibility determinations to reduce the burden on state agencies, providers, and families. • Improve coordination across social safety net programs. • Provide more ways to refer eligible families for the programs they need.
USE CASE #4: BARRIERS FOR INSURED FAMILIES ACCESSING HEALTH SERVICES, ESPECIALLY IN RURAL AREAS		
<p>State health policymakers and program administrators need information on which children ages birth to five years (based on location, family income, etc.) are receiving which health services, and what barriers might inhibit access to those services.</p>	<p><i>Not addressed in this iteration of the ECIDS plan, given available time and resources. Future expansion of the ECIDS plan would include this and additional use cases.</i></p>	<p>Use Case #4 would help fulfill the vision for the ECIDS developed by stakeholders and executives: to identify and reach children and families who should be but are not receiving services. Participating agencies would be able to identify and potentially address barriers to accessing health care, particularly in rural areas.</p>

Use Case	Agencies Involved	Value Proposition
USE CASE #5: MEASURE LONG-TERM EFFECTS OF EARLY CHILDHOOD EDUCATION ON ACADEMIC AND BEHAVIORAL OUTCOMES		
<p>Stakeholders, executives, and the Governor want to better understand the impact of early childhood education on future academic and behavioral outcomes, to improve service delivery and measure the benefits of state investments in early childhood education programs. Use Case #5 asks how early childhood education affects future outcomes, and how those outcomes may differ by geographic location and demographic characteristics of children and their families.</p> <p>According to the <i>OKFutures Needs Assessment—Preliminary Draft for Review</i>, CAP Tulsa—a nonprofit organization with a mission to help young children in lower-income families grow up and achieve economic success—is leading a landmark study that links universal pre-kindergarten and Head Start child-level data with K-12 school data, in an effort to understand the effects of pre-kindergarten and Head Start on outcomes later in life. The ECIDS presents an opportunity to adapt this approach to be used statewide.</p>	<p><i>Not addressed in this iteration of the ECIDS plan, given available time and resources. Future expansion of the ECIDS plan would include this and additional use cases.</i></p>	<p>Use Case #5 would help fulfill the vision for the ECIDS: to evaluate long-term outcomes of early learning, such as school readiness and behavioral and health outcomes. It would also address feedback from stakeholders and nearly every executive that measuring outcomes is a high priority. Participating agencies will be able to:</p> <ul style="list-style-type: none"> • Better understand the relationship of their programs to child outcomes, even when those outcomes are tracked by other agencies. • Measure the benefits of their program investments. • Use the information to improve outcomes for children.

2.3 Challenges and Opportunities

In addition to addressing the use cases described in Table 2-1, an ECIDS provides multiple benefits to participating agencies and stakeholders:

- **Improved productivity:** integrating all early childhood data cuts down on the time it takes to analyze that data, because employees would no longer have to build connections from scratch when they need to perform analysis or run a report. For example, the Department of Education (OSDE) frequently needs to match its data with child-level data from the Department of Human Services (OKDHS). A linked data system like the ECIDS would reduce the effort OSDE would need to invest in this time-intensive process. In another example, many participating agencies would need access to the most recent physical addresses for early childhood service providers. By investing in building and maintaining a common dataset of physical addresses, individual agencies would not have to recreate this dataset on their own.
- **Accessibility and collaboration:** an ECIDS is designed to make data accessible to users throughout the enterprise, rather than limiting access to just a few. It also enables more collaboration across agencies, departments, and geographies. Good data governance practices would ensure that even with more widespread usage, the data are used appropriately.
- **Consistency:** linking data in an ECIDS creates the opportunity to develop a single source of truth. A single source of truth is a concept in data management whereby everyone in an enterprise bases decisions on the same data from the same sources. Having no single source of truth can cause inefficiencies and inaccuracies in data output, leading to wasted time and resources. Worse still, the lack of a single source of truth may cause inaccuracies that lead to poor decision-making and adverse impacts on programs serving children. For example, if two agencies produce analyses of similar scope—such as determining how many children are eligible for early childhood education programs—but use data from different sources, the analyses may return different outcomes. This discrepancy could call into question the validity of both analyses. A single source of truth would encourage consistency in data output and make it easier for users to investigate and address discrepancies across analyses of similar scope.
- **Virtuous cycle:** as the ECIDS becomes more useful to participating agencies by providing them with information that enables better decision-making, stakeholders will be motivated to develop new use cases and generate more data to support those use cases. This creates a virtuous cycle by which the ECIDS becomes more valuable to and valued by participating agencies.

Unfortunately, many states have had difficulty realizing the promise of an ECIDS. This is because the traditional model for integrating and using data has multiple challenges: it takes time and effort to link disparate data, the process of requesting and accessing data can be laborious, and once states receive the data, its quality can be unreliable.

The emergence of cloud computing has enabled a more efficient and effective approach to integrating and using early childhood data. The following section describes how cloud technology addresses many of the challenges of the traditional model for integrating data.



TIME AND EFFORT TO INTEGRATE DISPARATE DATA

Integrating data to conduct analyses currently takes substantial time, effort, and resources. This is because administration of early learning in Oklahoma is fragmented, as it is in most states. Early childhood data systems are managed by multiple agencies, all of which have mandates that go beyond early childhood education. Other data that would be useful to integrate with early childhood data—such as data on juvenile justice—also involve disparate data systems and agencies.

Integrating data from disparate systems requires resolving differences among uncoordinated data before it can be used for analysis. For example, agencies and data systems may have different ways of identifying the same child or program, update data at different frequencies, or use incompatible processes. These differences need to be addressed before the data can be integrated.

Cloud technology creates an opportunity to shorten the time between accessing and using data in decision-making. Current cloud technology, with its faster processing speeds, almost limitless storage space, and automated processes, allows for more efficient cleaning, transformation, and integration routines.

For example, the cloud offers algorithms that identify and match data from multiple data sources. These algorithms would make it faster and easier to match data elements—such as physical addresses for service providers—from disparate data systems. Using these algorithms would not eliminate the need for oversight, but it would reduce (or, in some cases, eliminate) the need to write and update code to handle these functions, which can be time consuming and inefficient.

Although cloud technology offers many advantages, Oklahoma's most important resource remains the technical and subject matter experts. Even with the cloud, Oklahoma will need human resources—both internal human resources and consultants—to write the programs that will match, clean, and reformat data and provide content expertise about the data

CLOUD COMPUTING AND DATA MANAGEMENT

Cloud computing uses a network of remote servers hosted on the Internet to store, manage, and process data, rather than an on-premise server or a personal computer. Cloud computing has had a major impact on how organizations manage their data. Cloud computing offers the following advantages:

- *It is remote, eliminating the time and expense of building and maintaining on-premise servers.*
- *It is more flexible and scalable than on-premise servers because it has virtually unlimited capacity, so when an organization needs more capacity, it is immediately available.*
- *It includes new technologies that automate processes like data matching, overcoming the need to do this time-consuming task manually.*
- *Cloud providers integrate with analytic tools, making it easier to do advanced analytics.*

sources as they are being integrated into a single source of truth. This combination of cloud technology, leveraged by talented staff, provides the most efficient path to a successful data product.¹⁴



FRUSTRATING DATA REQUESTS

Agencies that hold early childhood data have a legal obligation to make sure only the right people have access to specific pieces of information; otherwise, they place their entire organization—not to mention the children and families represented in the data—at risk. To meet this obligation, current data governance models focus on establishing legal and regulatory processes governing extraction, transformation, and use of specific data fields. Furthermore, because of the fragmented nature of early childhood education, data governance in Oklahoma involves multiple agencies with different data management processes. This translates to a lengthy set of steps for accessing data.

What typically happens is a user goes through all of the steps to request and acquire data, only to encounter anomalies in the data—for example, the description of a data element does not match the actual data—that require them to go back and redo all of the data request steps. This can create a circular process that further delays the ability to use the data.

A more efficient approach allows the user to confirm which data they need *before* completing data requests. Cloud technology makes it relatively fast and easy to pull raw data in different file formats into one place. The user could assess the raw data first, determine which data files they need, and then submit data requests, thereby avoiding a lengthy and frustrating circular process.



UNRELIABLE DATA QUALITY

This is not unique to Oklahoma, the field of early learning, or even government in general. Most data—even those used by the private sector—have quality issues. A study by the Harvard Business Review published in 2017 found that, on average, 47 percent of newly-created data records have at least one critical error, and less than three percent of the sample was considered of “acceptable” quality.¹⁵ A separate study by Harvard Business Review estimates that “knowledge workers” (i.e., people who need and use data to do their jobs) waste 50 percent of their time hunting for data, finding and correcting errors, and seeking confirmation for data they do not trust.¹⁶ The authors of this study hypothesized that many of the errors associated with poor data quality can be attributed to data being managed by different business units than those that are analyzing and using the data in decision-making. In other words, the people managing the data do not understand how the data will be used and are not directly impacted by poor data management practices.

Cloud technology allows for improved data management routines in two ways. First, cloud technology offers a range of tools that provide automated and consistent processes for integrating data from disparate sources. Once data is in the cloud, data engineers and data scientists can use the tools best

¹⁴ See Section 4.2.7 for more information about internal staff capacity requirements.

¹⁵ <https://hbr.org/2017/09/only-3-of-companies-data-meets-basic-quality-standards>, accessed July 17, 2019.

¹⁶ <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>, accessed July 17, 2019.

suited to their specific needs. For example, data engineers can use advanced processing tools—such as Spark/DataBricks—to do the heavy lifting of extracting and transforming data, and data integration tools—such as Azure Data Factory—to provide an automated and consistent process for integrating data from disparate sources. By using these tools, data engineers can ensure data quality and consistency through automation. Similarly, data analysts and scientists can access analytic tools such as SQL, Python, and Tableau.

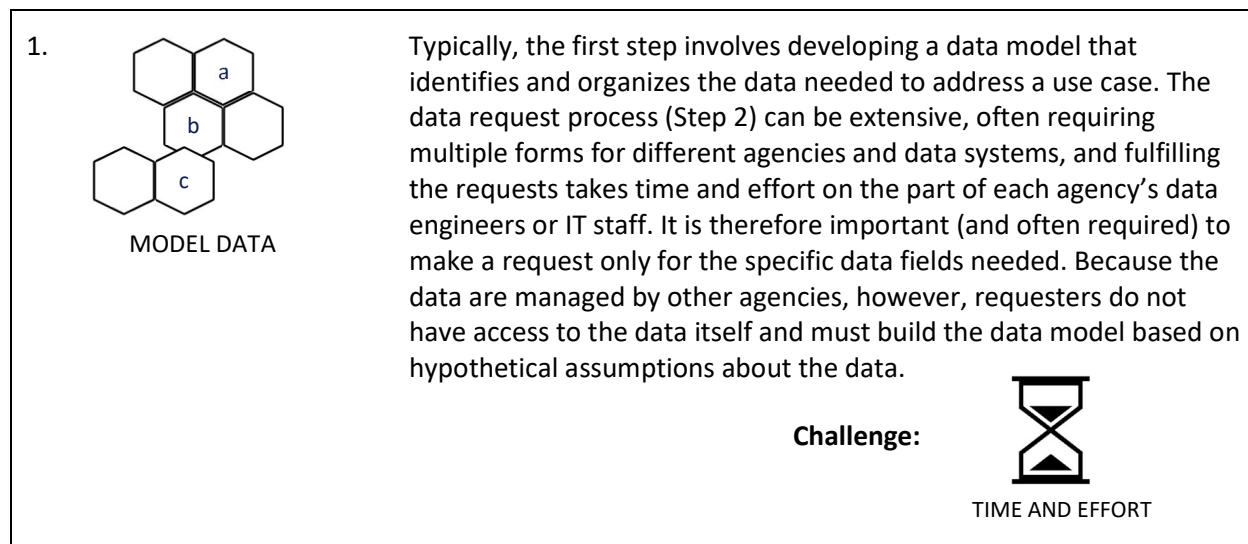
Second, by centralizing raw data, cloud technology makes it easier for data engineers and data scientists to work together to clean and transform data. This results in higher quality data than when data engineers and data scientists work in silos. If they find data anomalies, data engineers and data scientists can collaborate to investigate the raw data, determine the root cause of the anomalies, and adjust the data integration process accordingly.

2.3.1 Oklahoma's Opportunity

Oklahoma is well-positioned to realize the promise of an ECIDS. This is partly because Oklahoma does not have an existing ECIDS and is not anchored to old infrastructure investments and practices. To illustrate how Oklahoma can leverage new technology and data integration practices, it is helpful to understand the traditional method of integrating data and compare it to new approaches used in the cloud.

The traditional approach to integrating data from multiple, disparate systems involves extracting data from their sources and loading them into what is called an enterprise data warehouse, which is a centralized repository for cleaned and transformed data. Figure 2-2 provides a simplified illustration of the steps in this process.

Figure 2-2 - Traditional data integration approach



2.



IDENTIFY DATA SOURCES AND
DATA REQUEST PROTOCOLS

Before requesting data, requestors need to identify the source(s) of the data and the protocols for requesting these data. Because requestors have to build a data model, they have to determine not just data sources but specific data fields to populate the model. Determining information at this level of granularity (i.e., literally every single data element) involves a tremendous amount of time and effort.

Challenge:



TIME AND EFFORT

3.



COMPLETE DATA REQUESTS

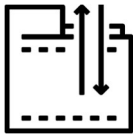
The requestor completes and submits data requests to the appropriate agencies. As described in Step 2, completing data requests for each separate field and data element is a laborious task.

Challenge:



FRUSTRATING DATA REQUESTS

4.



EXTRACT DATA

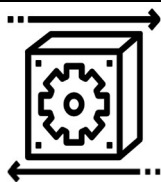
The agency who owns the data reviews a data request. Once a request has been approved, a data engineer prepares the data for extraction. This data engineer, who works for the agency that holds the source data, is unlikely to have a relationship with the requestor's program or know much about the use case. As such, they probably won't have the context to know if the requestor made correct assumptions about the data (Step 1) and requested the data elements they actually need. Furthermore, as mentioned earlier, because the data engineer is not the end user of the data, they may not be aware of data quality issues that affect the use case.

Challenge:



POOR DATA QUALITY

5.



CLEAN AND TRANSFORM
DATA

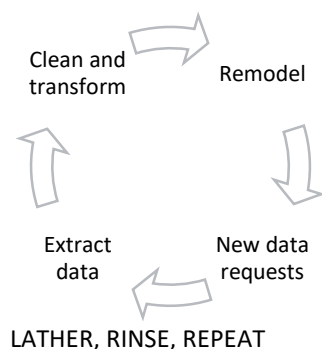
When all the requested data are extracted to a staging area, the requestor reviews the data and determines what needs to be done to match and integrate the data. This may involve another data engineer cleaning and reformatting data and addressing anomalies, such as null fields. Again, if the data engineer has limited understanding of the use case, this could compromise data quality. Alternatively, the requestor could attempt to clean and reformat it themselves, but that also takes a long time and a lot of effort.

Challenge:



POOR DATA QUALITY

6.



After doing some cleaning and transforming, the requestor may realize they need more or different data to answer the use case, and/or some of the data are not what was expected, unusable, or have been updated since it was received. The requestor has to go back to Step 1 and redo the data model, and then submit new data requests. This circular process usually happens multiple times, until the requestor is satisfied they have the data needed to address the use case. This can feel like an endless cycle, because by the time the process is completed, a new set of issues may emerge and the requestor will have to embark on the process yet again.

Challenges:



TIME AND EFFORT



FRUSTRATING DATA REQUESTS

7.



LOAD DATA INTO WAREHOUSE

A data engineer loads the cleaned, transformed data into an enterprise data warehouse.

8.

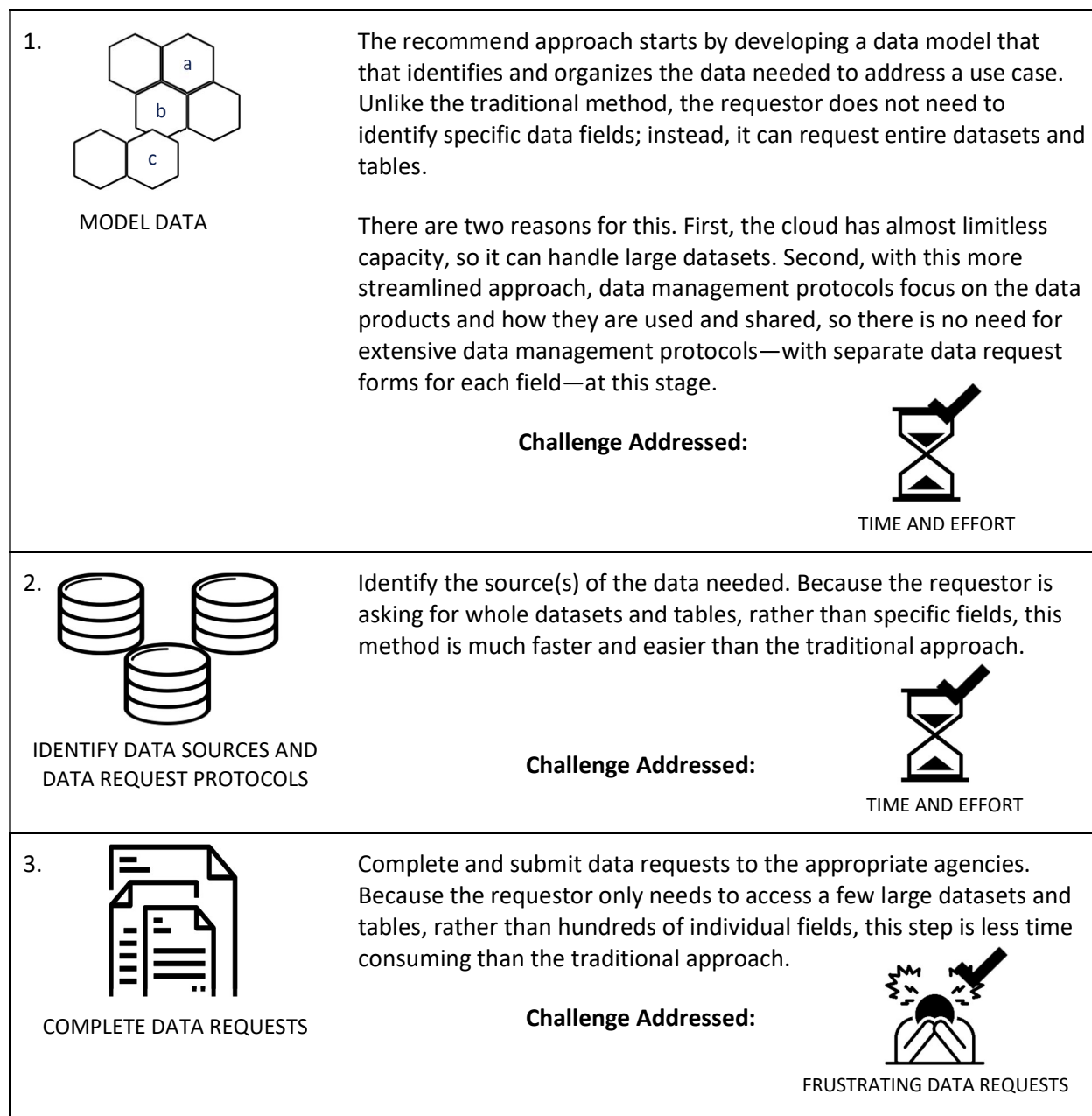


ANALYZE DATA

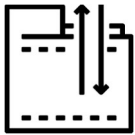
Data is ready to be analyzed by a data scientist and used in reports and other data products.

Our recommended approach streamlines this process. Figure 2-3 provides a simplified illustration of this approach.

Figure 2-3 - Recommended approach to data integration



4.



EXTRACT DATA

The agency reviews the request. Once approved, a data engineer prepares the data for extraction. This takes less time and effort than the traditional method because data are extracted in their raw, native format.

Challenge Addressed:



TIME AND EFFORT

5.



LOAD DATA

The agency loads the datasets and tables in their native form into a data lake, which is a repository for raw (i.e., not yet cleaned or transformed) data in the cloud.

6.



CLEAN AND TRANSFORM
DATA

Review the data and determine what needs to be done to integrate it. Cloud storage offers automated tools for cleaning and transforming all kinds of data, speeding up the process and ensuring more reliable data quality. And because data are stored in their native format, if there are problems with the data it is easier to investigate and address the problems than if the data are stored in its transformed state. Finally, because it is stored in its native state, the data in the cloud update automatically when the source data updates.

Once the data are loaded into the data lake, experts who transform the data (e.g., data engineers) and those who analyze it (e.g., data scientists), work in consultation with experts on the data subject matter to clean and transform data. This results in higher quality data than what is yielded by the traditional data integration approach, in which data engineers and data scientists work in silos.

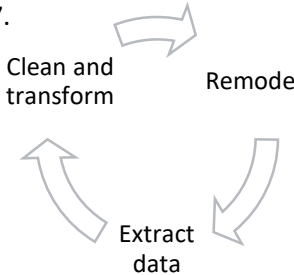


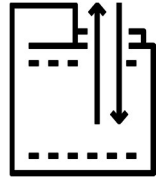


Challenges Addressed:



TIME AND EFFORT



POOR DATA QUALITY

<p>7.</p>  <p>Clean and transform</p> <p>Remodel</p> <p>Extract data</p> <p>LATHER, RINSE, REPEAT</p>	<p>As with the traditional data integration method, the requestor may realize they need more or different data to answer the use case. If this happens, they will need to redo the data model. Because the data request process is much more streamlined, the cycle will go much faster and encounter fewer bottlenecks. In fact, because the requestor extracted whole datasets, they may already have the data they need in the data lake.</p> <p>Challenge Addressed:</p>  <p>TIME AND EFFORT</p>
<p>8.</p>  <p>ANALYZE DATA</p>	<p>Data is ready to be analyzed and used in reports and other data products.</p>
<p>9.</p> 	<p>Data management is focused on ensuring the data are being used appropriately and shared in a way that adheres to data privacy standards and regulations. The cloud automates many of these approval processes, making it faster and easier to get authorization to use and share data products.</p> <p>Challenges Addressed:</p>  <p>TIME AND EFFORT</p>  <p>FRUSTRATING DATA REQUESTS</p>

2.4 Purpose of this document

This document combines the Data Inventory, Data Integration Plan, and Data Governance Plan into one ECIDS plan.

The Data Inventory, Data Integration Plan, and Data Governance Plan include specific design elements that leverage cloud technology and our recommended approach to data integration to minimize or overcome many of the challenges described in Section 2.3. For example, the data inventory emphasizes data systems and tables, rather than more granular data fields. The data integration plan is designed to leverage cloud technology to help drive down costs and increase flexibility and functionality. Finally, the data governance framework protects data privacy and security while making it faster and easier for stakeholder agencies to analyze and use data in decision-making.

The ultimate goal of this document is to provide a buildable design for an interagency data system and a state-level governance system. And finally, the stakeholder engagement associated with development of this document—particularly the data governance framework—is intended to foster shared ownership of the ECIDS by ensuring key stakeholders understand the value proposition and the resources they need to invest in this project.

Sections 2.4.1-2.4.3 below describe what is included in each of these sections.

2.4.1 Data Inventory

The data inventory is a list of datasets needed for each priority use case and a description of their contents, source, owner, and elements we need to know or address before extracting and/or transforming data. This data inventory is not exhaustive; its scope is broad enough to address priority Use Cases #1, 2, and 3 and incorporate additional data sets and use cases in the future. The data inventory:

- Describes priority use cases.
- Assesses high-level feasibility of these use cases.
- Identifies and describes the datasets needed for each priority use case.
- Prioritizes each dataset by importance (e.g., 3=critical to data product, 2=important, but not absolutely necessary to data product, 1=nice to have).
- Identifies the source of each dataset to the extent possible given time constraints (i.e., the owner, location of data, and access requirements).
- Describes each data asset (i.e., content type, record series, metadata, etc.)
- Identifies challenges or barriers that need to be addressed before extracting data.

2.4.2 Data Integration Plan

Data integration is the process of combining data from multiple and disparate sources into meaningful information. The data integration plan describes the 3Si/Foresight team’s approach to extracting the data from the Oklahoma Departments of Health, Human Services, and Education, among other potential data sources, and the steps we would take to combine the data and prepare it for use in data products and analyses. The data integration plan:

- Assesses readiness for automated data synchronizing routines to extract data from existing systems, combine and merge provider records into master lists, and load data into ECIDS.
- Assesses options for data transportation (moving data from its source to its destination).
- Describes recommendations for the technical ownership, storage, and management of data.
- Assesses options for security of individual-level data for each agency.
- Describes initial analytic outputs the ECIDS will generate for the use cases and the resources, tools, and types of data integration that will be required to support these initial data products.
- Recommends processes for developing and maintaining data products in the ECIDS.

2.4.3 Data Governance Plan

The Data Governance Framework is an interagency structure for managing and maintaining the ECIDS and collecting, managing, and reporting the integrated data. The data governance framework:

- Documents Oklahoma’s existing data governance frameworks.
- Describes the structure for agencies to manage and make decisions about data. The framework delineates composition and membership, roles and responsibilities, hierarchy, and issue resolution.
- Provides a structure for overall management of ECIDS.
- Provides a method for adding new agencies to the governance structure in the future.
- Identifies updates to Oklahoma’s laws and regulations necessary to support the proposed governance framework.
- Describes personal information policies for each dataset (where possible as time and budget allows).
- Identifies necessary modifications to existing interagency agreements and/or new agreements.
- Proposes privacy and security protocols that meet federal and state requirements and balance improved access to data with ensuring data is used for legitimate educational purposes only.

3.Data Inventory

3.1. Introduction and Background

This data inventory is a description of the data assets related to early childhood education and outcomes that are available to Oklahoma. The data inventory provides the details needed to locate, request, extract, transform, and use data to answer the priority use cases described in Section 2.2.2. The data inventory also identifies potential data gaps that would need to be addressed to answer the priority use cases. As noted later in this section, this preliminary data inventory will need to be expanded and updated over time as new information becomes available, including assessment of actual data for quality and completeness.

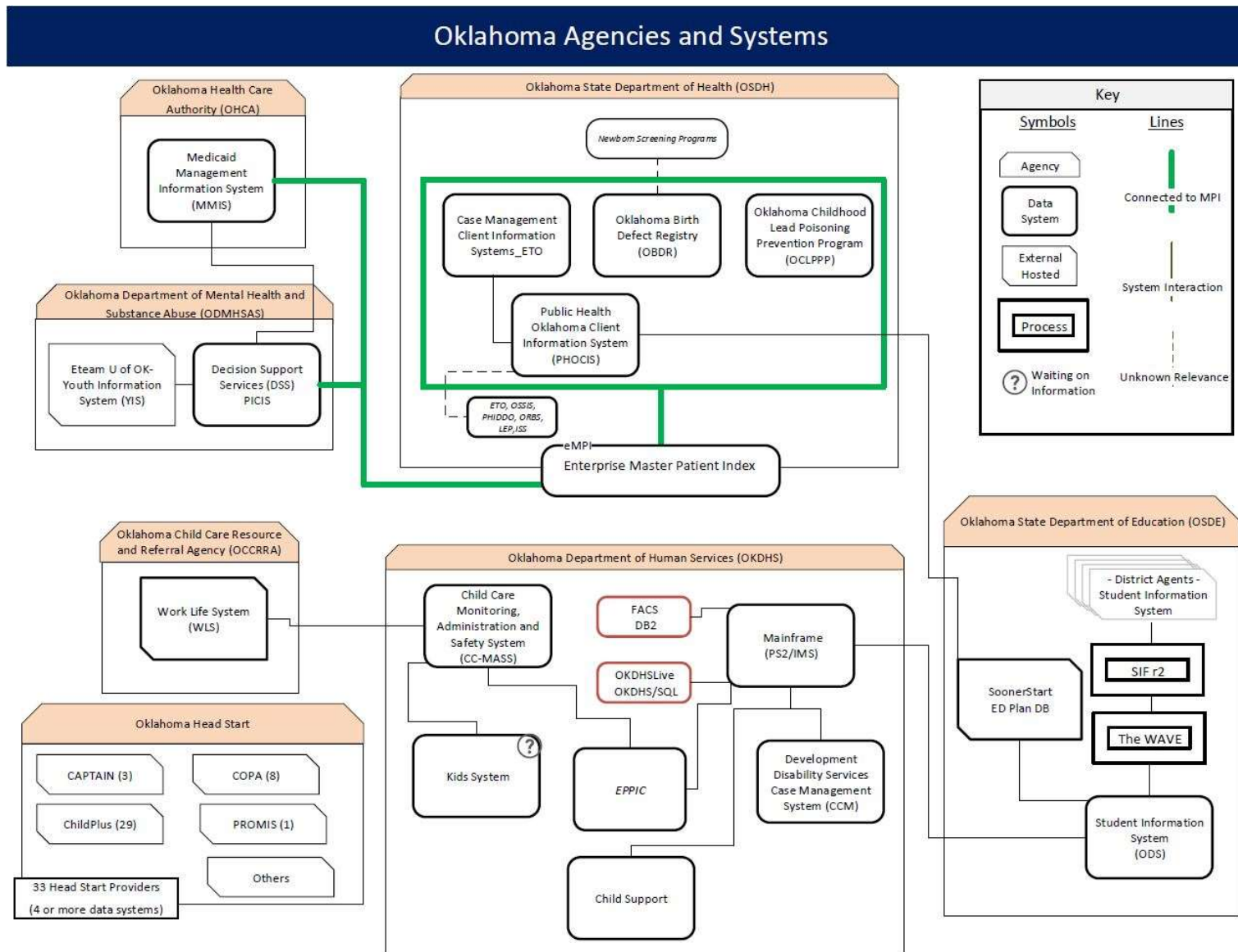
The data inventory includes basic information about each data asset, including its name, description of its contents, data source, the frequency of updates, who owns and manages the data, and other relevant details. The data inventory also describes how each data asset relates to the priority use cases.

The data inventory informs our recommendations for data integration, data transport, and data security described in Sections 4 and 5 of this document. When Oklahoma stakeholders are ready to develop new use cases, the data inventory will help identify available data assets and those that would need to be developed.

3.1.1. Data Landscape Map

Figure 3-1 (on the following page) is a data landscape map that shows the data systems that could provide early childhood education data and how they relate to each other. The landscape map represents what exists and is known today, but it is not intended to capture every data system or table in Oklahoma that could relate to early childhood education. *[Note: The data landscape map shows only the systems identified as potentially useful to Use Cases #1-3. The systems map will likely need to be expanded to include additional systems relevant to Use Cases #4-5 and other use cases over the course of ongoing ECIDS development.]*

Figure 3-1- Oklahoma early childhood education data landscape map



3.2. Methodology

As illustrated in the data landscape map above, numerous data systems exist in Oklahoma that are in some way connected to early childhood education. Not all of these data systems are directly related to the priority use cases described in Section 2.2.2. To avoid “boiling the ocean” and investing excessive time and resources in investigating data systems that are tangentially related to the ECIDS, the 3Si/Foresight team—with agreement from OPSR—adopted an approach of successive inquiries. Each successive inquiry went into greater depth and was more targeted to the data systems that are relevant to the priority use cases.

For the first inquiry, 3Si/Foresight worked with OPSR to identify the agencies and data systems that seemed relevant to the broad scope of the ECIDS. We prioritized the agencies and data systems based on the following criteria:

- Tier 1 – These agencies own data or manage systems or infrastructure that are critical to Oklahoma’s prioritized use cases. 3Si interviewed all Tier 1 agencies.¹⁷
- Tier 2 – These agencies own data that may support the priority use cases. Agencies can also be Tier 2 if they do not own data but can provide valuable perspective on the use of data (for example, they aggregate data from other agencies and therefore understand some of the challenges and limitations of the data). 3Si/Foresight interviewed Tier 2 agencies to the extent possible given the time available.
- Tier 3 – These agencies do not own data with a direct connection to the use cases but may own data or systems that, when combined with other data, would expand ECIDS usability. For example, the Oklahoma Office of Juvenile Affairs data combined with the child program data could allow Oklahoma to connect early childhood education with behavioral outcomes. Given time constraints in producing this data inventory, 3Si/Foresight did not interview Tier 3 agencies. That said, agencies identified as Tier 3 are engaged in supporting development of the ECIDS and should continue to be engaged once ECIDS plans are more developed. Appendix C.1 lists the agencies related to early childhood education by tier status. 3Si/Foresight met with each Tier 1 agency and some Tier 2 agencies and requested system documentation. We confirmed the data systems they own and used a template to organize high-level information on each data system and, where appropriate, relevant tables.

Next, we identified the data needed for each priority use case. Based on the information collected in the initial inquiry, 3Si narrowed the list of data systems to those most relevant to the priority use cases. In many cases, only a subset of data systems owned by Tier 1 agencies warranted a more detailed inquiry

¹⁷ Note: In Oklahoma, there are four primary data reporting tools and eight total data reporting tools used by 34 Head Start (HS) and Early Head Start (EHS) grantees across 78 counties. Based on combinations of a) tools used, b) number of counties served, and c) specific populations served (HS/EHS, tribal/nontribal, urban/rural), 3Si and Chapin Hall selected a representative subset of grantees for data inventory interviews. Appendix C.1. lists these grantees.

(for instance, 3Si only took inventory of six of the total 18 systems maintained by OSDH, based on early input from OPSR and OSDH data managers).

For the successive inquiry, we collected detailed information only on the target systems deemed necessary to answer the priority use cases. The 3Si/Foresight team identified which data systems are needed for each priority use case and any gaps in data availability. Based on this analysis, 3Si/Foresight assessed the feasibility of each priority use case.

The following sections summarize the findings of this data inventory and provide considerations for next steps.

3.3. Data Inventory Content and Findings

3.3.1. General Inventory of Agencies and Systems

The preliminary data inventory focused on systems-level information for Tier 1 agencies. See Appendix C.2 for a summary of the data inventory for each Tier 1 agency. While 3Si interviewed some Tier 2 agencies, the purpose of those interviews was to uncover context to support the data inventory approach; Tier 2 interviews did not reveal any data systems to be directly relevant to the priority use cases. As such, this data inventory does not list or summarize interviews with Tier 2 agencies.

3Si broke its data inventory into five areas:

- Basic system information
- Data history
- Technical information
- Usage
- Tables information

The system-level information helped us identify systems and system owners within an agency. System-level information provides the elements required to design the data integration plan (discussed throughout Section 4) such as vendor name, type of system, what data systems were connected to other systems, the process by which data was collected, and any data issues or years of missing data.

The table-level information collected for each Tier 1 system focused on the five primary content areas: family, child, provider, program, and assessments. This information helped us identify data systems that were relevant to the use cases, conduct additional investigation to determine the population covered within those systems (e.g., all children birth through age five or some subset of those children), and confirm data elements necessary to address the priority use cases. Appendix C.2 contains a summary of this system-level inventory of Oklahoma agencies and systems.

3.3.2. Data Needed to Fulfill Use Cases

This section describes the data needed to fulfill priority Use Cases #1, 2, and 3.

Needed Data

Population

Identification of the entire birth to five population is necessary to determine the potential unmet demand for services and ensure the ECIDS is scalable as it incorporates new data sources and serves new use cases. Failure to identify the entire population may limit the usefulness of results. For example, some data systems track only those children and families who have applied for and/or are receiving service, which could lead to a significant undercount of unserved or underserved children. Defining the entire birth to five population as completely as possible will help to establish a common frame of reference across use cases and over time as the ECIDS continues to expand.

Program Eligibility

Identification of unserved and underserved populations requires defining which children birth to five population are eligible for which programs, based on family income, work requirements, disabilities and other eligibility criteria. Segmenting the population based on program eligibility may require inclusion of additional data sources and/or inference to accommodate limited granularity of available data. For example, Oklahoma may have to rely on data that captures the licensed capacity for child care rather than the actual number of children served. Despite this initial limitation, data granularity is expected to improve over time as the ECIDS incorporates new data sources.

Some of the programs for Use Cases #1, 2, and 3 overlap and can be used to categorically determine eligibility for other programs associated with additional use cases. For instance, SoonerStart is included in Use Case #1 and is also the primary program of interest for Use Case #2. As another example, TANF is a criterion for Head Start eligibility in Use Case #1 and is also a program in Use Case #3. Considerations for determining eligible populations for Use Cases #1-3 are outlined below.

Use Case #1

3Si/Foresight selected the programs to consider as part of Use Case #1 based on multiple factors:

- Programs in Use Case #1 include an early learning component.
- Based on input from stakeholders and due to their alignment with the *OKFutures Needs Assessment—Preliminary Draft for Review*, we included several large, state-funded programs: Universal Pre-K, SoonerStart, and Subsidy Child Care.
- Though not state-funded, we included Head Start because it is a critical early learning program in Oklahoma.
- 3Si also recommends including all licensed child care in this use case, even though some licensed child care is neither state funded nor strictly early learning. Including all licensed child care may be useful in assessing the extent to which populations of interest for this use case are being served, especially because many licensed providers serve children with and without subsidies.

Eligibility indicators vary across these programs, so (to the extent possible) data modeling will need to account for the overlap/intersection between these eligibility requirements. Appendix C.3.2 summarizes eligibility requirements by category for each program. Appendix C.3.3 summarizes specific eligibility requirements by program.

Use Case #2

Based on stakeholder input and the draft Needs Assessment, 3Si/Foresight developed Use Case #2 based on the need to:

- Understand the gap in children who are referred to SoonerStart but do not receive services or drop out of receiving services.
- Assess what factors could be preventing children in need of early intervention services from receiving these services. Potential barriers could include the inability to track down guardians, guardians choosing to opt out, children who cannot be located, overly restrictive program guidelines, and/or the impact of budget cuts.

Use Case #3

3Si/Foresight selected social safety programs to consider as part of Use Case #3. The programs included for this use case have been approved by OKDHS with the understanding that other programs (both with OKDHS and other agencies) could be added at a later time.¹⁸

Similar to Use Case #1, the eligibility indicators vary across these safety net programs and data modeling will need to account for the overlap/intersection between these eligibility requirements. Appendix C.3.2 summarizes eligibility requirements by category for each program.

Children Served, Unserved, or Underserved

As the target programs often serve some of the same children, it will be necessary to determine a distinct count of children receiving any combination of these services. These distinct counts, in combination with the eligible population, will allow Oklahoma to quantify unserved or underserved children by child/family characteristics (e.g., geographic location, program eligibility, race, etc.).

In addition to knowing which children are served or unserved by these programs, data on the quality, duration, or dosage of these services may be relevant to each use case. 3Si has identified provider quality ratings, in particular, as a key data component. Several specific data elements will be needed to understand who is unserved and dosage of those who are served by SoonerStart in Use Case #2, including information on SoonerStart referrals and opt-outs, the number of visits administered, information on screenings (including exemptions to the formal screening process), and the customized service plans (Individualized Family Service Plan or Individualized Education Plan) for each child, which vary significantly in time to completion.

3.3.3. Feasibility of Use Cases

This section describes the current availability of data to answer priority use cases, including gaps in data availability that would limit feasibility or present implications for the data modeling, transportation, or integration. We also identify any remaining questions that will need to be answered to inform an

¹⁸ Although the Food Distribution Program on Indian Reservations (FDPIR) and WIC are not OKDHS programs, they were identified as social safety net programs with eligibility requirements that matched or aligned with OKDHS safety net programs.

assessment of use case feasibility. Appendix C provides a detailed summary of our assessment of data availability for each use case.

This analysis is limited by the time and resources available to investigate all potential data sources. We summarize the potential risks and mitigations below in Section 3.3.5.

As explained earlier, OPSR directed 3Si/Foresight to focus this data inventory on Use Cases #1-3, given their foundational nature and immediate relevance to primary stakeholders of Oklahoma's ECIDS. Investigation of data availability and feasibility for Use Cases #4-5 will start during the implementation phases of this project.

FEASIBILITY OF USE CASE #1: GAP ANALYSIS FOR EARLY CHILDHOOD EDUCATION

Use Case #1 appears feasible, despite some gaps in data as described below, because 3Si can use data proxies to temporarily make up for these gaps. While these gaps can be mitigated through the data modeling methodology described below in Section 4.2.5, the mitigation strategies will introduce their own limitations in terms of detail and accuracy of the resulting data model. For this reason and others, developing useful data analysis and data products for this use case early on will create a compelling justification for the continued improvement of data sources and related data model and analysis over time.

FEASIBILITY OF USE CASE #2: GAP BETWEEN REFERRALS AND RECIPIENTS OF EARLY INTERVENTION SERVICES FOR CHILDREN WITH DISABILITIES

Use Case #2 appears feasible. There are readily available data on developmental delay diagnoses, SoonerStart referrals, and service delivery. All referrals, services received, communication attempts, and interaction with parents are tracked.

FEASIBILITY OF USE CASE #3: ELIGIBILITY DETERMINATION FOR SOCIAL SAFETY NET PROGRAMS

Use Case #3 appears feasible, with the caveat that data proxies will need to be developed to fill in data gaps similar to those described in Use Case #1. Eligibility requirements for the safety net programs are very similar, and some overlap with other programs within OKDHS and OSDH. With a unified Master Person Index across agencies, recipients could be identified as eligible for other programs in other agencies.

For example, WIC is a program within OKDHS and has income levels that align with other OKDHS programs such as TANF and SNAP. In turn, those receiving SNAP who have eligible children or are pregnant would be eligible for WIC. Like Use Case #1, eligibility criteria for these programs will have some gaps in population and income data, but creating a data model that determines program eligibility across agencies will mitigate these data gaps.

This use case may also support OKDHS in establishing an agency MPI that links its data across OKDHS programs, and eventually links to data across OSDH programs. OKDHS has taken initial steps to establish an MPI, but this effort is still in its infancy.

