Justice Reinvestment Initiative (JRI) Site-Based Awards

Arkansas Action Plan

Section I: Grantee Snapshot

In July 2015, Arkansas Governor Asa Hutchinson, with the support of the legislature, requested assistance from The Council of State Governments (CSG) Justice Center to use a Justice Reinvestment approach to assist the state with developing data-driven policy options designed to reduce the prison population, contain corrections spending, and reinvest in strategies that can reduce recidivism and increase public safety.[1]

In support of this work, the Arkansas legislature enacted Act 895 to establish the bipartisan Legislative Criminal Justice Oversight Task Force.[2] Key stakeholders worked together to develop evidence-based policies aimed at reducing the prison population, improving efficiencies in the delivery of critical programming and treatment, and increasing public safety.[3] As a result, the 2017 Criminal Justice Efficiency and Safety Act was passed, codifying the policy framework developed by the task force.[4]

Justice Reinvestment in Arkansas has proven to be a successful model, and the State has proven its commitment to the process through statutory and budgetary support to JRI efforts. However, Arkansas' criminal justice system continues to face challenges. From 2009 to 2014, the 3-year recidivism rate in Arkansas increased from almost 16%, from 41.6% to 57.4%, with most recidivism occurring in less than 18 months.[5] The capacity to evaluate the multitude of factors that influence crime and recidivism in Arkansas is crucial for effective, evidence-based resource allocation and planning.

Accordingly, the State of Arkansas has committed to increased data-sharing, integration, and collaboration, and recidivism reduction. Arkansas Governor Asa Hutchinson has identified recidivism reduction as one of two primary goals of the Restore Hope initiative in Arkansas. Restore Hope has since worked to increase cross-agency collaboration and has created a working group to focus specifically on the issues of data-sharing and integration across agencies in the justice system. The Governor's and the State's commitment on this front was further evidenced with the enactment of Act 1282 in 2015, which created a task force to determine best practices for the state to achieve an efficient system for providing open data and transparency in government. Furthermore, findings from the comprehensive assessment of the justice system by the CSG in 2016 identified, among other things, the lack of a statewide

[1] Arkansas - CSG Justice Center - CSG Justice Center
[2] Arkansas Act 895 of 2015
[3] Justice Reinvestment in Arkansas, Discussion of Policy Options and Reinvestments, 2016
[4] Act 423, codified in Ark. Code. Ann. Title 5 & Title 10
[5] Arkansas Department of Correction Recidivism Study, 2018

mechanism for data sharing between county, local, and state criminal justice agencies as a key challenge to crime prevention and recidivism reduction in the State[6].

As a result of the initiatives discussed above, two key pieces of legislation were passed in 2017, affirming that the State and the justice community, would place a special emphasis on increased interagency collaboration and data-sharing.

First, the previously mentioned Criminal Justice Efficiency and Safety Act[6] issued a legislative recommendation that a statewide database be implemented and maintained as a repository for county jail offender data, and further that local and regional detention facilities would implement and collect data on behavioral health screenings. This latter recommendation is significant in that it ties directly to the need to effectively triage and divert offenders demonstrating behaviors indicative of mental health crises to Crisis Stabilization Units (CSUs) established following Justice Reinvestment recommendations from 2015.

Second, the Arkansas General Assembly passed Act 912[7], (Exhibit A) which created the Arkansas Data and Transparency Panel and the position of Chief Data Officer for the State of Arkansas, and further issued several mandates for that position including the facilitation of open data exchange between state agencies and the implementation of data warehousing, business intelligence, and master data management solutions across state data systems. Following the enactment of Act 912, several member agencies of the justice community in the State worked with the newly appointed CDO's office to draft and execute a standing Public Safety Interagency Data Exchange Agreement in November 2017 (Exhibit B). This agreement confirms the commitment of all involved agencies to the safe open exchange of pertinent offender information for the greater public good.

a. Main Points of Contact

	Name	Title	Organization	Phone	Email
Project POC	Robert McGough	Deputy State Chief Data Officer / Chief Analytics Officer	Arkansas Division of Information Systems	501.682.4191	robert.mcgough@arkansas.gov
Financial POC	Lasonja Evans	Accounting Operations Manager	Arkansas Department of Corrections	501.529.5392	lasonja.evans@arkansas.gov

---

[6] Council of State Governments. 2016. Legislative Criminal Justice Oversight Task Force: Report and Recommendations. https://csgjusticecenter.org/wp-content/uploads/2017/01/JR-in-AR_Final-Report.pdf
[7] Codified at Ark. Code Ann. §25-4-126
https://www.arkleg.state.ar.us/Acts/FTPDocument?path=%2FACTS%2F2017R%2FPublic%2F&file=912.pdf&ddBienniumSession=2017%2F2017R&_ga=2.72427664.264961465.1655223681-1982510181.1646430358

b. Project Title
   Development of a Criminal Justice Data Hub for Interagency and Cross Sector Sharing, Notifications, and Analysis

c. JRI Partner Agencies and Organizations

   Three categories of JRI Partner Agencies and Organizations will be involved in the Arkansas JRI implementation: 1.) Core Implementation Team, 2.) Domain Expertise Advisors, and 3.) Stakeholder Representatives. The first category is comprised of individuals that will serve as the Core Implementation Team. These members, their related area of expertise, and contact information is provided below.

Core Implementation Team Members	Background/Expertise	Contact Information
Jerry Bradshaw Director **Division of Community Correction**	Jerry Bradshaw began his career with the Arkansas Department of Correction in February 1992 as a Parole Officer with the Little Rock Parole Office. He advanced to serve as a Probation/Parole Manager and then as the Deputy Director of Residential Services for Arkansas Community Correction. In that role, he oversaw the State of Arkansas' five Community Correction Centers and the Technical Violator Program. In July 2016, Mr. Bradshaw was hired as the Deputy Director of Parole and Probation Services for Arkansas Community Correction. In November 2019, the Board of Corrections named Mr. Bradshaw the Director of the Division of Community Correction. Mr. Bradshaw has a Bachelor of Arts degree in Criminal Justice/Psychology from the University of Arkansas at Little Rock.	501-837-3001 jerry.bradshaw@arkansas.gov

Robin White Information Systems Manager **Department of Corrections**	Robin White has spent the last 8 years working strictly with DOC's offender management system, first on the help desk level and now as the project manager overseeing its maintenance and development.	501-515-0269 robin.white@arkansas.gov
Melissa Gibbons **Information Systems Coordinator** **Department of Corrections**	Melissa Gibbons has spent 18 years with DOC, including 10 years as a Probation/Parole Officer prior to switching over to Research and Planning, where she specializes in probation/parole data and data analytics.	501-682-9529 melissa.ashcraft@arkansas.gov
Tabrina Bratton Quality Improvement & Program Evaluation Administrator **Department of Corrections**	Dr. Bratton has her PhD in Criminal Justice from the School of Criminal Justice and Criminology at the University of Arkansas Little Rock. She conducts applied research for the Arkansas Department of Corrections. Her work has been published in *Sociological Inquiry*, *Criminal Justice Studies*, and the *International Journal of Arts and Humanities*.	501-590-1913 tabrina.bratton@arkansas.gov
Bryan Rodgers Software Support Analyst **Department of Corrections**	Bryan Rodgers has a background in mathematics and statistics, and has 4 years experience working with corrections data and statistical analysis. He provides experience with both data linkage and analysis, as well as with DOC practices.	870-267-6252 bryan.rodgers@arkansas.gov
Robert McGough Deputy State Chief Data Officer Chief Analytics Officer **Transformation and Shared Services, Division of Information Systems**	Robert McGough is the Chief Analytics Officer for the State of Arkansas with primary responsibility for managing the Arkansas data hub and statewide longitudinal data system. Robert has over 20 years of experience in government data management and analytics, holds a Master of Science in Information Quality, and is an Information Science PhD candidate conducting research to advance the field of longitudinal data analysis.	501-682-4191 robert.mcgough@arkansas.gov

Carrie Anderson ARData Chief Policy Officer **Transformation and Shared Services, Division of Information Systems**	Carrie Anderson has 15 years of experience working in data analysis, program management, and policy development within and across state agencies. Her background includes serving as a core team member for the state's pilot Crisis Stabilization Unit (CSU) program. She holds master's degrees in public policy and education.	501-682-1453 carrie.anderson@arkansas.gov
Heather Saco ARData Program Manager **Transformation and Shared Services, Division of Information Systems**	Heather Saco has five years of experience managing regulatory and non-regulatory programs for the State of Arkansas. She provides experience coordinating resources, funding, and personnel.	501-683-6966 heather.saco@arkansas.gov

Beyond the Core Implementation Team, this project will the seek the expertise and guidance of a second tier of stakeholders categorized as Domain Expertise Advisors. These partners will provide valuable domain-specific expertise and cross agency perspectives on service delivery. The third tier of partners will be comprised of additional representatives from stakeholders that have a more general interest and role in implementation.

The state will rely on the feedback of these three diverse groups throughout the project implementation process. A primary role of these stakeholder groups will be participating in design thinking exercises to develop use cases such as data dashboards or specific notification processes that may be developed with the JRI data systems infrastructure improvement. These stakeholders are familiar with the challenges that arise from lack of data integration, inconsistent metrics, and notification deficiencies and they are eager to provide tangible ideas for data products and resources that could be developed once infrastructure challenges are addressed.

The larger domain expertise and general stakeholder groups will include representation from:
- Governor's Office
- Health and Human Services
- Workforce Services
- Higher Education, Career Pathways Initiative
- Public Safety
- Parole Board
- Office of the Courts
- Restore Hope

Section II: Description of the Problem

The criminal justice system in Arkansas faces serious challenges. From 2012 to 2015, Arkansas had the fastest growing prison population in the nation and was projected to continue to rise.[8] Corrections spending was increasing at an alarming rate, and corrections facilities were at capacity with county jails holding a growing number of inmates awaiting transfer to prison.[9]

The data inform decision-making processes across the various criminal justice agencies in Arkansas represent important information collected at different stages of the offender's trajectory through the system. At each of these stages (arrest through reentry), there may be multiple agencies interacting with an offender, and in doing so, creating, and using data related to that offender.