GAPS IN DATA NEEDED TO FULFILL USE CASES AND MITIGATIONS

3Si/Foresight has identified several gaps and temporary mitigations from the preliminary data inventory:

1. *The data inventory identified potential gaps in available data to adequately define the entire population and demographics of children birth through five for all three use cases.*
 - The MPI will be limited to the children present in the data systems feeding the MPI. Initial estimates suggest that approximately 65 to 80 percent of children ages birth to five years living in Oklahoma are currently included in the MPI. The 20 to 35 percent of children not included in the MPI are largely children born outside of Oklahoma who are not receiving services through the OSDH.¹⁹ In early 2020, additional Vital Statistics records will be loaded into the MPI, so this data gap should reduce significantly. As the MPI incorporates additional data sources, we assume the proportion of children excluded from these data will shrink even further.
 - Some key demographics—such as household income, parent work status, and race—may be less available, which will limit the ability of the ECIDS to model program eligibility in some circumstances. 3Si/Foresight is not able to assess the completeness of these data with specificity, given the preliminary nature of this data inventory, but the findings suggest that available sources of demographic data will cover only a subset of the total child/family population. This finding presents two challenges. First, the data model may need to incorporate new data sources to help address these gaps. Second, when adding new data sources, there is a risk of conflicting information (e.g., two systems present different values for household income, one of which will need to be prioritized), so the data model will need to account for and address these conflicts. Section 4.2.6 discusses these approaches in greater detail.
2. *Gaps in available child-level data may require incorporation of aggregate data sources, such as census or other publicly-available data.*
 - If there is only limited child- or family-level data to define birth to five population and determine program eligibility, ECIDS data will need to be supplemented with sources of aggregate data. For example, incomplete child-level data on family income could be supplemented with data from the American Community Survey (ACS), which provides aggregated estimates of household income at varying levels of geographic granularity. This approach has some limitations. For example, incorporating aggregate data may allow stakeholders to identify geographic locations where eligible children and families are unserved or underserved, but it will not allow for child-level determinations of program eligibility. Section 4.2.6 of this report describes how aggregate data could be incorporated into the ECIDS to address each of the three use cases.
 - Beyond the limitations noted above for child *population* data, not all *services* are available with a level of granularity that will allow for child-level data matching. Specifically, OCCRRA does not have access to records of actual children receiving private (unsubsidized) licensed child care. The available data only includes licensed capacity and *preferred* capacity for child

¹⁹ 3Si/Foresight were unable to interview representatives of systems containing tribal-operated ECCE programs. As such, the project team was not able to assess whether these programs feed directly into the MPI, which would have an impact on which children are or are not present in the MPI.

care providers. As a temporary mitigation it will be necessary to use inference to determine which children are being served or unserved by licensed child care, based on child and family characteristics and reasonable assumptions about which families will be able to afford unsubsidized child care. The data model will also need to make assumptions to estimate the overlap of licensed child care with other services to estimate the distinct counts of children served across multiple programs. This finding presents significant implications for the data integration approach; Section 4.2.6 of this document discusses concrete examples of how to advance these use cases.

3. *Incorporation of additional programs into the MPI and ECIDS will gradually improve the completeness of available data and lower the reliance on external/aggregate data sources.*
 - Child-level data may replace data from aggregate sources over time. For example, early ECIDS implementation is likely to model Head Start data based on currently available aggregate counts of children served, using an approach similar to the example in the previous bullet for licensed capacity of child care providers. There is child-level data for Head Start enrollment, but it is owned by the 33 individual Head Start programs that serve Oklahoma, using at least eight third-party vendor solutions for data entry and storage (see Appendix C.6 for details).²⁰ Many Head Start programs work across multiple vendor systems simultaneously. Over time, with sufficient coordination with local Head Start programs and vendors, child-level data may be integrated into the ECIDS, precluding the need for this aggregate “plug” in the data model and improving overall data accuracy. The short- and long-term approaches to the ECIDS data model will consider the need for flexibility in swapping these aggregate data sources for child-level data over time.

3.3.4. Summary of Use Case Feasibility

Table 3-1 summarizes the feasibility of each use case, which agencies will be required to provide data, and selected next steps. Data availability at this level is summarized as:

- “High,” indicating good availability of data.
- “Medium,” meaning that some adjustments or workarounds will be required to model the data to fulfill the use case, but these adjustments or workarounds are within the control of ECIDS managers.
- “Low” would indicate unfeasibility without significant improvements to available data (though none of the assessed use cases fall into this category).

²⁰ 20 Head Start and 13 Tribal Head Start programs.

Table 3-1 - Summary of feasibility by use case

Use Case	Feasible?	Data Availability	Agencies Required	Next Steps
1. Gap analysis for early childhood education	Yes	Medium	<ul style="list-style-type: none"> • ODMHSAS • OHCA • OKDHS • OSDE • OSDH • Head Start programs 	<p>Some gaps in child-level data will require additional data sources and tailored solutions in the data model and integration plan. Next steps will include:</p> <ul style="list-style-type: none"> • Identification of additional data sources that can supplement ECIDS data. • Development of a data model to support this use case given the current state of data availability. • Establishment of short- and long-term plan for this use case, including the anticipated phase-out of external data sources.
2. Gap between referrals and recipients of early intervention services for children with disabilities	Yes	High	<ul style="list-style-type: none"> • OSDE • OSDH • OKDHS 	<p>There are no gaps in the data from referrals to receiving services. Lack of funding has led to lower levels of service per child overall and use of contractors to manage some of the case load.</p> <p>To account for these complexities, the data model could expand to facilitate:</p> <ul style="list-style-type: none"> • Comparison of child service delivery and outcomes before and after program budget cuts (in 2009) • Comparison of service plans relative to actual services provided • Comparison of referral trends and frequency of coordinators' active recruitment activities

3. Eligibility determinations for social safety net programs	Yes	Medium	<ul style="list-style-type: none"> • ODMHSAS • OHCA • OKDHS • OSDE • OSDH • Head Start programs 	Gaps for income data will be the biggest issue for finding eligible children that do not already participate in one of the safety nets programs. In addition to the mitigations provided above (for Use Case #1) to address this gap, a child's participation in one program will often imply eligibility for other safety net programs they are not receiving. For example, a child under five on TANF will qualify for WIC.
4. Barriers for insured families accessing health services	<i>[TBD – out of scope for these deliverables]</i>			
5. Long-term effects of early childhood education on academic and behavioral outcomes	<i>[TBD – out of scope for these deliverables]</i>			

[Note: Future work with OK ECIDS stakeholders will be necessary to assess the feasibility of Use Cases #4-5, along with other use cases identified going forward.]

3.3.5. Risks and Mitigations

The accuracy and completeness of this data inventory is highly dependent on access to accurate information and the availability of key stakeholders to participate in interviews. These and other key risks and proposed mitigations are as follows:

ENSURING DATA QUALITY AND/OR COMPLETENESS

3Si/Foresight's assessment is wholly dependent on the quality of existing documentation and/or data managers' ability to provide such assessments. Due to the limited scope of this project, 3Si/Foresight was not able to directly assess issues of quality or completeness of data, as such an assessment would require access to actual data for the purposes of pulling record counts, confirming values are within expected ranges, and testing across fields for internal consistency. Inaccessible, incomplete, or otherwise poor-quality metadata and other data descriptors could contribute to a data model design and integration plan that will need to be updated once assessments can be made on actual data.

3Si is mitigating this risk by producing a data integration plan and data models that are flexible and can be adjusted on an ongoing basis as new information is obtained. Section 4.2 discusses this flexibility and scalability in more detail. Oklahoma should also account for the lack of complete information in future planning; for example, it should build time to assess actual data into the timeline for developing an ECIDS during its implementation.

ESTABLISHING ENTIRE POPULATION OF CHILDREN BIRTH TO FIVE

As noted above, limitations of child-level data to define an entire population of children ages birth to five, including the child- and family-level demographics needed to establish program eligibility, could impact the ability of the ECIDS to address Use Cases #1-3.

3Si is mitigating this risk by presenting a data model and integration plan that leverage readily available aggregate data, and providing considerations for how to optimize data quality under this approach. Section 4.2.6 further discusses these considerations.

ADDRESSING CONFLICTING DATA FROM DISPARATE SOURCES

Data systems often have overlapping data elements that must be reconciled to ensure the ECIDS is populated with the most accurate data possible. For example, multiple data systems contain family income data, but each data system may use a different approach to grouping family income data.

This issue is already addressed to a large extent by the MPI, which addresses redundant data in its matching algorithm (for instance, prioritizing the most recent record for the place of child residence). As the data model becomes increasingly complex, the rules for prioritizing and reconciling these data sources will become increasingly important. Factors to inform this process are data recency, granularity, and reliability, among others. 3Si has further mitigated this risk by noting redundancy in source data to simplify data integration (see Appendix C for additional details).

INCOMPLETE INFORMATION FOR THIS DELIVERABLE

3Si/Foresight performed this data inventory with the best available information within the constraints of this project's limited scope. The project team did not have access to actual agency data, instead relying on documentation and metadata as well as the expertise of system owners. Eventually, the actual data will need to be assessed, and invariably, new insights will emerge. These new insights could necessitate adjustments to the proposed approach for data transportation, integration, and modeling. These new insights are unlikely to change the direction of 3Si/Foresight's feasibility assessment, however, given that other data systems that present implications for the data model and integration plan—such as Licensed Child Care—are likely to be unaffected by this incomplete information.

3Si has mitigated this risk with a flexible data model and integration plan.

3.4. Future Work to Expand and Refine Data Inventory

While this data inventory signals that Use Cases #1-3 are feasible, details regarding implementation are limited by the scope of this engagement. During implementation of this ECIDS plan, Oklahoma will need to expand this data inventory.

With that said, this section provides a preview of the process to expand and refine the data inventory. First, each data system owner should assemble the most up-to-date documentation on each of the systems currently in scope.²¹ In some cases, this documentation does not readily exist and will need to be created by the system owners.

This report's current data inventory does not include data from some key agencies (notably, information is lacking from tribal-run programs and services). Addressing these gaps in the data inventory should be a high priority.

As noted elsewhere, the data inventory does not assess actual data for quality and completeness. Assessment of actual data for quality and completeness is another important next step that will inform more detailed mapping (at the table and field levels) of data elements.

This process will be iterative, given the complexity of the ECIDS and the likelihood that its development will occur in phases. In particular, Oklahoma will likely continue to add data systems not currently in the data inventory, as has already happened several times during the short course of this project. For example, OKDHS maintains the OKDHSLive system, which stores records of Oklahomans who apply for OKDHS benefits and services. Although 3SI did not initially include the OKDHSLive data system in the data inventory because its data appeared to be redundant with other OKDHS systems, the use cases present a need to better understand which families are applying for services but deemed ineligible. As such, 3SI added the OKDHSLive system to this data inventory to fill in this information that would otherwise be missing.

Similarly, the assessment of feasibility and prioritization of these use cases will likely be iterative, as will any corresponding decisions around data modeling and integration. To manage these iterations efficiently, dedicated resources at each agency will need to be prepared to support these ongoing efforts (as discussed in Section 4.2.7).

²¹ Documentation would include entity diagrams, data dictionaries, etc.

4. Data Integration Plan

4.1. Introduction and Background

Data integration is the process of combining data from several different sources into a unified format so it can be analyzed and used in decision-making. By bringing together data from multiple sources—and multiple agencies—integrated data allows for new and more complex questions to be answered. Even within agencies, integrating data across source systems can produce new insights that aid in better operational and policy management.

Without unified data, analysis and reporting may involve requesting data from multiple sources and copying, reformatting, and matching it before any analysis can be done. This process can take a lot of time and effort, which data integration is designed to mitigate or solve.

This data integration plan proposes the use of cloud technology as the platform upon which Oklahoma integrates data across agencies. As discussed in Section 2.3 of this deliverable, cloud technology provides an affordable way to integrate and store data and is rapidly becoming the standard for which data analysis and processing is conducted. Private and public institutions are migrating more of their infrastructure to the cloud for reasons explained throughout this section.

4.1.1. Data Integration Preparation

Successful development of an ECIDS is not just about which technology is used; it is also about the mindset of the leadership and staff of each agency participating in the ECIDS. Cloud technology is not a panacea, and success will rely on participating agencies—both data developers and consumers of data—working together effectively. In other words, if the participating agencies' cultures do not support collaboration on data integration, there is no technological solution to overcome this problem.

In the experience of the 3Si/Foresight team, staff usually embrace the idea of moving operations to the cloud. Despite the positive outlook, however, an agency needs to be proactive and purposeful in supporting staff with new technology adoption. It is important to establish the value proposition, which may be different for each contributing agency. Continuously reinforcing *why* the agency is participating in an ECIDS is critically important to ensure staff understand and work in support of the ECIDS. Creating open channels of communication is important, as is ensuring staff who are involved in the ECIDS have sufficient time and availability to engage in the work, ask questions, and collaborate effectively. Showing results early, both major and minor, will also promote information sharing and build momentum as the integration progresses.

The following guiding principles, which are based on a preliminary analysis of Oklahoma's current infrastructure, best practices, and the 3Si/Foresight team's experience and expertise, will help Oklahoma executive and agency staff prepare for the significant data integration entailed with launching its ECIDS:

1. Leadership and staff of each participating agency understand and actively support the vision for and goals of data integration and the value proposition to their organization of participating in data integration.

2. Team members tasked with contributing to and performing data integration understand the work required is an ongoing process of building and sustaining an analytic culture; in other words, it is not a “one and done” project.
3. Each participating agency designates an ECIDS point person to serve as an internal champion/owner with the mandate and authority to refine and adjust priorities to advance the agency’s use of an ECIDS, in coordination with other agency stakeholders.
4. Participating agencies are proactive and flexible to ensure they foster the skill sets necessary to contribute to and benefit from an ECIDS. Collecting for, contributing to, and using data from an ECIDS requires specific technical and analytic skill sets. To maximize the value of an ECIDS for all participating agencies and stakeholders, each participating agency needs to develop and/or obtain these skills in some capacity, through hiring, internal staff training, or consulting resources.

Fortunately, Oklahoma is well-positioned to create a successful ECIDS. The 3Si/Foresight team has observed that participating agencies are well on their way to implementing many of the guiding principles described above. As noted in Section 2, Oklahoma has made substantial progress for data integration. Senior leadership—at the Cabinet and agency leadership levels—have already endorsed the idea of a large data integration project. Establishing a Master Person Index (MPI) represents a significant step toward an ECIDS, along with the creation of data committees such as DISCUSS.²² Oklahoma agencies are already working in the cloud: for example, OMES has significant deployments within Microsoft Azure (discussed in Section 4.2) with plans for more integration in the near future. These examples represent important first steps for agencies coming together, integrating their respective data, and creating new and valuable insights that not only benefit young children but any resident who receives or needs services and support from their government.

4.2. Data Integration Plan Options and Recommendation

4.2.1. Choosing a Cloud provider

The 3Si/Foresight team believes Oklahoma would be best served by formally committing to using one of the top three cloud service providers (CSPs). The three dominant CSPs are Amazon Web Services (AWS), Microsoft (Azure), and Google Cloud Platform (GCP), with a combined 58 percent of the market share. In general, all three CSPs provide similar services, quality, and price. Since Oklahoma is already using many

²² The Department of Health has built a Master Person Index (MPI), which currently pulls data from three agencies (OSDH, OHCA, ODMHSAS) across 20 different source systems. It deploys an algorithm to determine which children are represented across the multiple systems and assigns them a unique identifier and stores the data in a consolidated index. As of today, the MPI is in production and is owned by OSDH with plans to incorporate data from all other agencies. The MPI is directed by a committee, Deliver Interoperable Solution Components Utilizing Shared Services (DISCUSS), with representatives across multiple agencies and an existing interagency data use agreement.

of Azure’s services, the 3Si/Foresight team recommends Oklahoma select Microsoft Azure as its CSP. The following paragraphs explain our reasoning.

Although AWS holds a commanding lead in market share as of Q4 2018, Azure and GCP are growing rapidly in this space (Figure 4-1). There are also dozens of CSPs in the marketplace. For this analysis, we focused on the three largest CSPs because they offer the most comprehensive range of building blocks with which to create an integrated data system. 3Si did not include other CSPs in our analysis—such as Alibaba, IBM, and Oracle—because they have significantly smaller product ecosystems and developer communities, which makes them less desirable options for Oklahoma.

Figure 4-1 - Worldwide cloud infrastructure spending and annual growth

Vendor	Q4 2018 (US\$ billion)	Q4 2018 Market share	Q4 2017 (US\$ billion)	Q4 2017 Market share	Annual growth
AWS	7.3	32.3%	5.0	32.2%	+46.3%
Microsoft Azure	3.7	16.5%	2.1	13.7%	+75.9%
Google Cloud	2.2	9.5%	1.2	7.6%	+81.7%
Alibaba Cloud	1.0	4.2%	0.6	3.5%	+73.8%
IBM Cloud	0.8	3.6%	0.6	4.2%	+27.6%
Others	7.7	33.8%	6.1	38.9%	+26.7%
Total	22.7	100.0%	15.6	100.0%	+45.6%



Source: Canalys Cloud Channels Analysis, February 2019

Jeff Bezos, Amazon’s CEO, has famously said “AWS had the unusual advantage of a seven-year head start before facing like-minded competition,” which has resulted in AWS’ platform having the most robust and diverse set of product and service offerings. Although AWS is the clear market leader, Azure and GCP are comparable platforms, having many products and services that overlap with AWS. This is reflected in their market share growth since 2017.

Since AWS, Azure, and GCP offer similar services, deciding which one to move forward with can be complicated. All three offer a government deployment designed for highly secure and sensitive information (see Section 4.2.2: Cloud Data Security). Cost is comparable across platforms. Another important consideration is whether Oklahoma is already using one CSP more than others.

COMPARING AWS, AZURE, AND GCP

In general, all three primary CSPs provide similar services, quality, and price. For Oklahoma’s scale of data, all three CSPs will offer comparable services at roughly the same cost for storage, computing resources [Central/Graphical Processing Units (CPU, GPU)], and data warehousing.

Where CSPs differentiate are in user experience, compatibility with on-premise infrastructure, and analytical toolsets.

Table 4-1 summarizes the 3Si/Foresight team’s assessment of key features for each CSP, based on our experience with and research on these three providers.

Table 4-1 - Decision matrix for cloud service providers²³

Feature	AWS	Azure	GCP
Pros	Offers the most robust platform with the most analytical tools.	Offers tools very similar to Oklahoma’s current data environment (SQL Server, Power BI).	Offers a powerful, platform that many engineers find easier to use than AWS.
	Huge community of developers.	Integrates with Active Directory for single-sign on (i.e., the user only has to log into Windows to access Azure).	Google “simplicity” is baked into the engineering experience.
	Leader in the industry.	Graphical User Interfaces provide easy navigation.	Provides good customer support, including collaboration on innovation.
		Currently on Oklahoma’s technology roadmap (OMES).	

²³ Some of the considerations in this table—such as compatibility with OK infrastructure—are necessarily constrained by the time and resource limitations of this project and could be further expanded and/or refined if necessary. Later in this section, we describe the limitations of this analysis.

Feature	AWS	Azure	GCP
Cons	<p>Requires strong engineering skill sets; not meant for average developers.</p> <p>AWS support lacks personalized service; mainly self-service.</p>	<p>Unlike AWS and GCP, which use Linux programming language, Azure uses Windows. AWS/GCP and Azure developers are not interchangeable.</p> <p>User manuals are sometimes outdated or incomplete.</p>	<p>BigQuery data warehouse is complex from a user perspective. Doesn't relate to SQL Server user experience.</p> <p>Limited product diversity.</p>
Point and Click Administration	Good	Excellent	Good
Learning Curve	Advanced	Intermediate	Intermediate
Compatibility with OK infrastructure	No Overlap	High Overlap	No Overlap
Analytical Tool Diversity	Excellent	Very Good	Good
Data Warehousing Capabilities	Excellent	Very Good	Good
Reliability (Uptime)	Excellent	Very Good	Excellent
Cost of Raw Storage (10TB)	\$500/Month	\$500/Month	\$500/Month
Cost of Database Storage (1TB)	1-2K/Month	1-2K/Month	Cost based on query usage
Machine Learning Algorithms	Excellent	Good	Very Good
Product and Developer Ecosystem	Excellent	Very Good	Good

As explained earlier, because of its seven-year head start, AWS is considered by many to be the most advanced for CSPs in many areas. The biggest complaint is that AWS requires the most advanced skill sets. The good news is that all three offer the same basic feature set, and developer communities for all three (AWS, Google, and Azure) have large pools of skilled engineers who support the platform.

Although GCP has less market share than AWS and Azure, Google is aggressively trying to capture more market share by offering unique services as part of its platform. Google has been known to provide very good consulting resources at affordable—or even free—rates. For example, Google helped San Joaquin County create a “Perfect Match” algorithm that identifies best matches for families and foster children.²⁴

It is our opinion that all three platforms contain the necessary set of functions to adequately support an Oklahoma ECIDS.

EXAMPLES OF CLOUD DEPLOYMENTS IN OKLAHOMA

Oklahoma has experience within the cloud, particularly Microsoft Azure. The 3Si/Foresight team interviewed Bo Reece, Chief Information Officer (CIO) of the Office of Management and Enterprise Services (OMES) regarding the existence of any cloud service providers and deployments with Oklahoma’s IT infrastructure. Mr. Reece indicated that the IT infrastructure leans heavily towards Microsoft-related products for Oklahoma’s data management and warehousing.²⁵ OMES resource familiarity with Azure could help reduce time and cost of linking these data systems to the cloud product. Mr. Reece also indicated that OMES has been moving forward with deploying more resources within Microsoft Azure. Overall, Oklahoma’s data infrastructure is primarily serviced by Microsoft products, both on-premise and cloud-based.

CONCLUSION

Given the overlap of current infrastructure and current contractual agreements, the 3Si/Foresight team’s analysis focused on determining whether there was a compelling reason to go with a CSP other than Azure. In our opinion, the advantage of being able to more easily transition to Azure outweighs any benefits of AWS’ Best in Class status or Google’s consulting services.

LIMITATIONS OF THIS ANALYSIS

3Si has extensive experience working with all three cloud service providers (AWS, Azure, GCP) and this analysis is based primarily on 3Si’s experience and feedback from the developer communities. Cost estimates are simplified to showcase similarities in major cost categories (storage and data

²⁴ See <https://cloud.google.com/match-engine/> for details on Google’s project.

²⁵ Some examples for how Oklahoma is using cloud services are as follows (information gathered from OMES): 26 instances of SQL Server are deployed within Azure.; roughly 120 Terabytes of data are current stored in Azure Blob storage; 24 licenses of Azure Power BI in use within 4-5 agencies; ODOT has an Oracle database deployed within Amazon Cloud; 5 MySQL servers are deployed within Azure; 61 virtual machines are deployed in Azure., and 104 Web Application Servers are deployed on Azure.

warehousing), but are not comprehensive and does not include other services that will incur cost. Cost comparisons are also based on products and services available to commercial users; 3Si/Foresight did not have access to government pricing for the three CSPs. Staffing costs to operate and manage the cloud are explicitly excluded from this analysis.

4.2.2. Cloud Data Security

For most organizations, data security is at the top of the list of business concerns. Data breaches can have catastrophic effects on business operations leading to lawsuits, lost profitability, reduced trust among customers, and negative publicity. Cloud security requires the same discipline, testing, due diligence, and good judgement as on-premise infrastructure.

One of the biggest benefits of the cloud is, ironically, its greatest risk: ease of set up. Within weeks, organizations can be set up in the cloud, collect and process data, and serve data back to consumers via web pages, applications, mobile products, etc. The momentum of getting up and running rapidly can lead to oversights that cause security issues. Despite all of the promise of the cloud, security in the cloud still relies mostly on usernames, passwords, processes, policies, and continued monitoring of threats.

COMPARING CLOUD RISK TO ON-PREMISE RISK

A common misconception is that on-premise infrastructure provides greater security against cybercrimes compared to CSPs. There is evidence that, increasingly, organizations that require the highest standards of security are relying on cloud computing resources, including:

1. U.S. Department of Defense: Microsoft signed a \$10 billion contract in October 2019 to host U.S. Department of Defense (DoD) data.²⁶ The Department of Defense has a standardized assessment and authorization process for CSPs to gain a DoD provisional authorization so they can serve DoD customers. Azure has demonstrated compliance with DoD security standards and received this authorization, reducing the time necessary for a DoD mission owner to authorize one of their systems for operation in Azure.²⁷
2. The U.S. Department of Veterans Affairs: The Access to Care site (<https://www.accesstocare.va.gov/>) is a cloud-based tool that provides a portal where veterans can obtain needed care, check wait times at specific locations, and provide feedback into the system for continuous improvement.
3. The National Institute of Health (NIH): NIH partners with GCP. The STRIDES (Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability) Initiative launched with Google Cloud as its first industry partner and aims to reduce economic and

²⁶ <https://www.nytimes.com/2019/10/25/technology/dod-jedi-contract.html>

²⁷ More information about DoD's Cloud Computing Security Requirements can be found here: https://rmf.org/wp-content/uploads/2018/05/Cloud_Computing_SRG_v1r3.pdf

technological barriers to accessing and computing on large biomedical data sets to accelerate biomedical advances.²⁸

CSPs are incentivized to produce products and services that are on the cutting edge in data security. They have the economies of scale to employ teams of cybersecurity experts who specialize in securing their services against all modern threats. Additionally, there are hundreds of third-party products that enable customers to deploy a comprehensive security architecture.

To maintain security and achieve cloud-level breadth, speed, and flexibility within an on-premise infrastructure, state governments would need to purchase and maintain substantial and expensive security hardware and software. They would also need to recruit and retain data security experts. Furthermore, state governments must navigate local and regional constraints (such as government salary bands) that may pose barriers to mirroring the same data security as CSPs.

GOVERNMENTS USING CLOUD SERVICE PROVIDERS

Because of the speed and cost savings cloud-based computing provides, governments—particularly the U.S. Government—are using CSPs more frequently. The Federal Risk and Authorization Management Program (FedRAMP) was created to provide a standardized approach to security assessment, authorization, and continuous monitoring of cloud products and services. FedRAMP's goal is to accelerate the adoption of secure cloud solutions, improve confidence in the security of cloud solutions, create baselines and standards, and ensure consistent application of security practices for the federal government. FedRAMP is part of the AWS, Azure, and GCP platforms and has worked with each to develop a secure set of services designed for governments. These three CSPs have extensive documentation on how to secure sensitive data, in transit

CAPITAL ONE HACK: A CAUTIONARY TALE

In July 2019, Capital One Financial was hacked by a former Amazon employee who obtained approximately 100 million accounts from credit card applications, 400,000 social security numbers and 80,000 bank account numbers. This hack wasn't a failing of the AWS platform; it was due to inadequate controls over which applications can access sensitive data. This type of attack is called Server-Side Request Forgery (SSRF). The hacker exploited a hole in a public-facing web application by tricking a web service into copying files from storage located behind the firewall to the hacker's account. In this case, the web service had too much access to data it never used, creating an opportunity for exploitation. Once the issue was identified, Capital One quickly change its data access policies. The hack could have been prevented through adequate data access policies: any public-facing application should have only the bare minimum data access. AWS has tools in place to monitor for these types of security holes; it is unclear why Capital One did not use them.

SSRF attacks like this one can occur with any application exposed to the web, even with on-premise infrastructure.

²⁸ A summary of the STRIDES cloud partnership can be found here: <https://commonfund.nih.gov/data>. Additional discussion is available here: <https://www.nih.gov/news-events/news-releases/nih-makes-strides-accelerate-discoveries-cloud>.

and at rest, within their cloud infrastructure. Local and state governments can benefit from this effort.

AWS, Azure, and GCP provide ample documentation to follow almost any compliance or certification, including the Health Insurance Portability and Accountability Act (HIPAA), Family Education Rights and Privacy Act (FERPA), Health Information Technology for Economic and Clinical Health Act (HITECH), and Protected Health Information (PHI). To confirm the documentation, 3Si reviewed examples of documentation on compliance and certification. For example, we reviewed an AWS white paper, *Architecting for HIPAA Security and Compliance on AWS*, that details security measures for each of their services that would store Health Insurance Portability and Accountability Act (HIPAA)-related data.²⁹ Similarly, we reviewed Google's standards for FERPA-related data.³⁰

3Si contacted Azure's sales team to discuss pros and cons along with pricing for the Government cloud premium services. While Azure's sales team did not provide exact pricing for certain products and services, they authorized 3Si to suggest a "25% increase over commercial pricing" as a good benchmark to use for cost estimates (note that because all data centers serving U.S. government clouds are based in the U.S., government pricing is typically higher than commercial pricing). For large scale projects (e.g., greater than \$1MM per month) the additional premium would be a concern. However, the additional security provided by a Government cloud service may provide a reasonable cost-to-benefit proposition because the cost for a relatively small scale for an ECIDS project should not be prohibitive, while the additional security is a critical benefit. Additionally, Azure offers free trials within their Government cloud offerings for customers to fully test the additional value prior to making any monetary investments. Therefore, it is 3Si's recommendation that Oklahoma consider the use of Government cloud offerings for the ECIDS project.

Cloud security is rapidly becoming a new field of its own with specific roles and responsibilities dedicated to securing platforms. The most common job title is a "Cloud Security Engineer," whose sole responsibility is to protect the organization's data. While larger enterprises, such as Salesforce, employ dozens of cloud security engineers to monitor their platform, smaller organizations try to incorporate security into existing roles. 3Si does not recommend this approach: given its importance, maintaining cloud security requires dedicated capacity and/or explicit delegation to an individual and/or team who is accountable for all cloud security matters. This includes keeping up to date with new threats, working with CSPs to monitor their platform, and to hold regular presentations on cybercrime threats so that end users stay current with the latest information.

TAKEAWAYS ON CLOUD SECURITY

As with on-premise infrastructure, the cloud is only as secure as the people, processes, and policies that manage it. The following are typical cloud security practices:

²⁹ The full white paper is available here:

https://d1.awsstatic.com/whitepapers/compliance/AWS_HIPAA_Compliance_Whitepaper.b97af4d3b25c49b662f203bccbbb624ec1ecb810.pdf.

³⁰ Google's guidelines for FERPA are available here: <https://cloud.google.com/security/compliance/ferpa/>.

1. **Security monitoring.** This is perhaps the most important cloud security practice. It means reviewing data usage and users on a frequent basis to identify anomalies that may suggest the potential for a data breach. Security monitoring is easier with an abundance of tools provided by CSPs. Proper training is strongly recommended for users of these tools to ensure security.
2. **Identity Access Management (IAM)** is critical to access management. It is important to have stringent data access policies that provide the bare minimum access to end users. IAM allows administrators to create, modify, and delete user identities to ensure users have appropriate access.
3. **Policy Management** involves authorizing which user groups have access to which data. Many data security issues are related to inadequate policies that inappropriately provide too much access to user groups. Following best practices means that user accounts should have the minimum access to data needed to do their job.
4. **Multifactor Authentication** involves using two methods to authenticate users. For example, multifactor authorization on a bank account requires the customer to enter their username and password, receive a text message with an access code, and then enter the access code to confirm their identity. This provides an extra layer of security in case the username and password are stolen. This is especially critical for root accounts, the highest privilege account access, which forces a user to retrieve a one-time password for each time they log in.
5. **IP Whitelisting** is another layer of security that specifies a range of IP addresses that can see data elements. Users must be within the organization's network to access these data. If a user tries to access data from home, they would have to log into the organization's network before they could access the data, even though it may be stored on the cloud.
6. **Prioritizing data security.** A common mistake is allowing software developers to secure their own applications instead of a structured security framework created by trained experts in cloud security. Developers of applications often build and deploy products under strict timelines and budgets. Sometimes, this means developers put deployment over security, leaving holes that can be exploited. It's imperative that data security is not sacrificed in favor of speed or cost savings.
7. **Encrypt data at all times.** Good practice requires all data to be encrypted during transit (i.e., while being transmitted over the internet) or while at rest in cloud storage. Only authorized individuals should hold the decryption keys.

NEXT STEPS

The next step will be for Oklahoma to assess existing capacity regarding cloud security expertise and/or existing data security protocols. Where there are gaps, such as the lack of a properly trained cloud security engineer, Oklahoma will need to determine whether the gaps can be filled using existing roles and processes or whether a new role must be created and hire accordingly. Additionally, Oklahoma will need to make a decision about whether to take 3Si/Foresight's recommendation to consider government cloud services.

4.2.3. Data Transportation

To use the cloud requires an efficient, repeatable, and scalable way to port data from source systems into the cloud. Oklahoma's data system landscape is diverse and includes some antiquated systems (e.g., COBOL, which is an older programming language that is being replaced or integrated with more modern coding languages).

These diverse data source systems will likely require different methods to transport into the cloud. That said, it would be inefficient to create a customized data transport method for each data system. Therefore, we recommend classifying data source systems into buckets with similar characteristics. From there, Oklahoma can use a simplified data transport method suitable to each classification.

Table 4-2 provides a simple classification system that can be used to segment source systems into four classes:

Table 4-2 - Source system classification

Classification	Examples of Systems and Features
Class 1: Database Management Systems	<ul style="list-style-type: none">• Applications using a relational database structure as a backend.• Location is on-premise.• System has Open Database Connectivity (ODBC) or Java Database Connectivity (JDBC) access to the SQL tables.• On-premise data warehouses built in SQL Server, Oracle, Postgres, MySQL, or comparable database systems.
Class 2: Static or slowly changing data, public, or Excel-based systems	<ul style="list-style-type: none">• Census data (e.g., American Community Survey, etc.).• In-house survey data.• Financial models built and maintained in Excel.
Class 3: Third Party Data Producers	<ul style="list-style-type: none">• Student Information Systems (Schools Interoperability Framework (SIF) and Zone Integration Servers (ZIS)).
Class 4: Cloud Based Applications	<ul style="list-style-type: none">• Any Software as a Service (SaaS) platform, such as Salesforce, QuickBooks Online, Intercom, and HubSpot (Appendix E provides a detailed explanation of SaaS platforms).

EXTRACTING DATA FROM SOURCE SYSTEMS

As discussed in Section 2.3 of this deliverable, the traditional approach to integrating data has been to extract it from its source, transform or clean it (e.g., fix problems or anomalies), then load it into a data

warehouse. While this method works, it tends to be slow and expensive, for reasons described in detail in Section 2.3. Due to the affordability and scalability of the cloud, it is faster, cheaper, and easier to extract all the data from source systems in its raw form (i.e., before it is transformed), make the raw data available to data consumers, and then transform and produce analytics on-demand using a variety of cloud-based tools. Modern applications provide built-in Application Program Interfaces (APIs) to enable transportation of raw data, and many third-party services sell products and services that automate this data transport method.

BUILD OR AUTOMATE DATA TRANSPORTATION

Oklahoma will need to decide, depending on each system integration, whether to build its data transportation infrastructure in-house, buy data transportation tools, or repurpose its existing infrastructure. Regardless of what Oklahoma decides, automating data transportation as much as possible is critical to remove bottlenecks and make the data available to consumers as fast as possible.

Considerations include:

- Building and maintaining a data transportation infrastructure from scratch will be expensive. It will require bringing in consultants or hiring skilled data engineers to build bridges between data sources and cloud systems. This is costly because the demand for data skill sets (engineering, analysis, science) is high. Companies competing for these resources offer high salaries and expensive perks to attract and retain talent. If OMES is selected to provide this service, another consideration is their limited resources and competing project priorities, both of which could significantly extend timelines and impact deadlines.
- Many third-party companies offer data transportation services, and this approach may be the fastest and cheapest way to transport data to the cloud. Where possible, these services can provide real value by automating data transportation between cloud-based platforms. However, some of these products require that data source applications are cloud-based, which is not the case for many of Oklahoma's data sources.
- A third approach is to repurpose the data transportation infrastructure already in place. It is likely that Oklahoma has invested in data transportation infrastructure that could be modified to automate movement of data sources to the cloud. An in-depth investigation of Oklahoma's existing data transportation infrastructure was not part of the scope of this project, so 3Si/Foresight is not able to confirm that Oklahoma's data transportation tools could be repurposed. The due diligence will occur during the first half of 2020, per the draft timeline that is included in Section 6.

Table 4-3 (on the following page) provides some guidelines to assist Oklahoma in deciding whether to build or buy a data transportation infrastructure.

Table 4-3 - Data transportation methods

Method	Best used when...
BUILD OR REPURPOSE	<p>The data sources fall into Class 1 or Class 2 (see Table 4-2 for definitions of these classifications).</p> <p>There is an established and consistently used data transportation tool such as SSIS, Informatica, Ab Initio, Talend, Pentaho, etc.</p> <p>Data producers and applications are specialized and either built in-house or designed for specific use cases.</p> <p>In-house staff can often augment current data transportation packages (if they exist) quickly and with little effort.</p>
BUY	<p>The data sources fall into Class 3 or Class 4.</p> <p>The source application provides public-facing APIs that allow for data integration (e.g., developer.salesforce.com provides access to all APIs for Salesforce).</p> <p>The source systems provide an automated data transportation as an additional product feature.</p> <p>There is a need for real-time data.</p>

RECOMMENDATIONS FOR DATA TRANSPORTATION

To develop recommendations for data transportation, 3Si analyzed the data systems needed for Use Cases #1-3 and classified each data system into one of the four categories described in Table 4-2. Based on our analysis, we conclude that each data system identified has a pathway for its data to be integrated in the cloud. Many of these pathways can be derived from cloud-based services that exist for that purpose. There appears to be nothing unique about Oklahoma's technical landscape that would hinder cloud integration for any identified system.

For systems in Class 1 (Database Management Systems), data engineers should use standard data integration services to integrate data into the data lake. All three CSP's offer these services. For example, Azure Data Factory's (ADF) data replication services are designed to simplify and streamline the process to transport data into the cloud; ADF has 90+ native integrations with other systems and documentation that guides users step-by-step on how to integrate source data in the cloud.

Only one data system, Work Life System, was classified as Class 2 (static or slowly-changing files). These data come in by email and are manually entered into another system. While it is possible to sync manually-entered output into the cloud, it may be better to store raw email attachments (or raw emails themselves) in the data lake and use cloud-based tools to process and clean that data. This could

eliminate manual processing steps. This approach also has the advantage of storing the raw data indefinitely for historical and reference purposes.

OSDE's Wave data system comes through a network of servers (called Zone Integration Servers), which integrate with all Student Information Systems across Oklahoma school districts. This is a Class 3 system (Third Party Data Producers) because it relies on a third party to do the heavy lifting of integrating many systems. This raw data can be synced directly with a data lake and any changes, new data, or deletions will automatically update within the cloud. OSDE is already ingesting this data into the Wave system and similar logic can be used to ingest the data into a cloud-based data warehouse. *(Note that OSDE is currently exploring options to replace the WAVE system.)*

We classified four data systems (eMPI, MMIS, Mainframe PS2, e-Childcare) as Class 4 integrations (where the vendor may provide cloud-based integrations as part of their services). Each of these vendors (IBM, ORION, DXC, and Conduet) publicize cloud-based analytics on their website as part of their service offerings. This capability usually means their source systems are cloud-based, so they should be able to use cloud integration services to send data directly to Oklahoma's cloud or provide a cloud-based channel to integrate with Azure Data Factory. While cloud-based services to integrate Class 4 systems will incur extra cost, they are usually more cost efficient than having internal staff build a solution. Additionally, Oklahoma can hold a vendor accountable to maintain these services with a standard service-level-agreement.

NEXT STEPS

The next step is to conduct an in-depth and comprehensive investigation of Oklahoma's existing data transportation methods and determine whether it is feasible for Oklahoma to repurpose some of its existing transportation infrastructure for the ECIDS. Additionally, the system classification should be confirmed, particularly with Class 4 systems, to determine whether the vendors can provide a cloud-based integration channel to substitute the need for Oklahoma to have to build a custom integration.

4.2.4. Data Storage: The Data Lake

All CSPs provide similar data storage services. This type of storage—called a “data lake”—has effectively limitless capacity and is able to store data of any type. As described in Section 2.3 of this document, a data lake is a centralized repository allowing storage of all types of data as-is (in other words, without first carrying out costly and time intensive data modeling and transforming processes.) The following section describes the advantages to the Oklahoma ECIDS of storing all types of data in a data lake.

STRUCTURED VERSUS UNSTRUCTURED DATA

Among the benefits of cloud storage is that it can store data in any form—including structured, semi-structured, and unstructured (i.e., raw) data.

Structured data are data that have been organized in a way that makes them easily searchable. Examples include SQL database tables, Excel worksheets, and comma delimited files.

Semi-structured data contain tags that delineate records or fields, but may still need to be organized to be searchable. Most web page data are semi-structured.

Unstructured (raw) data have no structure or organization and data objects have no relationship to other data objects. Examples include media files, emails, and database backup files.

A key advantage of storing all types of data is that users can access more data from more sources in less time, which leads to better, faster decision-making. Once all the data are available in a data lake, users are able to produce data products on demand. Data analysts, researchers, and BI teams (see Section 4.2.7: Data Integration Roles and Responsibilities) can sift through the lake and determine which elements they need. Processes to collect data—such as surveys—can automatically send data to the data lake, so it can be used for reports or an analytical process.

As stated above, data lakes can store any type of data—including structured, semi-structured, and unstructured data (see sidebar for an explanation of these data types). Transforming unstructured or semi-structured data into structured data—which is required for most analysis—can be laborious and expensive and is often a major bottleneck in accessing data. The concept behind the data lake is to extract data out of the source systems in its rawest form, centralize it in the lake, then structure it as needed.

DATA LAKES VS. DATA WAREHOUSES

Data lakes and data warehouses are not interchangeable terms. Data lakes store large pools of raw data where the usage is yet to be defined. Thus, raw data is more accessible and malleable for preliminary analyses and storage is relatively inexpensive. Data warehouses store subsets of raw data that have been cleaned, processed, and structured for a specific purpose. Data warehouses require purchase of a software application (i.e., Oracle, Teradata, DB2), and, if created on-premise, the hardware (i.e., servers, disks, etc.) to run the software application and store the data.

Before the advent of the cloud, many organizations invested heavily in creating “enterprise data warehouses” that provided comprehensive data for all departments of the organizational “enterprise.” Since creating an enterprise data warehouse involves identifying, cleaning, and structuring all data upfront, there is a substantial time and cost associated with its development. Additionally, an enterprise data warehouse is less flexible if and when the enterprise’s data needs shift. Changing the structure of a data warehouse can be onerous, and when change happens, end users often must wait for data to become available in the warehouse.

A data lake does not replace a data warehouse. Oklahoma will still need to structure data in some type of data warehouse for reporting and analysis. The benefit of separating data lakes and data warehouses is that data can be moved from the lake into a warehouse on an as-needed basis (i.e., on-demand), which reduces the need to clean and prepare data that is not immediately relevant or may never be used at all. For example, case management systems typically collect case notes within their system. Many organizations struggle to use this data, however, due to its lack of structure and random text. As such, most data warehouses exclude case notes. If processed correctly, however, case notes can be useful. By using cloud tools that specialize in extracting insights from text, the data can be loaded into a separate data warehouse designed specifically for text insights. This type of data warehouse would be considerably smaller and easier to build and maintain than an enterprise data warehouse. This application of a cloud-based data warehouse is typically called a “data product” and is discussed more in Section 4.2.6 below.

DATA REDUNDANCY

Another key advantage of cloud storage is data redundancy. Cloud service providers provide automatic redundancy by replicating the data automatically across their data centers. This removes the need for a separate backup process or a comprehensive disaster recovery plan.

STREAMLINED DATA SHARING AGREEMENTS

Data lakes provide a centralized location to store all raw data. This allows data sharing agreements to be simplified, reducing the need to create multiple data sharing agreements and providing user access to individual users or groups on specific systems. Because all the data are consolidated, a simple change to the data sharing policy—with the governing body’s approval—can provide users access to the data they need quickly. The data can update automatically, allowing for more continuous and real-time analytics.

Additionally, preliminary conversations with the heads of Oklahoma state agencies indicate willingness to incorporate their data to the cloud to produce more valuable insights across the agencies. This broad motivation will further streamline data sharing agreements by removing any questions around where and how the data will be stored.

COST SAVINGS AND SCALABILITY

Another major difference between lakes and warehouses is cost. Storage in an enterprise data warehouse can be expensive, when considering the database license, the hardware, and the people to manage the database (i.e., Database Administrators). A data lake, on the other hand, is designed specifically for low-cost storage. CSPs charge based on usage and there are multiple options within a data lake to archive data automatically, achieving greater cost efficiencies. There is no database application cost and analysts can utilize whatever tool they prefer. Python, for example, is an open source (free) analytics language that is wildly popular in usage today among data scientists and data engineers.

The biggest advantage of the data lake is scalability. CSPs offer effectively limitless storage, meaning a customer can store as much data as it wants, as long as it is willing to pay for that storage. A data warehouse will be limited to the hardware it is running on and, oftentimes, the licensing of the software. Within a CSP’s data lake, there is no need to procure and manage new hardware with increased data storage needs, which produces cost savings related to people and infrastructure.

DATA RETENTION AND ARCHIVING

While there will be raw data elements that seem irrelevant to store at the time, some of these data may provide channels to accessing more useful data. Furthermore, the cost of storing low-usage data is relatively minimal. If 20% of the data stored will never be used but the other 80% of the data would otherwise be unavailable, the cost of storing useless data is, literally, a small price to pay.

From a data security standpoint, it is important to set timelines for archiving and deleting data as part of data management policies. Archiving data that is not being used also reduces costs, as archival storage is often less expensive than other forms of storage. CSPs provide options to manage data retention automatically: customers can set parameters for sending data to archival storage and/or deletion. Cloud tools that monitor which data are being used and by whom can provide the people responsible for setting data retention policies the information they need to set appropriate data retention parameters.

TAKEAWAYS ON DATA STORAGE

As described in Section 2.3 of this document, prior to cloud storage organizations would create on-premise data warehouses that consolidated structured data for the purposes of reporting and analytics.

While this methodology worked, it was expensive and had limitations on the types and volume of data it could store. Cloud storage provides a cost-effective solution to storing massive amounts of data, of any type, to a centralized location (the data lake).

With the ability to store vast amounts of raw data comes the responsibility to manage and maintain it. Good practice requires all data to be encrypted while “at rest” in cloud storage with the keys held securely by authorized individuals. Raw data may contain sensitive information, so having a strict policy that manages access to this data is mandatory. Data retention policies should automatically archive or destroy sensitive data after a set time period. This approach can also save money and reduce the opportunity for data breaches. Finally, using cloud tools that monitor data usage can provide policy makers the appropriate information to set and manage data retention policies accordingly.

ESTIMATED COSTS

As stated before, cloud storage for an ECIDS is widely available and cost efficient. Cloud technology is a fixed monthly cost, because the service automatically upgrades technology as needed. This eliminates the need to replace depreciated equipment after a period of time (and incur a new upfront investment for updated equipment), as would be the case with on-premise storage. Pricing for the cloud is linear, meaning it maintains the same per-unit price regardless of the number of units purchased (for example, \$100 per month for five terabytes (TB) of storage would increase to \$200 per month for 10TB of storage).³¹

STORAGE AND COST MODELING: AN EXAMPLE USING DATA FROM THE MPI AND OSDE’S SIS

3Si’s preliminary analysis indicates that the highest volume of data will be derived from the MPI and OSDE’s Student Information Systems (SIS).

MPI: *Oklahoma stakeholders who maintain the MPI estimate that MPI data represents approximately 60 million records totaling 60 gigabytes (see Appendix D for a table of MPI data elements and size estimates).*

SIS: *3Si has not obtained volume and size estimates, but we believe the SIS would generate a high volume of data—perhaps up to 500 records per student. Based on Oklahoma’s Pre-K through 12th grade population, which was slightly under 700,000 students in the 2018-19 school year, this would produce up to 350 million records annually.*

Combined MPI and SIS data volume: *Based on these assumptions, the total data from the MPI and SIS would contain 410 million records totaling around 440 GB. Assuming an initial upload of 440GB of data, a monthly upload of 100GB of data, and users accessing data objects 3,000 times per month, the total cost for 12 months would be between \$300 and \$500 depending on exact usage requirements. The table below contains one estimate using Azure’s current pricing model.*

Azure Data Lake Storage Estimated Costs			
Usage	Quantity	Price per Unit	Monthly Charge
Initial Upload of MPI and SIS Data	440 GB	\$ 0.039	\$ 17.16
New Data Uploaded Per Month	100 GB	\$ 0.039	\$ 3.90
Reads Per Month	3000 Reads	\$ 0.004	\$ 12.00
Monthly Total			\$ 33.06
Estimated Annual Total*			\$ 396.72

Based off Azure’s pricing model: <https://azure.microsoft.com/en-us/pricing/details/data-lake-storage-gen1/>

³¹ This is not an estimate of data volume for the ECIDS; it is intended to demonstrate the pricing model of cloud technology.

Storage costs vary depending on how accessible the data needs to be for quick retrieval and analysis. Critical data that needs to be available for fast retrieval incurs the highest storage cost. On the other end of the spectrum, data that does not require fast retrieval can be stored at significantly lower costs. Best practices recommend that customers set lifecycle policies on data elements to archive or automatically delete files and folders after a specified period of time. Additionally, it is generally considered a best practice to compress data files to maintain free space. Compression can reduce file size by 50% to 90%, with the equivalent reduction in cost.³²

The example in the sidebar on the previous page demonstrates how one might estimate the amount of storage required by an ECIDS. This estimate is only illustrative; to create an actual budget would require more information about how the data will be used. Additionally, this example does not include other cloud services such as SQL Data Warehouse, DataBricks, or Data Factory, which will incur separate charges for usage.

NEXT STEPS

Oklahoma should take three additional steps when estimating storage costs for the ECIDS. First, Oklahoma should consider for what purposes the data will be used (i.e., point in time vs. period of time analyses). The purpose of the use case will determine the frequency of raw data updates from the different systems, whether these updates will produce a new instance for the child or overwrite previous records, etc. Second, Oklahoma should estimate the anticipated volume of data to determine the amount of storage space that will be needed. Finally, Oklahoma should calculate costs of storing the estimated volume of data in the desired format. These steps will allow optimal and cost-effective use of the cloud.

3Si has included preliminary, high-level assumptions that account for these considerations already; for example, three independent third parties have provided estimates of Oklahoma's future data capacity needs. These estimates consider factors that impact the amount of storage space that will be needed. Section 6 provides high-level estimate scenarios of storage size and cost, based on Oklahoma's current and future data capacity needs.

4.2.5. Data Management and Processing

Organizations invest substantial effort to collect, clean, and categorize data. Therefore, a sound data management plan is essential to distill valuable insights and reap the benefits of this investment. The data management plan will articulate analysis strategies that will derive value from the raw data, and identify subject matter experts who can support accurate and useful output by interpreting nuances within the data.

DATA MANAGEMENT BEST PRACTICES

The best analyses are derived from carefully-defined use cases. Raw data must be available to analysts so they can interpret data complexities and recommend a structure to answer each specific use case.

³² https://techterms.com/definition/file_compression

These analysts must have access to and proficiency with modern toolsets used to profile data quickly, and sufficient computing resources (i.e., processors and disk space in the cloud).

The following example—based on Use Case #1 (What and where are the gaps in early childhood education services and quality)—illustrates how Oklahoma may collect and profile the raw data to assess how best to structure, clean, and analyze it. For this use case, because data is sourced from many unrelated systems, it will be fragmented (i.e., actual data will likely differ from metadata descriptors, many datasets will contain errors, and some data will be missing). On the plus side, Oklahoma will likely identify additional datasets that can be leveraged. Given these factors, it makes sense to profile the raw data first, rather than try to structure it without understanding what is there and what is missing (see Section 2.3 for a more detailed discussion of why this approach makes sense).

CSPs offer a variety of tools to profile data and assess how best to structure, clean, and analyze it. These tools offer greater flexibility for collecting and merging data from disparate systems by removing the arduous task of having to model multiple data structures to simply collect data from multiple systems. For example, databases called “NoSQL” offer the ability to store and append new data within a flexible environment. In a NoSQL environment, one table would be created using an existing index (such as the MPI) and any data that matches that index (in this example, the child record) would be merged in its native format. Each child record, sometimes referred to as a “collection,” would contain all the raw information from multiple systems on a specific child, which can be processed into structured table (in a SQL database) for use within business intelligence tools and general reporting.

The reality is that early learning data is often messy. Unlike private industries that are able to structure their business operations to produce consistent datasets, the social sector usually relies on a variety of disparate systems to operate, resulting in unpredictable data. Therefore, it is important to use flexible tools that work with the data as-is rather than requiring effort to structure the data before even knowing what is available.

The next section looks at tools and models that can be put into place as part of a sound data management strategy.

DATA PROCESSING TOOLS AND MODELS

There are many possible models for data management and processing in the cloud. All CSPs provide an abundance of tools for these functions, and massive third-party marketplaces offer software products from independent vendors. With so many options, users and developers can deploy a vast array of tools and services, but this can create a cluttered environment that can become difficult to manage over time. It is therefore good practice to institute policies that provide consistency across the enterprise. The ECIDS’ governing body should provide a framework for approving data management and processing tools.³³

³³ “Governing body” in this context refers to either the Interagency Board, the Center, or both. Section 5: Data Governance Plan introduces and discusses these entities.

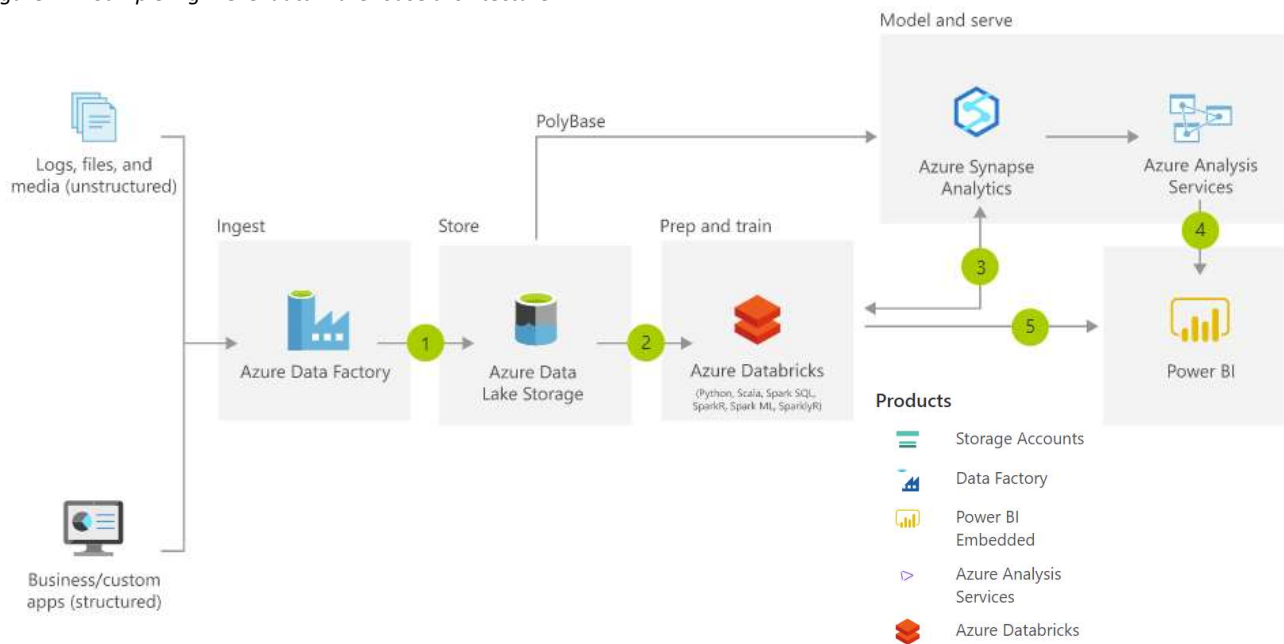
3Si recommends distributing responsibility for choosing products and services. The governing body would take recommendations from users. If the governing body feels the product/service should be part of the core set offered across Oklahoma, they would approve and fund the usage going forward. If not, the agency could still use the product/service, but would have to pay for it themselves. Many cloud products are “pay as you go” or “pay per usage”, both of which allow users to explore tools without making a major financial commitment or signing a long-term contract.

Approval policies could allow for a modest amount of experimentation, so users can recommend new tools to the governing body. For example, Azure offers a service called the Azure Data Factory (ADF) that streamlines and automates data processing. This service was modeled after a commonly used Microsoft tool called SSIS (SQL Server Integration Services), which is currently used in Oklahoma. Agencies may want to explore the opportunity to replace or augment SSIS with ADF, and trying out ADF can be done without much cost. Furthermore, agencies could experiment with ADF on their own timeline.

To illustrate how this could work, consider a scenario in which two different agencies in Oklahoma are processing data using two different tools (e.g., SAS and Python programming languages). The governing body may initially approve the usage of both languages, but may eventually decide that all agencies should use the same language for consistency and cost savings. At that point, the governing body would select one language that it would support across all agencies—for example, Python, which as of 2019 has become the most common language to use in data analytics due to its open source availability (i.e., zero cost) and large developer community. At that point, any agency wishing to continue using a different language would need to fund and manage it themselves.

To manage and process data, CSPs provide recommendations for common data lake/warehouse architectures. For example, the figure below illustrates Azure’s “Modern Data Warehouse” architecture.

Figure 4-2: Sample high-level data warehouse architecture



Source: Microsoft’s pricing page: <https://azure.microsoft.com/en-us/pricing/calculator/>, accessed 11/11/2019.

Table 4-4 illustrates the diversity of cloud products and services, with examples of products available on the Microsoft Azure platform.^{34,35}

Table 4-4 - Cloud products on Microsoft Azure

Product/ Service	Product/Service Description	Usage
Azure Synapse Analytics	A fully managed, elastic (flexible) data warehouse with security at every scale at no extra cost.	Reporting
Azure Databricks	A fully managed and collaborative analytics platform optimized for Azure.	Analysis
PowerBI	A business analytics service providing consumers with dashboarding and visualization capabilities	Reporting and Analysis
Azure Data Factory	A data integration service to automate data movement and transformation.	Data transportation and processing
Azure Analysis Services	Enterprise-grade analytics engine as a service.	Analysis
Azure Data Lake Storage	Highly scalable, low cost data repository to store raw data in its native format until it is needed	Storage

DATA GOVERNANCE

Section 5 of this document discusses data governance in depth. However, it is important to note the strong link between governance and data management and processing. An efficient governance process dictates how data is managed and processed. To illustrate this, take the following example:

Two separate teams have produced analyses that display children and programs; however, each analysis is slightly different and returns results that are similar but differ enough to raise concern around the validity of each. Which analysis is the correct one?

This is a very common scenario in everyday analytics. To resolve it, the governing body must understand how the data was managed from the start. This includes understanding which sources were used, how the raw data was profiled and processed, what assumptions were made about the data, and how the data were standardized to produce a common result. Reaching this understanding will entail consultation with subject matter experts who will be most familiar with relevant business rules and interpretation of source data (Section 4.2.7 describes these and other staff roles).

³⁴ Note that Azure Synapse Analytics was rebranded from Azure SQL Data Warehouse on November 7th, 2019.

³⁵ For a full menu of products available on the Azure platform, see <https://azure.microsoft.com/en-us/product-categories/analytics/>.

Ideally, the governing body would have been involved in these decisions prior to generating the analysis, but this is not always the case. Regardless, the governing body determines which analysis is to be used after a review of how the data was managed and processed from source to analysis. A strong governing body can ensure that data management and processing is carried out efficiently and effectively.

4.2.6. Initial Analytics and Data Products

Generally speaking, there are two ways to create analytic and data products. The traditional method—often referred to as the “Waterfall” method—involves planning the entire project upfront before starting to develop the analytic or data product. A newer approach—called “Agile”—tackles development of analytic and data products in iterative sprints (usually two weeks). With the Agile approach, development teams focus on quick deliverables. Agile strives to create a “draft” analytic or data product—called a Minimum Viable Product or MVP—as fast as possible so that users can try it out and provide feedback for the next iteration.³⁶ This continuous cycle creates a rapid development culture that assumes and welcomes changes during product development.

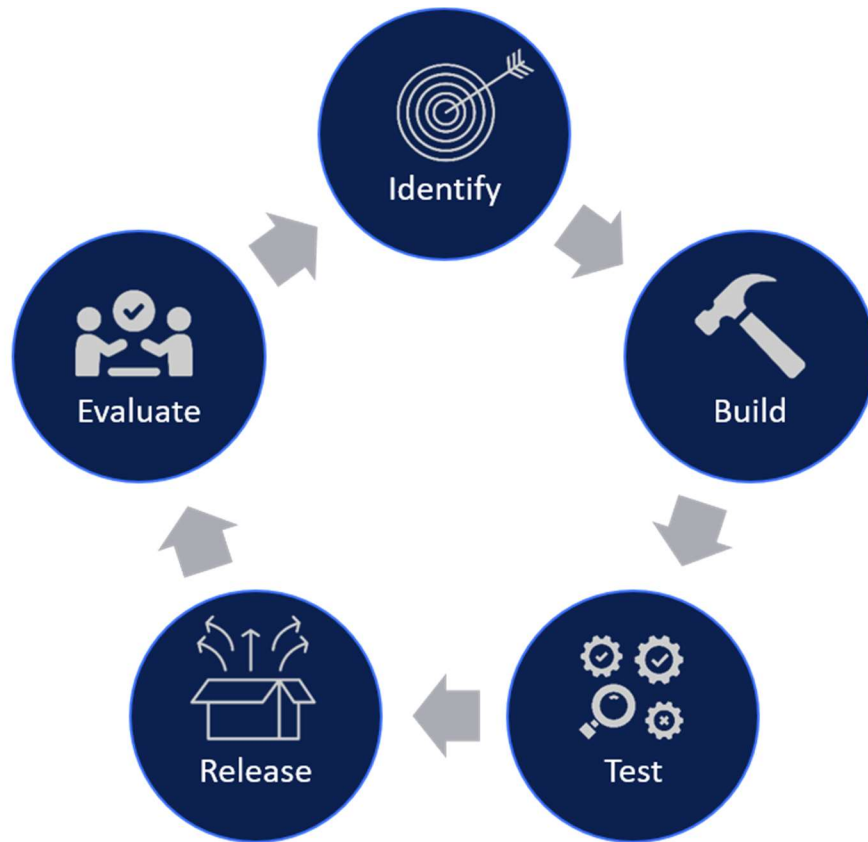
The 3Si/Foresight team recommends the Agile approach for creating initial analytics and data products, for two key reasons:

1. For complex data integration projects, it is important to demonstrate value quickly to participating agencies and other stakeholders. This builds support and momentum for expanding the use and usability of ECIDS. Traditional enterprise data warehouses, which take a long time to produce analytics and data products, are often criticized for over-promising and under-delivering value. The Agile method will allow Oklahoma to deliver valuable analytics and data products faster and demonstrate early wins.
2. Early childhood data is messy and complicated, so trying to plan an analytic product prior to modeling and interacting with the data itself is inefficient at best and impossible at worst. An Agile methodology—which allows for modeling and building a product based on real data rather than a conceptual plan—is well suited for early childhood data.

Figure 4-3 (on the following page) shows the Agile development process:

³⁶ It is important to note that the draft analytic or data product should be developed in the secure cloud first and then refined until it is considered ready to be vetted as an end product by the appropriate data governance protocols. Draft products should never be released outside of the secure cloud.

Figure 4-3 - Agile methodology



Questions to ask:

- What should we measure or analyze?
- What is the value of this analysis to my organization and its stakeholders?



The goal is to build a Minimum Viable Product (MVP)—a “draft” analytic product—as fast as possible (usually about two weeks) so users can try it and provide feedback.



Questions to ask:

- Is the MVP working as intended?
- What needs to be fixed/improved in the next iteration?



Questions to ask:

- How do we know the analytic product is final and ready for release?



Questions to ask:

- How could our development process work better?
- What do we know now that would have been useful then?

To illustrate how Agile works, take the following example:

A 3Si client needed to produce new data products in a short amount of time (approximately two months). If they used the Waterfall methodology, they would have spent much of those two months meeting with stakeholders to capture all the project requirements. Instead, the client and 3Si used the Agile methodology. First, we met with a small subset of stakeholders to brainstorm seven use cases. Next, we worked with the client to rank the use cases according to priority and feasibility. Based on these rankings, we selected the three highest ranked use cases and delayed the four remaining use cases. This prioritization simplified the data product, which translated into faster delivery of a usable data product.

Next, the Agile team (the client and 3Si) worked in two-week sprints. The short duration motivated developers to make progress toward the final product (e.g., a use case) and avoid distractions that might hinder that progress. Testing occurred parallel to development; in other words, the Agile team coded and tested all within the same two-week sprint. To ensure the sprint is moving along as planned, it is common to hold brief “stand ups” (meetings) several times a week to make sure the Agile team was aligned, unblocked, and progressing forward. For this effort, we held 15-minute stand-up meetings three times per week. Once the sprint was finished, we met to evaluate the process and product and understand what went well and did not go well, so the next sprint could incorporate that feedback.

Due to the high-speed nature of Agile development, it is critical that the one who is managing the sprint development cycle is experienced in this type of workflow. This individual, commonly referred to as a “scrum master,” leads the development team and is accountable to drive towards a product release. For this particular MVP, 3Si acted as the scrum master and the client acted as the product owner responsible for defining and prioritizing the use cases.

This example demonstrates the following:

1. Complex data gathering, analysis, and presentation can be developed and presented within eight weeks, compared to the typical 12-month duration under a Waterfall methodology.
2. The evaluation step is very important and provided a channel for team members to communicate.
3. Based on the feedback in the evaluation meeting, the team concluded that Agile development is more fun.

MPI EXPANSION IS REQUIRED

A critical step in developing initial analytic and data products is to leverage the MPI. However, as explained in Section 3, the MPI only includes select data sources and does not provide all the data needed to address the priority use cases. To better address these use cases, Oklahoma would need to expand the MPI to include OKDHS, the OSDE, OCCRRA, and Head Start (see Section 3.1.1: Data Landscape Map).

Alternatively, there may need to be a separate matching process that incorporates other raw data in the data lake. For example, the MPI may contain health-related data on children birth to five, but not education-related data for these same children. To match these data, Oklahoma could construct a process similar to the current MPI in the cloud; it would match common fields (e.g., Name, Address, Date of Birth, Social Security Number) from both systems to connect the data and make it useful.

Pending Oklahoma's decision on which CSP it will use, 3Si/Foresight will explore options to expand the MPI in coordination with ECIDS development, in collaboration with OSDH and other key stakeholders. 3Si expects this work to begin in 2020 (see draft project timeline in Section 6).

GETTING STARTED

For each of the priority use cases described in Section 2.2.2, the process starts with transporting data to the cloud. Pending Oklahoma's approval of a data governance plan, each agency will have a landing zone in the data lake to store raw data for all source systems owned by the agency. We recommend that data owners create snapshots of data, specific to the priority use cases, and upload these snapshots to their landing zone in the data lake. This approach will allow:

- Each agency to introduce staff to the new technology.
- Construction of data contributor agreements based on priority use cases.
- Construction of matching processes to supplement the MPI (as needed).
- Engineering of the data to accommodate required certifications (HIPAA, FERPA, etc.).

The next step is to build out more robust and automated data transportation pipelines. This could include repurposing existing data transportation processes (e.g., SSIS packages) or creating new processes using tools like Azure Data Factory, which automate extractions from the data source to the data lake. While Oklahoma is constructing data transportation pipelines, other important tasks can move forward as follows:

- Merge de-identified data across agencies into a consolidated database.
- Develop data models for analytics and reporting.
- Test and deploy visualization tools.
- End users can test products with real data and provide feedback.

Using Agile methodology and real data from the lake, Oklahoma will be able to develop data products quickly, provide value, and collect early feedback. Developing MVPs with real data and gathering feedback from end users (within appropriate data governance protocols) would also expose potential gaps in data. These learnings can be fed back to some source system owners to improve the data source, which will, improve the data in the data lake.³⁷ This continuous quality improvement cycle will produce better results in a shorter amount of time.

The following describes an initial data product for Use Cases #1-3.

³⁷ Some systems are limited to what agency data can be input, especially with older legacy systems like Mainframe/PS2.

MVP for Use Cases #1-3

As noted in Section 3.3.2, the data needed for these use cases fall into several broad categories, which are thematically similar across use cases. While data are readily available for each of these categories, these data will not be complete given the current limitations of source data systems. Though it is not possible to issue a full set of recommendations until more extensive assessments are completed (beyond the scope of this engagement) preliminary mitigations are presented below.

Establishing the entire birth to age five population

As discussed above, the first critical step is to model the entire birth to age five population in order to ensure fidelity of the data model to the use case; this approach is particularly necessary in order to identify unserved and underserved populations. While most children are likely to be identified by existing health records (i.e., birth and vaccination registries, among other sources), relevant demographics will not be available for all of these children. Notably, family income and parent/guardian work status, which are essential for determining program eligibility, are likely to be unavailable for a significant fraction of children. Similarly, demographic data on children who have recently moved to Oklahoma from another state or country may not be available.³⁸

As explained in Section 3, 3Si recommends tracking available demographic data for all children and imputing these demographics based on best available data where necessary. One known source for this information is the American Community Survey (ACS), which provides aggregated estimates on household income level and parent work status for children under age six. These data are typically available at various geographic levels. Use of aggregate data will present inherent limitations on future use cases. For example, aggregate data could not be used for child-level analysis. Therefore, these fields would need to be identified in the data model to prevent confusion or mislabeling as true child-level data.

Oklahoma can use rules to mitigate the otherwise arbitrary assignment of demographics to children with missing data. For example, children served by a program with specific eligibility requirements may be more likely to have certain demographic characteristics (e.g., family income levels that fall within the program eligibility requirements). Such an approach will allow Oklahoma to match, with reasonable confidence, aggregated data on family income with children eligible for certain programs. These constraints can be applied at the aggregate level as well. For example, Oklahoma may be able to make inferences about children with missing data on household incomes by comparing aggregate Head Start data with known demographics of children and applying those distributions to children with missing data. This process is imperfect and would need to be noted in any reporting or interpretation of the data.

³⁸ This population may include children who have moved out of state who no longer need to be considered, though some of the mainframe programs (SNAP, Child care) have a closure code or an indicator that a child moved out of state.

Defining eligibility for specific programs for each child

Program eligibility varies significantly by program. It will be necessary for the data model to support various definitions of program eligibility. In some cases, this will entail designating a program eligibility requirement as a binary input (for example, the input would be a 0 if the child is not eligible for SoonerStart and a 1 if the child is eligible). Or, eligibility for one program may categorically determine eligibility for related programs. In other instances, the data model will need to allow for non-binary input values (e.g., 90 percent of federal poverty level as opposed to “low-income” or “not low-income”), which can then be used to determine program eligibility. As program requirements may change over time, the data model should be built to accommodate such changes.

As eligibility requirements overlap for many programs, children may be eligible for multiple programs at once. In support of the approach described above to identify the levers available to decision-makers, it will be necessary to identify this overlap. For example, some children eligible for Subsidy Child Care will also be eligible for Head Start, while others will be eligible for only one or the other. This example would result in three eligibility segments:

1. Children meeting Head Start requirements but not Subsidy Child Care, because their parents do not meet the work requirements.
2. Children meeting the requirements for Subsidy Child Care but not the more-restrictive income requirements for Head Start.
3. Children meeting the requirements of both programs.

A cogent and actionable analysis of unserved and underserved children will require similar segmentation of all children based on eligibility requirements across all relevant programs.

Determining which children have been served by which combinations of these programs

Where possible, the data model should be based on child-level data linked through the MPI and provider-level data (e.g., quality ratings). Such linking will allow Oklahoma to identify from which programs a child has received service, including overlapping services of multiple programs. Incorporation of service-level details at the child level will enable determinations of dosage, drop-out or opt-out (e.g., children referred for SoonerStart whose parents decline this service), or completion of services. This modeling at the child, program-detail, and provider levels overcomes a typical challenge of early childhood data by allowing for distinct (rather than duplicated) counts of children served by combinations of programs.

While this approach represents the core value proposition of an ECIDS, the limited availability of some child-level data will likely require modeling of service overlap through other means. As an example, the best estimate of children served by private licensed child care is the licensed capacity of the providers. This capacity, along with estimates for typical vacancies, would inform counts of children served at a point in time. Similar to the case noted above for inferring child demographics, this method poses significant limitations to any child- or case-level uses, which would have to be noted in any downstream analysis.

Oklahoma could mitigate these limitations through clearly defined and sensible business rules. For example, it may be reasonable to assume that children receiving private (unsubsidized) child care are unlikely to also receive Head Start. This is because Head Start has strict family income eligibility requirements, so it is reasonable to assume that private child care payments are cost-prohibitive for a

family who is eligible for Head Start. This solution, imperfect as it may be, could serve as the best available estimate, in accordance with the general principle noted above to model what is known as best as possible, despite limitations of data availability.

Visualizing the results

3Si/Foresight recommends that Oklahoma use an interactive data tool (IDT)—such as Tableau—in the cloud that will enable analysis of children eligible for and receiving services, including racial, ethnic, family income, and other data that would be useful to ECIDS stakeholders. Data will need to be properly structured in order to feed an interactive data tool. The IDT should include the ability to perform community-level data aggregations, comparisons across regions, and descriptive elements to provide context and guide data interpretation. It should also allow users to view localized information regarding population characteristics and highlight the potential need for additional early childhood services.

The data categories found within the IDT should be arranged thematically according to the Eligibility-Access-Service-Impact (EASI) framework:

- **Eligibility** – Demonstrates the need for early childhood services in a given area (e.g., service demand, risk factors based on population data).
- **Access** – Demonstrates the supply of service providers (e.g., number of service providers, provider characteristics).
- **Service** – Demonstrates child/family use of services (e.g., enrollment numbers, based on a subset of ECIDS data).
- **Impact** – Demonstrates the effect of services on children/families over time.

Analytic outputs and capabilities of the IDT

A primary theme for Use Case #1 is to adequately model eligibility and service for early learning programs and determine the difference between the two. Note that there is an implicit assumption that those eligible for services should receive these services; therefore, an eligible population who is not served by these programs would be considered a “gap” in service. This may or may not always be the case, but additional modeling could consider program uptake as an additional assumption.³⁹ In accordance with the EASI framework, Oklahoma would calculate service gaps based on aggregate totals for each geography (e.g., zip code, county, etc.) across Oklahoma. The gaps identified based on this landscape analysis could also be associated with the program eligibility of those children who are unserved (who are often eligible for multiple similar or complimentary programs), identifying potential levers for policy makers and early childhood advocates to better reach these populations.

An additional element of gap analysis would be to identify extreme gaps and/or significant barriers to families seeking these services. One such method is a statistical predictive model that identifies *extreme access deserts*. These access deserts are clusters of geographies (e.g., zip code, census tract, etc.) with

³⁹ “Uptake” is the proportion of eligible children or families who use the service for which they are eligible.

service coverage (measured as children served and capacity to serve children) that is atypically low at statistically significant levels. While it may be an insurmountable challenge to address all unmet need across Oklahoma, identifying extreme access deserts could help policymakers and providers develop strategies to overcome barriers to service for certain communities, prioritize service expansion, and make other informed decisions that will help Oklahoma meet the needs of children and families. Appendix F discusses the methodology to determine extreme access deserts.

4.2.7. Data Integration Roles and Responsibilities

Note: This section is under development, as it is dependent on Oklahoma’s data governance approach. Pending Oklahoma’s approval of the data governance plan, a complete description of data integration roles and responsibilities can be created.

Table 4-6 lists some common roles that should comprise the team that builds and maintains an integrated data system such as the ECIDS. This list is not exhaustive.

Table 4-6 - Data integration roles and job titles

Role	Job Title(s)	Description of duties
Stewardship	Data Stewards	<ul style="list-style-type: none"> • Responsible for using an organization’s data governance structure to ensure fitness of data elements, both content and metadata (i.e., Stewards will know what data to use, what NOT to use, and why, and how to use the data). • Manages and maintains user roles and data access policy for the data lake. • Provides a main point of contact for analysts and researchers to access data in the data lake. • Maintains records that allow users to validate the accuracy of data.

Role	Job Title(s)	Description of duties
Programming / Engineers	Data Engineers, Software Developers, Business Intelligence Engineers	<ul style="list-style-type: none"> Writes code that extracts and loads data from source systems to data lake.⁴⁰ Builds and tests data transport pipelines between data systems and the data lake. Manages and operates scheduling of data transportation through Windows or Linux scheduling software. Deploys resources in the cloud and scales infrastructure as needed. Tests code and maintains source control (e.g., GitHub). Stays current with technology platforms (e.g., new cloud features).
Analysis and Reporting	Data Analysts, Business Analysts, Product Analysts, Business Intelligence Analysts	<ul style="list-style-type: none"> Writes code to transform unstructured data from data lake into relational databases (e.g., SQL). Builds automated visualizations, dashboards. Enriches analysis by integrating data with other data sources, such as survey data. Works with research departments within agencies to create data extracts for a specified analytical purpose. Ad hoc reporting and statistical analysis in consultation with agency Research departments
Research, Statistics, Predictive Analytics	Statisticians, Data Scientists, Researchers, SME	<ul style="list-style-type: none"> Works with agency Research departments to design and conduct program evaluations, policy analyses, and other research tasks Writes code focused on statistical research (e.g., SAS, Python, STATA, R). Builds and maintains analytic tools used by government agencies, such as predictive models.⁴¹
Cybersecurity	Data Security Analyst	<ul style="list-style-type: none"> Creates data access policy (who can see what data). Monitors data access and usage. Stays current with evolving threats to cloud security. Works with cloud service providers to ensure the highest safety standards.

⁴⁰ As described in Figure 2-3 in Section 2.3.1, a data engineer will still need to prepare data for extraction from the data source to the data lake. This job takes less time and effort and minimizes the potential for data quality issues associated with the traditional method because data are extracted in their raw, native format, eliminating the need to clean and transform the data first.

⁴¹ Note that, though many of the roles are yet undecided, the roles of these staff relative to the agency staff analysts that will be data users are described in the data governance section.

These roles are rarely centralized and often distributed across agencies, which can lead to siloed work streams and misaligned priorities. Section 5: Data Governance Plan discusses our recommended approach for centralizing management of these functions, to ensure that everyone working on the ECIDS is well-coordinated and aligned around the same goals.

4.3. Key Considerations for Data Integration

Oklahoma stakeholders need to decide on a cloud service provider (CSP) before the project team can finalize data integration plans. Also, 3Si recommends that Oklahoma inventory existing internal capabilities skill sets that could be deployed to implement an ECIDS. It is highly likely that many of the necessary skill sets exist but are distributed across many agencies. Having a consolidated view will expose gaps in resources, gaps in required skill sets, and opportunities to leverage resources, and will provide valuable information for head-count management and up-coming hiring cycles. Additionally, conducting this assessment could assist in resource planning and may even provide opportunities for underused staff.

Section 6 provides high level costs, timelines, and hiring considerations associated with this integration plan.

5.Data Governance Plan

5.1. Introduction

For Oklahoma’s ECIDS to produce the desired impact, the state needs a governance structure that can support interagency data use on an ongoing basis. The success of the ECIDS will require a continued focus on the collective benefit of data sharing and the value to the entire system of having broad agency participation. It will also require a design that supports collaboration among agencies so each agency is advancing its own mission while contributing to the broader field. Leadership from the Governor’s Office and individual agencies will be needed on an ongoing basis to sustain momentum.

No data governance structure is perfect. Through stakeholder engagement we identified key objectives of Oklahoma’s agency leaders; we have combined that information with our knowledge of best practices in other states and the national literature on data governance to develop this proposal. Our proposal is meant to represent the best possible starting point for Oklahoma’s efforts to manage its ECIDS, with the expectation that the structures will need to evolve over time to reflect the State’s lived experience.

5.2. Data Governance Structure

5.2.1. Data Governance Objectives

The overarching goal of the ECIDS itself is to improve service to children and families. To do that, the ECIDS will need an effective governance system. Through the literature review and stakeholder engagement, we have identified six objectives for data governance. These objectives guided our proposal:

- *Focus on child and family outcomes:* The system must view all of its activities through the lens of how they improve outcomes for children and families.
- *Usefulness:* The data system must be useful to end users—with an emphasis on program outputs and impact rather than program inputs and processes—while streamlining and reducing governance and time spent on data extractions, loading, and modeling as much as possible and within appropriate legal and administrative parameters.
- *Action Orientation:* The governance structure must orient toward meeting the operational needs of the Governor’s Office and state agencies. The data governance structure focuses on providing quick and efficient access to data for the Governor’s Office and state agencies, so the data can be used to drive decision-making. It will also emphasize maximizing the use of its existing capabilities at any given time, rather than waiting for additional capacity to be built before acting.
- *Expandability:* The governance structure will be launched with a set of committed state agencies who see the value of partnership. Over time, that structure may be expanded to include other agencies. The system may also begin accepting data from partners outside of state government.
- *Inside-Outside Collaboration:* While the oversight of the system is a core state government function, many states have chosen to include external stakeholders in their governance structure. Oklahoma agency leaders expressed interest in this possibility. In particular, the system will seek to partner with Oklahoma’s tribes and Head Start providers.

- *Inspiration:* Ideally, the new data system will allow agencies to fundamentally change the way they operate by providing them with information that is more comprehensive and useful than any information they had previously (with a focus on outcomes), delivered at a faster speed. This has significant implications for the way agencies operate and deliver services, both individually and collectively. Agencies will need to evaluate their capacity to use the new system on an ongoing basis and expand as needed to improve their operations. To do this, agencies must consider who will be the end users of the data.

5.2.2. Success Factors for the Governance Model

Ultimately, the success of the ECIDS will be determined by whether it provides a return on its investment to participating agencies. The initial setup cost of the ECIDS and its annual carrying cost should allow the State to make better decisions about its resources, so it can more efficiently and effectively provide high-quality services to the children and family who need it most. For the ECIDS to support those child outcomes, however, certain factors must be present in the governance structure, regardless of its exact design:

- *Leadership.* As with any large or long-term project, executive leadership and buy-in to the vision for the ECIDS is critical to its success. Establishing a group of committed leaders, a shared vision, and the mission's objectives early on are necessary to move forward. State leaders need to understand at all times why the ECIDS project is needed, what impact it is having, and what it will take to continue the work; the leaders within the ECIDS governance structure need to communicate these issues with the Governor, legislative leaders, and other top policymakers on an ongoing basis, including bringing new policymakers up to speed. The model must leverage state agency leadership and buy-in to take advantage of the institutional knowledge that will help to address subsequent considerations below.
- *Communication.* Any data governance model should prioritize communication across its various stakeholder groups, particularly between executive leadership and those who work more directly with the data. One common challenge is that high-level leaders do not communicate adequately with the staff and stakeholders who have deep content expertise in the management and use of data. Establishing and maintaining communication channels between these groups should happen early.
- *Hierarchy.* Similarly, agency leaders need to anticipate possible roadblocks and work with staff at multiple levels of their agencies to dismantle or avoid them before they become bigger problems. If certain entities are known to struggle with issues of hierarchy that may stymie the work—such as power structures that prevent agency staff from sharing honest feedback with leadership—identifying and either planning around or working to address these bottlenecks will be necessary and more difficult as time goes on. Taking the time early on to address communication across hierarchical levels within an agency will facilitate later work, when it becomes more technical and task-oriented.
- *Knowledge.* Data literacy means staying informed about current data issues, the questions that data can answer, and the basics of what data can, can't, should, and shouldn't do. It is easier to communicate the worth of a data infrastructure project to those who have data literacy. Assessing data literacy—particularly in executive leadership—in the early stages and taking steps to expand

that knowledge as needed will inform the types and makeup of the governing bodies and highlight potential future barriers to the progression of the work. The ultimate choice of a data governance model may be informed by the overall degree of available stakeholder knowledge and data literacy.