These interactions result in a variety of different data points, representing a variety of sometimes distinct, and sometimes redundant, pieces of information, and are captured and stored in a variety of sometimes connected, but most often independent, systems. Although there is some limited data exchange between criminal justice agencies occurring, the majority of the decision-making at the department-level takes place with little to no information gleaned from external data sources. Those decisions that do benefit from externally sourced data insights usually do so only in a timeframe limited, ad-hoc, incidental fashion.

Significant investment in proprietary agency systems has resulted in a current state wherein each agency's vested interests in their own system has made migration to a shared platform an almost intractable solution. The diverse 'ecosystem' of data sources and systems represents the Arkansas justice community's greatest potential asset and its' greatest barrier to meaningful collaboration and progress.

Similar to the disparate mechanisms of data collection and storage the measurement of performance and evaluation of crime reduction outcomes is compartmentalized and most often narrowly scoped to the particular mission of the department, agency, or entity engaged in the line of inquiry. This challenge is compounded by the fact that different, and often competing, goals and objectives both within and across components of the justice system result in distinct definitions of important metrics.

As an example, the annual measurement and reporting of recidivism for parolees, is conducted by both the Arkansas Division of Correction (ADC) and the Arkansas Division of Community Correction (ACC). It may seem intuitive to assume that two agencies, using data from the same database, guided by the same statutory definition of recidivism[10], would arrive at similar conclusions. Nuances in operational definitions, however, consistently yield divergent findings. Further, although there is some limited intermingling of data between divisions for the purposes of outcome measurement, this is done solely with the intent of describing an outcome as it relates to ADC or ACC recidivism cohorts, not assessing

---

[8] Arkansas's Justice Reinvestment Approach: Enhancing Local Mental Health Services for People in the Criminal Justice System, The Council of State Governments Justice Center

[9] Urban Institute, Justice Reinvestment Initiative, Arkansas

[10] Arkansas Code Annotated § 16-1-101 defines recidivism as "…a criminal act that results in the re-arrest, reconviction, or return to incarceration of a person with or without a new sentence during a three-year period following the person's release from custody."

the impact of one division's activities on another or other, more substantive cross-division, inter-agency, or system-wide strategic planning.

Acknowledging the opportunity for more substantive data use, the Arkansas legislature passed Act 939 in 2019 (Exhibit C). This Act outlies, as one of the duties of the Data and Transparency Panel, that the state ***shall develop a unified longitudinal system that links existing siloed agency information for education and workforce outcomes to continuously conduct a business systems assessment to: (A) Help the leaders of this state and services providers develop an improved understanding of individual outcomes resulting from education and workforce pipelines in Arkansas; (B) Identify opportunities for improvement by using real-time information; and (C) Continuously align programs and resources to the evolving economy of this state.*** [11] The goal of the JRI award to integrate criminal justice data into a statewide unified longitudinal system for continuous analysis and improvement aligns well with this recent legislative mandate.

Since the state's original grant proposal submission and with the codification of Act 939 there have been additional maturations of the data sharing landscape in Arkansas. Multiple agencies have recognized the need for cross-agency data sharing and the state has identified several new opportunities for the integration of criminal justice data:

Workforce Data Quality Initiative (WDQI)
The state has increased cross agency data collaboration through the Workforce Data Quality Initiative (WDQI). The WDQI a collaborative partnership at the federal level between the Departments of Labor (DOL) and Education (DOE). The WDQI supports the development of, or enhancements to, longitudinal administrative databases that integrate workforce and education data. The JRI team could leverage this work to consider the workforce development needs of justice involved individuals and ways to support those needs through data infrastructure and products. The implementation of the WDQI and JRI provides a unique opportunity to develop data products and/or service delivery models that could connect justice-involved individuals to workforce services that support successful reentry reducing overall recidivism.

Russellville and Ready for Life Pilot Program
Under the Arkansas "Ready for Life" initiative, the state is currently expanding capabilities to include longitudinal education and workforce data and products from pre-K through career to support more granular, mature, and actionable support for evidence-based policy, strategy, and citizen-facing data products. Through this initiative, a pilot program is underway in Russellville, Arkansas. This pilot program will leverage a network of community-based workforce development systems and employers to connect unemployed and underemployed individuals to training and career coaching. A focus of this work is on the employment, hiring, and retention success of those with prior justice involvement. Further, through the integration of criminal justice and workforce data the pilot seeks to gather employment outcome data of this

---

[11] Codified at Ark. Code Ann. §25-4-127
https://www.arkleg.state.ar.us/Acts/FTPDocument?path=%2FACTS%2F2019R%2FPublic%2F&file=936.pdf&ddBienniumSession=2019%2F2019R&_ga=2.96003103.264961465.1655223681-1982510181.1646430358

population to support and scale permanent, sustainable hiring demand and retention for people with criminal records.

Applied Analytics Training Program
The state has recently established and delivered an applied data analytics training program for developing analytic capacity within state agencies. This aligns with the JRI goal of improving criminal justice data-use infrastructure by building capacity within the state for increased interagency, interstate, and cross-sector communities of practice.