- *Representation.* A data governance model that relies too heavily on executive leadership or higher-level decision-makers and knowledge-seekers can collapse under the weight of its own ambition. The knowledge of those who regularly work with the data, understand relationships between agency departments and/or higher-level entities, and will be impacted most by how the data are used is critical to any model of data governance. While multi-agency or multi-entity approaches run the risk of “too many cooks in the kitchen,” a single-agency approach may encounter problems with a lack of redundancy should agency leadership or legislative mandates change. Both lateral (across entities/stakeholder groups) and vertical (across roles and hierarchies) representation are necessary components of data governance.

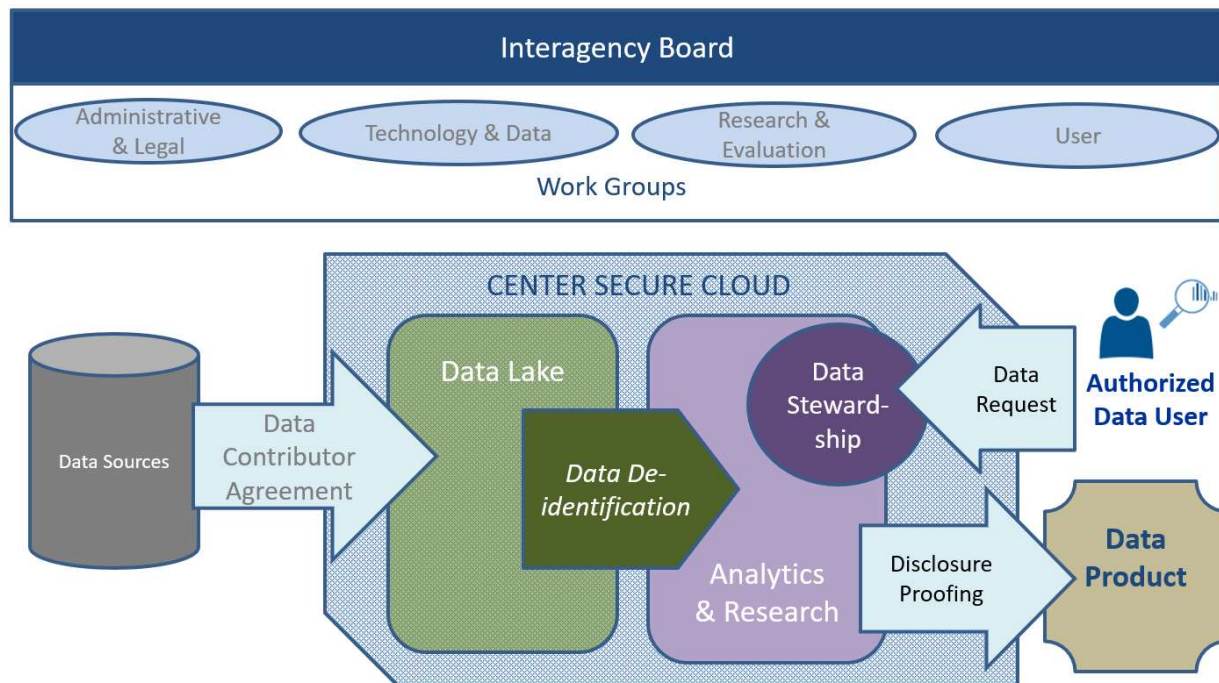
5.2.3. Data Governance Framework

The data governance framework will include three critical components:

1. A high-level interagency board of empowered agency representatives that provides leadership for the governance structure (referred to here as the Board).
2. A lean work group structure that integrates critical agency expertise to facilitate policy development.
3. A “Center” that provides centralized administrative capacity and technical infrastructure.

Figure 5-1 (on the following page) visualizes this proposed data governance structure.

Figure 5-1 - Proposed data governance structure



The Board will be the leading decision-maker for the governance structure and have oversight of the ECIDS. The work groups will have deep expertise in essential components of running the system and communicate regularly to ensure interagency collaboration and inform Board decision-making. The Center will have the capacity to execute critical functions of the ECIDS.

While their roles are described in greater detail below, an important concept underlying the entire design is the *action orientation* identified in 5.2.1. In the Center, staff with specific expertise and dedicated capacity handle most of the day-to-day operations, so the Center will be able to provide rapid response to constituent needs without constantly seeking Board approval. The Board will have a more intensive role in establishing the Center and maintain an ongoing role in setting a direction and priorities for the ECIDS, but it should not be involved in day-to-day decision-making or it will become a roadblock. Similarly, the work groups have a critical role in advising on Board and Center decisions, but they should not become a bottleneck that impedes the Center's ability to respond to the needs of the Governor's Office and participating agencies. Certainly, there will be specific projects that are so complex and difficult that the Board (and work groups) will need to get directly involved, but those should be rare; if Board action is required on too many decisions that have operational urgency, the system will fail.

To that end, the governance structure will be defined by a formal governance agreement that codifies the following:

- The system's goals, requirements, functions, and expectations.
- Agency participants, and their rights and responsibilities.
- The establishment of the interagency board.
- Responsibilities of the interagency board—including its frequency of meeting and its manner of acting.

- The work group structure—including their responsibilities, manner of acting, and the process for restructuring the work groups.
- The role of the Center.
- The process for designating an entity as the Center Administrator.
- How the Center will be overseen and funded.

This proposal lays out a framework for the agreement. Note that if the agencies agree to this framework, there is remaining work to resolve certain issues. The following subsections discuss the elements of the intergovernmental agreement framework. Appendix G includes a draft interagency agreement consistent with this framework.

THE INTERAGENCY BOARD

The Interagency Board will include a chair appointed by the governor and one representative of each agency that is participating in the ECIDS. The governing agreement will give the Board specific powers to make policy and determine priorities for the ECIDS, and agencies will be bound by these decisions while retaining the right to authorize the use of their own data for specific projects. As noted above, the Board will play an intensive role in establishing the Center and provide an ongoing mechanism for agencies to hold the Center accountable. The Board should stay informed of the Center's activities and meet regularly to adjust policies and priorities based on new information. It will have formal channels of communication with the work groups to ensure that work group input shapes the system's operation.

Key responsibilities housed at the Board level will include:

- Defining the vision of the ECIDS: providing a clear articulation of the ECIDS' purpose, the intended benefits to all members and the general public, and the unique opportunity that cross-agency data sharing delivers. Much of this should be defined in the initial governance agreement.
- Adopting standards and business rules for the operation of the ECIDS based on recommendations of the working groups or the Center.
- Defining three basic categories of data use, so the Center does not have to seek Board permission before every use (The sidebar provides more detail on these three categories):
 - Data uses that are always permitted.

DIFFERENT CATEGORIES OF DATA USAGE

We anticipate four major kinds of usage of the shared data infrastructure, each of which will have their own defined rules:

- *Cabinet-driven priorities, including questions of importance to the Governor and the Cabinet.*
- *Agency-driven use, either with their own data (which they can do as often as they like without permission from others) or in combination with data from other agencies (which will include streamlined processes for interagency approval).*
- *Use driven by third-party intermediaries and key stakeholders, such as universities, philanthropies, or non-profits in the field.*
- *Public-facing data for parents and providers.*

Clear definitions of and parameters for data usage will allow data usage requests to be approved efficiently. We anticipate three basic categories of usage:

- 1. Data uses that are always permitted.*
- 2. Uses that are never permitted, including uses forbidden by privacy laws.*
- 3. Uses that may be permitted under some circumstances. These uses will require case-by-case resolution, and the governing body will establish procedures for resolving cases as quickly as possible. Many university and cross-agency research projects will fall in this category.*

- Data uses that are never permitted.
 - Data uses that are potentially permitted but require specific approval; the Board will then define the process for seeking approval.
- Setting a common agenda for data use, including priority use cases and/or a research agenda.
- Driving the implementation of Cabinet-identified priorities through the shared data infrastructure.
- Developing an administration agreement with the entity housing the Center (later in this section, we discuss considerations for selecting an entity to house the Center). This agreement will address important areas for which the Center would have management responsibility subject to the interagency Board's oversight and policy direction, with a defined input role for the work groups. Issues to be addressed in the Center's project scope include:
 - Defining security expectations for the system.
 - Defining privacy rules for the system based on relevant privacy laws—both federal and state—and overseeing compliance with those rules.
 - Defining the processes by which data will be collected from participating agencies and partners.
 - Overseeing and setting policy for the use of the secure, cloud-based data lake.
 - Setting policies for data maintenance and retention.
 - Maintaining data quality standards and a plan for continuously improving data quality. The Center will also explore the possibility of common definitions for specific terms that support consistent data use, while recognizing that it may not be possible to harmonize definitions completely given conflicting federal requirements.
 - A budget for the Center, and timelines for any infrastructure buildout for which the Center is responsible.
 - Defining access protocols to determine who will be allowed access to shared data. This includes establishing and overseeing an appeals process for data requests that are denied.
- Providing oversight of the Center's operations in accordance with the scope of work in the Center's administration agreement, and ensuring that the Center operates in compliance with board policy.
- Approving the individual who will serve as the Center's Executive Director.
- Providing for transparency in governance decision-making by holding public meetings and ensuring that materials relating to the governance structure are publicly available.
- Defining accountability for member agencies about adherence to data governance policies and timely participation in Board- and Center-related decision-making.
- Establishing a process for new agencies or non-agency partners to join the ECIDS.
- Establishing a process for adding board members, which may occur when new agencies or other partners join the ECIDS. The Board may wish to consider establishing a maximum size.

Because the Board is a policy-setting body, it will define policies that will be executed by the Center, as discussed further below. Many of these activities might need to be conducted in close collaboration with the Office of Management and Enterprise Services (OMES). Many of these activities will be supported and informed by work groups, as discussed below.

Representatives to the governing Board must be empowered to act on behalf of their agencies, to allow for rapid decision-making. Ideally this will mean the personal involvement of the agency head, but if the agency head is not available it should be a trusted top deputy who reports directly to the agency head

and understands the full scope of the agency's programmatic goals. When the Board is being established, the interagency agreement should set this standard for all participating agencies and this expectation should be clearly communicated by the Governor and understood by agency leadership. Simply put, the Board will not function successfully if this expectation is not met.

THE WORK GROUP STRUCTURE

For the governing board to be successful it will need to rely on the expertise of a wide range of stakeholders to provide insight and guidance to the Board and the Center in an advisory role. These key stakeholders include:

- Agency early childhood program staff
- Agency staff who are responsible for managing data systems (data stewards), for expertise in technology and data security
- Agency staff who are responsible for research and analysis
- Agency legal staff
- Key external users of data, who include:
 - Legislators (and legislative staff)
 - Researchers
 - The early childhood provider community
 - Advocacy groups focused on early childhood and related fields
 - Advocacy groups focused on vulnerable populations
 - Advocacy groups focused on data privacy
 - Other data end users

To systematize the engagement of these key stakeholders, the Board should establish work groups that provide expertise and insights on the full range of the Board's responsibilities. The work groups should be focused enough that all members will have something to contribute to most of the issues within its responsibility. The 3Si/Foresight team surveyed agencies about four proposed work groups; the agencies responded with recommendations for staff members who would participate in each of those work groups. Appendix H includes the survey and Appendix I includes a table summarizing the responses to the survey.

The four work groups proposed in the survey included:

- **Administrative and Legal:** The board will need advice and guidance on the development, operations, and administration of the ECIDS. This advice should come from people who work directly on issues of data security and legal agreements concerning data.
- **Technology and Data:** The Board will need support from leaders with expertise on the technical infrastructure and data management itself, including knowing what technology and data are available, and how data is collected, defined, and used in IT systems (data stewardship). These experts will make recommendations about curating data from the data sources and informing data access and use. This conversation should include agency data engineers, business analysts, and data stewards.
- **Research and Evaluation:** The Board will need methodological advice and guidance on questions of shared research inquiry. This conversation should include university partners, external

researchers, and other stakeholders that would work directly with the resultant ECIDS data on an analytical level. In combination with representatives for users and advocates, these stakeholders will be involved in helping to set the research priorities for the ECIDS.

- **User:** The Board will need advice about the user experience from external stakeholders and members of the community who have a vested interest in both the collection and use of the data and any resultant research, analysis, and reports (including legislators and legislative staff). This conversation should provide a more outside-in perspective on the work being done by the ECIDS team.

As mentioned earlier, the Board should establish communication channels with the work groups to ensure they exchange information regularly, and that their recommendations are elevated to the Board level. The Board can establish protocols for work group communication that allow for flexibility while complying with Oklahoma’s Open Meetings and Open Records Act. The Center will provide staff support to the work groups (e.g., scheduling meetings, taking notes, etc.).

Note that state agencies may end up with multiple staff involved in the ECIDS governing structure: one on the Board and one or more on each of multiple work groups. Agencies will be responsible for internal communication to ensure their staff are coordinating and providing consistent feedback. Ultimately the work of the ECIDS will only succeed if each agency is able to tap the expertise of staff with different areas of focus.

THE CENTER

The Center’s Responsibilities

The Board will enter into an agreement for the creation of the Center; Appendix J includes a draft Center Administration Agreement. Key responsibilities of the Center include:

Table 5-1 - Responsibilities of the Center

Responsibility	Brief Description
Designated Administrator	The system governing agreements will designate the Center as the authorized representative of the Board and participating agencies. The Center will establish and maintain all system components and manage the system on a day-to-day basis. The Center’s development of all system components will be subject to Board oversight and approval. The Center’s administration of all system components will remain subject to Board policy and oversight.

Responsibility	Brief Description
Project Proposal and Data Use Agreement	<p>The Center will define the process for seeking data use approval, including specifications for submitting a proposal. This process will include a streamlined and accelerated process for submitting and approving “exploratory” usage requests to access data for the purpose of determining the scope and viability of a project without the export of any data. Proposals for other non-exploratory projects will include identifying the required datasets and the expected outcomes of the project.</p> <ul style="list-style-type: none"> • The Center will develop a set of standard data use agreements for different project types and potential users will be expected to agree to these terms.⁴² In some cases, Institutional Review Board (IRB) approval must be obtained before carrying out research. • The Center will develop a web-based workflow and defined timelines for agency review and approval, with the goal of providing rapid responses to legitimate usage requests. • As noted in the objectives, the Board will identify certain uses that are “pre-approved.” The Center will review proposals and have authority to approve any proposals that meet those criteria without further steps. • As noted in the objectives, the Board will identify certain uses that are not permitted. The Center will pre-screen proposed data uses to ensure that they are potentially permitted before moving them on to agencies. • The Center will provide ongoing monitoring of each approved proposal, and the system will provide visibility to the participating agencies on the status of data use requests and the utilization of the agency’s data within the system.
Data Contributors	<p>Each agency contributing data to the system will enter into a Data Contributor Agreement with the Center specifying what data will be submitted, the schedule of submission, and the secure method of submission. Appendix K includes a draft of the Data Contributor Agreement. The Data Contributor Agreement will include a streamlined process to add or modify data in the future. The Center will ensure ongoing compliance with data submission requirements. In accordance with Board policy, new contributors can join the system in the future, potentially including contributors from outside state government.</p>

⁴² University research projects will be expected to reference the university’s Institutional Review Board.

Responsibility	Brief Description
Establish a “de-identification system”	The Center will move the current matching process for the Master Person Index into the cloud, and implement a matching process that results in de-identified data with names Social Security numbers, and any agency-specific identifiers masked. The Center will provide secure management of this matching process, and Board policy will specify the limited circumstances under which names and Social Security numbers may be legally accessed within an “Analytics and Research Hub” (described below). This is the first tier of de-identification within the Center to ensure that all data used for research and analytics has the most sensitive identifiers removed, while still allowing other information—such as demographic information and birthdates—to be used for valid analyses. The second tier of de-identification, Disclosure Proofing (described below), ensures that it is never possible to infer the identity of an individual using a dataset released from the Center, even when that dataset is combined with other accessible data.
Establish an Analytics and Research Hub with core data for data use projects	The Center will create a repository of data for analytics and research. The Analytics and Research Hub will consist of de-identified data that is ready for approved uses. Subject to Board policy and only to the extent authorized for an approved project, personally identifiable data may be moved into the Analytics and Research Hub. Board policy will define how the Center retains all technical documentation (meta-data, inventory, definitions, etc.).
Data Stewardship	The Center will establish data stewardship processes and workflows for agency data stewards to authorize access to that individual agency’s data and monitor the data’s utilization for any approved project. The data stewardship components of the system will also include reporting on data utilization to each agency.
Workspace and Tools	The Center will provide workspace and tools—ideally cloud-based—to use for analysis of data within the secure Analytics and Research Hub so that there is not a need to export the data outside of the system during the analytics process.
Data User Training and Authorization	All users of the system must receive training before using the system. Board policy will define criteria for approval, and the Center will oversee the credentialing process.
Disclosure Proofing and Data Projects	Prior to any data being released from the system, the Center will ensure that it is “disclosure-proofed” to ensure that it does not include personally identifiable data or small cell sizes. Board policy will define any review and comment periods needed to facilitate this analysis.
Security Policies	The Center will be responsible for ensuring data security and privacy in accordance with Board policy.

The Center will hire the staff and contractors needed to successfully execute these functions. In addition to these responsibilities the Center will be responsible for providing several capacities essential to the success of the ECIDS:

- **Communication.** The ECIDS will only thrive if policymakers and the public understand its value and see benefit from its work on a regular basis. The Center will be responsible for ensuring this communication takes place.
- **Legal.** Managing data use agreements and privacy laws requires legal capacity, which the Center will be responsible for providing (on its own or through contracting). For the ECIDS to succeed it is essential that the lawyers involved maintain a strong focus on the ECIDS' operational needs. In other states, agency lawyers have prevented data sharing and use that is allowable by law and does not jeopardize individual privacy. If data is restricted in this manner Oklahoma will not be able to realize the full benefit of the ECIDS.
- **Administrative.** The Center will be responsible for managing procurement, human resources, and other administrative functions, potentially in collaboration with other agencies.
- **Analytic.** While individual agencies will need their own analytic capacity to focus on agency-specific needs, having centralized analytic capacity provides benefit to the Cabinet, the public, and the early childhood field by allowing for the ongoing analysis of critical issues informed by the data. Note that the Center may develop this capacity in-house or choose to partner with others (such as universities) to develop it.

The Center will represent a new form of capacity to support a new kind of technical infrastructure, and will need to be prepared to act rapidly to support the ongoing needs of its partner agencies. It will play a critical role in establishing a new culture of data use in Oklahoma state government.

Where the Center will be Housed

Numerous agencies noted that a close association the Cabinet could provide the Center with leadership, political capital, and other benefits. The primary danger of this approach is that the Center would be too closely identified with the Cabinet or end up with a focus that is too narrow. The Cabinet should be careful to guard against these concerns in establishing the Center.

In particular, it is essential that the Center be seen as a “neutral” organization that provides support to its partner agencies while maintaining trust and acting in their interest, rather than as an organization that uses agency data to increase the Cabinet’s leverage over those agencies. The Center will be most successful if it creates a platform that agencies are eager to use for their own purposes on an ongoing basis; a strong customer service orientation toward its partner agencies is a critical element of the Center’s success. More work is needed with the Cabinet to determine the best location to host the Center.

The Center will also need to establish a relationship with OMES. OMES has indicated from a preliminary view that it supports the concept of the center, so long as the Center’s infrastructure and technology solutions are fit appropriately into the larger context of Oklahoma data initiatives. More work will be needed to establish the proper role for OMES in partnering with the Center.

How the Center will be Funded

Outreach to agencies yielded differing opinions on how best to structure ongoing funding for the Center (and the system as a whole). The two primary modes of payment are centralized funding, such as a

dedicated line item in the state budget, and agency cost-sharing, where each agency contributes to the overall costs of the Center. The feedback from stakeholders on both approaches was mixed.

Table 5-2 - Comparison of centralized funding source to agency cost sharing

	Advantages	Disadvantages
Centralized Funding	<ul style="list-style-type: none"> • May benefit from the political capital of the Governor's Office. • Makes it much easier for agencies to participate and plan for their participation. 	<ul style="list-style-type: none"> • Centralized funding for data systems may be one of the first things to go if the legislature is looking to make budget cuts, unless the legislature has been well educated in the value proposition of the ECIDS.
Agency Cost-Sharing	<ul style="list-style-type: none"> • Agencies may have a greater sense of ownership and responsibility for the success of the system. • If the legislature cuts funding for the ECIDS, agencies may be able to step in to provide funding. 	<ul style="list-style-type: none"> • If one or more agencies pulls out, there may be unpredicted cost spikes for the remaining agencies.

In the stakeholder engagement, no consensus emerged about which approach would be preferable. A hybrid of the two funding strategies—which mixes funding directly from agencies with funding from the state budget—provides advantages of both approaches and may therefore be the most sustainable. It is essential for the State to finalize a long-term sustainability strategy, as states without such a strategy have seen work grind to a halt when federal grants or one-time buildout funding runs out.

- For the work to be sustained there must be a favorable ratio of cost-to-utility. Agencies will not want to continue paying for capacity that does not help them achieve their goals.
- Relatedly, the Center cannot be perceived to be a profit center for its host. There is a perception that some previous data efforts ended up generating revenue for host agencies that was used for purposes beyond the data system. The Center must maintain a narrow focus on its mission and use all of its resources toward that mission, including any support for its budget provided directly from agencies.
- If agencies are being billed for data services, they need their contribution to be predictable and stable. If the cost fluctuates from year to year, that makes it very difficult for agencies to manage.

In other states, ongoing maintenance funds for shared data infrastructure have come from a variety of sources, but most states seem to use state appropriations as a primary funding source, typically through a lead agency. Initial startup funds for many extant ECIDS systems were sourced through Early Learning Challenge – Race to the Top federal grants.⁴³ Similarly, the 2009 federal Statewide Longitudinal Data Systems (SLDS) Grant program provided initial funding for broader Statewide Longitudinal Data Systems.

⁴³ https://www.acf.hhs.gov/sites/default/files/ecd/intergration_of_early_childhood_data_final.pdf

Additional funding through new federal grants is also common, and we identified funding sources such as user fees, third-party grants, and other entity contributions (Kentucky is a notable example of this disparate funding stream being codified in law). Some examples include:

- North Carolina: initial funds through Early Learning Challenge – Race to the Top grant; subsequent funds through state appropriation, new grants such as the federal PDG B-5 grant.
- Oregon: Since its integrated data systems is based in its Office of Forecasting, Research and Analysis, the vast majority of ongoing funds are through state appropriation; specifically, through Shared Services, which is based on revenues received from other parts of Oregon Department of Human Services and Oregon Health Authority for purchased services.
- Kentucky: funded through combination of state appropriation, federal grants, user fees, and other grants and contributions from public agencies or other entities.⁴⁴
- Maryland: Maryland Longitudinal Data System Center received initial funding through 2009 SLDS grants; ongoing support is primarily through state appropriation with additional funding from federal grants.
- Virginia: Virginia Longitudinal Data System received initial funding through 2009 SLDS; ongoing support is primarily through state appropriation to the State Council of Higher Education for Virginia.⁴⁵

It appears there may have been federal funding for data integration efforts via the Affordable Care Act and Medicaid—primarily for eligibility purposes—going back to 2012-2016.⁴⁶ Medicaid does not appear to be a primary or even secondary funding stream for any current data integration systems or maintenance efforts.

All of these issues require further work to resolve. Oklahoma is understandably focused on drawing down federal and foundation funds to support the buildout of the ECIDS, which is the correct strategy. Because those funds will not cover the entire cost of the buildout nor the ongoing carrying costs, the issue of funding for the system should be addressed at the outset.

Summary

Our proposal is for the ECIDS to be managed by an Interagency Board, which will draw on the expertise of work groups and oversee the management capacity provided by the Center. The Board, work groups, and Center will have tightly defined responsibilities and maintain an action orientation to meet the needs of participating agencies. Establishing capacity to manage the ECIDS will require an interagency agreement that reshapes the relationship among child-serving agencies and accounts for the important role of OMES. Further conversation is needed to solidify the exact shape of these relationships and the State's approach to paying for the needed capacity.

⁴⁴ <https://law.justia.com/codes/kentucky/2013/chapter-151b/section-151b.132/>

⁴⁵ <http://vlids.virginia.gov/archive/vlds-2016-general-assembly/>

⁴⁶ https://www.acf.hhs.gov/sites/default/files/ecd/intergration_of_early_childhood_data_final.pdf

5.2.4. Changes to Laws and Regulations

Our analysis of Oklahoma law is that the proposed structure can be implemented without any statutory changes. It is possible that operational issues will arise that may warrant consideration of potential statutory change, depending on where the Center is housed; at this time, however, no statutory changes are needed to begin the process of establishing the Board and the Center.

5.2.5. Privacy Laws and Regulations

One important legal issue facing the Center will be laws governing data privacy and security, which this ECIDS Plan references in numerous places. The Center's operations will be subject to the data privacy and security laws that govern the use of data by the participating agencies. The State of Oklahoma Multi-Agency Data Sharing Agreement includes an Appendix that inventories Oklahoma data privacy, security, and confidentiality laws applicable to all of the signatory agencies as of the date of the Agreement's execution (March 2017).⁴⁷ Foresight's review of these laws identified three primary areas of Oklahoma statutory compliance that need to be accounted for in the Center's development and operations:

1. The restrictions on the release of data and out-of-state hosting under the Student Data Accessibility, Transparency, and Accountability Act of 2013.
2. The general restrictions applicable to all State government agencies on furnishing information indexed by Social Security number.
3. The restrictions on the utilization of Medicaid data.

This review should be treated as an initial analysis for review by Oklahoma-licensed attorneys in the agencies with oversight of these issues.

1. Student Data Accessibility, Transparency, and Accountability Act of 2013 (70 O.S. 3-168)

[Release of Data](#)

Under the federal FERPA law, Oklahoma's ECIDS partners may disclose any student data required for the system without parent consent using FERPA's audit or evaluation of Federal or State education programs exception. FERPA's regulations permit "authorized representatives" (such as researchers or "State and local educational authorities") to access student's education records in connection with an audit or evaluation of Federal or State supported education programs. The U.S. Department of Education's FERPA regulations broadly define audit and evaluation. The department interprets

⁴⁷ We are not aware of any other Oklahoma data privacy laws that have passed since that time, but that issue should be reviewed by attorneys in the relevant agencies.

“evaluation” to “include all manner of studies, assessments, measurements, appraisals, research, and other efforts, including analyses of statistical or numerical data derived from education records.”⁴⁸

In addition to federal law, there are state laws that govern the release of data. According to Oklahoma’s Student Data Accessibility, Transparency, and Accountability Act of 2013 (SDATAA) (70 O.S. 3-168), the OSDE must develop criteria for approval of research and data requests from state and local agencies, the State legislature, researchers, and the public. (70 O.S. 3-168 C.2.c.) The SDATAA provides that under these criteria, “unless otherwise approved by the State Board of Education to release student or de-identified data *in specific instances*, the Department may only use aggregate data in the release of data in response to research and data requests.” *Id.* (Emphasis added.)

Under this provision, OSDE’s provisioning of data to the Center through the Data Contributor Agreement will require the State Board of Education’s approval of that agreement. However, for the reasons described below, State Board of Education approval should not be needed for each individual project within the Center, as student or de-identified data will never be “released” from the Center environment.

The Center and OSDE should take the reasonable position that de-identified and student data are never “released” when utilized for a project within the Center’s Research and Analytics Hub. The term “release” is not defined within the SDATAA, and it does not appear that the Oklahoma courts have interpreted the term in the SDATAA context. Merriam-Webster defines “release” as “to set free from restraint, confinement, ... to relieve from something that confines, burdens....”⁴⁹ Within the Center environment, de-identified and student data will never be set free from restraint or confinement; they will remain subject to strict data stewardship, workflow and access management, and data security and privacy controls. Data will only be “released” once they have been fully disclosure-proofed as aggregate data, which does not require State Board approval under the SDATAA.

Out-of-state Hosting

Under the SDATAA, “[u]nless otherwise approved by the State Board of Education, the State Department of Education shall not transfer student or de-identified data deemed confidential ... to any federal, state or local agency or other organization/entity outside of the State of Oklahoma...” except for several permitted exceptions. (70 O.S. 3-168 C.3). One exception is for OSDE contracts that “govern databases, assessments, special education or instructional supports with an out-of-state vendor.” *Id.*

The proposed cloud services of the Center will include hosting on out-of-state servers, even if the Center is administered by an Oklahoma-based entity. As previously noted, the State Board of Education will need to approve OSDE’s Data Contributor Agreement due to the release of OSDE data necessary for the

⁴⁸ There are additional considerations under federal law relating to the use of Head Start data. Because at this time the state does not store any Head Start data and it would not be included in the launch of the Center, we have not fully analyzed those additional considerations.

⁴⁹ <https://www.merriam-webster.com/dictionary/release>

Center's operations. To ensure out-of-state hosting is authorized under STADAA, the Agreement should specify that out-state-hosting will occur within the Center's technical environment.

2. Furnishing Social Security numbers

Under Oklahoma law generally governing State agencies, "[n]o state agency, board, commission or other unit or subdivision of state government may furnish any information indexed by social security number unless required by law or specifically authorized to do so by the holder of said social security number." (74 O.S. 74-3113) Similar to the analysis of release, this Section raises the question of the meaning of "furnish," which is not defined in statute, and it does not appear that the Oklahoma courts have interpreted the term "furnish" in the context of this provision. Merriam-Webster defines "furnish" as to "provide what is needed," a much broader definition than release.⁵⁰ Providing a researcher with a research-ready data set for analysis within the Center would presumably meet the definition of "furnish," and therefore such datasets should not be indexed by Social Security number. As a result, Center personnel (either staff or contractors) should manage all data involving Social Security numbers, and only provide data users with datasets that are indexed by the ID numbers generated through the Master Person Index system.

3. Using Medicaid Data

Under Oklahoma law governing the Oklahoma Health Care Authority, "All applications and records concerning any applicant or recipient under the Medicaid Program shall be confidential and shall be open to inspection only to persons duly authorized by the Oklahoma Health Care Authority, this state, or the United States, *and for purposes directly related to plan administration.*" (63 O.S. 63-5018, emphasis added). Under this same statutory provision, "[p]urposes directly related to plan administration" are limited to establishing eligibility, determining the amount of medical assistance, providing services to recipients, conducting or assisting with an investigation or prosecution, or civil or criminal proceedings in relation to the administration of the State Medicaid Program.

The Center's Data Contributor Agreement with the Oklahoma Health Care Authority will need to designate the Center Administrator as a "duly authorized" entity to support plan administration. Of the "purposes directly related to plan administration" allowed by law, research and analysis within the Center will likely need to be deemed in furtherance of "providing services to recipients." Any research project involving Medicaid data should include a description directly tying that research to the provision of Medicaid services. For example, the research scope could describe how its findings can be used by Department of Health staff to improve program delivery and reach.

5.2.6. Changes to Interagency Agreements

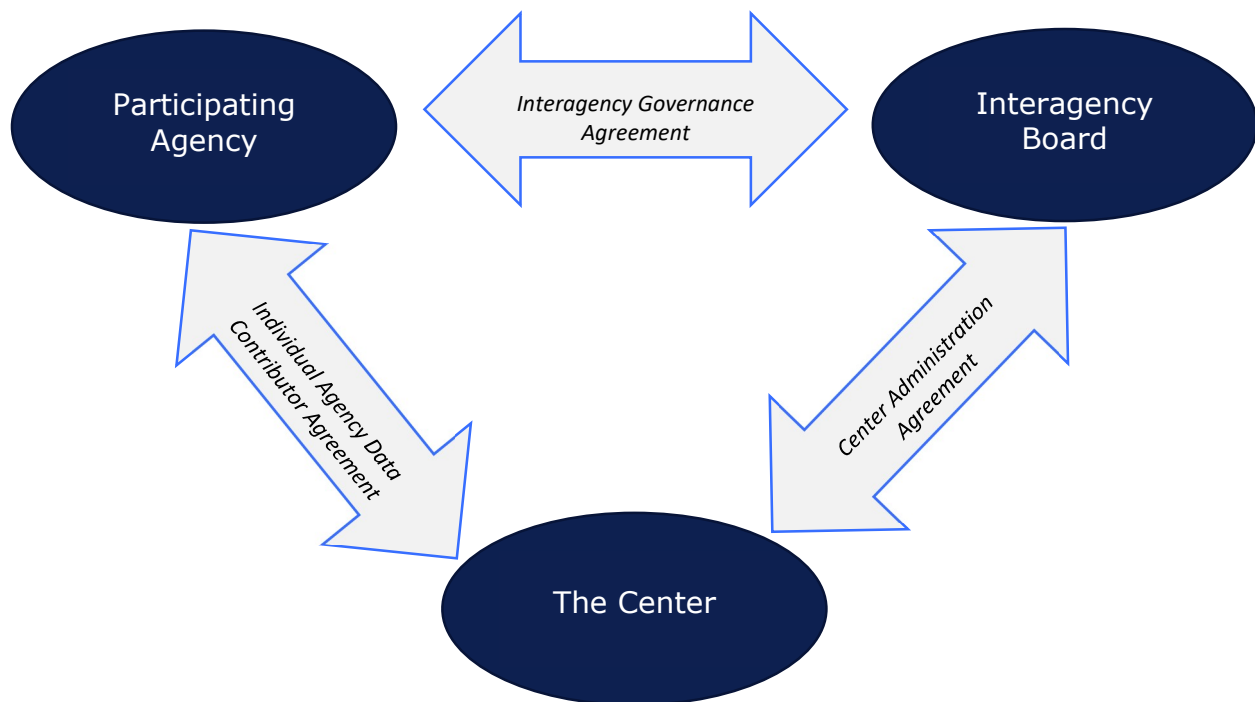
Oklahoma has an existing data use agreement, which was developed through the DISCUSS Committee. That agreement can continue to be used as needed, but new agreements should become the primary agreements dictating the use of interagency data.

⁵⁰ <https://www.merriam-webster.com/dictionary/furnish>

Figure 5-2 shows the relationship between the agreements and entities involved in ECIDS governance. The proposed structure uses three different kinds of agreements:

- The master interagency agreement creating the Board and detailing its responsibilities.
- An administrative agreement between the Board and the Center, which lays out the Center's responsibilities.
- Data Contributor Agreements, which are the agreements between the Center and each individual agency detailing the process by which the agency will submit data to the Center.

Figure 5-2 - Relationship among agreements and entities



5.3. Methodology for Developing Data Governance Framework

5.3.1. Stakeholder Engagement

Our stakeholder engagement included multiple conversations with relevant agencies. In particular, in late July and early August of 2019 we held individual conversations with each of the following agencies:

- Oklahoma Department of Human Services
- Oklahoma Department of Mental Health and Substance Abuse Services
- Oklahoma Department of Rehabilitation Services
- Oklahoma Health Care Authority
- Oklahoma Office of Juvenile Affairs
- Oklahoma State Department of Education
- Oklahoma State Department of Health

We also met with Steve Buck, the Cabinet Secretary for Early Childhood, and collaborated closely with OPSR, which manages the PDG-B5 Grant funding the work. Our analysis also draws on recommendations from a 2018 report on the status of Oklahoma’s work to integrate its early childhood data.

We initiated each agency meeting with an email to the agency head and any other agency personnel who had been part of the ECIDS development discussions. Prior to each meeting, we sent the agency a set of guiding questions. Appendix L includes the guiding questions.

We presented an interim summary of the proposal to a group of agency leaders in Oklahoma City on August 26, 2019. The purpose of this presentation was to acquire sufficient buy-in from agency leaders to inform Oklahoma’s application for a Renewal Grant in the Preschool Development Grant-Birth to Five program. At that time agency leaders expressed their support for the proposed direction, and OPSR submitted a Renewal Grant application based on that approach.

Following the submission of the Renewal Grant application agencies, we surveyed agencies about their preferences for the creation of work groups. We distributed the survey to agencies in November 2019 (Appendix H) and collected responses in December 2019 (Appendix I).

5.3.2. Review of Data Governance Models

In Appendix M, we provide examples of each of several different kinds of governance models. These include models led by:

- State agencies
- A central executive
- Independent not-for-profits
- Universities
- The private sector
- A hybrid model

Each of these approaches has different benefits. In Oklahoma, the Governor’s Cabinet is focused on models in which the central executive takes an active leadership role, and our proposed structure reflects that focus. While the proposed structure draws from several of the models discussed in Appendix M, it does not look exactly like any of them. Overall, the plan is most closely modeled on Kentucky and Maryland, which have created independent executive-led integrated agencies; Appendix M.2 describes these models in greater detail. We made modifications to the specific structures used in those states to reflect the Oklahoma context.

5.4. Critical Issues for Data Governance

At our presentation of this framework to Oklahoma state government leaders on August 26, we received confirmation that key constituents are comfortable with the overarching frame of the governance system design. Since then we have worked to resolve additional issues, as reflected in this final plan. The primary outstanding issues include resolving where the Center will be hosted and its relationship with OMES, and the proposed method for funding the Center. Progress has been made toward resolving these issues, and further engagement with the Cabinet will be needed to finalize plans addressing those issues.

6.Future Work Necessary to Implement Oklahoma’s ECIDS

6.1. Introduction

This section outlines the immediate next steps to so implement an ECIDS, including an implementation timeline, a draft hiring plan, and projected costs over a five-year period. While there are still unknowns, the Center and the Board, with support from the State and guidance from the agencies involved, will adjust this plan going forward to account for evolving circumstances or new information. Strong and consistent Center and Board leadership can help ensure successful execution and maintenance of an ECIDS, even with changes to the ECIDS plan.

6.2. Draft Implementation Timeline

Figure 6-1 shows a preliminary implementation timeline. While this timeline is subject to change depending on external factors—such as availability of funding and agency adoption of agreements—the indicated phases and sequences of work should remain the same.

Initially, the timeline entails creating a Board, identifying a workgroup structure, and creating a temporary Center in the first half of 2020. Appointing an Executive Director should happen in the second half of 2020. Data transfer and analysis will begin in 2020, with initial MVPs expected by 2021. The initial MVPs will demonstrate the usefulness of the ECIDS for core stakeholders—including data contributors and funders—which should help build the case for continued investments to sustain the system.

By the end of 2021, the Center will hire a full complement of staff and move from a temporary to permanent home. In 2022, the Center will continue to work on the priority use cases and develop additional use cases. By the end of 2022, hiring for the Center will be complete, and a long-term sustainability plan will take effect. Sustainable ECIDS operations and maintenance will then continue in 2023 and beyond.

This timeline assumes that the Governor or his Cabinet approves the creation of a permanent Center, and Oklahoma obtains the funds to establish and run a permanent Center.⁵¹ These assumptions notwithstanding, the 2020-2024 draft timeline includes the major milestones outlined in the figure below.

⁵¹ The estimated funds needed take into consideration Center expenses; transitional Center support; outsourced Center support; Azure expenses; data transport, storage and maintenance; and legal support.

Figure 6-1 - Draft implementation timeline

2020		
BOARD <ul style="list-style-type: none"> Establish Board Convene work groups Contract with agency to serve as temporary Center until permanent Center can be established Hire Center Executive Director Serve as final authority over ECIDS and Center 	CENTER <ul style="list-style-type: none"> Hire outsourced services (contractor) Implement necessary legal frameworks and data agreements (in coordination with Board) Carry out initial steps of ECIDS implementation, including setting up cloud and assessing existing cloud security expertise Begin developing MVPs for Use Case #1 	OUTSOURCED SERVICES <ul style="list-style-type: none"> Backfill temporary Center capacity until permanent staff can be hired Support technical aspects of ECIDS development, including data transport, storage, and maintenance Support development of MVPs for Use Case #1
2021		
BOARD <ul style="list-style-type: none"> Continue serving as final authority over ECIDS and Center Work with Center to determine appropriate mix of in-house and outsourced resources 	CENTER <ul style="list-style-type: none"> Hire and train core staff Complete transition from temporary to permanent Center by the end of 2021 Develop data products for Use Cases #1, 2, and 3 	OUTSOURCED SERVICES <ul style="list-style-type: none"> Continue supporting data transport, storage, and maintenance Support development of data products for Use Cases #1, 2, and 3
2022		
BOARD <ul style="list-style-type: none"> Continue serving as final authority over ECIDS and Center Work with Center to finalize sustainability plan to continue Center operations 	CENTER <ul style="list-style-type: none"> Develop and implement sustainability plan to continue Center operations Incorporate more data from partner agencies into ECIDS Continue developing data products for Use Cases #1, 2, and 3; add Use Cases #4 and #5, plus additional use cases 	OUTSOURCED SERVICES <ul style="list-style-type: none"> Continue supporting data transport, storage, and maintenance Support development of data products
2023-24		
BOARD <ul style="list-style-type: none"> Continue serving as final authority over ECIDS and Center Add new agencies as appropriate 	CENTER <ul style="list-style-type: none"> Continue implementing sustainability plan 	OUTSOURCED SERVICES <ul style="list-style-type: none"> Continue supporting data transport, storage, and maintenance Support development of data products

6.3. Draft Hiring Plan

During the first half of 2020, Oklahoma should identify a temporary agency to serve as the interim Center. Each agency will also need to identify and appoint representatives to the Board (the Governor should identify a Chair). The Board will need to create work groups that allow access to agency expertise for decision-making and to facilitate policy development.

This temporary Center will need to contract an outsourced organization to oversee technical aspects of the project and augment Center capacity until the Center hires permanent staff. The temporary Center, under the direction of the Board, will also recruit and hire an Executive Director for the Center.

In the second half of 2020, the Board will need to hire an Executive Director of the Center. The Executive Director will manage outsourced subcontractor services, including data transportation, storage, and maintenance and development of MVPs for use cases #1, 2, and 3.

During 2021, the Executive Director will fill roles in the Center. The sidebar provides a high-level overview of these roles, Appendix N (Center Roles and Responsibilities) provides more detail, and Appendix O (Recruitment Challenges and Strategy) recommends a hiring strategy.⁵²

During this period the Director will also determine the appropriate mix between in-house employees and outsourced services, and how an outsourced service organization should be integrated with the Center's permanent staff.⁵³ This mix of resources will be responsible for continuing to transport, store, and maintain data from partner agencies to increase data use. They will also develop data products associated with priority Use Cases #1, 2, and 3. Permanent staff

CENTER DATA ANALYTICS ROLES AND RESPONSIBILITIES

The following will need to be adapted to reflect who the Center can hire within a competitive market, which contractors the Center retains, and how the two fit together to comprise a team.

Executive Director

Oversee analytics and ensure outputs align with Center and stakeholder priorities.

Analytics Director

Set direction and priorities.

Analytics Manager

Recruit analytics team; manage tasks; plan and manage Agile sprints.

Data Engineer

Move data to and from systems; clean and format data.

Business Intelligence Engineer

Perform analysis using SQL; merge and clean data; build self-service tools; generate reports.

Data Analyst

Perform analytics with Excel and SQL; create reports using self-service tools.

Research Scientist

Build and monitor predictive models; publish research findings

⁵² Note: These roles do not include data stewards, as this role may exist within participating agency staff.

⁵³ A discussion of outsourceable roles and services is in Appendix O (Recruitment Challenges and Strategy).

will also assist the Director to carry out the ECIDS implementation plan.

Hiring for the Center will continue through 2022, and all remaining FTEs onboarded by the end of that year. With a permanent team in place, the Center will turn its attention to preparing and implementing a sustainability plan for ongoing operations, development and maintenance of the ECIDS, and to continue running the Center. The Center will operate according to this plan from 2023 onward.

6.3.1. Recruitment Challenges and Mitigations

Hiring in the technology field presents unique challenges compared to other industries. First, salary expectations can be high, due to intense market demand for technology skills. Oklahoma can try to leverage a generous benefits package, remote work/flex time, and job security to help compete the private sector.

Second, it can be difficult to identify ideal candidates for specific roles—for example, the responsibilities of a ‘data analyst’ can vary greatly across organizations. Defining the role with as much specificity as possible will help filter out candidates who do not possess the knowledge and skills required for the position. Similarly, credentials and qualifications do not necessarily reflect talent and fit. While some candidates offer impressive educational backgrounds, other criteria may be more predictive of success in a technology related role. We recommend administering technical tests as part of the interview process and focusing on candidates who demonstrate the ability to address technical problems with efficiency and accuracy. See Appendix O (Recruitment Challenges and Strategy) for a more comprehensive discussion of recruitment challenges and mitigations.

6.4. Estimated Five-year Implementation Costs

3Si generated estimates of the five-year implementation costs of an ECIDS with the understanding that the budget is preliminary and OPSR will continually adjust and refine its specifics. Within this context, the section below highlights several key points.

Figures 6-3 and 6-4 below respectively summarize the high- and low-cost estimated annual costs by expense category. See Appendix P (Preliminary Five-Year Cost Estimate for Oklahoma ECIDS) for details on Center and other ECIDS expenses. These estimates draw on a large body of data and research. We forecast that the ECIDS will cost between \$2.9-3.9M in Year One, \$4-5.2M in Year Two, \$4.1-5.3M in Year Three, \$3.1-4.2M in Year Four, and \$3.1-4.1M in Year Five onward.

Figure 6-3 - ECIDS implementation cost projection (High estimate)

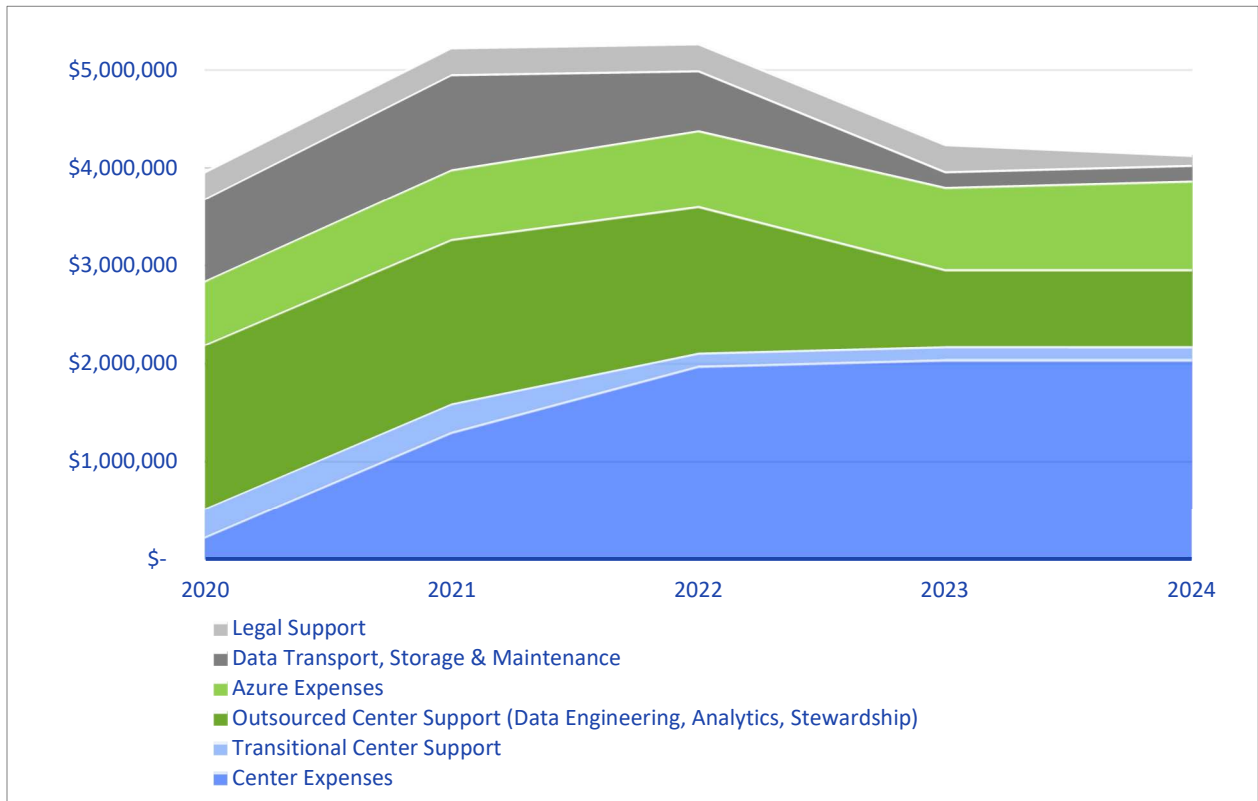
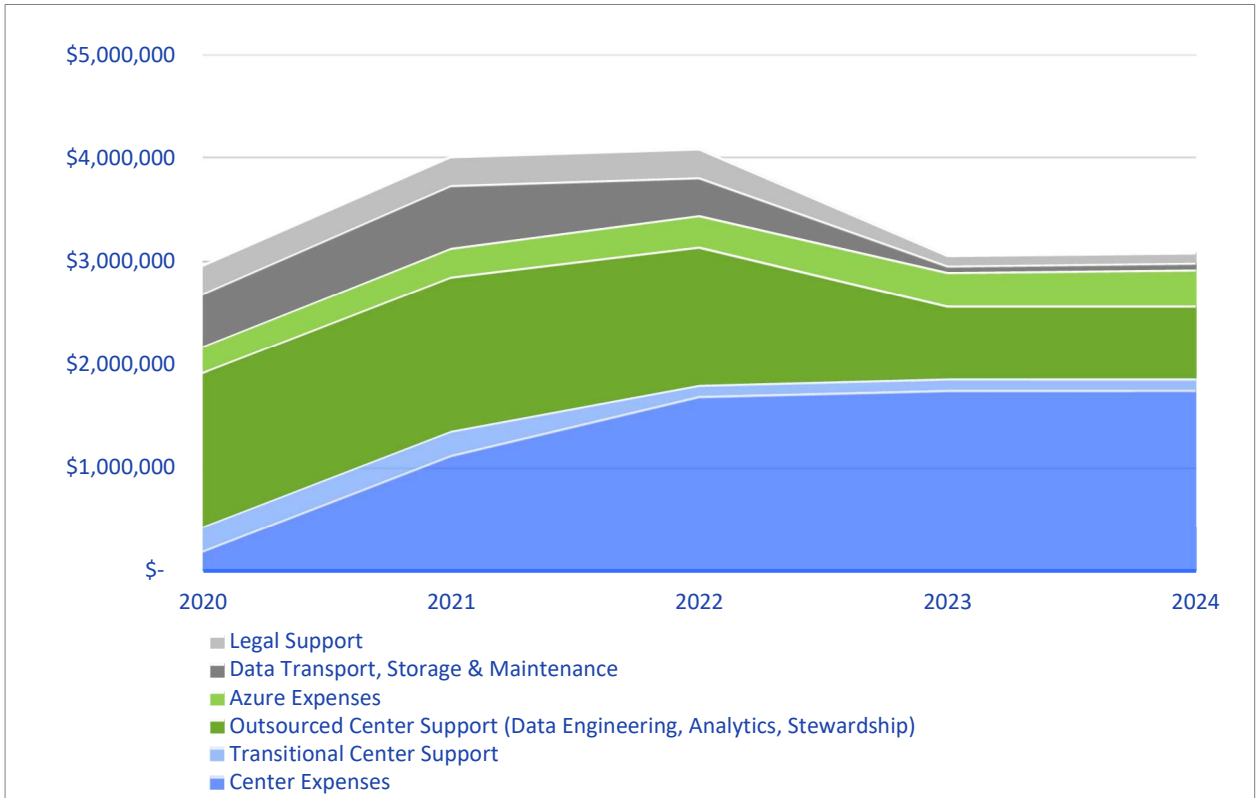


Figure 6-4 - ECIDS implementation cost projection (Low estimate)



Appendix Q provides expense categories and descriptions.

Projections are based on private market labor and costs to maximize the probability of success for the project. Private market compensation levels are considerably higher than typical government salary bands; the project budget skews higher because of this assumption. However, it is the project team's opinion that compensation in this case is highly correlated with, and predictive of, successful project outcomes. If Oklahoma recruits and compensates high-level talent, the odds are greater that the project will be successful. Conversely, if Oklahoma constrains its compensation to below-market rates within a state government salary schedule, the odds of failure increase significantly. Oklahoma's higher education system has shown a willingness to pay for talent when the occasion demands, and this occasion demands that approach.⁵⁴

6.5. Next Steps and Considerations

As described in the timeline above, we recommend that in the first half of 2020 Oklahoma establish three components of governance:

1. A high-level interagency board of empowered agency representatives that provides leadership for the governance structure (referred to here as the Board)
2. Work groups that integrate critical agency expertise to facilitate policy development
3. A "Center" that provides centralized administrative capacity and technical infrastructure

To create a Board, Oklahoma must decide appropriate representation from the agencies involved. Oklahoma must also decide on the composition of the work groups. Appendix I (Results of Agency Work Group Survey) includes the names of nominated staff members suitable for work groups.

In 2020, once Oklahoma establishes an ECIDS governance framework and protocols, the Center and its contractors will begin transferring, storing, and analyzing data. To do this, Oklahoma must first select a cloud service provider and set up the appropriate cloud security and data security protocols. Section 4 discusses these topics in more depth. If these initial milestones hold, Oklahoma can expect the production of initial MVPs by 2021.

6.5.1. Project funding and ownership

The Governor's office, OPSR, and other state agencies with vested interests in this work will need to secure project funding.⁵⁵ Significant funding will, of course, be necessary to start this work, but this funding must also be reliable to justify these initial investments as part of a long-term sustainable system. Such funding could come from a combination of state, federal, philanthropic or private sources.

⁵⁴ <https://oklahomawatchdata.org/university-salaries>

⁵⁵ In December 2019, Oklahoma learned that they would not receive renewed funding through the Preschool Development Grant Birth through Five (PDG B-5) program. While this development does not change the overall recommended plan to create and maintain an ECIDS, it may affect the project's timeline, hiring plan, and budget.

Regardless of the funding scenario, Oklahoma must decide on the location of the Center. OPSR and 3Si/Foresight have considered several agencies or entities within the state (for example, OMES or OPSR) as possible temporary locations for the Center while the other pieces fall into place to support long-term sustainability of the ECIDS. 3Si/Foresight recommend Oklahoma rely on OPSR as a manager or transitional agent for the Center, perhaps in coordination with another state institution should another agency or entity with capacity indicate its interest. Finalizing this decision will be dependent on available funding, interest, and capacity, among other factors.

Conclusion

Oklahoma is well-positioned to effectively implement a mixed delivery system. The state has a strong foundation in early childhood education, has identified the data linkages needed to build the ECIDS, and has convened a wide and diverse array of stakeholders to develop a thorough plan for implementation and maintenance. The state's next steps will enable ECIDS implementation, which capitalizes on the momentum it has built and support it has garnered over the past several years and will enable an increased ability to improve early childhood opportunities for all young Oklahomans.

Appendices

Appendix A: OK Early Learning Stakeholders

The following organizations are engaged in Oklahoma's state data systems work:

- George Kaiser Family Foundation
- Oklahoma Commission on Children and Youth
- Oklahoma Department of Human Services
- Oklahoma Department of Mental Health and Substance Abuse Services
- Oklahoma Department of Rehabilitation Services
- Oklahoma Head Start State Collaboration Office
- Oklahoma Health Care Authority
- Oklahoma Office of Juvenile Affairs
- Oklahoma Office of Management and Enterprise Services
- Oklahoma Partnership for School Readiness
- Oklahoma State Department of Education
- Oklahoma State Department of Health

In addition, the following organizations have a stake in development of an ECIDS:

- American Institutes for Research
- Community Action Project of Tulsa County (CAP)
- Inasmuch Foundation
- Oklahoma Department of Commerce
- Oklahoma Department of Labor
- Oklahoma Department of Libraries
- Oklahoma Institutions of Higher Education
- Oklahoma State Regents for Higher Education
- Potts Family Foundation

Appendix B: System and Data Inventory Templates

Basic System information

- Agency/Org. Name
- Data System Name
- System Abbreviation
- Subsystems
- System Vendor Name
- System Owner
- Subject Matter Expert
- Database Backend Access
- Description of data in system
- How are these data collected?
- Are any data in this system used to connect to other data systems, either within the agency or externally? If yes, please list known data system connections.

Additional comments

- Data history
- Annual cycle for data collection. (Select from drop down menu)
- First year of data available
- Most recent year of data available
- Frequency of record updates
- Frequency of new records
- Years with known data issues (e.g., missing, erroneous, or unusable data)
- Additional comments

Technical Information

- Database name
- Primary RDBMS (Relational Database Management System)
- Primary access client(s)
- Additional comments

Usage

- Key users of this data
- Describe reporting capabilities

Priority Content Area

- What Programs are stored in your system
- Family Data Available and Table Names
- Child Data Available and Table Names
- Provider Data Available and Table Names
- Program Data Available and Table Names
- Provider/Program/Child Data Available and Table Names
- Program/Assessment/Child Data Available and Table Names

Mapping to MPI fields

- Guardian
- Mother Maiden Name
- Mother Name
- Mother DOB
- Father Name
- Phone
- Address
- Child Name
- DOB
- Plural Birth Flag
- Gender
- SSN

Appendix C: Data Inventory

C.1 Agencies and Data Systems by Tier Status

Table C.1.- Agencies and data systems by tier status

Tier 1
Oklahoma Child Care Resource and Referral Agency (OCCRRA)
Oklahoma Department of Mental Health and Substance Abuse Services (ODMHSAS)
Oklahoma Health Care Authority (OHCA)
Oklahoma Department of Human Services (OKDHS)
Office of Management Enterprise Services (OMES)
Oklahoma State Department of Education (OSDE)
Oklahoma State Department of Health (OSDH)
Oklahoma Head Start (33 grantees in total – see representative sample below)
Tier 2
Bureau of Indian Education (BIE)
Oklahoma Department of Rehabilitation Services (DRS)
Oklahoma's Data Integration Project Contractors (Early Childhood Data)
Governor's Interagency Council on Homelessness (GICH)
Oklahoma Commission on Children and Youth (OCCY)
Oklahoma Department of Commerce (ODOC)
Oklahoma Association of Community Action Agencies (OKACAA)
Oklahoma Works (OKWORKS)
Tier 3
Center for Early Childhood Professional Development (CECPD)
Oklahoma Health Information Network (OKHIN)
Oklahoma Professional Development Registry (OPDR)
Oklahoma Office of Juvenile Affairs (OJA)

C.2 System-Level Inventory of Data Systems

[Note: Tables C.2-1 through C.2-7 include acronyms that were listed by data system owners during the interviews. Wherever possible, 3Si has conducted research and followed up with data system owners to clarify the acronyms. See Acronym section of this report for definitions.]

Table C.2-1 - System-level inventory of Oklahoma State Department of Health Data Systems

Basic Information for Tier 1 Data Systems				
Data System Name	Description of Data in System	Data Availability	Data Elements Available for Priority Use Cases	Use Case
Enterprise Master Person Index (eMPI)	The eMPI (or MPI) is the centralized person mapping system. It matches individual records from all Department of Health Systems and Health Care Authority Data Systems, to ensure the data of each person is correct and consistent throughout the enterprise regardless of which system is being updated.	The Data from the Oklahoma State Immunization Information System (OSIIS) is updated nightly; all other systems are updated by batch processing as needed. Currently OSIIS does not have Vital Records data beyond August 2017. There are no other known data issues and all records with issues have all be marked in the system. They estimate that this system currently includes 65 to 80 percent of children under five, although this proportion is expected to significantly increase as new records are loaded into the system in the near future.	<ul style="list-style-type: none"> • Address • Child date of birth • Child race • Child ethnicity 	1, 2, 3
Case Management Client Information Systems (Systems_ETO)	The Case Management Client Information Systems is an online database for Home Visiting Programs to monitor performance outcomes. It also provides assessment data.	Home Visiting staff enter the data daily. There are no known data issues reported by system managers. This system reports only having less than one percent (2,000) of children under five.	<ul style="list-style-type: none"> • Family income • Child address • Child date of birth 	1, 2, 3
Oklahoma Birth Defects Registry (OBDR)	The OBDR's primary purpose is to identify opportunities to prevent birth defects through educational outreach and optimize early detection of birth defects.	The data are collected daily from medical records at all birthing hospitals. The staff may refer clients to SoonerStart. There are no known data issues reported by system managers. This system is limited to children with disabilities at birth born in Oklahoma.	<ul style="list-style-type: none"> • Child address • Child date of birth • Child disabilities • Child race • Child ethnicity • Child gender 	1, 2

Basic Information for Tier 1 Data Systems				
Data System Name	Description of Data in System	Data Availability	Data Elements Available for Priority Use Cases	Use Case
Oklahoma Childhood Lead Poisoning Prevention Program (OCLPPP)	The OCLPPP provides screening and testing for lead exposure for eligible children six to 72 months of age and follow-up for children with blood lead levels that are 10 µg/dl or greater. OCLPPP may refer clients to SoonerStart and other programs. This system will also provide data on clients referred to SoonerStart and helps determine the gap in those that were referred to the program and did not enroll.	The data are batch uploaded daily to the agency. The staff may refer clients to SoonerStart and other programs. There are known data issues reported by system managers. This system is limited to only those children who have received a blood lead test.	<ul style="list-style-type: none"> • Child address • Child date of birth • Child disabilities • Child race • Child ethnicity • Child gender 	1, 2
Public Health Oklahoma Client Information System (PHOCIS)	The PHOCIS is used by county health departments for client registration, appointments, and billing purposes. The system captures program-specific data from a number of programs supporting child and adult health.	<p>Program staff update the data daily. Some data are loaded from the Inventory Supplies System and OSIIS and other data systems (ETO, PHIDDO, ORBS, LEP, ISS). The staff may refer clients to SoonerStart. There are known data issues that have been documented and are still being corrected.</p> <p>This system currently reports to have 38 percent (102,000) of children are represented in this system.</p>	<ul style="list-style-type: none"> • Family income • Child address • Child date of birth • Child disabilities • Child race • Child ethnicity • Child gender • Program type [Temporary Assistance for Needy Families (TANF), Supplemental Security Income (SSI)] 	1, 2, 3

Table C.2-2 - System-Level Inventory of Oklahoma child care resource and referral agency data systems

Basic Information for Tier 1 Data Systems				
Data System Name	Description of Data in System	Data Availability	Data Elements Available for Priority Use Cases	Use Case
Work Life System (WLS)	The Work Life System (WLS) is a web-based system that contains resources for families looking for child care. The users can search for child care by location, quality rating, provider type, and some limited programs. There are no program or family data stored in this system.	Every six months, regional staff call providers directly to collect updated data and enter the data manually. The staff also manually verify and update quality ratings. The data system is maintained offsite by Work Life Systems. OCCRRRA does not have direct access to the backend and needs to request any data they cannot access themselves through existing reports from Work Life. WLS currently contains data on 2,917 providers.	<ul style="list-style-type: none"> • Child care provider • Child care provider address • Child care provider quality • Child care provider capacity • Child care provider preferred capacity • Child care provider licensing • Child care provider type 	1

Table C.2-3 - System-level inventory of Oklahoma Department of Mental Health and Substance Abuse Services data systems

Basic Information for Tier 1 Data Systems				
Data System Name	Description of data in system	Data Availability	Data Elements Available for Priority Use Cases	Use Case
Personal Information Client Information System (PICIS)	PICIS contains the behavioral treatment information that connects to the Medicaid Management Information System (MMIS) for claims processing. PICIS also contributes to the MPI. The data records in the system are updated frequently as data is uploaded. The claims data are processed once per week. The most current processed data is one week behind.	<p>The data are entered into the system by batch (90 percent) and manually entered by behavior facility staff. The claims data is processed once per week. The current processed data is one week behind. There are no known data issues reported by system owners. However, the collected data set has grown and some data fields exist in newer records that may not exist in historical records. For example, the collected Adverse Childhood Experiences (ACE) scores are not present in all historical records.</p> <p>This system reports to have served seven percent (19,066) of children birth to five years old in 2018.</p>	<ul style="list-style-type: none"> • Child address • Child date of birth • Child race • Child ethnicity • Child gender • Child language • Program type (Foster Care) 	1, 3

Table C.2-4 - System-level inventory of Oklahoma Health Care Authority data systems

Basic Information for Tier 1 Data Systems				
Data System Name	Description of data in system	Data Availability	Data Elements Available for Priority Use Cases	Use Case
Medicaid Management Information System (MMIS)	<p>The MMIS contains member information, coverage information, relationship status, and medical home (Primary Care Physician). Data flows back and forth from partners like the Oklahoma Department of Human Services (OKDHS). For example, a member/potential member can go to OKDHS and apply/update eligibility information, but the data is housed on the MMIS tables.</p> <p>MMIS contributes to the MPI and is also connected to PICIS for claims processing.</p>	<p>Data are batch loaded/updated at different frequencies based on triggers by members, case workers, and other systems. There were no known data issues reported by system managers.</p> <p>This system reports to have 68 percent (178,424) of children under age five.</p>	<ul style="list-style-type: none"> • Family income • Child date of birth • Child race • Child ethnicity • Child gender • Child language 	1, 3

Table C.2-5 - System-level inventory of Oklahoma Department of Human Services data systems

Basic Information for Tier 1 Data Systems				
Data System Name	Description of Data in System	Data Availability	Data Elements Available for Priority Use Cases	Use Case
Child Care Monitoring, Administration and Safety System (CC-MASS)	CC-MASS contains child care provider/program demographics and provider quality monitoring data. The replicated data from CC-MASS is housed at the Office of Management Enterprise Services (OMES) and is accessed by other OKDHS systems, such as EPPIC and Kids System.	<p>The provider and program data are updated in real-time by the child care licensing staff update. The provider quality monitoring visits are synced by licensing staff daily and show in CC-MASS within two hours of successful sync. There were no known data issues reported by system managers.</p> <p>This system contains the total number of licensed child care providers.</p>	<ul style="list-style-type: none"> • Child care provider • Child care provider address • Child care provider quality • Child care provider capacity • Child care provider licensing • Child care provider type • Child care dates of service 	1
Mainframe PS2/IMS	The mainframe (PS2/IMS) is the backbone of the Adult and Family Services (AFS) eligibility system. All data elements used to determine eligibility/benefits are in the mainframe. The Mainframe PS2 is used in combination with other systems to manage the data collection, processing the data, and storing the data. Data from DB2 and OKDHS/SQL (OKDHSLive web app) feed into the PS2 and should be considered as a data source as some of the information in those systems do not get integrated into PS2.	<p>The data are continually updated through multiple channels. The data are archived after two years. There were no known data issues reported by system managers.</p> <p>This system reports to have 31 percent (82,000) of children under age five.</p>	<ul style="list-style-type: none"> • Family income • Child address • Child date of birth • Child disabilities • Child race • Child ethnicity • Child gender • Child language • Homelessness • Program enrollment [Foster Care, TANF, Supplemental Nutrition Assistance Program (SNAP), Free and Reduced-Price Lunch, SSI] 	1, 2, 3

Basic Information for Tier 1 Data Systems				
Data System Name	Description of Data in System	Data Availability	Data Elements Available for Priority Use Cases	Use Case
CCM	CCM is the case management software that contains the official case records for individuals served by OKDHS services, including data on service authorizations, contact notes, documents, medical information, demographics, financial information, employment history, etc.	Most of the data comes from the PS2 Mainframe.	<ul style="list-style-type: none"> • Child address • Child date of birth • Child disabilities • Program enrollment (SSI) 	1, 2, 3
Kids System	<i>[To date, we have not received response to our request for data on this system.]</i>			
Oklahoma Support Information System (OSIS)	This Child Support System manages the non-custodial and custodial child support payments and child support cases information	The data are entered by case workers in real time and automation from Oklahoma Employment Security Commission (OESC). There are known data issues reported by system managers. This system is limited to only those children who have a child support case. The children in this system are from Mainframe PS2.	<ul style="list-style-type: none"> • Family income • Family Address • Child support payments • Child Date of Birth • Child Race • Child Ethnicity 	1, 2, 3
EPPIC	The EPPIC Child Care system contains all the information needed for providers to be paid correctly for child care services. It is a Time and Attendance system for Child Care. The information is case specific and gives authorization information for each child receiving child care and also each provider providing care.	All child data comes from the PS2 Mainframe. This data represents child care payments/enrollment.	<ul style="list-style-type: none"> • Child address • Child date of birth • Program Type (Center/Home) • Provider address • Provider license contract number • Provider financial contract number • Provider Quality Rating • Dates of service • Amount paid 	1

Table C.2-6 - System-level inventory of Oklahoma State Department of Education data

Basic Information for Tier 1 Data Systems				
Data System Name	Description of Data in System	Data Availability	Data Elements Available for Priority Use Cases	Use Case
Student Information System (SIS)	The SIS is the Student Informational System that contains all student data from all public schools.	<p>The data is collected in real-time from all the school districts using a standard data transfer. Only data stored in each district's main system is part of the transfer. Some data may be inconsistent across districts if they store some data in district-specific systems. There are known data issues that have been identified and documented by the agency.</p> <p>This system reports to have 74 percent of Oklahoma's four-year olds enrolled in public pre-school programs.</p>	<ul style="list-style-type: none"> • Family income • Child address • Child data of birth • Child disabilities • Child race • Child ethnicity • Child gender • Child language • Homelessness • Program enrollment (Pre-K, SoonerStart, IDEA, Title I, FRL) • K-12 data (enrollments, grades, end-of-year assessments, etc.) 	1, 2, 3
SoonerStart ED Plan DB	The SoonerStart system is used to capture client information, frequency and duration of client visits, client encounters, and for billing purposes.	Staff update the data manually on a daily basis, although it usually lags from day to day. There are known data issues. This system is limited to three-year olds with a disability.	<ul style="list-style-type: none"> • Family income • Child address • Child date of birth • Child disabilities • Child race/ethnicity • Child gender • Child language 	1, 2, 3

Table C.2-7 - System-level inventory of Oklahoma Head Start data

Basic Information for Tier 1 Data Systems				
Data System Name	Description of Data in System	Data Availability	Data Elements Available for Priority Use Cases	Use Case
ChildPlus	ChildPlus is the preferred vendor and data system for the majority of Head Start programs in Oklahoma. It is a flexible system that is able to capture child-level and family-level data, as well as child outcome and professional development data. It sufficiently meets the requirements of federal Head Start reporting.	Availability varies across programs. ChildPlus is utilized for intake, so at a minimum, data are collected at time of Head Start application. Manual and automated record updates occur on a daily basis. Data issues exist, but profundity varies across programs. See Appendix [#] for details on alternative and supplemental systems.	<ul style="list-style-type: none"> • Family income • Child address • Child date of birth • Child disabilities • Child race • Child ethnicity • Child gender • Child language • Foster Care • Homelessness • Provider Address • Provider Slots • Provider Capacity • Provider License • Provider Type 	1, 2, 3

C.3 Program Eligibility Requirements

Note: No eligibility requirements exist for enrollment in unsubsidized licensed child care.

Table C.3-1 - Use Case #1 program eligibility requirements for selected Early Childhood Education programs, Use Case #2 program eligibility requirements for SoonerStart programs

Program	Universal Pre-K	Subsidy Child Care	Head Start	SoonerStart
Child Age	✓	✓	✓	✓
Family Income		✓	✓	
Family Work/School Status		✓		
Temporary Assistance for Needy Families (TANF) Participation			✓	
Foster Care			✓	
Homelessness			✓	
Supplemental Security Income (SSI)			✓	
Disabilities			✓	✓
Developmental delay/physical or mental condition			✓	✓

Table C.3-2 - Use Case #3 program eligibility requirements for social safety net programs

Program	WIC	TANF	Subsidy Child Care	SoonerCare	Free/ Reduced Lunch Meals	FDPIR	LIHEAP
Child Age	✓	✓	✓	✓	✓		
Family Income	✓	✓	✓	✓	✓	✓	✓
Family Work/School Status			✓				
Temporary Assistance for Needy Families (TANF) Participation	✓				✓		
Supplemental Nutrition Assistance Program (SNAP) Participation	✓				✓		
Foster Care					✓		
Disabilities				✓		✓	
SoonerCare (Medicaid)	✓						
FDPIR					✓		
Head Start					✓		
Live on a Reservation						✓	

C.3.3: PROGRAM ELIGIBILITY REQUIREMENTS BY PROGRAM

Table C.3.3 - Program eligibility requirements by program

Program	Eligibility Requirement	Description
Subsidy Child Care	Age	Children under the age of 13 years of age, or under 19 if incapable of self-care or under court supervision who needs assistance paying for child care.
	Income Level	The federal income eligibility threshold cannot exceed 85 percent of the state median income per family size. The income eligibility limits vary based on family size, and household income.
	Guardian: Working /Enrolled in School	Parents or guardians are at work, in training, or receiving an education.
Head Start Program	Age	Children 3 to 5 years of age.
	Income Level	<p>Children from families whose income is below the national poverty level defined by the U.S. Department of Health and Human Services.</p> <p>Head Start programs may enroll up to 10 percent of children from families that have incomes above the Poverty Guidelines.</p> <p>Programs may also serve up to an additional 35 percent of children from families whose incomes are above the Poverty Guidelines, but below 130 percent of the poverty line if the program can ensure that certain conditions have been met.</p>

Program	Eligibility Requirement	Description
	Eligibility Regardless of Income	Children in foster care, homeless children, and children from families receiving public assistance (Temporary Assistance for Needy Families or Supplemental Security Income) are also eligible for Head Start and Early Head Start services regardless of income.
	Disability	Federal law says that each Head Start program must reserve at least 10 percent of their enrollment for children with disabilities.(Regardless of income) In order for your child to get special education services from a Head Start program, she must be eligible based on a federal law called the “Individuals with Disabilities Education Act” (IDEA).
SoonerStart	Age	Part C: Birth to 36 months, Part B Children 3 to 5 years of age.
	Development delay/physical or mental condition	<p>Exhibit a delay in their developmental age compared to their chronological age of fifty percent or score two standard deviations below the mean in one or more of the following areas or in a sub-domain of one of the areas: cognitive, physical, communication, social or emotional, or adaptive development;</p> <p>Exhibit a delay in their developmental age compared to their chronological age of twenty-five percent or score one and one-half standard deviations below the mean in two or more of the following areas or in a sub-domain of two or more of the following areas: cognitive, physical, communication, social or emotional, or adaptive development; or</p> <p>Have a diagnosed physical or mental condition that has been identified as having a high probability for a developmental delay.</p>

Program	Eligibility Requirement	Description
WIC	Age	Children up to age 5 are eligible.
	Income Level	Must meet income guidelines or be receiving SNAP or TANF
SNAP	Age	Children under 18. Adults aged 18 to 50 who do not have children and are not pregnant can only get SNAP benefits for 3 months in a 3-year period unless they are working or participating in a work or workfare program.
	Income Level	To be eligible, the maximum gross monthly income is 130% of the federal poverty level. Be receiving disability-related assistance or benefits.
TANF	Age	Biological, adoptive or other relative child(ren) under the age of 18 living in the home.
	Income Level	Low income
Subsidy Child Care	Age	Children under the age of 13 years of age, or under 19 if incapable of self-care or under court supervision who needs assistance paying for child care.
	Income Level	The federal income eligibility threshold cannot exceed 85 percent of the state median income per family size. The income eligibility limits vary based on family size, and household income.
	Guardian: Working /Enrolled in School	Parents or guardians are at work, in training, or receiving an education.

Program	Eligibility Requirement	Description
SoonerCare (Medicaid)	Age	Individuals 65 and older, Children under 19 and pregnant women, and certain adults 19 or older who have minor dependent.
	Income Level	Must meet income guidelines or Individuals who are blind or who have disabilities
Free/Reduced Lunch Meals	Age	Children who attends school (high school or under)
	Income Level	Income is within the income guidelines
FDPIR The Food Distribution Program on Indian Reservations	Income Level	Households are certified based on income standards set by the Federal government
	Live on a Reservation	American Indian and non-Indian households that reside on a reservation and households living in approved areas near a reservation or in Oklahoma that contain at least one person who is a member of a Federally-recognized tribe
LIHEAP The Low-Income Home Energy Assistance Program	Income Level	The income limits are set at 125% of the federal poverty level (FPL)

C.4 Availability of Data Needed to Fulfill Use Cases #1-3

Tables C.4-1 through C.4-28 below map each of the existing data systems to the data elements needed for Use Cases #1, 2, and 3, at a high level. This process takes place in several steps:

1. Identification of available data to establish the individual components of the 0-5 population and indicators of program eligibility
 - a. Age
 - b. Location of residence
 - c. Demographics (race, ethnicity, gender)
 - d. Language
 - e. Family Income
 - f. Family work/school enrollment status
 - g. TANF
 - h. Foster Care
 - i. Homelessness
 - j. Supplemental Security Income (SSI)
 - k. Disabilities
 - l. Live on Reservation
 - m. FDPIR
 - n. SNAP
 - o. SoonerCare
2. Identification of available data to determine which children are served by the services identified for the following use cases:

Use Case #1

- a. Children served by Universal Pre-K, SoonerStart, Head Start, Subsidy Child Care, and Licensed Child Care
- b. Provider quality and characteristics

Use Case #2

- a. Children referred to SoonerStart and receive services, denied services, opted out of the services, or graduated from the services.

Use Case #3

- a. Children served by WIC, TANF, SSI, SNAP, LIHEAP, Free lunch, FDPIR, Child Support, and Subsidy child care

Synthesis of the overall completeness of data to feasibly fulfill the use case for each of the selected programs, based on the combined availability of data to describe each component of program eligibility and service

- b. Overall ability to model children eligible for each program
- c. Overall ability to model children served by each program

For each agency and associated system (table rows), the table indicates the extent to which those data are sufficient to meet the overall need of the use case. Any gaps in needed data are highlighted, though a holistic view across the many existing systems is necessary to determine the overall presence of gaps and implications for feasibility of the use case.

In Tables C.4-1 through C.4-28 below, data coverage is conveyed visually by *Harvey balls*,⁵⁶ which show the completeness of available data needed to represent the population for each data component. Note that this assessment does not address administrative feasibility, which is discussed as a consideration in Section 3.

In many cases, the Harvey ball indicates the extent to which available data represent the entire population of children under 5, but in other cases it is appropriate to represent a subset of this population, for example, *Children under 5 served by Subsidy*. Note that not all subsets are represented—for example, having a data point on *Children unserved by Subsidy* is unnecessary—it is sufficient to define only those served by the program for this component of the use case (remember all children, including those unserved, are accounted for under the population and eligibility components of our data mapping).

To assess data availability, 3Si evaluated the estimated population of children under 5 in each of the systems and determined the percentage coverage relative to the complete needed population. For example, the MMIS system owner reports to have 68 percent (178,424) of children under 5. This is a fairly large proportion and consequently is depicted by the $\frac{3}{4}$ Harvey ball. In contrast, Head Start is a relatively small program and reports to have 5 percent (14,313) of children under 5. This is a low proportion and therefore received the \bigcirc Harvey ball.






Summarizing data availability at the component level (e.g., income, homelessness, etc.) requires some extrapolation, since multiple systems often report similar data. It is known that there will be some overlap in the children represented in multiple systems and the extent of overlap is unknown. For instance, if one data source represents 25 percent of the population and another represents 50 percent, is the overall coverage 25 percent, 50 percent, or 75 percent? 3Si took this into consideration in the summary section of each table. While the specific overlap across these systems is unknowable without extensive analysis, the highest estimated coverage in the table will represent a minimum level of overall coverage listed in the summary for each table.

Summarizing data availability to model eligibility for an entire program is even more subjective, as the importance of different components of eligibility data vary depending on the specifics of the program. For instance, while comprehensive data on homelessness seems generally unavailable, this population only represents a small fraction of expected children qualifying for Head Start (and furthermore, many homeless children will be eligible based on other criteria, like family income, so will still be accounted for

⁵⁶ Harvey balls convey a directional sense of data availability, rather than implying greater (and false) precision with a more granular metric, like a percentage, and was deemed most appropriate for the cursory nature of this exercise.

in the data model). For this reason, the overall summary of available data to model the Head Start eligible population will not necessarily be limited by the low availability of a specific component. Each case is subjective, and 3Si welcomes input on this imperfect process.

Legend: Assessment of estimated data coverage

Completeness of needed population represented in system				
				
Little or no data on the needed population	Only some data on the needed population	Significant data on the needed population, but with some gaps	Comprehensive data on the needed population	Data availability unknown without further information or analysis

DATA AVAILABILITY BY COMPONENTS OF POPULATION AND PROGRAM ELIGIBILITY

Table C.4-1 - Systems with age data and detail on population coverage (Population: Children 0-5)






Age			
Agency	System	Estimated Data Coverage	Notes
OSDH	Enterprise Master Person Index (MPI)		The MPI is connected to all OSDH, ODMHSAS, and OHCA systems. They estimate this includes 65-80 percent of children under 5 (including potential duplicates). This number is predicted to increase substantially when the remaining Vital Records data is added to the MPI process.
OKDHS	Mainframe PS2		This system reports to have 31 percent (82,000) of children under 5. This data element is captured for all children present in this system.
OSDE	SIS		This system reports to have 74 percent of Oklahoma's four-year-olds enrolled in public Pre-K Programs.
Head Start	ChildPlus, alternate & supplemental program-level systems		Head Start is a relatively small program and reports to serve 5 percent (14,313) of children under five.
Summary			It is likely that the combined systems represent all children present in the associated agencies but do not represent the full population of children under five.

Table C.4-2: Systems with location data and detail on population coverage (Population: Children 0-5)






Location			
Agency	System	Estimate Data Coverage	Notes
OSDH	Enterprise Master Person Index (MPI)		The MPI is connected to all OSDH, ODMHSAS, and OHCA systems. They estimate this includes 65-80 percent of children under 5 (including potential duplicates). This number is predicted to increase substantially when the remaining Vital Records data is added to the MPI process.
OKDHS	Mainframe PS2		This system reports to have 31 percent (82,000) of children under 5. This data element is captured for all families and children present in this system.
OSDE	SIS		This system reports to have 74 percent of Oklahoma's 4-year olds enrolled in public Pre-Programs.
Head Start	ChildPlus, <i>alternate & supplemental program-level systems</i>		Head Start is a relatively small program and reports to serve 5 percent (14,313) of children under 5.
Summary			It is likely that the combined systems represent all children present in the associated agencies but do not represent the full population of children under 5.