 Arkansas Health and Opportunity for Me program (ARHOME)
In 2022, the Arkansas Health and Opportunity for Me program (ARHOME) began as the new Medicaid Expansion program under the federal Affordable Care Act. The ARHOME program uses Medicaid dollars to buy private health insurance for citizens between 19 and 64 years with a household income below 138% of FPL. Beyond providing health insurance, the program also includes targeted efforts to better serve subpopulations through Life 360 HOMEs. Related to the goals of the Arkansas JRI are the Success Life360 HOMEs which provide specialized services for justice involved young adults, both those formally in custody as minors at the Division of Youth Services and those formerly incarcerated as adults.[12]

Concurrent to these recent opportunities for cross-agency data sharing, the Arkansas Division of Information Services (DIS) has gained a richer understanding and increased technical capabilities related to data integration. Recent areas of growth in knowledge and capacity include expertise in Master Data Management, entity resolution, privacy guardrails, and successful data governance structures. Through the JRI implementation, DIS will continue the work of infrastructure improvements through the lens of cross-agency efforts for successful reentry, recidivism reduction, and diversion.

---

[12] https://humanservices.arkansas.gov/divisions-shared-services/medical-services/healthcare-programs/arhome/arhome-advisory-panel-information/ , https://afmc.org/health-care-professionals/arkansas-medicaid-providers/policy-and-education/webinars/arhome-webinar-july-2021/

Section III: Program and Research Model

To address the barriers and deficits discussed above, we propose the incorporation of criminal justice data to the existing statewide data hub to create a comprehensive statewide justice information integration and analytics platform. This platform will serve as a central hub for data federation and integration, data enrichment, compliance enforcement, master data management, and data governance. The planning, implementation, and maintenance of this project will be facilitated through a collaborative effort by a project coordination team and a steering committee comprised of key stakeholders from the justice community in Arkansas, in coordination with a technical assistance provider designated by the awarding federal agency.

The proposed platform scope is based on focused research of the data needs of the Arkansas criminal justice system and other Arkansas public sector services. The approach was developed in consultation with Arkansas IT leaders, industry analysts, and successful statewide data sharing and analytics programs in other states.

The key components of the proposed platform (Figure 1) and accompanying program include the following:

- Standardized policies and agreements for data sharing and federation
- Services-oriented messaging architecture
- Metadata management (data asset inventory, data dictionaries)
- Entity resolution service and master data management
- Analytics capabilities
- Data governance
- Authenticated tiered access
- Data quality management
- Business rules engine and notification service
- Data virtualization (federated views across all participating data sources)

Figure 1. Proposed criminal justice integrations and data hub platform design
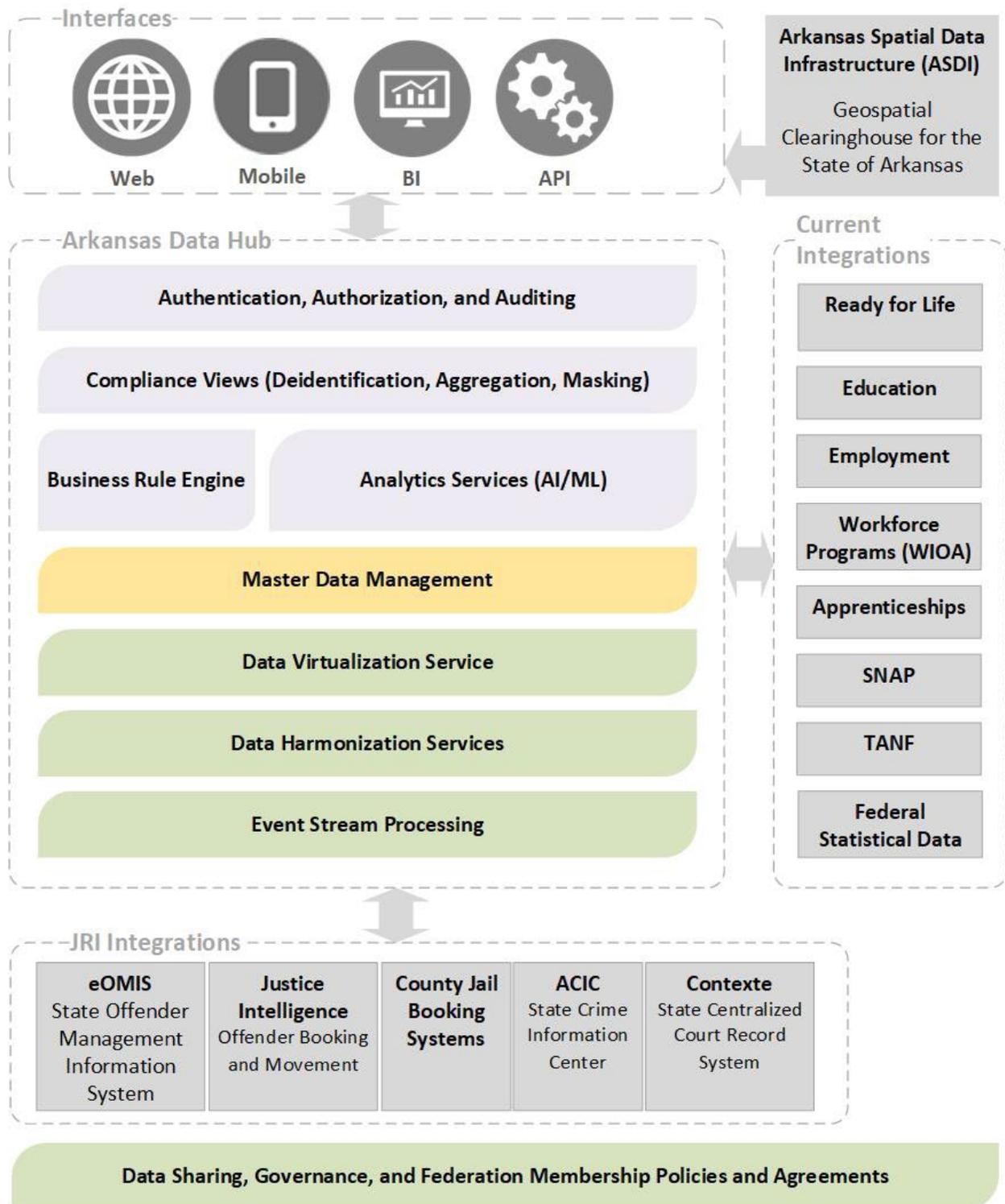
Proposed and potential applications for this platform include:

- Information Sharing. Provide shared access to comprehensive offender status and history via web application, mobile application, or web service APIs.
- Notifications. Augment the notification mechanisms currently in place to incorporate high-priority notifications for offenders with increased violence propensity and automated notifications for parole/probation officers based on caseload assignment.
- Collaboration. System-to-system workflow automation, common case notes, and user-to-user communication interface.
- Case Management. Provide a mechanism to collect and share data from community-service providers for offender outcomes (programming, treatment, support).

The proposed criminal justice data integration platform will be developed, implemented, and evaluated using the five-step JRI model:

Step 1 - Engage Stakeholders

To engage the necessary stakeholders, the proposed project will leverage the JRI Partner Agency Groups (Core Implementation Team, Domain Expertise Advisors, and Stakeholder Representatives). Representation on the Domain Expertise and Stakeholder groups will be solicited from each of the criminal justice departments, agencies, and organizations participating as stakeholders for this project. Participating entities will utilize the standard Arkansas Data Sharing Agreement. The JRI Core Implementation Team will act as a steering committee for platform development, implementation, and administration.

The Core Implementation Team will be responsible for organizing quarterly stakeholder meetings that include all the JRI partner agency groups. The partner agency groups will work together to develop the best methods for continuous communication and engagement as the groups move through subsequent steps.

Step 2 - Analyze Data and Identify Drivers

During action plan development, the steering committee and a broad group of additional community stakeholders will be involved in identifying and refining the systems and data to be integrated and the applications, notifications, and analytic models hypothesized to inform, support, or address key decision points or processes relevant to reducing crime and recidivism.

Step 3 - Develop Innovative Research-Based Responses

During the planning phase, the core implementation team will develop a business data dictionary to define key data elements, classifications, or taxonomies relevant to the shared platform. The business data dictionary will be accessible via the shared platform, and we plan to work with participating agencies to incorporate agency-specific metadata, with the intent of producing a business data translation reference. In addition to establishing important definitions for the platform, we anticipate that engaging in this process, and providing this resource, will facilitate greater collaboration and insight by building a common vocabulary for use throughout the Arkansas justice community.

Step 4 - Implement Responses

The project team will use an iterative implementation methodology in which the results of stakeholder feedback from each implementation phase will be incorporated into the planning process for the next phase.

Step 5 - Measure Outcomes

We will conduct strategy mapping sessions as part of comprehensive action planning to develop and implement outcome and performance measures in support of continuous monitoring, improvement, and data-driven performance management. The comprehensive, integrated view of interagency criminal justice data will make possible currently unavailable metrics such as re-arrest and sequential intercept diversions.

In addition, we will conduct pre and post implementation Information Quality Assessments. This will both inform our understanding of the current baseline availability of timely, accurate, and usable data from the perspective of various stakeholders and allow us to measure the outcomes of infrastructure improvement. The project team will also collect user data and surveys to assess the usefulness of data products. The team will seek to continually understand if stakeholders are satisfied with access to data and find it to be useful and actionable. Finally, while it is a challenge to make a causal connection directly from data infrastructure improvement to complex outcomes such as recidivism, the project team will consider focusing research on pilot sites and using statistical controls to better understand the impacts of data usage.

Theory of Change

Following the principles for decision-making proposed by the National Institute of Corrections we anticipate that the proposed system could improve public safety through harm reduction through the following mechanisms:

- Providing members of the justice community with access to more accurate, contextually relevant, timely information with which to make evidence-based decisions.
- Cross-agency collaboration allowing for better post incarceration service delivery.
- System-wide analytic maturity improvement through the collection, analysis, and use of integrated data.

From a more behavioral perspective, we anticipate that the ability to access and leverage a more holistic view of the offender will enable members of the justice community, specifically those tasked with supervision planning and case management, to more effectively develop individualized strategies aimed at promoting and sustaining prosocial behavior. The reentry process is one that has been described from a 'systems' perspective, or one that emphasizes the interconnected roles and responsibilities of the different agencies intersected by a given offender as they move from arrest to adjudication to incarceration to reintegration. Effective case management in reentry further requires the coordination and delivery of a variety of services and supports in the community, ranging from substance abuse treatment to employment.

Section IV: Training and Technical Assistance

We have identified potential training and technical assistance needs related to gathering and integrating stakeholder feedback to inform a Logic Model aligned with our broader Theory of Change. We could benefit from Targeted Technical Assistance that could assist with this.

With regard to development or implementation of our initiative the BJA and the JRI TTA team could assist with infrastructure and technical support through Peer-to-Peer Exchanges with other state agencies that are working on cross-agency data integration and/or the development of cross-agency notification systems. We are also specifically interested in Meetings or Subject Matter Experts, including academic researchers and practitioners, that could provide input on evidence-based policies and practices focused on intercept diversion and recidivism reduction.

Section V: Sustainability

Following an 18-month feasibility and cost study (Exhibit D) on the development of a statewide data hub, Arkansas initiated the program under the Arkansas Chief Data Officer and Data Transparency Panel. The statewide data hub currently supports education, workforce, and human services administrative data for operational and analytical use cases.

 Per direction set by Arkansas executive agencies, the legislature, and the Governor's Office, the state seeks to incorporate more data and programs into the statewide data hub, using a centralized service model for maintaining the infrastructure required to serve the needs of multiple state agencies and partners.

A central-service model establishes a standardized approach for data governance, providing all suppliers of data assurance needed to share their respective data. Adopting a cost-recovery model to replace the current contract-based funding model positions the state to support integration of additional administrative data into the statewide data hub beyond an initial 3-year grant award. This centralized service model will also streamline the process for criminal justice agencies sharing administrative data through use of the standardized Arkansas Data Sharing Agreement. This centralized model and the statewide data hub provides a supportive environment for JRI related data and data products to self-sustain beyond the initial grant. In addition, the goals of the JRI grant are well aligned with other statewide initiatives and legislative mandates. This also supports long-term sustainability of JRI supported projects.

The cumulative result of the executive initiatives and legislative activities discussed in previous sections has been to catalyze a broad, cross-agency effort aimed at data sharing, transparency, and collaboration. By leveraging the technical expertise and funding mechanisms made available through cooperative partnerships like that being offered by BJA in the JRI, we hope to capitalize on this unique moment in Arkansas' history to bring about a statewide, cultural paradigm shift in the way we use data to inform the business of state government and improve the lives of our citizens.

Section VI: Timeline of Activities

Please see attached Goals, Objectives, Deliverables, and Timelines

Section VII: Supplemental Materials Checklist

The following supplemental materials will be attached with the final submission:

☐ **Documentation of executive support and commitment of agency resources to the project.**

☐**Letters of commitment from external agencies or organizations** that are expected to participate in the project. In developing the action plan, it may be necessary to secure additional letters of commitment.

☐**Contract or memorandum of understanding with the research partner (if applicable)** that clearly delineates the role and responsibilities of the research partner, if applicable. This document should establish the authority of the research partner to access agency data, interview personnel, and monitor operations that are relevant to the evaluation of the initiative. Depending on the initiative, it also might include confidentiality agreements and how the partner will protect personally identifiable information.

Goals, Objectives, Deliverables, and Timelines

YEAR	QUARTER	GOAL STATEMENT	OBJECTIVE	DELIVERABLE
2022	Q3	Engage stakeholders	Identify and build relationships with criminal justice stakeholders	Stakeholders identified and group initiated
2022	Q3	Analyze data and identify drivers	Hold design-thinking workshops with identified stakeholder group to identify causes, symptoms, and scope of the problems	Design-thinking workshops completed, problems and responses identified
2022	Q3	Establish scope of data integration activities	Execute data sharing agreements between service provider and departments/agencies sharing administrative data	Signed Arkansas Data Sharing Agreement
2022	Q4	Provide shared access to comprehensive criminal justice data	Initiate OMIS database integration (Offender data)	OMIS data integrated into statewide data hub (Offender data)
2022	Q4	Provide shared access to comprehensive criminal justice data	Initiate OMIS database integration (Offender data)	OMIS data integrated into statewide data hub (Prison data)
2022	Q4	Provide shared access to comprehensive criminal justice data	Initiate OMIS database integration (Offender data)	OMIS data integrated into statewide data hub (Supervision data)
2023	Q1	Provide shared access to comprehensive criminal justice data	Initiate ACIC database integration	ACIC data integrated into statewide data hub
2023	Q1	Provide shared access to comprehensive criminal justice data	Initiate County Jail Booking database integration	County Jail Booking data integrated into statewide data hub
2023	Q2	Provide shared access to comprehensive criminal justice data	Initiate Justice Intelligence (Booking and Movement) integration	Justice Intelligence data integrated into statewide data hub

2023	Q3	Provide shared access to comprehensive criminal justice data	Initiate Contexte (Court Records) Integration	Contexte data integrated into statewide data hub
2023	Q4	Provide shared access	Develop user interface for self service access and analysis	User interface operational
2024	Q1	Train users and other stakeholders on use of statewide data hub	Develop data hub training course for criminal justice administrative programs	Data hub training course ready to deliver
2024	Q2	Train users and other stakeholders on use of statewide data hub	Conduct data hub training event	Training complete
2024	Q3	Document and provide report on final outcomes of JRI efforts	Craft final report	Final report completed