Table C.4-3: Systems with race, ethnicity, and gender data and detail on population coverage (Population: Children 0-5)






Race, Ethnicity, and Gender			
Agency	System	Estimated Data Coverage	Notes
OSDH	Enterprise Master Person Index (MPI)		The MPI is connected to all OSDH, ODMHSAS, and OHCA systems. They estimate this includes 65-80 percent of children under 5 (including potential duplicates). This number is predicted to increase substantially when the remaining Vital Records data is added to the MPI process.
OKDHS	Mainframe PS2		This system reports to have 31 percent (82,000) of children under 5. This data element is captured for all children in this system.
OSDE	SIS		This system reports to have 74 percent of Oklahoma's 4-year-olds enrolled in public Pre-K Programs.
Head start	ChildPlus, <i>alternate & supplemental program-level systems</i>		Head Start is a relatively small program and reports to have 5 percent (14,313) of children under 5.
Summary			It is likely that the combined systems represent all children present in the associated agencies but do not represent the full population of children under 5.

Table C.4-4: Systems with language data and detail on population coverage (Population: Children 0-5)






Language			
Agency	System	Estimated Data Coverage	Notes
OSDH	Enterprise Master Person Index (MPI)		The MPI is connected to all OSDH, ODMHSAS, and OHCA systems. They estimate this includes 65-80 percent of children under 5 (including potential duplicates). This number is predicted to increase substantially when the remaining Vital Records data is added to the MPI process.
OKDHS	Mainframe PS2		This system reports to have 31 percent (82,000) of children under 5. This data element is captured for all families and children in this system.
OSDE	SIS		This system reports to have 74 percent of Oklahoma's 4-year olds enrolled in public Pre-Programs
Head Start	ChildPlus, <i>alternate & supplemental program-level systems</i>		Head Start is a relatively small program and reports to have 5 percent (14,313) of children under 5.
Summary			It is likely that the combined systems represent all children present in the associated agencies but do not represent the full population of children under 5.

Table C.4-5: Systems with income data and detail on population coverage (Population: Children 0-5)








Income			
Agency	System	Estimated Data Coverage	Notes
OSDH	Case Management Client Information Systems_ETO		This system reports to have less than 1 percent (2,000) of children under 5.
	Public Health Oklahoma Client Information System		Currently 38 percent (102,000)of children are represented in this system.
	SoonerStart		This system is limited to only 3-year-olds with a disability.
OHCA	Medicaid Management Information System		This system reports to have 68 percent (178,424) of children under 5.
OKDHS	Mainframe PS2		This system reports to have 31 percent (82,000) of children under 5.
Head Start	ChildPlus, <i>alternate & supplemental program-level systems</i>		Head Start is a relatively small program and reports to have 5 percent (14,313) of children under 5.
Summary			Some of the children will overlap across systems and not all children under 5 are represented in this system. Of the children in the systems, income might not be included in the family record. We can conclude that the systems do not include income data for all children under 5. It will be necessary to look at other sources for income data.

Table C.4-6: Systems with family work status/school enrollment data and detail on population coverage (Population: Children 0-5)





Working/Enrolled in School			
Agency	System	Estimated Data Coverage	Notes
OKDHS	Mainframe PS2		This system reports to have 31 percent (82,000) of children under 5. We can estimate that the percentage is even lower for this data element because not all of the working/enrolled in school population are receiving services within this system.
OSDH	Case Management Client Information Systems_ETO		This system reports to have less than 1 percent (2,000) of children under 5.
Head Start	ChildPlus, <i>alternate & supplemental program-level systems</i>		Head Start is a relatively small program and reports to have 5 percent (14,313) of children under 5.
Summary			There is a lack of data for working/school enrollment for parents of children under 5. This is a needed eligibility component for subsidy child care. It will be important to find other data sources to establish the gap of eligible children not being served.

Table C.4-7: Systems with TANF data and detail on population coverage (Population: Children with TANF)





TANF			
Agency	System	Estimated Data Coverage	Notes
OSDH	Public Health Oklahoma Client Information System (PHOCIS)		Currently 38 percent (102,000)of children are represented in this system.
OSDH	Case Management Client Information Systems_ETO		This system reports to have less than 1 percent (2,000) of children under 5.
OKDHS	Mainframe PS2		All families and children receiving TANF are recorded in this system (100 percent coverage)
Summary			It is possible to derive 100 percent from Mainframe PS2 once the use of PHOCIS is verified.

Table C.4-8: Systems with foster care data and detail on population coverage (Population: Children in foster care)





Foster Care			
Agency	System	Estimated Data Coverage	Notes
OKDHS	Mainframe PS2		All families and children that are receiving foster care are recorded in this (100 percent coverage).
ODMHSAS	PICIS		This system reports to have 7 percent (19,066) of children under 5 in 2018.
Head Start	ChildPlus, alternate & supplemental program-level systems		Head Start is a relatively small program and reports to have 5 percent (14,313) of children under 5.
Summary			We know we can get 100 percent from Mainframe PS2 but we need to verify the use of PICIS.

Table C.4-9: Systems with homelessness data and detail on population coverage (Population: All children)





Homelessness			
Agency	System	Estimated Data Coverage	Notes
OKDHS	Mainframe PS2		Currently 31 percent of children are represented in this system. We estimate the percentage will further decrease as the system is expanding categories of "homelessness". The new values will be captured for all SNAP households by October 2020. Although homelessness is an important element to track, this is a low eligibility indicator and will not have much impact on determining program eligibility.
OSDH	Case Management Client Information Systems_ETO		This system reports to have less than 1 percent (2,000) of children under 5.
Head Start	ChildPlus, <i>alternate & supplemental program-level systems</i>		Head Start is a relatively small program and reports to have 5 percent (14,313) of children under 5.
Summary			There is uncertainty on how to capture the number of homeless children under 5.

Table C.4-10: Systems with SSI data and detail on population coverage (Population: Children in SSI)





SSI			
Agency	System	Estimated Data Coverage	Notes
OSDH	Public Health Oklahoma Client Information System		Currently 38 percent (102,000) of children are represented in this system
OKDHS	Mainframe PS2		This system reports to have 31 percent (82,000) of children under 5. This data element is captured for all children present in this system.
Head Start	ChildPlus, <i>alternate & supplemental program-level systems</i>		Head Start is a relatively small program and reports to have 5 percent (14,313) of children under 5
Summary			It is likely that the combined systems represent all children present in the associated agencies but do not represent the full population of children under 5.

Table C.4-11: Systems with data of children with disabilities and detail on population coverage (Population: Children with disabilities)










Disabilities			
Agency	System	Estimated Data Coverage	Notes
OSDH	Oklahoma Birth Defect Registry		This system is limited to children with disabilities at birth born in Oklahoma.
OSDH	Oklahoma Childhood Lead Poisoning Prevention Program		This system is very limited to only children who have received a blood lead test.
OSDH	Public Health Oklahoma Client Information System		Currently 38 percent (102,000)of children are represented in this system.
OSDH	Case Management Client Information Systems_ETO		This system reports to have less than 1 percent (2,000) of children under 5.
OHCA	Medicaid Management Information System		This system reports to have 68 percent (178,424) of children under 5.
OKDHS	Mainframe PS2		This system reports to have 31 percent (82,000) of children under 5. There would need to be further assessment to determine whether all children with disabilities have a record in this system.
OSDE	SIS		This system reports to have 31 percent (82,000) of children under 5. There would need to be further assessment to determine whether all children with disabilities have a record in this system.
Head Start	ChildPlus, alternate & supplemental program-level systems		Head Start is a relatively small program and reports to have 5 percent (14,313) of children under 5. Head Start requires that at least 10 percent of their enrolled children have a disability.
Summary			It is unclear whether all children with disabilities are registered in a least one of the above systems. There may be children under 5 with disabilities that are on private insurance and not reported to the state. This is an important eligibility requirement for Head Start and further assessment might be useful.

Table C.4-12: Systems with data on children served by select programs






Children Served			
Agency	System	Estimated Data Coverage	Notes
OKDHS	Mainframe PS2		This system has all children receiving Subsidy Child Care.
OSDE	SIS		This system has all children enrolled in Pre-K in the public-school system.
Head Start	ChildPlus, alternate & supplemental program-level systems		The providers have all children being served in the Head Start program.
Licensed Child Care	OCCRRA		OCCRRA only has data available on provider licensed capacity, which contains no child-level records of which children are served by licensed child care.
Summary			Without the data for children enrolled in licensed child care, we do not have information for all 0-5 children served.

Table C.4-13: Systems with data on providers of select programs serving children







Providers and Provider Characteristics			
Agency	System	Estimated Data Coverage	Notes:
OKDHS	Mainframe PS2		Subsidy programs have data on the quality rating and other characteristics of subsidy providers.
OSDE	SIS		All providers are assessed. Assessments of school-based Pre-Ks may be different from those of child care and Head Start. This system has other school demographics for Pre-K providers.
Head Start	ChildPlus, alternate & supplemental program-level systems		Head Start programs have quality ratings and other demographics for all Head Start providers.
OKDHS	Child Care Monitoring, Administration and Safety System (CC-MASS)		This system has the total number of licensed child care programs.
OCCRRA	Work life system (WLS)		This online system hosts child care providers and has quality ratings and other provider demographics including preferred capacity.
Summary			There are likely some child care providers that are not in one of the above systems and that do not have a quality rating.

Table C.4-14: Systems with data on providers of select programs serving children




SoonerCare Medicaid			
Agency	System	Estimated Data Coverage	Notes
OSDH	MMIS		This system reports to have less than ?? percent (178,424) of children under 5.
OKDHS	Mainframe PS2		This system reports to have 31 percent (82,000) of children under 5. This data element is captured for all children present in this system.
Summary			It is likely that the combined systems represent all children present in the associated agencies but do not represent the full population of children under 5.

Table C.2-15: Systems with data on providers of select programs serving children








FDPIR			
Agency	System	Estimated Data Coverage	Notes
Tribal	Unknown	?	We were not able to identify the systems used by the tribe and this should be inventoried at a later time.
OKDHS	Mainframe PS2		This system reports to have 31 percent (82,000) of children under 5. This data element is captured for all children present in this system.
Summary			It is likely that the combined systems represent all children present in the associated agencies but do not represent the full population of children under 5.

Table C.4-16: Systems with data on providers of select programs serving children

Living on a Reservation			
Agency	System	Estimated Data Coverage	Notes
Tribal	Unknown	?	We were not able to identify the systems used by the tribe and this should be inventoried at a later time.
OSDH	Enterprise Master Person Index (MPI)		The MPI is connected to all OSDH, ODMHSAS, and OHCA systems. They estimate this includes 65-80 percent of children under 5 (including potential duplicates). This number is predicted to increase substantially when the remaining Vital Records data is added to the MPI process.
OKDHS	Mainframe PS2		This system reports to have 31 percent (82,000) of children under 5. This data element is captured for all families and children present in this system.
OSDE	SIS		This system reports to have 74 percent of Oklahoma's 4-year olds enrolled in public Pre-Programs.
Head Start	ChildPlus, alternate & supplemental program-level systems		Head Start maintains tribe-exclusive programs funded via a separate mechanism from non-tribal programs
Summary			It is likely that the combined systems represent all children present in the associated agencies but do not represent the full population of children under 5.

OVERALL DATA AVAILABILITY TO MODEL ELIGIBLE AND SERVED CHILDREN, BY PROGRAM

Table C.4-17: Overall availability of data to model eligibility for and service by Universal Pre-K




Universal Pre-K			
Eligibility	Category	Priority for Eligibility Component	Estimated Data Coverage
	Age	High	
	<i>Eligibility Summary: As the coverage of the MPI increases with new loaded records and with the integration of new source systems, the 4-year old population should be well defined. Some analysis may be limited by lack of complete demographic data.</i>		
Served	<i>Served Summary: Records of all children served are in the SIS systems.</i>		

Table C.4-18: Overall availability of data to model eligibility for and service by SoonerStart






SoonerStart			
	Category	Priority for Eligibility Component	Estimated Data Coverage
Eligibility	Age	High	
	Disability	High	
	<i>Eligibility Summary: Data availability is generally high for these eligibility requirements. However, further assessment will be needed to assess the ability of existing systems to capture the complete population of children with qualifying disabilities. Other data sources may be required to establish the overall population in need and any associated gaps in services.</i>		
	<i>Served Summary: Records of all children served are in the OKDHS systems.</i>		
Served	<i>Served Summary: Records of all children served are in the OKDHS systems.</i>		

Table C.4-19: Overall availability of data to model eligibility for and service by Head Start










Head Start			
	Category	Priority for Eligibility Component	Estimated Data Coverage
Eligibility	Age	High	
	Income	High	
	Disability	Medium	
	Foster	Low	
	TANF	Low	
	Homeless	Low	
	SSI	Low	
	<i>Eligibility Summary: Many of the needed data components are available, but the gap in income data and representation of all children in the MPI will require other data sources to establish the full eligible population (and any associated gaps in service). While available data on children experiencing homelessness is likely not comprehensive, this is expected to be a relatively small population (low 'weight') that will likely be captured by other data sources (e.g., income, SSI, etc.)</i>		
Served	<i>Served Summary: Records of all children served are in the Head Start program systems. Integration of the 33 separate systems may pose administrative challenges.</i>		

Table C.4-20: Overall availability of data to model eligibility for and service by Subsidy Child Care






Subsidy			
	Category	Priority for Eligibility Component	Estimated Data Coverage
Eligibility	Age	High	
	Income	High	
	Working/Enrolled in School	High	
	<i>Eligibility Summary: Gaps in available data for both family income and work/school enrollment status will likely require other data sources in order to establish the full eligible population (and any associated gaps in service).</i>		
Served	<i>Served Summary: Records of all children served are in the OKDHS systems.</i>		

Table C.4-21: Overall availability of data to model eligibility for and service by Licensed Child Care




Licensed Child Care			
	Category	Priority for Eligibility Component	Estimated Data Coverage
Eligibility	Age	High	
	<i>Eligibility Summary: As the coverage of the MPI increases with newly loaded records and with the integration of new source systems, the full population by age should become well defined. Some analysis may be limited by lack of complete demographic data.</i>		
Served	<i>Served Summary: Child-level data is unavailable. The availability of licensed capacity data will impose significant implications for the feasibility of including this program in the use case.</i>		

Table C.4-22: Overall availability of data to model eligibility for and service by WIC








WIC			
	Category	Priority for Eligibility Component	Estimated Data Coverage
Eligibility	Age	High	
	Income Level	High	
	TANF	High	
	SNAP	High	
	SoonerCare	Medium	
	<i>Eligibility Summary: Data availability is generally high for these eligibility requirements. However, Gaps in available data for family income will likely require other data sources in order to establish the full eligible population.</i>		
Served	<i>Served Summary: Records of all children served are in the OKDH systems.</i>		

Table C.4-23: Overall availability of data to model eligibility for and service by SNAP






SNAP			
	Category	Priority for Eligibility Component	Estimated Data Coverage
Eligibility	Age	High	
	Income Level	High	
	<i>Eligibility Summary: Data availability is generally high for these eligibility requirements. However, Gaps in available data for family income will likely require other data sources in order to establish the full eligible population.</i>		
	<i>Served Summary: Records of all children served are in the OKDHS systems.</i>		
Served	<i>Served Summary: Records of all children served are in the OKDHS systems.</i>		

Table C.4-24: Overall availability of data to model eligibility for and service by TANF





TANF			
	Category	Priority for Eligibility Component	Estimated Data Coverage
Eligibility	Age	High	
	Income Level	High	
	<i>Eligibility Summary: Data availability is generally high for these eligibility requirements. However, Gaps in available data for family income will likely require other data sources in order to establish the full eligible population.</i>		
	Served	<i>Served Summary: Records of all children served are in the OKDHS systems.</i>	

Table C.4-25: Overall availability of data to model eligibility for and service by SoonerCare






SoonerCare			
	Category	Priority for Eligibility Component	Estimated Data Coverage
Eligibility	Age	High	
	Income Level	High	
	Disability	High	
	<i>Eligibility Summary: Data availability is generally high for these eligibility requirements. However, Gaps in available data for family income will likely require other data sources in order to establish the full eligible population.</i>		
Served	<i>Served Summary: Records of all children served are in the OKDH systems.</i>		

Table C.4-26: Overall availability of data to model eligibility for and service by Free and Reduced Lunch (FRL)










FRL			
	Category	Priority for Eligibility Component	Estimated Data Coverage
Eligibility	Age	High	
	Income Level	High	
	TANF	Medium	
	SNAP	Medium	
	Foster Care	Medium	
	FDPIR	Medium	
	Head Start	Medium	
	Eligibility Summary: Data availability is generally high for these eligibility requirements. However, Gaps in available data for family income will likely require other data sources in order to establish the full eligible population.		
	Served Summary: Records of all children served are in the OKDE systems.		

Table C.4-27: Overall availability of data to model eligibility for and service by FDPIR








FDPIR			
	Category	Priority for Eligibility Component	Estimated Data Coverage
Eligibility	Age	High	
	Income	High	
	Live on Reservation	High	
	<i>Eligibility Summary: Data availability is generally high for these eligibility requirements. However, Gaps in available data for family income will likely require other data sources in order to establish the full eligible population. It is also unknown if all reservation addresses are stored in the other systems.</i>		
Served	<i>Served Summary: Records for some of the children served are in the OKDHS systems. Data inventory of tribal systems is required.</i>		?

Table C.4-28: Overall availability of data to model eligibility for and service by LIHEAP

LIHEAP			
	Category	Priority for Eligibility Component	Estimated Data Coverage
Eligibility	Income	High	
	<i>Eligibility Summary: Data availability is generally high for these eligibility requirements. However, Gaps in available data for family income will likely require other data sources in order to establish the full eligible population.</i>		
	<i>Served Summary: Records of all children served are in the OKDHS systems.</i>		
Served			

C.5: Availability of Data Needed to Map to OKDH MPI

Table C.5-1: Availability of data needed to map to OKDH MPI

MPI Identifier	Presence in agency systems					
	OSDH	ODMHSAS	OHCA	OCCRRA	OKDHS	OSDE
Guardian	Generally	No	Yes	No	Generally	Yes
Mother Maiden Name	Generally	No	Yes	No	Sometimes	No
Mother Name	Yes	No	Yes	No	Generally	Yes
Mother DOB	Generally	No	Yes	No	Sometimes	No
Father Name	Generally	No	Yes	No	Generally	Yes
Phone	Generally	No	Yes	No	Sometimes	No
Address	Yes	Yes	Yes	No	Generally	Yes
Child Name	Yes	Yes	Yes	No	Generally	Yes
DOB	Yes	Yes	Yes	No	Generally	Yes
Plural Birth Flag	Yes	No	Yes	No	No	No
Gender	Yes	Yes	Yes	No	Sometimes	Yes
SSN	Sometimes	No	Yes	No	Sometimes	Yes

Table C.5-2: Detailed availability of data needed to map to OKDH MPI

Agency Name	OSDH				ODMHSAS	OHCA	OCCRRR	OKDHS					OSDE
Data System Name	Systems_ETO	OBRD	PHOCIS	Ed Plan DB	PICIS	MMIS	WLS	CC-MASS	Mainframe PS2	CCM	OSIS	EPICC	SIS
System present in MPI	✓	✓	✓	✓ (only for children receiving SoonerStart service)	✓	✓							
Guardian		✓	✓	✓		✓			✓		✓	✓	✓
Mother Maiden Name		✓	✓	✓		✓			✓				
Mother Name	✓	✓	✓	✓		✓			✓		✓	✓	✓
Mother DOB	✓	✓	✓	✓		✓					✓	✓	
Father Name		✓	✓	✓		✓			✓		✓	✓	✓
Phone	✓	✓	✓	✓		✓			✓		✓		

Agency Name	OSDH				ODMHSAS	OHCA	OCCRRRA	OKDHS					OSDE
Data System Name	Systems_ETO	OBRD	PHOCIS	Ed Plan DB	PICIS	MMIS	WLS	CC-MASS	Mainframe PS2	CCM	OSIS	EPICC	SIS
Address	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓
Child Name	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓
DOB	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓
Plural Birth Flag	✓	✓	✓	✓		✓							
Gender	✓	✓	✓	✓	✓	✓			✓		✓		✓
SSN			✓			✓			✓		✓		✓
System present in MPI	✓	✓	✓	✓ (only for children receiving SoonerStart service)	✓	✓							
Guardian		✓	✓	✓		✓			✓		✓	✓	✓

Agency Name	OSDH				ODMHSAS	OHCA	OCCRRRA	OKDHS					OSDE
Data System Name	Systems_ETO	OBRD	PHOCIS	Ed Plan DB	PICIS	MMIS	WLS	CC-MASS	Mainframe PS2	CCM	OSIS	EPICC	SIS
Mother Maiden Name		✓	✓	✓		✓			✓				
Mother Name	✓	✓	✓	✓		✓			✓		✓	✓	✓
Mother DOB	✓	✓	✓	✓		✓					✓	✓	
Father Name		✓	✓	✓		✓			✓		✓	✓	✓
Phone	✓	✓	✓	✓		✓			✓		✓		
Address	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓
Child Name	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓
DOB	✓	✓	✓	✓	✓	✓			✓	✓	✓		✓

Agency Name	OSDH				ODMHSAS	OHCA	OCCRRA	OKDHS					OSDE
Data System Name	Systems_ETO	OBRD	PHOCIS	Ed Plan DB	PICIS	MMIS	WLS	CC-MASS	Mainframe PS2	CCM	OSIS	EPICC	SIS
Plural Birth Flag	✓	✓	✓	✓		✓							
Gender	✓	✓	✓	✓	✓	✓			✓		✓		✓
SSN			✓			✓			✓		✓		✓

C.6: Distribution of Head Start Programs and System Utilization

In Oklahoma, there is one primary data reporting tool and eight supplementary and/or alternative data reporting tools used in 34 Head Start (HS) and Early Head Start (EHS) grantees across 78 counties.

Below you will find tables that illustrate the geographic and technical spread of HS and EHS grantees in Oklahoma.

Table C.6-1: Unique counties and grantees by reporting tools

Primary and secondary data reporting tools	Total counties	Total grantees
CAPTAIN	4	2
<i>No secondary data reporting tool</i>	3	1
Microsoft Office (Word, Excel, Access)	1	1
ChildPlus/ChildPlus.net	74	31
<i>No secondary data reporting tool</i>	71	25
ADP	1	1
Teaching Strategies Gold	3	1
Procare	1	1
COPA	3	1
Salesforce	1	1
CORE Advantage	1	1
KidReports	1	1
Cornerstone	1	1
MAPS-Bluemark	1	1
COPA (Child Outcome, Planning, and Administration/Assessment)	7	2
ChildPlus	7	2
PROMIS (Program Resources and Outcomes Management Information System, Cleverex)	4	1
Teaching Strategies Gold	4	1
Grand Total	78	34

Table C.6-2: Grantees by county

County	Total Head Start grantees	Total Early Head Start grantees	Total grantees
Adair	2	2	2
Alfalfa	1	1	1
Atoka	2	2	2
Beaver	1	1	1
Beckham	3	1	3
Blaine	2	1	2
Bryan	2	2	2
Caddo	1	1	1
Canadian	2	1	2
Carter	2	1	2
Cherokee	2	2	2
Choctaw	2	2	2
Cimarron	1	1	1
Cleveland	2	3	3
Coal	2	2	2
Comanche	2	1	2
Cotton	1	1	1
Craig	1	1	2
Creek	2	1	2
Custer	2	1	2
Delaware	2	3	4
Dewey	2	1	2
Ellis	1	1	1
Garfield	1	1	1
Garvin	1	1	1
Grady	1	1	1
Grant	1	1	1
Greer	1	1	1
Harmon	1	1	1
Harper	1	1	1
Haskell	2	2	2
Hughes	2	1	2
Jackson	1	1	1
Jefferson	1	1	1
Johnston	2	1	2
Kay	3	1	3
Kingfisher	2	1	2
Kiowa	2	1	2

County	Total Head Start grantees	Total Early Head Start grantees	Total grantees
Latimer	2	2	2
Le Flore	1	1	1
Lincoln	3	2	4
Logan	2	3	3
Love	1	1	1
Major	1	1	1
Marshall	1	1	1
Mayes	2	2	2
McClain	1	1	1
McCurtain	2	2	2
McIntosh	2	1	2
Murray	2	1	2
Muskogee	2	2	2
Noble	3	1	3
Nowata	2	2	2
Okfuskee	2	1	2
Oklahoma	2	2	3
Okmulgee	2	1	2
Osage	2	1	2
Ottawa	1	1	2
Pawnee	1	1	1
Payne	2	3	3
Pittsburg	2	2	2
Pontotoc	2	1	2
Pottawatomie	2	1	2
Pushmataha	2	2	2
Roger Mills	2	1	2
Rogers	2	2	2
Seminole	1	2	2
Sequoyah	2	2	2
Stephens	1	1	1
Texas	1	1	1
Tillman	1	1	1
Tulsa	4	4	5
Wagoner	2	2	2
Washington	1	1	1
Washita	2	1	2
Woods	1	1	1
Woodward	1	1	1

Table C.6-3: Counties served by grantee and EHS/HS service

Grantee	Total counties with Early Head Start	Total counties with Head Start
Big Five Community Services, Inc.	23	23
Central Tribes of The Shawnee Area, Inc.	4	4
Cherokee Nation	9	9
Cheyenne & Arapaho Tribes	0	8
Chickasaw Nation	0	4
Choctaw Nation of Oklahoma	10	10
Community Action Agency of Oklahoma City and OK/CN Counties	1	1
Community Action Development Corporation	6	7
Community Action Project of Tulsa County, Inc.	1	1
Community Action Resource and Development, Inc.	6	6
Crossroads Youth & Family Services, Inc.	4	3
Delaware Tribe of Indians	1	0
Delta Community Action	3	3
Green Country Behavioral Health Services, Inc.	1	1
Inca Community Services, Inc.	4	4
Iowa Tribe of Oklahoma, Inc.	4	0
Jay Public Schools	0	3
Ki Bois Community Action Foundation, Inc.	8	8
Kickapoo Tribe of Oklahoma	0	3
Kiowa Tribe of Oklahoma	0	2
Little Dixie Community Action Agency, Inc.	3	3
Muscogee Creek Nation	0	6
Native American Coalition of Tulsa	1	1
Northeast Oklahoma Community Action Agency, Inc.	3	0
Osage Nation	0	1
Otoe-Missouria Tribe	0	2
Ponca Tribe of Oklahoma	0	2
The Seminole Nation of Oklahoma	1	1
Southwest Oklahoma Community Action Group, Inc.	4	4
Sunbeam Family Services, Inc.	1	0
Tulsa Educare	1	0
United Community Action Program, Inc.	8	8
Washita Valley Community Action Council, Inc.	2	2
Wewoka Public Schools	0	1

Appendix D: Example of Data Elements and Sizes for Cost Exercise

Table D-1: MPI data elements and sizes for a single MPI record and for 60 million records

MPI Field	Data Size (Bytes)	Data size for 60 million records (bytes)
Guardian	100	6,000,000,000
Mother Maiden Name	100	6,000,000,000
Mother Name	100	6,000,000,000
Mother DOB	100	6,000,000,000
Father Name	100	6,000,000,000
Phone	100	6,000,000,000
Address	100	6,000,000,000
Child Name	100	6,000,000,000
DOB	100	6,000,000,000
Plural Birth Flag	25	1,500,000,000
Gender	25	1,500,000,000
SSN	50	3,000,000,000
Total Bytes	1,000	60,000,000,000 (0.06 terabyte)

Note: These data elements were selected based on conversations with Oklahoma stakeholders who maintain the MPI and are intended for illustrative purposes only.

Appendix E: More Information about Cloud Service Providers

Cloud computing can be categorized into the following three buckets: Infrastructure as a Service (IaaS), Software as a Service (SaaS), and Platform as a Service (PaaS).

- Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP) are examples of **IaaS** that can replace on-premise infrastructure.
- Salesforce, Dropbox, and GoToMeeting are examples of **SaaS**. They use Cloud Service Providers (CSPs) to deliver browser-based applications that are managed by a third-party vendor. For SaaS applications, users are not responsible for any hardware, software, or scaling issues and customers pay for service on demand.
- **PaaS** is intended for developers by providing frameworks that are specific to use case needs. For example, Heroku provides simplified mobile application development that allows for one application to service both iPhone and Android based phones. There are many of these PaaS frameworks and a few related to the data integration use cases. Cloudera, Hortonworks, and MapR are examples of PaaS solutions that are frameworks for data engineering and analytics. Their services can be deployed on-premise or within CSPs and developers can use their products to streamline Hadoop⁵⁷ processing and analytics. 3Si considered Cloudera as part of this integration plan due to their ability to provide a one-stop-shop for data storage, processing, administration, and governance but has excluded them from further analysis due to the lack of Hadoop ecosystem within Oklahoma data infrastructure for which Cloudera is mostly specialized in.

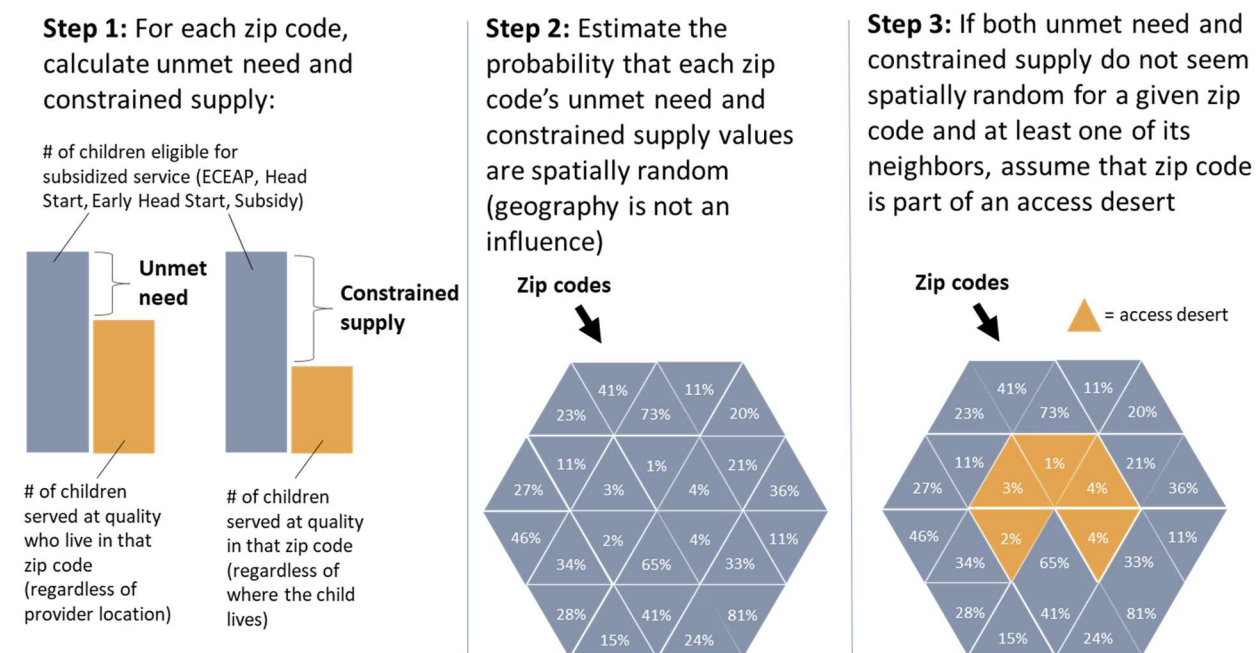
⁵⁷ Hadoop is an open-source software framework for storing data and running applications on clusters of hardware. It provides storage, processing power and the ability to handle virtually limitless concurrent tasks. Cloudera's open source platform distributes Hadoop and related projects.

Appendix F: Methodology for Determining Access Deserts

Access deserts are an application of a geostatistical method called Local Moran's I, a specific type of Local Indicator of Spatial Association (LISA). LISAs examine the influence (or lack of influence) of geography on a group of contiguous geographic areas of the same type. Access desert methodology occurs in several steps, summarized below in Figure F. While access deserts are the result of two distinct LISA analyses, each using a different measure, consider first the example of only unmet need⁵⁸.

Each geographic unit (say, for now, zip code) has an unmet need value and, collectively, these values comprise a distribution of values. For each zip code, LISA assesses the likelihood that its unmet need value is spatially random (that geography is not an influencing factor). This is done by comparing the primary zip code's unmet need value to that of its neighbors (defined in this analysis as the zip codes that are immediately contiguous to the primary zip code) and assigning a probability to the primary zip code's unmet need value that it would occur at random. If only 10 percent of unmet need values in the overall distribution are as high as that of the primary zip code, the primary zip code's value has a 10 percent chance of occurring. This analysis is done for each zip code—each zip code in turn is treated as primary and its unmet need value is assigned a probability score.

Figure F: An Overview of Access Desert Methodology



Assigning these probabilities based on only one map, however, would not create a robust analysis. So, for each zip code, the primary value is fixed and all other values are permuted randomly to create other

⁵⁸ Unmet need: the difference between the number of children in the zip code eligible for subsidized service and the number of children served at quality who live in that zip code (regardless of where that child is served).

hypothetical maps. Say for now that this is done 999 additional times, this would allow for comparison of the primary zip code's unmet need value to 1,000 different sets of neighbors which, in turn, allows for a much more robust estimation of the probability that each zip code's unmet need value occurs at random.

There are five possible outcomes for each zip code in a LISA analysis:

1. **Not significant:** The zip code's unmet need value falls within the range of probabilities that suggest it is likely to have occurred by chance.
2. **High-low:** The zip code's unmet need value is high enough to suggest it is unlikely to have occurred by chance but the same is not true of any of its neighbors.
3. **Low-high:** The zip code's unmet need value is low enough to suggest it is unlikely to have occurred by chance but the same is not true of any of its neighbors.
4. **Low-low:** The zip code's unmet need value is low enough to suggest it is unlikely to have occurred by chance and the same is true of at least one of its neighbors.
5. **High-high:** The zip code's unmet need value is high enough to suggest it is unlikely to have occurred by chance and the same is true of at least one of its neighbors.

In the last of these cases (high-high), the zip code is part of a spatial cluster. A cluster of high values is colloquially called a "hot spot." If a zip code is part of a hot spot, both for unmet need and constrained supply (after running a LISA analysis using constrained supply as the non-spatial measure), that zip code is considered to be part of an access desert.⁵⁹

3Si is grateful to Erin Hardy, a Fellow at the Institute for Child, Youth, and Family Policy (ICYFP) at the Heller School for Social Policy and Management at Brandeis University, for her advice on how to apply LISA to this type of analysis.

⁵⁹ Constrained supply: the delta between the number of children in the zip code eligible for subsidized service and the number of children served at quality in that zip code (regardless of where the child lives).

Appendix G: Draft Interagency Agreement

MULTI-AGENCY AGREEMENT FOR THE GOVERNANCE OF THE OKLAHOMA EARLY CHILDHOOD INTEGRATED DATA SYSTEM

THIS MULTI-AGENCY AGREEMENT (this “Agreement”) is entered into as of the Effective Date by and among the following entities:

- a. The following State of Oklahoma agencies (each, an “ECIDS Agency”, and collectively, the “ECIDS Agencies”):
 - i. The Oklahoma State Department of Education (“OSDE”);
 - ii. The Oklahoma State Department of Health (“OSDH”);
 - iii. The Oklahoma Health Care Authority (“OHCA”);
 - iv. The Department of Human Services (“DHS”);
 - v. The Office of Juvenile Affairs (“OJA”);
 - vi. The Oklahoma Department of Mental Health and Substance Abuse Services (“ODMHSAS”);
 - vii. The Department of Rehabilitation Services (“OKDRS”);
 - viii. The Oklahoma Department of Corrections (“DOC”); [?] [*Party to Multi-Agency DSA*]
 - ix. The Oklahoma Commission on Children and Youth (“OCCY”); [?] [*Party to Multi-Agency DSA*] and
 - x. The Office of Management and Enterprise Services (“OMES”); and [?]
- b. The Office of the Governor [?]

(each, a “Party”, and collectively, the “Parties”) under authority as provided by law. In consideration of the mutual covenants set forth below, the Parties agree as follows:

I. RECITALS.

- A. In recent years, the Parties have undertaken various efforts to integrate and utilize data from multiple ECIDS Agencies to analyze and improve early childhood service delivery within the State of Oklahoma.
- B. Recognizing the power of integrated data across the ECIDS Agencies and the opportunity to improve upon prior efforts, the Parties commissioned and adopted an Oklahoma Early Childhood Integrated Data System Framework dated [_____, 2019] (“Data Governance Framework”) which sets forth objectives, success factors, and a proposed framework for the development and administration of a State of Oklahoma early childhood integrated data system (“ECIDS”).
- C. Through this Agreement, the Parties seek to establish a governance structure to implement the Data Governance Framework and address the following overall

requirements, functions, and expectations for the ECIDS (collectively, the “ECIDS Expectations”):

- i. The primary focus of the ECIDS must be to make consolidated, longitudinal data and analysis available to improve the delivery of services and outcomes for Oklahoma children and families;
- ii. The ECIDS must be useful to end users, with an emphasis on program outputs and impact, while streamlining and reducing governance and time spent on data extractions, loading, and modeling as much as possible within the parameters of applicable law and this Agreement;
- iii. The governance of the ECIDS must remain oriented toward meeting the operational needs of the Office of the Governor and the ECIDS Agencies;
- iv. The governance of the ECIDS must ensure that data is absolutely secure in accordance with industry leading security standards, and that the operations of the ECIDS are in compliance with all applicable privacy laws;
- v. The ECIDS must provide centralized capacity and technical infrastructure designed to promote efficiency and the active use of data;
- vi. The ECIDS must support various usage scenarios, including Cabinet-driven priorities, ECIDS Agency-driven usage and priorities, uses by third-party intermediaries and key stakeholders such as universities and nonprofits, and the provisioning of public-facing data;
- vii. The ECIDS technical infrastructure must be designed to utilize industry leading technologies, and be administered and supported to continually improve the technical infrastructure over time;
- viii. The governance of the ECIDS must engage and be informed by external stakeholders who are collaborating with the State to improve early childhood services, including Head Start providers, Native American tribal nations, universities, nonprofits, and others; and
- ix. The governance of the ECIDS must strive to fundamentally improve the operations of the ECIDS Agencies by providing them with information that is comprehensive, useful, timely, and accessible.

D. Each of the Parties has the legal basis to enter into this Agreement under Oklahoma law.

E. The Parties wish to work cooperatively to effectively govern the ECIDS, carry out and comprehensively address the ECIDS Expectations, and otherwise fulfill the terms of this Agreement.

II. DEFINITIONS.

Whenever used in this Agreement, all capitalized terms have the meanings specified herein. For reference, a table of capitalized terms is included in Exhibit A.

III. **PURPOSES, SCOPE, AND TERM.**

A. **Purposes of Agreement.** This Parties are entering into this Agreement for the purposes of:

- i. Establishing a governance structure to develop and administer the ECIDS in a manner that addresses the ECIDS Expectations;
- ii. Authorizing the establishment and administration of a center that will provide centralized administrative capacity and technical infrastructure for the ECIDS in accordance with this Agreement and the oversight and direction of the Board (the “Center”);
- iii. Committing each of the ECIDS Agencies to contribute to the ECIDS data on early childhood programs, services, workforce, and providers (“ECIDS Data”) and participate in the governance of the ECIDS; and
- iv. Ensuring the use of the data within the ECIDS supports interagency policy development and analysis in a secure manner consistent with applicable federal and Oklahoma law.

B. **Scope and Compliance With Law.** This Agreement supplements, and does not supplant, any existing intergovernmental agreements between or among the Parties. This Agreement does not limit the Parties from entering into agreements separate from this Agreement involving ECIDS Data, provided any such agreement does not conflict with this Agreement. All ECIDS Data accessed, shared, used, or disclosed pursuant to this Agreement will only be accessed, shared, used, and disclosed in a manner permitted by applicable law including, without limitation, those laws and regulations specified in Exhibit B.

C. **Term.** The initial term of this Agreement will be ten (10) years commencing upon the full execution of this Agreement by all of the Parties (the “Effective Date”). Upon the expiration of the initial term, this Agreement will, without further action by the Parties, automatically renew for consecutive one-year periods unless and until three (3) or more ECIDS Agencies submit a notice of termination to all of the other Parties no less than one hundred eighty (180) days prior to the expiration of the then-current term. The “Term” includes the initial term and any renewal terms occurring pursuant to this Section.

IV. **ECIDS BOARD AND WORK GROUPS.**

A. **Responsibilities.** The ECIDS will be governed and overseen by a governing board established in accordance with this Agreement (the “Board”). The Board’s responsibilities include, without limitation:

- i. Defining the vision and mission of the ECIDS in a manner consistent with the ECIDS Expectations, and providing a clear articulation of its purpose and the intended benefits to the ECIDS Agencies and the general public;
- ii. Adopting standards and business rules for the operation of the ECIDS based on recommendations of the Work Groups and the Center;

- iii. Defining the categories of and standards for ECIDS Data use, which will include uses that are always permitted, uses that are never permitted, and uses that require specific approvals through procedures defined by the Board;
- iv. Adopting and maintaining a common agenda for ECIDS Data use, including priority use cases and a research agenda;
- v. Driving the implementation of Cabinet-identified priorities through effective utilization of the ECIDS;
- vi. Selecting the Center Administrator in accordance with Section V of this Agreement;
- vii. Overseeing the administration of the Center and ensuring compliance by the Center Administrator with this Agreement, the Center Administration Agreement, Board policies, and applicable law; and
- viii. Defining accountability for ECIDS Agencies for adherence to the ECIDS Expectations, the requirements of this Agreement, and decisions of the Board.

B. Membership. The Board will consist of the following members:

- i. A representative of the Office of the Governor, as appointed by the Governor, who will serve as the non-voting chairperson of the Governing Board (“Chair”), and who will not be counted for determining a quorum of the Governing Board; and
- ii. The chief executive officer, or Eligible Designee, of each of the ECIDS Agencies. For purposes of this Agreement, “Eligible Designee” means a senior-level staff person at an ECIDS Agency with executive decision-making authority for the ECIDS Agency, and who is fully authorized to take action on behalf of the ECIDS Agency at meetings of the Board.

C. Manner of Acting and Meetings. The affirmative vote of a majority of members of the Board, not including the Chair, is required for any action of the Board. Notwithstanding the foregoing the Board may, by resolution, establish an executive committee consisting of a subset of members of the Board, and delegate to the executive committee such decision-making authority as the Board determines appropriate. No more than two additional representatives of an ECIDS Agency, other than its member designated pursuant to Section IV.B, may participate in meetings of the Board, provided an ECIDS Agency will only have one vote on all matters before the Board. All meetings of the Board, and notices thereof, will comply with the Oklahoma Open Meeting Act (25 O.S. §§ 301 – 314). The Board will establish an annual schedule of meetings that provides for no less than two (2) meetings each year of the Term. Special meetings may be called at the discretion of the Chair.

D. Work Groups. The Board may, from time to time, establish, charge, and select the members for one or more standing or ad hoc work groups to advise and support the Board in the administration and performance of, and planning for, the ECIDS and the Center (each, a “Work Group”). The membership of Work Groups and the chairs or co-chairs thereof shall be determined by resolution of the Board, and may include both ECIDS Agency staff and external representatives. Work Groups will meet on an

as-needed basis at the discretion of each Work Group's chair or co-chairs. Work Groups will only advise the Board and will not have decision-making authority.

V. CENTER ESTABLISHMENT AND ADMINISTRATION.

- A. Establishment.** Subject to the availability of funding, the Board will establish the Center to develop and operate centralized administrative capacity and technical infrastructure for the purpose of addressing the ECIDS Expectations substantially in accordance with the scope description in Exhibit C.
- B. Center Administrator.** In accordance with applicable law and procedures determined by the Board, the Board will select an entity to establish and administer the Center (the "Center Administrator") in accordance with this Agreement, the Center Administration Agreement (as defined in Section V.C), and the directives of the Board. The Center Administrator shall be considered a duly authorized representative of the ECIDS Agencies to the fullest extent permitted under applicable federal and Oklahoma law for accessing and utilizing ECIDS Data and for the other purposes and functions contemplated by this Agreement.
- C. Center Administration Agreement.** The ECIDS Agencies will either enter into a joint agreement with the Center Administrator, or designate an ECIDS Agency to enter into an agreement with the Center Administrator on behalf of the Board (in either case, the "Center Administration Agreement"). The duties of the Center Administrator as specified within the Center Administration Agreement shall include, without limitation:
1. Developing and administering the Center in accordance with this Agreement and a Center development plan approved by the Board;
 2. Defining and overseeing the following Center policies and processes, subject to Board approval and oversight:
 - a. Security expectations;
 - b. Privacy rules;
 - c. Data collection policies and procedures;
 - d. Policies for the use of on-line infrastructure;
 - e. Policies for data maintenance retention;
 - f. Data quality standards and plans for continuous improvement of data quality;
 - g. Data stewardship processes and workflows;
 - h. Data request and access protocols, including an expedited process for exploratory projects for the purposes of determining the scope and viability of a potential project;
 - i. Processes and procedures for disclosure-proofing of data prior to the release of any data from the Center;
 - j. Policies for ECIDS Agency review and commenting on research and reports prior to publication;
 - k. Maintenance of records of disclosures of any data; and
 - l. Other policies and processes deemed necessary by the Board;

- 3.Preparing an annual budget and plan for the Board’s approval;
- 4.Hiring an Executive Director, subject to Board approval;
- 5.Hiring employees and entering into contracts as necessary to carry out the development plan and in accordance with the approved budget;
- 6.Participating in audit, compliance, and security review processes defined by the Board;
- 7.Ensuring the orderly transition and protection of data in the event the Center Administrator’s status is terminated; and
- 8.Other duties and responsibilities as may be approved from time to time by the Board.

D. Center Funding. *[TBD – either Centralized Funding approach, agency cost-sharing, or hybrid]*

E. Center Executive Director. The Center Administration Agreement shall provide for the designation of an individual to serve as the executive director of the Center (“Executive Director”). The designation of the Executive Director shall be subject to the prior approval of the Board through such procedures as the Board may establish. The Center Administration Agreement shall provide for an annual review of the Executive Director’s performance by the Board, and shall further provide that the Executive Director may be removed and a new Executive Director named whenever a majority of the Board deems such action to be in the best interests of the ECIDS, subject to reasonable provisions for the winding down of the Center’s services and payment through the date of termination.

F. Termination of Center Administrator. The Center Administration Agreement shall provide that the Center Administrator may be terminated and a new Center Administrator may be designated by the Board whenever it determines the best interests of the ECIDS would be served thereby.

VI. OBLIGATIONS AND RIGHTS OF ECIDS AGENCIES.

- A. Obligations.** Each ECIDS Agency agrees to:
- i. Align and coordinate its early childhood data initiatives with the ECIDS;
 - ii. Ensure the continuous appointment and participation of a Board representative authorized to act on behalf of the ECIDS Agency;
 - iii. Ensure appropriate staff participation in Work Groups;
 - iv. Submit data to the Center Administrator and fully participate in the Center’s services in accordance with an agreement form and procedures, schedule, and requirements adopted by the Board;
 - v. Perform timely reviews and take timely action on all decisions necessary for the development and administration of the Center in accordance with schedules and plans approved by the Board;
 - vi. Cooperate with the Center Administrator to improve the quality of ECIDS Data submitted;

- vii. Ensure the training of appropriate staff on data use, stewardship, and analysis within the Center;
- viii. Take such further actions as are necessary to ensure the Center Administrator can access and maintain ECIDS Data under applicable law; and
- ix. Seek funding in support of the ECIDS, provided that ECIDS Agency funding contributions are subject to appropriation.

B. Rights.

- i. Each ECIDS Agency shall be the sole determiner of the rights of access to personally identifiable ECIDS Data it has submitted to the Center by another ECIDS Agency or external party in accordance with the data stewardship processes and workflows managed by the Center and approved by the Board.
- ii. In accordance with procedures approved by the Board, an ECIDS Agency may require the destruction of its data within the Center if the ECIDS Agency's data is used in a manner that violates this Agreement or applicable law.

VII. ADMISSIONS AND WITHDRAWALS OF AGENCIES.

- A. Admissions.** The Board may approve the admission of a new ECIDS Agency upon the agency's entering into a "New Agency Admission Agreement" in the form of Exhibit D. Upon the effective date of the New Agency Admission Agreement, such agency shall be deemed an ECIDS Agency for all purposes of this Agreement and the Board. Following Board approval, the ECIDS Agencies hereby authorize the Office of the Governor to enter into a New Agency Admission Agreement with the new ECIDS Agency on behalf of all Parties.
- B. Withdrawals.** An ECIDS Agency may withdraw from this Agreement upon 180 days prior written notice with the written authorization of its chief executive and the Governor. Upon withdrawal of any ECIDS Agency, the Agreement will continue in full force and effect for all other ECIDS Agencies. Data submitted by a withdrawing ECIDS Agency relating to any project still pending within the Center as of the date of withdrawal may continue to be used for that project.

VIII. DISPUTE RESOLUTION.

The Governing Board is responsible for resolving any disputes that may arise with respect to this Agreement or a Party's required performance hereunder through a collaborative process. If the Governing Board is unable to resolve any dispute, the dispute shall be resolved by the Governor or Governor's designee, which resolution shall be binding on the ECIDS Agencies.

IX. GENERAL PROVISIONS.

- A. Amendment.** An amendment to the Agreement may only occur following authorization by the Governing Board and the written approval of each of the Parties.

- B. Notices.** All notices or other correspondence required to be given pursuant to this Agreement shall be sent by regular or electronic mail to each Party's member of the Governing Board.
- C. Entirety.** This Agreement, together with the Exhibits attached hereto, constitutes the entire agreement among the Parties with respect to the subject matter hereof, and supersedes any other negotiations, agreements, or communications, whether written or oral, that have been made by any Party.
- D. Governing Law.** This Agreement shall be governed by and construed in accordance with the laws of the State of Oklahoma.
- E. Severability.** In case any provision in this Agreement is held to be invalid, illegal, or unenforceable, the validity, legality, and enforceability of the remaining provisions shall not be affected.
- F. Authority to Execute.** Each Party represents and warrants to the other Parties that this Agreement has been duly authorized, executed, and delivered by and on behalf of each such Party, and constitutes the legal, valid, and binding agreement of said Party.
- G. Counterparts.** This Agreement may be executed in several counterparts, each of which shall be an original, and all of which shall constitute one and the same instrument. For purposes of this Agreement, a facsimile copy or a scanned printable document format (pdf) of a Party's signature shall be sufficient to bind such Party.
- H. Recitals and Exhibits Incorporated.** The recitals in Section I and the following Exhibits attached hereto are hereby incorporated into this Agreement by this reference and expressly made a part of this Agreement.

Exhibit A	Table of Defined Terms
Exhibit B	Controlling Laws and Regulations
Exhibit C	Center Scope Description
Exhibit D	New Agency Admission Agreement Form

[Signature Pages Follow]

**MULTI-AGENCY AGREEMENT FOR THE GOVERNANCE OF THE
OKLAHOMA EARLY CHILDHOOD INTEGRATED DATA SYSTEM**

SIGNATORY AUTHORITY

Approved and authorized on behalf of each of the Parties.

OKLAHOMA STATE DEPARTMENT OF EDUCATION:

By: _____

Title: _____

Date: _____

[Add signature block for all Parties]

EXHIBIT A

TABLE OF DEFINED TERMS

Term	Section Defined	Page
Agreement	Preamble	1
Board	IV.A	3
Center	III.A	3
Center Administration Agreement	V.C	5
Center Administrator	V.B	4
Chair	IV.B	4
DHS	Preamble	1
Data Governance Framework	I.B	1
DOC	Preamble	1
ECIDS	I.B	1
ECIDS Agencies	Preamble	1
ECIDS Agency	Preamble	1
ECIDS Data	III.A	3
ECIDS Expectations	I.C	1
Effective Date	III.C	3
Eligible Designee	IV.B	4

Executive Director	V.E	5
New Agency Admission Agreement	VII.A	7
OCCY	Preamble	1
ODMHSAS	Preamble	1
OHCA	Preamble	1
OJA	Preamble	1
OKDRS	Preamble	1
OMES	Preamble	1
OSDE	Preamble	1
OSDH	Preamble	1
Party; Parties	Preamble	1
Term	III.C	3
Work Group	IV.D	4

EXHIBIT B

CONTROLLING LAWS AND REGULATIONS

[Incorporate/Update Applicable Confidentiality Laws and Regulations by Party – from the Multi-Agency Data Sharing Agreement. Request Parties to do this.]

EXHIBIT C

CENTER SCOPE DESCRIPTION

The Center’s responsibilities will include the following:

Responsibility	Brief Description
Designated Administrator	The Center will serve as the authorized representative of the Board and the agencies for establishing and maintaining all system components and managing the system on a day-to-day basis. The Center’s development of all system components will be subject to Board oversight and approval, and the Center’s administration of all system components will remain subject to Board policy and oversight.
Project Proposal and Data Use Agreement	<p>The Center will define the process for seeking data use approval, including specifications for submitting a proposal. This process will include a streamlined and accelerated process for submitting and approving “exploratory” usage requests for the purpose of accessing the data within the Center to determine the scope and viability of a project without the export of any data. Proposals for other non-exploratory projects will include identifying the required datasets and the expected outcomes of the project.</p> <ul style="list-style-type: none">• A set of standard data use agreements for different project types will be developed and potential users will be expected to agree to these terms. In some cases, Institutional Review Board (IRB) approval must be obtained before carrying out research.• The Center will develop a web-based workflow and defined timelines for agency review and approval, with the goal of providing rapid responses to legitimate usage requests.• The Board will identify certain uses that are “pre-approved.” The Center will review proposals and have authority to approve any proposals that meet those criteria without further steps.• The Board will identify certain uses that are not permitted. The Center will pre-screen proposed data uses to ensure that they are potentially permitted before moving them on to agencies.• The Center will provide ongoing monitoring of each approved proposal, and the system will provide visibility to the participating agencies on the status of data use requests and the utilization of the agency’s data within the system.

Responsibility	Brief Description
Data Contributors	Each ECIDS Agency will enter into an agreement with the Center specifying what data will be submitted, the schedule of submission, and the secure method of submission. This agreement will include a streamlined process to add or modify data in the future. The Center will ensure ongoing compliance with data submission requirements. In accordance with Board policy, new contributors can join the system in the future, potentially including contributors from outside state government.
Establish a de-identification system	The Center will move the current matching process for the Master Person Index into the cloud, and implement a matching process that results in de-identified data with names Social Security numbers, and any agency-specific identifiers masked. The Center will provide secure management of this matching process, and Board policy will specify the limited circumstances under which names and Social Security numbers may be legally accessed within an “Analytics and Research Hub” created within the cloud (as further described below). This is the first tier of de-identification within the Center to ensure that all data used for research and analytics has the most sensitive identifiers removed, while still allowing other information -- such as demographic information and birthdates -- to be used for valid analyses. The second tier of de-identification, Disclosure Proofing (described below), ensures that it is never possible to infer the identity of an individual using a dataset released from the Center, even when that dataset is combined with other accessible data.
Establish an Analytics and Research Hub with core data for data use projects	The Center will create a repository of data for analytics and research, the Analytics and Research Hub. This will consist of de-identified data that has been processed through the de-identification vault and is prepared for access for approved data use project activities. Subject to Board policy and only to the extent authorized for an approved project, personally identifiable data may be moved into the Hub for a particular project. The Center will retain all technical documentation (meta-data, inventory, definitions, etc.), and Board policy will define how it will do so.
Data Stewardship	The Center will establish data stewardship processes and workflows for agency data stewards to authorize access to that individual agency’s data and monitor the data’s utilization for any approved project. The data stewardship components of the system will also include reporting on data utilization to each agency.
Workspace and Tools	The Center will provide workspace and tools, ideally cloud-based, to use for analysis of data within the secure Analytics and Research Hub so that there is not a need to export the data outside of the system during the analytics process.
Data User Training and Authorization	All users of the system must receive training before using the system. Board policy will define criteria for approval, and the Center will oversee the credentialing process.

Responsibility	Brief Description
Disclosure Proofing and Data Projects	Prior to any data being released from the system, the Center will ensure that it is “disclosure-proofed” to ensure that it does not include personally identifiable data or small cell sizes. Board policy will define any review and comment periods needed to facilitate this analysis.
Security Policies	The Center will be responsible for ensuring data security and privacy in accordance with Board policy.

The Center will hire the staff and contractors needed to successfully execute these functions. In addition to these responsibilities the Center will be responsible for providing several capacities essential to the success of the ECIDS:

- **Communication.** The ECIDS will only thrive if policymakers and the public understand its value and see benefit from its work on a regular basis. The Center will be responsible for ensuring this communication takes place.
- **Legal.** Managing data use agreements and privacy laws requires legal capacity, which the Center will be responsible for providing (on its own or through contracting).
- **Administrative.** The Center will be responsible for managing procurement, human resources, and other administrative functions, potentially in collaboration with other agencies.
- **Analytic.** While individual agencies will need their own analytic capacity to focus on agency-specific needs, having centralized analytic capacity provides benefit to the Cabinet, the public, and the early childhood field by allowing for the ongoing analysis of critical issues informed by the data. The Center may develop this capacity in-house or choose to partner with others (such as universities) to develop it.

The Center will represent a new form of capacity to support a new kind of technical infrastructure, and will need to be prepared to act rapidly to support the ongoing needs of its partner agencies. It will play a critical role in establishing a new culture of data use in Oklahoma state government.

EXHIBIT D

NEW AGENCY ADMISSION AGREEMENT FORM

NEW AGENCY ADMISSION AGREEMENT TO THE MULTI-AGENCY AGREEMENT FOR THE OKLAHOMA EARLY CHILDHOOD INTEGRATED DATA SYSTEM

THIS NEW AGENCY ADMISSION AGREEMENT (this “New Agency Agreement”) is entered into as of [DATE] by [AGENCY NAME] (“New Agency”) and the Office of the Governor, on behalf of the other Parties to the Multi-Agency Agreement for the Governance of the Oklahoma Early Childhood Integrated Data System (“Multi-Agency Agreement”).

Whenever used in this New Agency Agreement, capitalized terms have the meanings specified herein or in the Multi-Agency Agreement.

I. RECITALS

- A. The ECIDS Agencies and Office of the Governor entered into the Multi-Agency Agreement for the purposes of:
 - i. Establishing a governance structure to develop and administer the ECIDS in a manner that addresses the ECIDS Expectations;
 - ii. Authorizing the establishment and administration of a Center that will provide centralized administrative capacity and technical infrastructure for the ECIDS in accordance with this Agreement and the oversight and direction of the Board;
 - iii. Committing each of the ECIDS Agencies to contribute data for the ECIDS and participate in its governance; and
 - iv. Ensuring the use of the data within the ECIDS supports interagency policy development and analysis in a secure manner consistent with applicable federal and Oklahoma law.
- B. Following approval of the Board, Section VII.A of the Multi-Agency Agreement authorizes the Office of the Governor to enter into a New Agency Admission Agreement on behalf of all Parties, whereupon the new agency shall be deemed an ECIDS Agency for all purposes of the Multi-Agency Agreement and the Board.
- C. New Agency seeks to fully participate in the ECIDS, the Board, and the services provided by the Center.

- D. The Board approved the admission of the New Agency pursuant to a duly authorized resolution included in the Board meeting minutes attached as Exhibit A to this New Agency Agreement.

II. ADMISSION OF NEW AGENCY

- A. Upon the effective date of this New Agency Agreement set forth in the preamble, New Agency is and shall be deemed an ECIDS Agency for all purposes of the Multi-Agency Agreement and the Board.
- B. New Agency agrees to be fully bound by the terms and conditions of the Multi-Agency Agreement, including, without limitation, the ECIDS Agency obligations specified in Section VI.A of the Multi-Agency Agreement.
- C. The Office of the Governor and the New Agency represent and warrant to the other party that this New Agency Agreement has been duly authorized, executed, and delivered by and on behalf of each such party, and constitutes the legal, valid, and binding agreement of said party.
- D. This New Agency Agreement may be executed in several counterparts, each of which shall be an original, and all of which shall constitute one and the same instrument.

Exhibit A: Board minutes authorizing New Agency Agreement

SIGNATORY AUTHORITY

Approved and authorized on behalf of the Office of the Governor and New Agency.

OFFICE OF THE GOVERNOR:

By: _____

Title: _____

Date: _____

[NEW AGENCY NAME]:

By: _____

Title: _____

Date: _____

Appendix H: Survey of Agencies Regarding Work Groups

Oklahoma Early Childhood Data System: Work Group Survey

As part of the development of Oklahoma’s Early Childhood Integrated Data System (ECIDS), the state is planning to establish an Interagency Board to administer the system. To support that Board, Oklahoma will develop collaborative work groups that include members from both state government and outside partners. Foresight Law + Policy (Foresight) and Third Sector Intelligence (3Si) are working with the Oklahoma Partnership for School Readiness (OPSR) and other state leaders to design the Interagency Board and its work group structure. This survey is designed to collect information that informs the work group development process.

The work groups will bring together experts on a range of important topics, providing key advice to Board-level leaders. The work groups should focus on identifying critical issues within their focus area, collaborative problem-solving on those issues, and making recommendations to the Board for formal action.

Collectively the work groups should include at least the following key stakeholders:

- Agency early childhood program staff.
- Agency staff who are responsible for managing data systems (data stewards), for expertise in technology and data security.
- Agency staff who are responsible for research and analysis.
- Agency legal staff.
- Key external users of data, who include:
 - Legislators (and legislative staff)
 - Researchers
 - The early childhood provider community.
 - Advocacy groups focused on early childhood and related fields
 - Advocacy groups focused on vulnerable populations
 - Advocacy groups focused on data privacy
 - Other data end users

In order to establish the work groups, Foresight and 3Si are surveying agencies that might participate in the ECIDS about potential agency staff and outside partners who might participate in these work groups. This information will be used to produce a recommended configuration of work groups that maximizes efficiency and minimizes administrative burden.

Note that state agencies may end up with multiple staff involved in the ECIDS governing structure: one on the Board and one or more on each of multiple work groups. Agencies will be responsible for internal communication to ensure their staff are coordinating and providing consistent feedback. Ultimately the work of the ECIDS will only succeed if each agency is able to tap the expertise of staff with different areas of focus.

It is our strong recommendation that the Board establish work groups with a process for updating their responsibilities over time; as the work evolves, the structure of the work groups should evolve with it. Our goal at this time is to propose initial work groups that actively engage the needed expertise in the right dosage, and help the Board establish the ECIDS as a successful enterprise.

The exact level of time commitment required for the work will vary among work groups, and will be dictated by the needs of the project as determined by the Interagency Board and participating agencies.

Typically the work will involve an intensive setup period in the first year, with meetings as frequently as monthly and potentially substantial work between meetings. After the initial period the work may settle into a less demanding cadence, with meetings bi-monthly or quarterly and less work in between meetings. While the setup work can be demanding, if done correctly it can produce a substantial payoff by reducing the amount of work needed in the future to maintain the system on an ongoing basis.

Please fill out the tables below for each topic area with the information needed to make recommendations for initial ECIDS work groups. Note that the agency can propose the same lead for multiple topics, or more than one participant for any topic (please insert more rows if there is more than one potential additional participant).

The state has identified several topics on which it would like to establish work groups. At this point, no determination has been made as to whether each topic will require its own work group, or whether a single work group may manage several topics. These topics include:

- **Administration and Legal:** The board will need advice and guidance on the development, operations, and administration of the ECIDS. This advice should come from people who work directly on issues of data security and legal agreements concerning data.

Operations	Recommended Individuals	Title	Contact (email/phone)
Agency Lead			
Add'l Participant(s)			

- **Technology and Data:** The Board will need support from leaders with expertise on the technical infrastructure and data management itself -- including knowing what technology and data is available, and how data is collected, defined, and used in IT systems (data stewardship). These experts will make recommendations about curating data from the data sources and informing data access and use. This conversation should include agency data engineers, business analysts, and data stewards.

Data	Recommended Individuals	Title	Contact (email/phone)
Agency Lead			
Add'l Participant(s)			

- **Research & Evaluation:** The Board will need methodological advice and guidance on questions of shared research inquiry. This conversation should include university partners, external researchers, and other stakeholders that would work directly with the resultant ECIDS data on an analytical level. In combination with representatives for Users and Advocates, these stakeholders will be involved in helping to set the research priorities for the ECIDS.

Research	Recommended Individuals	Title	Contact (email/phone)
Agency Lead			
Add'l Participant(s)			

- **Users and Advocates:** The Board will need advice about the user experience from external stakeholders and members of the community who have a vested interest in both the collection and use of the data and any resultant research, analysis, and reports (including legislators and legislative staff). This conversation should provide a more outside-in perspective on the work being done by the ECIDS team.

Users	Recommended Individuals	Title	Contact (email/phone)
Agency Lead			
Add'l Participant(s)			

- **Other topics:** If there are other topics you believe the work groups should address, please describe them here and then list potential contacts:

Users	Recommended Individuals	Title	Contact (email/phone)
Agency Lead			
Add'l Participant(s)			

Please send your completed form to Elliot Regenstein (elliott.regenstein@flpadvisors.com). Thank you for your support of the design work for Oklahoma's ECIDS!

Appendix I: Results of Agency Work Group Survey

Note: OKHCA proposed an additional topic, “Initiatives Related to Health Improvement,” and proposed that it include Shelly Patterson, Director (Shelly.Patterson@okhca.org).

Agency	Admin/Legal	Tech/Data	Research/Eval	Users/Advocates
Education	Brad Clark, General Counsel (Brad.clark@sde.ok.gov)	Erik Friend, Chief Data Officer (Erik.friend@sde.ok.gov)	No response	No response
	Lori Murphy, Assistant General Counsel (Lori.murphy@sde.ok.gov)			
Health	Molly Clinkscales, Staff Attorney (mollyc@health.ok.gov)	Betsy Gloyne, eMPI Manager (BetsyG@health.ok.gov)	Evaren Page, Director of Science and Institutional Review Board (IRB) (EvarenP@health.ok.gov)	John Corpolongo, SoonerStart Director (john@health.ok.gov)
	Kim Heaton, Staff Attorney (kim.heaton@health.ok.gov)	Becki Moore, Informatics Director (BeckiM@health.ok.gov)	Robert Morey, HIPAA Privacy and Security Program Officer RobertXM@health.ok.gov	Gina Richardson, SoonerStart Assistant Director (Gina@health.ok.gov)
	Nicole Nash, Staff Attorney (NicoleN@health.ok.gov)			
OKHCA	Jillian Welch, Deputy General Counsel (Jillian.Welch@okhca.org)	Linh Conley, System Analyst (Linh.Conley@okhca.org)	Angie Brannen, Sr. Research Analyst (Angie.Brannen@okhca.org)	No response
		LaShonda Phillips, System Analyst (LaShonda.Phillips@okhca.org)		

Agency	Admin/Legal	Tech/Data	Research/Eval	Users/Advocates
OKDHS	Molly Green, Information Security Administrator (Molly.green@okdhs.org)	Helen Goulden, Deputy Director (AFS) (Helen.goulden@okdhs.org)	Jennifer Dalton, Research and Evaluation Administrator (Jennifer.dalton@okdhs.org)	TBD
	Ken Kever, Assistant General Counsel (Kenneth.Keever@okdhs.org)	Wanda Threatt, Programs Manager (CSS) (Wanda.threatt@okdhs.org)	TBD, Researcher IV (lead researcher) (position currently being recruited)	
OJA	Kevin Clagg, Deputy Director Finance/Admin (Kevin.clagg@oja.ok.gov)	Len Morris, I.T. Strategist (Len.morris@oja.ok.gov)	Paul Shawler, Chief Psychologist (Paul.Shawler@oja.ok.gov)	Amanda McLain, Program Manager (Amanda.mclain@oja.ok.gov)
	Rachel Holt, Chief Operating Officer / Senior General Counsel (Rachel.holt@oja.ok.gov)	Cheryl McNair, Data Analyst (Cheryl.mcnair@oja.ok.gov)	Len Morris, I.T. Strategist (Len.morris@oja.ok.gov)	David McCullough, Program Manager (David.mccullough@oja.ok.gov)
ODMHSAS	Tracy Leeper, Decision Support Policy Analyst (Tracy.leeper@odmhsas.org)	Austin Ralstin, Sr. Decision Support Analyst (Austin.ralstin@odmhsas.org)	Austin Ralstin, Sr. Decision Support Analyst (Austin.ralstin@odmhsas.org)	Audra Haney, Sr. Manager Infant and Early Childhood Mental Health (Audra.haney@odmhsas.org)
DRS	No response	No response	Lyuda Polyun, Process Improvement Administrator (LPolyun@okdrs.gov)	No response

Appendix J: Center Administration Agreement Key Terms

CENTER ADMINISTRATION AGREEMENT FOR THE OKLAHOMA EARLY CHILDHOOD INTEGRATED DATA SYSTEM

KEY TERMS

Parties:

- Center Administrator selected by the ECIDS Governing Board (“CA”)
- Either all ECIDS Agencies through a joint agreement, or an ECIDS Agency designated by the Board to enter into and oversee this Agreement

Agreement Phases:

- Three phases to the Agreement:
 - Phase I: CA prepares and submits to Board a “Development Plan” that includes detailed scope and specifications for the Center, all proposed subcontractors, all proposed Board policies and processes impacting Center administration (addressing all topics in Section V.C of the Multi-Agency Agreement), detailed schedule for implementation, detailed budget for Phase II development not to exceed an amount pre-approved by the Board, and detailed annual budget for Phase III operations not to exceed an amount pre-approved by the Board. Phase I is concluded upon the Board’s approval of the Development Plan.
 - Agreement will include date by when Development Plan must be completed and submitted for Board review.
 - Agreement will include timelines and process for Board review of the plan and resolving any comments, questions, or objections.
 - CA will enter into Data Contributor and Center Participation (“DCCP”) Agreements with all ECIDS Agencies during Phase I.
 - Board will adopt all policies impacting Center administration during or shortly after the conclusion of Phase I.
 - Phase II: CA develops the Center in accordance with the approved Development Plan, tests and verifies performance of all Center components, and performs initial training of ECIDS Agencies’ staff.
 - Agreement will require completion of the Center in accordance with the Board approved schedule, subject to delays caused by the Board or ECIDS Agencies or other causes outside of the reasonable control of the CA.
 - Agreement will include processes for verifying performance of all Center components by the Board or the Board’s designee.
 - All ECIDS Agencies must make initial data contributions to the Center during Phase II in accordance with DCCP Agreements.
 - During Phase II, the CA submits 1st year operational budget and plan that adheres to operational budget in Development Plan.
 - Phase III: Operations and administration of the Center by the CA.
 - CA submits annual operations budget and plan for each year of Phase III.

- CA annually submits and proposed changes to Board policies and processes for Center administration.
- Agreement will include provisions for budget increases if new ECIDS Agencies join, or other operational changes approved by the Board.

Term:

- Term of the Agreement will extend from date of execution through the expiration of Phase III, subject to early termination rights of Board (described below under Default and Termination)
- Agreement term should provide for at least 5 years of Phase III operations, with automatic annual renewals thereafter unless action is taken by the Board to terminate the Agreement.

Compensation to CA:

- Compensation to CA will consist of:
 - Phase I fee for completion and delivery of the Development Plan, with payment milestones specified in Agreement
 - Phase II fee for completion of the development of the Center, with payment milestones specified in the Agreement
 - Phase III annual payments in accordance with an approved operations budget

Staffing and Contracting:

- During Phase I, the CA will hire an Executive Director for the Center, subject to the Board's approval.
- The Development Plan will include a staffing and contracting plan for the Center during Phase II and Phase III.
- CA and Executive Director will make all hires, subject to Board oversight (but not approval).
- All Subcontractors must either be approved in the Development Plan, or subsequently approved by the Board.

Project Meetings and Reports:

- The Agreement will specify the frequency of meetings and reports between the CA and ECIDS Agencies/Board during all three Phases.
- Phase II should include monthly updates on progress.
- Phase III should include updates for all Board meetings.

General Obligations of CA:

- Throughout each phase, the CA will perform the services in compliance with this Agreement, the DCCP Agreements, and all applicable laws.
- The CA will participate in audit, compliance, and security review processes reasonably required by the Board.

- The CA will comply with other duties and responsibilities reasonably required from time to time by the Board, provided that any duties or responsibilities not contemplated by the Development Plan may lead to an increase in the CA's compensation.
- The CA will adhere to all terms and requirements of the DCCP Agreements, including, but not limited to, the data usage, stewardship, protection, and security requirements.
- The CA will maintain appropriate insurance coverages for each Phase as specified in the Agreement, including business liability, worker's compensation (including employer's liability), and cyber liability coverages.

General Obligations of ECIDS Agencies

- The ECIDS Agencies will cooperate with the CA in the performance of its services under this Agreement.
- The Board and ECIDS Agencies will adhere to the timelines for approval of the Development Plan and Phase II deliverables specified in the Agreement.
- The ECIDS Agencies will maintain and comply with DCCP Agreements with the CA during Phase III.

Intellectual Property

EITHER:

- To the extent feasible, CA commits to release all software and intellectual property developed under the Agreement as open source. If the software and intellectual property cannot be made available as open source, CA grants a license to the ECIDS Agencies to use the software for purposes of the Center.

OR

- CA owns intellectual property relating to the Center, but grants to ECIDS Agencies a nonexclusive, nontransferable, revocable license to ECID Agencies to access and use the Center's system components during the term of the Agreement.

OR

- CA owns intellectual property created independently of this Agreement. ECIDS Agencies own intellectual property created under this Agreement, but CA retains right to use for this project and other projects.

Confidential Information

- Each party agrees to maintain and protect the confidentiality of the other party's confidential information.
- CA provided the right to redact proprietary and confidential information from any request to access the information through the Oklahoma Open Records Act.

Hold Harmless and Insurance

- *Hold Harmless*: CA agrees to defend, indemnify, and hold harmless the ECIDS Agencies for claims to the extent resulting from its negligence or willful misconduct, except to the extent resulting from agency data collection processes, agency disclosures of data, an agency's breach, or an agency's negligent acts or omissions.
- *Consequential Damages*: CA not responsible for consequential damages.
- *Limitation of Liability*: CA's liability limited to direct damages actually incurred, or the fees actually received over the prior 12-month period.
- *Insurance*: CA agrees to maintain business liability, worker's compensation (including employer's liability), and cyber liability insurance coverages in appropriate amounts, naming the ECIDS Agencies as additional insureds.

Default and Termination

- Agreement includes dispute resolution procedures involving Board representatives and Executive Director of the Center.
- Board has right to authorize termination if it determines the best interests of ECIDS would be served thereby, provided CA must be provided opportunity to consult with Board to avoid such a termination.
- Board has right to terminate for default, following notice and cure opportunity.
- CA has right to suspend services for failure to pay or other failures to perform by ECIDS Agencies.
- Agreement addresses orderly transition and protection of data and the disposition of software and equipment in the event Agreement is terminated.

General Provisions

- Agreement includes standard set of general provisions similar to those included in DCCP Agreement form.

Appendix K: Data Contributor Agreement

DATA CONTRIBUTOR AND CENTER PARTICIPATION AGREEMENT FOR THE OKLAHOMA EARLY CHILDHOOD INTEGRATED DATA SYSTEM

THIS DATA CONTRIBUTOR AND CENTER PARTICIPATION AGREEMENT (this “Agreement”) is entered into as of the Effective Date by and between the [*Agency Name*] (“Agency”), and [*Center Administrator Entity Name*] (“Center Administrator”) (each, a “Party”, and collectively, the “Parties”) under authority as provided by law. In consideration of the mutual covenants set forth below, the Parties agree as follows:

I. RECITALS.

- A. Recognizing the power of integrated data across multiple State of Oklahoma agencies involved in early childhood service delivery and the opportunity to improve upon prior efforts to integrate and utilize data, certain State of Oklahoma agencies commissioned and adopted an Oklahoma Early Childhood Integrated Data System Framework dated [_____, 2019] (“Data Governance Framework”) which sets forth objectives, success factors, and a proposed framework for the development and administration of the Oklahoma early childhood integrated data system (“ECIDS”).
- B. In order to establish a governance structure to implement the Data Governance Framework and establish and specify the overall requirements, functions, and expectations for the ECIDS, ten State of Oklahoma agencies, including the Agency, (collectively, the “ECIDS Agencies”) and the Office of the Governor entered into that certain Multi-Agency Agreement for the Governance of the Oklahoma Early Childhood Integrated Data System with an effective date of [_____, 202__] (“Multi-Agency Agreement”).
- C. Pursuant to Section V of the Multi-Agency Agreement, the ECIDS Agencies and the Center Administrator entered into that certain Center Administration Agreement with an effective date of [_____, 202__] (“Center Administration Agreement”) pursuant to which the Center Administrator agreed to develop and administer a Center providing centralized administrative capacity and technical infrastructure for the ECIDS in accordance with the Multi-Agency Agreement and the Center Administration Agreement.
- D. Under Section VI.A of the Multi-Agency Agreement, the Agency agreed to submit data to the Center Administrator through a data contributor agreement and in accordance with the procedures, schedule, and requirements adopted by the Board.
- E. The Board approved the form of this Agreement on [_____, 20__].

- F. The Parties wish to enter into this Agreement in order to:
- i. Effectuate the intent and requirements of the Multi-Agency Agreement to ensure the Center Administrator can effectively develop and administer the Center;
 - ii. Establish the Center Administrator as the authorized representative of the Agency for the purposes of receiving, maintaining and administering Agency data and the performance of other related services; and
 - iii. Facilitate research on and the audit and evaluation of education, human services, and public aid programs for, or on behalf of, the Agency in a manner permitted by applicable Oklahoma and federal law.

II. **DEFINITIONS.**

Whenever used in this Agreement, all capitalized terms have the meanings specified herein or referred to in Exhibit A.

III. **DATA CONTRIBUTION AND USAGE.**

- A. **Designation as Authorized Representative.** The Agency hereby designates and recognizes the Center Administrator as the authorized administrator of the Center and as the authorized representative of the Agency for the purposes of the Center Services. The Agency acknowledges and agrees that the Center Services will:
- i. Facilitate the Agency's audit or evaluation of various state and federal programs, and support the Agency's administration of public funds and provision of services to its constituents;
 - ii. Enable studies for, or on behalf of, schools, school districts, early childhood providers, and the Agency in a manner permitted by applicable Oklahoma and federal law; and
 - iii. Otherwise support the performance of the Agency's official duties.
- B. **Participation in Center Services.** The Agency agrees to participate in the Center Services in accordance with this Agreement, the Multi-Agency Agreement, and the expectations established from time to time by the Board. The Center Administrator grants to the Agency a nonexclusive and nontransferable license, during the term of this Agreement, to access and use the Center Services in compliance with the terms and conditions of this Agreement. The license provided to the Agency in this paragraph may be suspended or revoked by the Center Administrator in the event the Agency fails to comply with the terms and provisions of this Agreement. The Center Administrator agrees to support the Agency's participation in the Center Services through the training and support activities described in Exhibit B.
- C. **Agency Data Submissions.**
- i. **Agency Data and Schedule.** For the purposes of establishing and performing the Center Services, the Agency agrees to submit to the Center Administrator the data set forth in Exhibit C (the "Agency Data") in accordance with the

submission schedule provided in Exhibit C. The Parties acknowledge that the description of Agency Data included in Exhibit C must be periodically updated to ensure inclusion of all necessary and appropriate data for the Center Services. Accordingly, Exhibit C may be amended by the agreement of the Agency Representative and the Executive Director of the Center in the form of Exhibit C-1 without the requirement of any further approvals by either Party.

- ii. Authorized Submission by Agency. The Agency represents and warrants that it has the legal right and authority to submit and provide the Agency Data to the Center Administrator for the purposes contemplated by this Agreement. The Agency is responsible for obtaining all necessary consents and otherwise complying with all applicable laws with respect to its submission of Agency Data to the Center Administrator. The Center Administrator shall not be liable for the accuracy, completeness, validity, or utility of Agency Data.
- iii. Secure Transmissions. The Agency will submit all Agency Data using secure transmission methods defined by the Center.

D. Data Usage and Stewardship.

- i. Utilization and Integration of Agency Data. The Center Administrator agrees to use the Agency Data solely for the purposes and extent as authorized by this Agreement. The Agency acknowledges and agrees that within the Center environment and for the sole purpose of performing the Center Services, Agency Data will be maintained and integrated with data from other ECIDS Agencies and other external data submitted to the Center for the purposes of De-identification Services and Authorized Projects performed strictly in accordance with this Section III.D and policies approved by the Board. External data may be received from both governmental and non-governmental entities in accordance with Board policy.
- ii. De-identification Services. Promptly upon the receipt of Agency Data, authorized Center Administrator staff will process and de-identify the Agency Data in accordance with Exhibit D (“De-identification Services”). The Agency acknowledges and agrees that the Center Administrator will combine Agency Data with data from other ECIDS Agencies for the purposes of De-identification Services. The Center Administrator will only utilize and provide access to De-identified Data for Authorized Projects unless the use of Personally Identifiable Data is necessary for a legitimate project purpose and is approved in advance by the Agency Representative.
- iii. Authorized Projects and Levels of Authorization. The Center Administrator will deploy a Data Stewardship Module that enables authorized Agency staff to manage data access requests and approval workflows and approve all

utilization of Agency Data for Authorized Projects. The Data Stewardship Module will include four levels of authorization:

- a. *Access to Non-Sensitive Data:* Non-Sensitive Data classified as such in accordance with Board policy may be made available by the Center Administrator to any party for any purpose, including through dashboards, downloadable files, and Application Programming Interfaces.
- b. *Agency Access:* Authorized staff of the Agency may access and utilize Agency De-identified Data within the Center at any time following the user's completion of an acknowledgement within the Data Stewardship Module of the confidentiality and security requirements applicable to such data.
- c. *Interagency Access:* Interagency Projects utilizing Sensitive Data must be approved within the Data Stewardship Module through procedures approved by the Board and shall be subject to the approval of each of the Agency Representatives whose Agency's data will be utilized.
- d. *External Party Access:* External Party Projects utilizing Sensitive Data must be approved within the Data Stewardship Module through procedures approved by the Board and shall be subject to the approval of each of the Agency Representatives whose Agency's data will be utilized. Further, the External Party representatives accessing Sensitive Data within the Center must be trained and authorized in accordance with Center procedures approved by the Board.

The Agency agrees that its Agency Representative is and shall be authorized to approve all projects in accordance with this Agreement.

- iv. Exploratory Projects, Authorized Project Agreements, and Disclosure-Proofing. Exploratory Projects occurring entirely within the secure Center environment will be subject to a streamlined and accelerated process for authorization approved by the Board. All Interagency and External Party Projects utilizing Sensitive Data must be performed in accordance with an Authorized Project Agreement in a digitalized form approved by the Board, with a simplified form for Exploratory Projects. All analysis and utilization of Sensitive Data for Authorized Projects will occur entirely within the secure Center environment. The Center Administrator will not permit the export of Sensitive Data from the Center's secure environment until and unless it has been fully Disclosure-Proofed. Disclosure-Proofing will be performed by the Center Administrator in accordance with industry leading practices and procedures approved by the Board.