Stricken language would be deleted from and underlined language would be added to present law.
Act 912 of the Regular Session

1	State of Arkansas	*As Engrossed: H3/15/17*
2	91st General Assembly	A Bill
3	Regular Session, 2017	HOUSE BILL 1793

4

5   By: Representatives McCollum, S. Meeks, *Sabin*

6

For An Act To Be Entitled

8   AN ACT TO CREATE A PANEL ON DATA TRANSPARENCY; TO

9   CREATE THE POSITIONS OF CHIEF DATA OFFICER AND CHIEF

10   PRIVACY OFFICER WITHIN THE DEPARTMENT OF INFORMATION

11   SYSTEMS; AND FOR OTHER REASONS.

12

13

Subtitle

15   TO CREATE A PANEL ON DATA TRANSPARENCY;

16   AND TO CREATE THE POSITIONS OF CHIEF DATA

17   OFFICER AND CHIEF PRIVACY OFFICER WITHIN

18   THE DEPARTMENT OF INFORMATION SYSTEMS.

19

20

21   BE IT ENACTED BY THE GENERAL ASSEMBLY OF THE STATE OF ARKANSAS:

22

23   SECTION 1. Arkansas Code Title 25, Chapter 4, is amended to add an

24   additional sections to read as follows:

25   25-4-126. Chief Data Officer and Chief Privacy Officer.

26   (a)(1) The Director of the Department of Information Systems shall

27   select an individual to serve as the Chief Data Officer of the Department of

28   Information Systems and the Chief Privacy Officer of the Department of

29   Information Systems.

30   (2) The Chief Data Officer of the Department of Information

31   Systems and the Chief Privacy Officer of the Department of Information

32   Systems shall not be the same person.

33   (b) The Chief Data Officer of the Department of Information Systems

34   *shall:*

35   *(1)(A) Provide master data management by facilitating*

36   *standardization, deduplication, sharing, and integration of critical data*

1   *between systems and state agencies.*

2   *(B) Master data management may include without limitation*

3   *a shared master citizen record;*

4   *(2) Establish and promote data architecture management by*

5   *developing an integrated set of specifications and documents that define the*

6   *blueprint for managing data resources;*

7   *(3) Provide data quality management by:*

8   *(A) Applying data quality management concepts and*

9   *practices that include without limitation:*

10   *(i) Policies;*

11   *(ii) Measurement;*

12   *(iii) Process improvement; and*

13   *(iv) Education; and*

14   *(B) Leveraging data warehouse, business intelligence, and*

15   *master data management solutions;*

16   *(4) Provide data governance by:*

17   *(A) Exercising authority, control, and decision-making*

18   *over the management of data assets; and*

19   *(B) Introducing accountability for data asset management*

20   *through formalized data stewardship;*

21   *(5) Support open data exchanges based on standardized and*

22   *published application programming interfaces that:*

23   *(A) Facilitate standardized access to data within,*

24   *between, or outside systems or state agencies; and*

25   *(B) Establish a data catalog of data housed, maintained,*

26   *and utilized by each state agency;*

27   *(6) Utilize business intelligence that includes without*

28   *limitation embedded business intelligence and advanced analytics that*

29   *maximize the value of data in this state in order to facilitate access to and*

30   *the analysis of data; and*

31   *(7) Direct and oversee the Data and Transparency Panel.*

32   *(c) The Chief Privacy Officer of the Department of Information Systems*

33   *shall:*

34   *(1) Oversee, develop, and implement methods to ensure that all*

35   *state agencies comply with federal and state laws governing the privacy and*

36   *access to protected data;*

1   *(2) Assure that the use of technology sustains and does not*
2   *erode privacy protections relating to the use, collection, and disclosure of*
3   *personal information;*
4   *(3) Assure that personal information contained in privacy act*
5   *systems of records complies with the Privacy Act of 1974, 5 U.S.C. § 552(a);*
6   *(4) Review regulatory proposals and procedures involving the*
7   *collection, use, and disclosure of personal information by the federal*
8   *government, state government, and local government; and*
9   *(5)(A) Prepare and submit an annual report to the Joint*
10 *Committee on Advanced Communications and Information Technology concerning*
11 *activities that affect privacy.*
12   *(B) Activities that affect privacy may include without*
13 *limitation complaints of privacy violations, implementation of the Privacy*
14 *Act of 1974, 5 U.S.C. § 552(a), and internal controls.*
15
16   25-4-127. Data and Transparency Panel — Creation — Duties.
17   (a) The Data and Transparency Panel is created within the Department
18 of Information Systems.
19   (b) The panel shall consist of the following members:
20   (1)(A) Three (3) appointees from the private sector who shall be
21 appointed as follows:
22   (i) One (1) appointee shall be appointed by the
23 Governor;
24   (ii) One (1) appointee shall be appointed by the
25 Speaker of the House of Representatives; and
26   (iii) One (1) appointee shall be appointed by the
27 President Pro Tempore of the Senate.
28   (B) Each appointee shall serve at the pleasure of his or
29 her appointer.
30   (C) The appointer of an appointee who vacates his or her
31 position on the panel shall fill the vacancy as required under this section;
32   (2) The Attorney General or his or her designee;
33   (3) The directors, or their designees, of the following
34 *departments:*
35   *(A) The Arkansas Crime Information Center;*
36   *(B) The Department of Arkansas State Police;*