IV. DATA PROTECTION, SECURITY, AND COMPLIANCE.

E. Protection and Security.

- i. Security Measures. The Center Administrator will restrict access to and maintain the security of Sensitive Data through appropriate administrative, technical, and physical security measures as more fully set forth in Exhibit E. All data transmissions between the Agency and Center Administrator will be encrypted and protected in accordance with Exhibit E.
- ii. Access Restrictions. The Center Administrator will limit access to Sensitive Data to those employees and contractors who reasonably need access to them in order to perform their responsibilities under this Agreement. The Center Administrator will instruct all persons having access to Sensitive Data on the use and confidentiality restrictions set forth in this Agreement. Each employee or contractor of the Center with access to Sensitive Data must execute a Security Pledge in a form approved by the Board, which shall be maintained and updated by the Center Administrator at all times.
- iii. Reporting Infractions. The Center Administrator will fully report to the Agency within one day of discovery any infraction of the data protection, security, and compliance provisions set forth in the Agreement. The report will include the nature of the infraction, what the Center Administrator has done or will do to mitigate any deleterious effect, and what corrective action the Center Administrator has taken or will take to prevent future similar infractions.

F. Compliance.

- i. Compliance With Relevant Laws. The Center Administrator will comply with the relevant requirements of Oklahoma and federal law relating to data use, privacy, security, and dissemination, including, without limitation, those laws specified in Exhibit B of the Multi-Agency Agreement. If and to the extent applicable, the Center Administrator will not disclose protected health information as defined in HIPAA other than as permitted or required by law and agrees to use appropriate safeguards to prevent use or disclosure of any protected health information.
- ii. Open Records Act Requests. The Agency will have all responsibility for responding to any request to access Agency Data through the Oklahoma Open Records Act, and will hold the Center Administrator harmless for any costs or expenses incurred by the Center Administrator as the result of a determination made by the Agency in response to an Open Records Act request.
- iii. Records of Access and Utilization. The Center Administrator will maintain records of access and utilization of all Sensitive Data received pursuant to this

Agreement. The records will record the names of the persons and organizations accessing the data, and the description of the data accessed.

- iv. **Contractors and Subcontractor Compliance.** The Center Administrator will ensure all of its contractors and subcontractors accessing Sensitive Data agree by contractual terms to all provisions of this Agreement pertaining to data protection, privacy, security, dissemination, and compliance.

V. HOLD HARMLESS AND INSURANCE.

- A. **Hold Harmless.** The Center Administrator agrees to defend, indemnify, and hold harmless the Agency against any and all claims, suits, damages, and causes of action to the extent arising out of the negligent performance of the Center Services or the Center Administrator's obligations under this Agreement, except to the extent that any such claim results from the Agency's data collection processes, the Agency's disclosure of data to the Center Administrator, the Agency's breach of this Agreement, or the Agency's negligent acts or omissions.
- B. **Consequential Damages.** In no event will the Center Administrator be responsible for indirect, consequential, special, exemplary, or incidental damages arising out of, or otherwise relating to, this Agreement or the Center Services.
- C. **Limitation of Liability.** The total liability of the Center Administrator with respect to the Center Services to all ECIDS Agencies shall be limited to the lesser of the Agencies' direct damages actually incurred, or the fees actually received by the Center Administrator for the Center Services over the prior twelve- (12-) month period.
- D. **Insurance.** The Center Administrator will, at all times during the term of this Agreement, maintain the business liability, worker's compensation (including employer's liability) and cyber liability insurance coverages in the amounts and in accordance with the Center Administration Agreement. The Agency will be named as an additional insured on all such coverages (except for the worker's compensation policy).

VI. TERM AND TERMINATION.

- A. **Term.** The term of this Agreement will commence on the date of signature by both Parties ("Effective Date") and, subject to any earlier termination as provided in Section VI.D, shall remain in full force and effect until:
 - i. the expiration of the Multi-Agency Agreement;
 - ii. the Agency's authorized withdrawal from the Multi-Agency Agreement; or
 - iii. the termination or expiration of the Center Administration Agreement.
- B. **Dispute Resolution.** In the event of any disputed matter arising under this Agreement between the Parties, the dispute shall be referred jointly to the Agency

Representative and the Executive Director of the Center for resolution. If such disputed matters remain unresolved after ten (10) days, the Parties may thereafter pursue such other remedies as are available under this Agreement or at law or in equity.

- C. **Suspension of Services.** The Center Administrator may suspend the availability of the Center Services to the Agency during any period in which the Agency fails to perform any agreement, obligation, duty, or provision of this Agreement.
- D. **Default and Termination.** The Agency may terminate this Agreement if the Center Administrator materially fails to observe or perform any agreement, obligation, duty, or provision of this Agreement, and the material failure continues for thirty (30) days after the Center Administrator's receipt of written notice from the Agency and the conclusion of the dispute resolution process set forth in Section VI.B.
- E. **Force Majeure.** Notwithstanding any other provision of this Agreement, the Center Administrator shall not be deemed in default or breach of this Agreement or liable for any loss or damages or for any delay or failure in performance due to circumstances beyond the Center Administrator's reasonable control, including, without limitation, acts of God, acts of government, flood, fire, earthquakes, civil unrest, acts of terror, strikes or other labor unrest or problems, computer, telecommunications, Internet service provider or hosting facility failures or delays involving hardware, software or power systems not within the Center Administrator's possession or reasonable control. Without limiting the generality of the foregoing, the Center Administrator shall not be liable to the Agency in any way for any failure or delay in the performance of the Center Administrator's obligations under this Agreement if such failure is caused, directly or indirectly, by the Agency's performance of, or failure to perform, its obligations under this Agreement.
- F. **Destruction of Data.** Upon termination for any reason, Sensitive Data provided by the Agency will be destroyed through a secure manner as directed by the Agency and meeting any Board requirements for destruction. Notwithstanding the foregoing, Sensitive Data being utilized for an Authorized Project as of the date of termination may continue to be utilized until the conclusion of that project.

VII. **GENERAL PROVISIONS.**

- A. **Cooperation.** Each Party agrees to use reasonable efforts to cooperate with the other Party with respect to providing the Center Services and commits to working in good faith with the other.
- B. **Amendment.** An amendment to the Agreement may only occur through the written approval of both Parties.
- C. **Subcontracting.** The Center Administrator may subcontract the Center Services in accordance with the Center Administration Agreement.

- D. Assignment.** This Agreement may not be assigned by the Agency without the written consent of the Center Administrator. The Center Administrator may only assign this Agreement to an entity receiving the assignment of the Center Administration Agreement in accordance with the terms and provisions of such agreement.
- E. Retention of Records.** The Center Administrator will maintain records pertaining to this Agreement, including information stored in databases or other computer systems, for a period of three (3) years from the termination of this Agreement. Records required to be maintained under this paragraph shall be available for review or audit by representatives of the Agency or other governmental entities with monitoring authority upon reasonable notice and during normal business hours.
- F. Independent Contractor.** The Center Administrator is acting as an independent contractor in its capacity under this Agreement. Nothing contained in this Agreement or in the relationship of the Agency and Center Administrator shall be deemed to constitute a partnership, joint venture, or any other relationship between the Parties except as is limited by the terms of this Agreement.
- G. Notices.** All notices or other correspondence required to be given to this Agreement shall be sent by regular or electronic mail as follows:
For the Agency: To the Agency Representative
For the Center Administrator: To the Executive Director
- H. No Third Party Beneficiaries.** Except as expressly provided herein, this Agreement is entered into solely between, and may be enforced only by, the Agency and Center Administrator. This Agreement will not be deemed to create any rights or causes of action in or on behalf of any third parties, or to create any obligation of a Party to any such third parties.
- I. Entirety.** This Agreement, together with the Exhibits attached hereto, constitutes the entire agreement among the Parties with respect to the subject matter hereof, and supersedes any other negotiations, agreements, or communications, whether written or oral, that have been made by any Party.
- J. Governing Law.** This Agreement shall be governed by and construed in accordance with the laws of the State of Oklahoma.
- K. Severability.** In case any provision in this Agreement is held to be invalid, illegal, or unenforceable, the validity, legality, and enforceability of the remaining provisions shall not be affected.
- L. Authority to Execute.** Each Party represents and warrants to the other Party that this Agreement has been duly authorized, executed, and delivered by and on behalf of the Party, and constitutes the legal, valid, and binding agreement of said Party.

M. Counterparts. This Agreement may be executed in several counterparts, each of which shall be an original, and all of which shall constitute one and the same instrument. For purposes of this Agreement, a facsimile copy or a scanned printable document format (pdf) of a Party's signature shall be sufficient to bind such Party.

N. Recitals and Exhibits Incorporated. The recitals in Section I and the following Exhibits attached hereto are hereby incorporated into this Agreement by this reference and expressly made a part of this Agreement:

Exhibit A	Definitions
Exhibit B	Center Services
Exhibit C	Agency Data and Submission Schedule
Exhibit C-1	Amendment to Agency Data and Submission Schedule
Exhibit D	De-identification Services
Exhibit E	Administrative, Technical, and Physical Security Measures

[Signature Page Follows]

**DATA CONTRIBUTOR AND CENTER PARTICIPATION AGREEMENT FOR THE
OKLAHOMA EARLY CHILDHOOD INTEGRATED DATA SYSTEM**

SIGNATORY AUTHORITY

Approved and authorized on behalf of each of the Parties.

AGENCY:

[*AGENCY NAME*]

By: _____

Title: _____

Date: _____

CENTER ADMINISTRATOR:

[*CENTER ADMININSTRATOR ENTITY NAME*]

By: _____

Title: _____

Date: _____

EXHIBIT A⁶⁰

DEFINITIONS

“Agency” is defined in the Preamble of this Agreement.

“Agency Data” is defined in Section III.C.1 of this Agreement.

“Agency Representative” means the Agency’s member of the Board designated pursuant to the Multi-Agency Agreement.

“Agreement” is defined in the Preamble of this Agreement.

“Authorized Project” means a project involving the utilization of Sensitive Data within the Center for research, analysis, audit, or evaluation purposes that has been approved by all Agencies whose Sensitive Data will be utilized through the Data Stewardship Module, and is performed pursuant to an Authorized Project Agreement.

“Authorized Project Agreement” means a digitalized agreement in a form approved by the Board among the Center Administrator, one or more ECIDS Agencies, and the Project Sponsor for the performance of an Authorized Project.

“Board” means the governing board of the ECIDS established pursuant to the Multi-Agency Agreement.

“Center” means the centralized administrative capacity and technical infrastructure for the ECIDS developed and operated pursuant to the Multi-Agency Agreement, the Center Administration Agreement, and the oversight and direction of the Board.

“Center Administration Agreement” is defined in Section I.C of this Agreement.

“Center Administrator” is defined in the Preamble of this Agreement.

“Center Services” means the services, supports, and training provided by the Center Administrator in furtherance of the operations and administration of the ECIDS, as more particularly described in Exhibit B of this Agreement.

“Data Governance Framework” is defined in Section I.A of this Agreement.

“Data Stewardship Module” means a software module and related policies and procedures of the Center that provides (i) management of dataset policies and Agency data steward identification and authorization rights, (ii) management of Sensitive Data access requests and Agency approval

⁶⁰ Note that normally each exhibit would begin at the top of a page; they have been compressed in this Appendix K to save space.

workflows, (iii) reports on Sensitive Data access and utilization, and (iv) records of disclosure of Sensitive Data.

“De-identification Services” is defined in Section III.D.2 of this Agreement.

“De-identified Data” means data resulting from the De-identification Services that is encrypted and does not include name, social security numbers, or any state agency identification numbers, but may include other information such as demographic information and birthdates that facilitate valid research and analytics within the Center.

“Disclosure-Proofed” means that is not possible to infer the identify of any individual through the analysis of a dataset, even when the data within the dataset is combined with other individual data.

“Disclosure-Proofing” means the process utilized by the Center to make data Disclosure-Proofed.

“ECIDS” is defined in Section I.A of this Agreement.

“ECIDS Agencies” is defined in Section I.B of this Agreement.

“Effective Date” is defined in Section VI.A of this Agreement.

“Exploratory Project” means an Interagency or External Party Project undertaken for the primary purpose of determining the scope or viability of a subsequent research or analytical project and that does not involve the Disclosure-Proofing of data for export from the Center.

“External Party” means any person or entity other than staff or contractors of an ECIDS Agency, the Office of the Governor, or the Center Administrator.

“External Party Project” means an Authorized Project where the Project Sponsor is an External Party.

“HIPAA” means the Health Insurance Portability and Accountability Act of 1996, as amended, and the regulations and rules promulgated thereunder.

“Interagency Project” means an Authorized Project where the Project Sponsor is an ECIDS Agency.

“Master Person Index ID” means the identification number generated through the ECIDS Agencies’ master person index system.

“Multi-Agency Agreement” is defined in Section I.B of this Agreement.

“Non-Sensitive Data” means data that does not contain any Sensitive Data, and does not allow the re-identification or reasonably potential inference of the identity of any individual therein, even when combined with other available data.

“Party” and “Parties” is defined in the Preamble of this Agreement.

“Personally Identifiable Data” means data that can be used, on its own or in combination with other available data, to identify an individual person and includes, but is not limited to, names, addresses, birthdates, social security numbers, Federal Identification Numbers, and direct personal identifiers used by an Agency for administrative purposes.

“Project Sponsor” means either an External Party or ECIDS Agency who is serving as the principal entity responsible for the undertaking, management, and completion of an Authorized Project.

“Security Pledge” means a pledge in a form approved by the Board that must be executed by any individual with access to Sensitive Data where the individual agrees to (i) maintain the confidentiality of the data, (ii) protect it from unauthorized disclosure and use, (iii) abide by all Board and Center requirements relating to data security and confidentiality, and (iv) immediately report any known incident threatening the security and confidentiality of the data.

“Sensitive Data” means Personally Identifiable Data or other data that is confidential within the meaning of any governing law, regulation, or Agency directive.

EXHIBIT B

CENTER SERVICES

[Provide a description of the services, support, and training to be provided by the Center Administrator that will be generally applicable across all ECIDS Agencies.]

EXHIBIT C

AGENCY DATA AND SUBMISSION SCHEDULE

Description of Agency Datasets to be Submitted for Center Services

System Name	General System Description	Initial Time Period of Dataset*	Description of Data Element Categories**	Submission Frequency

*Unless otherwise specified, data will be provided from the initial time period through the most currently available data

** May specify “all”, or describe certain categories of data to be included

EXHIBIT C-1

AMENDMENT TO AGENCY DATA AND SUBMISSION SCHEDULE

**AMENDMENT # _____ AMENDING EXHIBIT C OF
THE DATA CONTRIBUTOR AND CENTER PARTICIPATION AGREEMENT FOR
THE OKLAHOMA EARLY CHILDHOOD INTEGRATED DATA SYSTEM**

REVISED Description of Agency Datasets to be Submitted for Center Services

The following table replaces and supersedes the table previously agreed to by the Parties:

System Name	General System Description	Initial Time Period of Dataset*	Description of Data Element Categories**	Submission Frequency

*Unless otherwise specified, data will be provided from the initial time period through the most currently available data

** May specify “all”, or describe certain categories of data to be included

AGENCY:

[*AGENCY NAME*]

By: _____

Title: _____

Date: _____

CENTER ADMINISTRATOR:

[*CENTER ADMININSTRATOR ENTITY NAME*]

By: _____

Title: _____

Date: _____

EXHIBIT D

DE-IDENTIFICATION SERVICES

[Provide a description of the de-identification services to be provided by the Center Administrator that will be generally applicable across all ECIDS Agencies.]

EXHIBIT E

ADMINISTRATIVE, TECHNICAL, AND PHYSICAL SECURITY MEASURES

[Provide a description of the administrative, technical, and physical security measures to be provided by the Center Administrator that will be generally applicable across all ECIDS Agencies.]

Appendix L: Governance Questions Shared with Agencies

The following text was sent to each agency prior to its stakeholder interview:

Oklahoma has received a federal grant through the Preschool Development Grant-Birth to Five program that is being used in part to support the design of an Early Childhood Integrated Data System (ECIDS) that can link data across agencies. The state has engaged Foresight Law + Policy and Third Sector Intelligence to develop a plan for the ECIDS, which must include a governance framework. To develop a plan for governance, Foresight will be conducting interviews with agency leaders in Oklahoma City. The questions below will form the primary basis for those discussions. If you have any questions about the interview, please contact Elliot Regenstein at Elliot.Regenstein@flpadvisors.com.

- Vision and Mission
 - What would you articulate as your vision of shared interagency data use? How do you see it potentially benefitting your agency?
 - How do you see your agency using information from the priority use cases that have been identified in this project?
 - What are the most significant benefits you see for your agency, and for the State, arising from an interagency governance structure?
- Composition and Membership
 - What agencies do you see as critical to an interagency governance structure?
 - The success of the governance structure will require empowered leaders from each agency, who can make decisions on behalf of the agency. In your agency, who are some candidates to fill that role?
 - In many states key leaders from outside government play some role in the interagency governance structure, often representing a cross-system perspective or the voice of end users. Do you have any concerns with this approach? If so, what are they and why? Does it make a difference if these individuals are in purely advisory roles?
 - Are there examples of interagency collaborations that you have seen as successful? If so, what made them that way?
 - What are some aspects of interagency collaboration that have been challenging? What made them that way?
 - What is most important to you about your agency's role?
- Roles and Responsibilities
 - Are you comfortable with an interagency governance body that includes your agency addressing the following issues:
 - Defining the system's purpose
 - Providing guidance for data use
 - Ensuring privacy and security
 - If you have any concerns about the governance body playing these roles, what are they?
 - Are there any other roles you would like to see the governing body play?

- Data Decisions
 - Are you comfortable with an interagency governance body that includes your agency making decisions about each of the following:
 - Setting a common data agenda, including priority use cases and/or a research agenda
 - Data collection – content and processes
 - Data linking processes
 - Data sharing processes
 - Data maintenance and retention policies and processes
 - The use of data
 - Data quality standards, including common definitions and business rules
 - Ownership issues relating to data
 - Access protocols, including appeals processes for data requests that are denied
 - Transparency in governance decision-making
 - Accountability for member agencies about adherence to data governance policies
 - If you have any concerns about the governance body playing these roles, what are they?
 - Are there any other issues you would like to see the governing body address?
- Committee Processes
 - Data governance bodies typically develop committee structures to address the full range of issues for which they are responsible. Sometimes those committees involve different staff than the staff on the primary governance board – for example, a policymaking committee, a management committee, and/or a committee of data stewards. Are you prepared for the possibility of having multiple staff involved in this work and coordinating internally?
- Sustainability
 - What do you recommend as a strategy to ensure that the governance structure is sustainable, with ongoing funding and access to adequate dedicated staffing?
 - Would you be willing to support the development of legislation to codify the role of the governance body?
 - Would you be willing to contribute a proportionate share to the shared costs of the governance structure, and participate in an interagency annual budgeting process?
 - How would you like to see stakeholders from outside of state government engaged in this work, potentially including as part of the interagency governance body or on its committees?

Note: The structure of this protocol and several specific questions are inspired by the Data Quality Campaign publication [Roadmap for Cross-Agency Data Governance](#). Other critical resources include:

- Data Quality Campaign, [The Art of the Possible: Data Governance Lessons Learned from Kentucky, Maryland, and Washington](#)
- Statewide Longitudinal Data System Grant Support Team, [Early Childhood Integrated Data Systems: Data Governance](#) (part of a larger [Early Childhood Integrated Data Systems Toolkit](#))
- Statewide Longitudinal Data System Grant Support Team, [Early Childhood Data Governance in Action! An Introduction](#) and [Early Childhood Data Governance in Action! Initial Steps to Establish Data Governance](#)

Appendix M: Examples of Data Governance Models

M.1 Different Models of Data Governance

IDS data governance models can be primarily categorized by the following criteria: the type of entity (or entities) leading the initiative; the source(s) of the data to be incorporated; and the level of data integration occurring within the Integrated Data System (IDS) itself (federated models often operate around a Master Person Index, or MPI, and minimize direct integration of the various source data). Below are categorized examples of data governance models being used at a variety of entities at various levels:

- **State agency-led** – these IDS initiatives are led by an established state government agency or multiple established state agencies working collaboratively.
 - *Multi-agency, integrated*
 - North Carolina – NC ECIDS (DHHS, DPI, Head Start, North Carolina Partnership for Children)
 - Rebecca Planchard, Hayley Young (hayley.young@dhhs.nc.gov)
 - <https://www.ecids.nc.gov/ecids/>
 - Iowa – Early Childhood IDS (DPH, Workforce, DOE, DHR, OED, DHS)
 - Heather Rouse (hlrouse@iastate.edu)
 - *Single-agency, integrated*
 - Oregon – Office of Forecasting, Research and Analytics (OFRA)’s OEDA, LDS, and ICS systems
 - Wesley Mouw (WESLEY.J.MOUW@dhsosha.state.or.us), Bobby Webber
 - <https://www.oregon.gov/DHS/BUSINESS-SERVICES/OFRA/Pages/index.aspx>
- **Executive-led** – these IDS initiatives are led by a state-level entity, typically under the Governor’s office or similar executive branch that has been established expressly for the purpose of integrating and managing state data.
 - *Single-agency, integrated*
 - Kentucky – Kentucky Center for Statistics (KYStats) (pulls data from DOE, KHEAA, and EWD, EPSB, CPSE)
 - Jessica Cunningham (jessica.cunningham@ky.gov)
 - <https://kystats.ky.gov/>
 - Maryland – MLDS Center (DOL, DOE, HEC, DOT)
 - mlds.center@maryland.gov
 - <https://mldscenter.maryland.gov/welcome-index.html>
 - *Single-agency, federated*
 - Indiana – Management Performance Hub (MPH)
 - Darshan Shah (dashah@mph.in.gov)
 - <https://mph.in.gov>
 - Virginia – VLDS (DOE, HEC, Employment Commission, DSS, Health Professions, Community Colleges, Aging and Rehab. Services)
 - Todd Masa (TodMasa@schev.edu)
 - <https://vlds.virginia.gov/about-vlds>

- **Independent not-for-profits** that centralize data from multiple sources – these IDS initiatives are led by a nonprofit entity that establishes distinct agreements with multiple government entities and other data sources, doing the integration work in-house and partnering with the source entity to provide research products.
 - Open Commons Consortium (cloud-based solutions, works with federal and university data)
 - <http://occ-data.org/>
 - Robert Grossman, Li Ko (li.ko@occ-data.org)
 - Chapin Hall at the University of Chicago (include state and city early childhood, K-12, human services, early intervention, nutrition and income maintenance program data)
 - <https://www.chapinhall.org/project/the-integrated-database-on-child-and-family-services-in-illinois/>
- **University-led** – these IDS initiatives are led by a university partner of one or multiple government entities; typically, these universities are publicly-funded and reside within the partner state.
 - *Multi-agency, integrated*
 - California – California Policy Lab (UC Berkeley and UCLA)
 - Evan White (evanbwhite@berkeley.edu)
 - Janey Rountree (janey@cpl.ucla.edu)
 - <http://www.capolicylab.org>
 - Rhode Island – Rhode Island Innovative Policy Lab (RIPL) (Brown)
 - *Single-agency, integrated*
 - California – Children’s Data Network (USC)
 - Emily Putnam-Hornstein (ehornste@usc.edu)
 - <http://www.datanetwork.org/research/chhs-annual-record-reconciliation/>
- **Hybrid models** – these IDS initiatives are led by a hybrid body comprised of multiple types of entities working in partnership. While many, if not all, IDS initiatives have executive committees that represent leadership from multiple entity types, the “primary” entity or entities is/are usually of a single type. In a hybrid model, the collaboration is codified at the highest possible level.
 - *Multi-agency, federated*
 - Arkansas Research Center (University/State collaborative entity) – lost DoE after 2013
 - Greg Holland (greg.holland@arkansas.gov)
 - <http://arc.arkansas.gov>
 - *Multi-agency, integrated*
 - Environmental Data Commons (NFP/University collaborative entity – OCC and University of Chicago)
 - *Single-agency, federated*
 - Washington – Washington State Institute for Public Policy (WSIPP) (Evergreen State College/state gov collaborative entity)
 - <https://www.wsipp.wa.gov/>

- **Private sector models** – these data governance models are not in partnership with any government-level entities; rather, these companies provide data governance solutions to purchasing parties and consulting services to those they contract with.
 - Erwin (proprietary) <https://erwin.com/solutions/data-governance/>
 - Palantir (proprietary) <https://www.palantir.com/palantir-gotham/>

M.2 Deeper Descriptions of Executive-led data governance models

The following are dossier-style descriptions of identified examples of executive-led data governance models. This list is not exhaustive, but details some of the diversity that may be found within this type of data governance model.

KYSTATS (KENTUCKY CENTER FOR STATISTICS)

Date established: 2012

Established by law (e.g., state statute, gubernatorial executive order): state statutes KRS 151B.131 - KRS 151B.134

Data integration approach: Fully integrated – data are housed with the agency and linked internally

Includes data from the following agencies: Kentucky Department of Education (KDE), the Council on Postsecondary Education (CPE), the Education Professional Standards Board (EPSB), the Kentucky Higher Education Assistance Authority (KHEAA), the Kentucky Education and Workforce Development Cabinet

Scope: P-20W; some evidence of SNAP integration and possible other health/human services programs

Funding: 2015 SLDS grant, WDQI grant, SNAP Data and Technical Assistance grant, Employment and Training Administration grant, Bureau of Labor Statistics Labor Market Information Cooperative Agreement Funded Programs grant, Preschool Development grant (ESSA); state appropriations; user fees (KRS 151B.132(8))

Point of contact: Jessica Cunningham (jessica.cunningham@ky.gov)

Website: <https://kystats.ky.gov/>

Description: The Kentucky Center for Statistics (KYSTATS) collects and links data to evaluate education and workforce efforts in the Commonwealth. This includes developing reports, responding to research requests, and providing statistical data about these efforts so policymakers, agencies, and the general public can make better informed decisions.

The primary focus of its output seems to be on labor, including unemployment rates, labor markets, and skill building. Reports are largely static documents, and data requests may be made by form submission. A data dictionary, policy support documents, data security inventory and other supporting products are available (<https://kystats.ky.gov/Reports/DataRequest>). The website also appears to be a portal for provider access and data input as well.

PII is used to perform initial record linkage; the data are then deidentified and the linked records are stored in a separate system from which reports and analyses are generated.

<https://kystats.ky.gov/About/Security>

MARYLAND LONGITUDINAL DATA SYSTEM (MLDS)

Date established: 2010 (MLDS); 2013 (MLDS Center as gov't agency)

Established by law (e.g., state statute, gubernatorial executive order): Yes, for data system

Data integration approach: Fully integrated – data are housed with the agency and linked internally

Includes data from the following agencies: Maryland Department of Labor, Licensing and Regulation; Maryland Higher Education Commission; Maryland State Department of Education; Maryland Department of Transportation, Motor Vehicle Administration; collects data directly from LEAs, community colleges, and public senior higher education institutions

Scope: P-20W

Funding: State funds with additional funding from federal grants

Point of contact: mlds.center@maryland.gov

Website: <https://mldscenter.maryland.gov/Aboutus.html>

Description: The Maryland Longitudinal Data System (MLDS) was established by state law in 2010. In July of 2013, the MLDS Center began operations as an independent unit of state government. The Center is overseen by a 12-member Governing Board. The Center has fifteen full and part-time positions and a partnership with the University of Maryland, School of Social Work which provides research services and houses the Center's headquarters. Staff of the Center are also located at the Maryland State Department of Education building in Baltimore.

It has a relatively transparent leadership model and system, with a clearly identified Data Governance Advisory Board, Research and Policy Advisory Board, and separate bylaws for each body. The primary sub-entities within the MLDS Center seem to be focused on data administration, and research. There is a public research agenda set by the governing board

(<https://mldscenter.maryland.gov/ResearchAgenda.html>), a publicly available data inventory

(<https://mldscenter.maryland.gov/DataInventory.html>) and data reporting standards

(<https://mldscenter.maryland.gov/DataReportingStandards.html>), including data suppression rules, and an electronic form for data requests.

Output includes interactive data visualization dashboards, short briefs and long-form reports, associated research, and workshops and webinars that have been facilitated by MLDS Center staff.

INDIANA DATA HUB (MANAGEMENT PERFORMANCE HUB)

Date established: July 2017

Established by law (e.g., state statute, gubernatorial executive order): unknown

Data integration approach: Federated – data are housed by agency partners, linked as needed by MPH; all data ultimately reported through MPH

Includes data from the following agencies: Unspecified, but intends to serve all IN gov't agencies

Scope: K-12, higher education, workforce, Medicaid and public safety; may expand scope for specific projects; also houses gov't transparency data

Funding: unspecified

Point of contact: Darshan Shah (dashah@mph.in.gov)

Website: <https://www.in.gov/mph/index.htm>

Description: MPH provides analytics solutions tailored to address complex management and policy questions enabling improved outcomes for Hoosiers. They empower their partners to leverage data in innovative ways, facilitating data-driven decision making and data-informed policy making.

This entity takes more of a start-up approach to data access. They house a public dataset portal (IN Data Hub) while also providing a channel for access to other agency data. While they also house IN government transparency data in the Indiana Transparency Portal, there is less transparency about how the MPH is formed or run.

Evidence of project-based integration projects with diverse datasets is featured on the front page, with contributions from the State Police, public health entities, the Indiana Business Research Center, IU Public Policy Institute, the Polis Center of IUPUI (Indiana University - Purdue University Indianapolis), and other disparate entities. The foundational data stock appears to be education and workforce development, formerly referred to as a distinct grouping (EWD).

VLDS (VIRGINIA LONGITUDINAL DATA SYSTEM)

Date established: 2009

Established by law (e.g., state statute, gubernatorial executive order): “The authority to perform this complex merge was provided in language from the Appropriations Act, which was due to expire with the biennium beginning July 1, 2010. In light of this, Delegate Robert Tata of Virginia Beach carried HB 7, which provides ongoing authority for such matches and included the Virginia Employment Commission within the authority.”

Data integration approach: Federated – data are housed by agency partners, data linked and deidentified by VLDS with no access to PII by VLDS

Includes data from the following agencies: Virginia Department of Education (VDOE), the State Council of Higher Education for Virginia (SCHEV), the Virginia Employment Commission (VEC), the Virginia Department of Social Services (VDSS), the Virginia Community College System (VCCS), the Virginia Department for Aging and Rehabilitative Services (DARS), and Virginia Department of Health Professions (DHP)

Scope: Originally P16; has since expanded to include social services, additional programs

Funding: 2009 Statewide Longitudinal Data Systems Grant

Point of contact: Todd Masa (TodMasa@schev.edu)

Website: <https://vlds.virginia.gov/about-vlds>

Description: VLDS (Virginia Longitudinal Data System) is a pioneering collaboration for Virginia’s future, giving the Commonwealth an unprecedented and cost-effective mechanism for extracting, shaping and analyzing partner agency data in an environment that ensures the highest levels of privacy. VLDS is comprised of several component technologies that support secure, authorized research addressing today's top policy and state program questions. VLDS is the result of a shared effort by several Virginia government agencies. Under the IT Initiatives category, VLDS won the 2013 Governor’s Technology Award in Cross-boundary Collaboration for the first-in-the-nation collaboration among four founding state agencies.

The agency seems to center itself around shared research questions and some governance processes, but due to the federated integration model there is still a good deal of information that is decentralized. Much of the external effort seems to come from the State Council of Higher Education for Virginia (SCHEV), which comprises some of the VLDS leadership (including Todd Masa). Research guides and data dictionaries are available, as well as agency-specific research agendas and research products that resulted from VLDS data (<https://vlds.virginia.gov/insights>). They also have a public data governance book (https://vlds.virginia.gov/media/1087/vlds_book_of_dg.pdf).

There is an emphasis on data privacy above and beyond the previous examples – this value seems to have informed their selection of a federated model. (<https://vlds.virginia.gov/privacy>).

Appendix N: Center Roles and Responsibilities

	Analytics Director	Analytics Manager	Data Engineer	Business Intelligence Engineer	Data Analyst	Research Scientist
Overarching Center responsibilities	Set direction and priorities	Recruit analytics team; manage tasks; plan and manage Agile sprints.	Move data to and from systems; clean and format data.	Perform analysis using SQL; merge and clean data; build self-service tools; generate reports.	Perform analytics with Excel and SQL; create reports using self-service tools.	Build and monitor predictive models; Publish research findings
Preliminary Center FTE Targets	One	One (to manage engineers and analysts)	Two	Two per domain (for example, Education, Health, and Human Services would require 6)	One-Two per domain (for example, Education, Health, and Human Services would require 3-6)	At least 1, but dependent on research needs of Center
Typical employment period ⁶¹	4-6 years	4-5 years	3-5 years	2-3 years	1-2 years	5+ years
Other titles	Director, Data Science	BI Manager Data Manager	Software Engineer Data Warehouse Engineer	BI Engineer (BIE) Data Scientist	Business Analyst Product Analyst Operational Analyst Marketing Analyst	Statistician

⁶¹ Based on Glassdoor estimates.

Appendix O: Recruitment Challenges and Strategy

Hiring tech workers in the public sector is a recognized and ongoing challenge that appears to be increasing in intensity. The biggest factors in this shortfall are a large and growing age gap⁶² and an inability to keep pace with private sector salaries.⁶³ Job fairs and the aid of nonprofit organizations that focus on public sector hiring may help to bridge that gap,⁶⁴ but ultimately restructuring the nature of government work and requisite positions might be necessary.⁶⁵ Oklahoma specifically has been losing tech sector jobs at a high rate,⁶⁶ while also employing a larger proportion of its citizenry in state government than most other states in the country.⁶⁷

To mitigate these factors, we recommend expanding the search for talent nationwide, which will widen the talent pool as well as attract a more diverse team. Additionally, offering the option to work remotely at least for some staff, may address a variety of recruitment and staffing challenges, including workforce retention,⁶⁸ diversity,⁶⁹ and reducing costs for workers.⁷⁰ To reach a wide range of candidates, the Center should take advantage of strategic relationships in Oklahoma and beyond, including with private sector partners, educational institutions, professional associations, and state agencies.

Compensation that is competitive with the private sector for engineering roles will be among the most difficult aspects to address in recruitment. Some government agencies attempt to appeal to prospective workforce by specifically calling out the benefits unique to government employment. A 2018 blog post for GovLoop, City of Boston, notes the unique advantages of working in public service, including the opportunity to impact society and hold influence in the public sphere, stating that “candidates are drawn to public service because they want to make a positive impact on the lives of city residents”.⁷¹ A recent NY Times article cites a “Techlash” among Millennials working in tech, suggesting that some may sacrifice salary to work for an organization or company they support.⁷² A similar framing could make up for a difference in compensation between public and private sector positions.

Strategy to Recruit for Roles at the Center

A critical aspect to hiring in tech involves differentiating the quality of the candidates interviewed, particularly within analyst roles where a resume may not accurately display the skills of a candidate. For

⁶² <https://www.nextgov.com/cio-briefing/2017/12/governments-struggle-hire-young-tech-talent-worse-you-thought/144225/>

⁶³ <https://www.accenture.com/us-en/insights/public-service/job-one-reimagine-state-government-workforce>

⁶⁴ <https://www.nextgov.com/cybersecurity/2017/11/hundreds-attend-federal-cyber-and-tech-hiring-fair/142336/>

⁶⁵ <https://www2.deloitte.com/us/en/insights/industry/public-sector/future-of-work-in-government.html>

⁶⁶ https://www.cyberstates.org/pdf/CompTIA_Cyberstates_2019.pdf

⁶⁷ <https://www.usatoday.com/story/money/economy/2018/06/01/states-where-the-most-people-work-for-government/35302753/>

⁶⁸ <https://www.gsb.stanford.edu/insights/why-working-home-future-looking-technology>

⁶⁹ <https://www.entrepreneur.com/article/333010>

⁷⁰ <https://www.tecla.io/blog/2019-remote-it-workers-stats-companies-should-know/>

⁷¹ <https://yello.co/blog/the-top-3-challenges-for-government-sector-recruiting/>

⁷² <https://www.nytimes.com/2020/01/11/style/college-tech-recruiting.html>

example, a candidate may have years of "Tableau Dashboard" experience without actually having transformed and analyzed actual datasets; instead they may have simply created graphical visualizations that leverage other more in-depth work of others. Conversely, a candidate may be highly skilled at analysis but not in creating dashboards and visualizations in Tableau or another data visualization tool. Both might have similar titles but would be capable of producing very different outputs.

In today's technology hiring, it is commonplace to install some type of case study or test that identifies the strengths and weaknesses of the candidate. For example, for Software Engineering roles, companies often ask candidates to write code in the appropriate language so that a committee of peers can review and critique candidate output. For roles related to Data Analysis, a case study could involve having candidates work through a business-related problem, using fictitious data, and code solutions to get to an answer. The goal of such tests is not to judge whether the candidate is correct or not, but to gain insight into how they solve problems, and whether they can maintain focus on improving business practices and functions. Many services that specialize in hiring people in the technology sector create and offer these types of tests, and thus can save tremendous time and effort to filter out candidates.

To give an example of the strategy described above, when hiring for tech roles, 3Si seeks candidates with a rich portfolio they can highlight during an interview, but also tests how curious the candidate is when it comes to technology self-learning. The best candidates tend to show the ability to experiment with technology to produce output of value. Many cloud computing experts learn via a combination of on-the-job experience and self-study. They are often not current or former employees of top technology companies; instead they are motivated to learn technology regardless of their professional and personal circumstances. These are candidates that have a demonstrated passion for technology—an ideal quality to help implement and maintain an ECIDS.

In addition to hiring for roles, sometimes organizations can tap existing staff to fill technology positions. To identify these outliers, it is important to raise awareness about cloud-based projects through internal marketing strategies, brown-bag meetings, and brainstorming sessions, all of which provide channels for staff to self-select into a specific recruiting and hiring process. From there, recruiters can assess the individual's motivations and abilities and see if the role—perhaps with some additional training—would be a good fit. Filling roles via existing staff can be advantageous in that it can provide a roadmap to convert and better retain staff. It also represents a risk in that an existing employee who converts and does not ultimately represent a good fit can be both unproductive and difficult to let go.

Outsourcing is sometimes the easiest way to acquire talent. While it is usually more expensive per hour than performing the work in-house, it is not necessarily more costly for a given project, depending on productivity, timelines, existing capacity, etc. An advantage to outsourcing is that it provides a short-term commitment and the ability to swap in and out specialized resources to build a team. For example, the Center might consider data engineers who specialize in data transportation, or BI analysts who combine data visualization, art, and overall story presentation. Both skill sets exist within contracting pools and Oklahoma should use them as needed.

A potential disadvantage is long-term continuity, assuming the Center can retain FTEs for a longer period than an individual with an outsourced service (which is not always the case). Because it can be challenging to find a complement of individuals who together have the skills needed for a given project, we recommend that the Center remain flexible and seek a hybrid combination of FTEs and contractors, as the market dictates, to secure and retain the necessary experience and expertise to enable the success of the ECIDS long-term.

Appendix P: Preliminary Five-year Cost Estimate for Oklahoma ECIDS*

	Low Estimate					High Estimate				
Component	2020	2021	2022	2023	2024	2020	2021	2022	2023	2024
Center Expenses	\$200,781	\$1,128,776	\$1,701,750	\$1,761,652	\$1,761,653	\$235,773	\$1,306,424	\$1,982,625	\$2,047,026	\$2,047,028
Transitional Center Support	\$237,467	\$237,467	\$109,200	\$109,200	\$109,200	\$291,900	\$291,900	\$134,400	\$134,400	\$134,400
Outsourced Center Support	\$1,495,000	\$1,495,000	\$1,339,000	\$707,200	\$707,200	\$1,677,000	\$1,677,000	\$1,495,000	\$785,200	\$785,200
Azure Expenses	\$252,000	\$277,200	\$302,400	\$327,600	\$352,800	\$648,000	\$712,800	\$777,600	\$842,400	\$907,200
Data Transport, Storage and Maintenance	\$513,000	\$603,000	\$363,000	\$63,000	\$63,000	\$837,000	\$972,000	\$612,000	\$162,000	\$162,000
Legal Support	\$276,000	\$276,000	\$276,000	\$102,000	\$102,000	\$276,000	\$276,000	\$276,000	\$276,000	\$102,000
Total Budget	\$2,974,248	\$4,017,443	\$4,091,350	\$3,070,652	\$3,095,853	\$3,965,673	\$5,236,124	\$5,277,625	\$4,247,026	\$4,137,828

Appendix Q: Expense Categories and Descriptions for Oklahoma ECIDS

Expense Category	Expense Description
Center Expenses	Compensation for Director, Manager, Data Engineer, BI Analyst, Data Steward positions, etc.
Transitional Center Support	Expenses associated with hiring for the Center team, set up and administration of contracts and data use agreements, and short-term management (first two years) of operations.
Outsourced Center Support	Expenses associated with analytical product design, data engineering and analytical needs and development (once the data is stored in the cloud). Outsourced center support will augment Center capacity during the first 2-3 years of operations, and perform more advanced data engineering design and development.
Azure Expenses	Estimated costs for running an Azure cloud-based analytical platform, including cloud infrastructure costs for transport, storage, security, and tools (Azure Databricks, Data Factory, Analysis Services, etc.).
Data Transport, Storage, and Maintenance	Engineering expenses associated with migrating to and storing data in the cloud. Duties include mapping data elements, scheduling jobs (to transport data), scripting (for complex coding to support transport, error handling, etc.). Most of these expenses are frontloaded in the first three years to account for the fact that most of the data transport will be performed up front (only changes in the data will be updated incrementally thereafter, which drastically reduces data transport fees).
Legal Support	Costs include initial development period (first few years) which includes setting up agreement policies, advising for hiring and general administration, protocol and policy development, etc. Also included are ongoing costs of general board consultation and support.