1   *(C)   The Department of Career Education;*

2   *(D)   The Department of Community Correction;*

3   *(E)   The Department of Correction;*

4   *(F)   The Department of Education;*

5   *(G)   The Department of Finance and Administration;*

6   *(H)   The Department of Health;*

7   *(I)   The Department of Higher Education;*

8   *(J)   The Department of Human Services;*

9   *(K)   The Department of Information Systems; and*

10   *(L)   The Department of Labor;*

11   (4)(A)   The Chief Data Officer of the Department of Information

12   Systems.

13   (B)   The Chief Data Officer of the Department of

14   Information Systems shall be the Chair of the Data and Transparency Panel.

15   (C)   The members of the panel shall select a vice chair

16   annually; and

17   (5)   The Chief Privacy Officer of the Department of Information

18   Systems.

19   (c)   The panel shall:

20   (1)   Perform a feasibility and cost study on the development of a

21   statewide data warehouse program;

22   (2)   Evaluate and identify data to be included in the statewide

23   data warehouse;

24   (3)   Determine and recommend procedures necessary for the

25   implementation of a statewide data warehouse;

26   (4)   Oversee a statewide data warehouse program implemented in

27   this state;

28   (5)   Evaluate and identify data that may be provided to the

29   public in accordance with data standards and specifications developed by the

30   Department of Information Systems; and

31   (6)   Provide annual reports to the Joint Committee on Advanced

32   Communications and Information Technology.

33   (d)(1)   The panel shall meet at least quarterly in each calendar year

34   at a time and place determined by the panel.

35   (2)   Special meetings may be called at the discretion of the

36   chair.

1   (e) Nine (9) members of the panel shall constitute a quorum to

2   transact the business of the panel.

3

4   25-4-128. Data and Transparency Panel — Records — Confidentiality.

5   (a) All records, reports, and other information obtained by the Data

6   and Transparency Panel shall be confidential unless approved for publication

7   in accordance with data standards and specifications developed by the

8   Department of Information Systems.

9   (b) A person, agency, or entity that furnishes confidential

10   information in good faith under this chapter is immune from criminal or civil

11   liability arising out of the release of the confidential information.

12

13

14

15   */s/McCollum*

16

17

18   *APPROVED: 04/05/2017*

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

PUBLIC SAFETY INTERAGENCY DATA EXCHANGE AGREEMENT

Agency	Arkansas Community Correction
Signatory	Sheila Sharp
Title	Director
Address	105 W. Capitol Avenue, Little Rock, AR 72201
Phone	(501) 837-6928
Email	sheila.sharp@arkansas.gov

Agency	Arkansas Department of Correction
Signatory	Wendy Kelley
Title	Director
Address	P.O. Box 8707, Pine Bluff, AR 71611
Phone	(870) 267-6200
Email	wendy.kelley@arkansas.gov

Agency	Arkansas Parole Board
Signatory	John Felts
Title	Chairman
Address	105 W. Capitol Avenue, Little Rock, AR 72201
Phone	(501) 682-3850
Email	john.felts@arkansas.gov

Agency	Arkansas Crime Information Center
Signatory	Brad Cazort
Title	Director
Address	322 S. Main St., Little Rock, AR 72201
Phone	(501) 682-2222
Email	brad.cazort@arkansas.gov

Agency	Arkansas Department of Information Systems
Signatory	Yessica Jones
Title	Director, Chief Technology Officer
Address	1 Capitol Mall, Little Rock, AR 72201
Phone	(501) 682-9990
Email	yessica.jones@arkansas.gov

I. Parties

The parties to this Public Safety Interagency Data Exchange Agreement (Agreement) are Arkansas Community Correction (ACC), Arkansas Department of Correction (ADC), Arkansas Parole Board (APB), Arkansas Crime Information Center (ACIC), and Arkansas Department of Information Systems (DIS).

II. Effective Date

The Agreement shall become effective on the date it is fully executed by ACC, ADC, APB, ACIC, and DIS. It shall remain in effect until it is specifically superseded or terminated in writing by ACC, ADC, APB, ACIC, or DIS as to its participation in the Agreement.

III. Authorities

This Agreement serves in furtherance of the responsibilities and authorities of all involved parties to engage in efficient and effective efforts aimed at promoting public safety through the incarceration, supervision, and rehabilitation of criminally-involved populations facilitated by evidence-based decision making, interagency data sharing, and cross-agency collaboration. Ark. Code Ann. § 9-28-203; Ark. Code Ann. § 10-3-3001; Ark. Code. Ann. § 12-27-103; Ark. Code Ann. § 12-27-125; and Ark. Code Ann. § 25-4-126.

IV. Purpose

The purpose of this Agreement is to establish a formal data exchange framework among ACC, ADC, ACIC, APB, and DIS with the intent of facilitating evidence-based and data-driven assessment, case management, and policy development in the interest of public safety through the sharing of historically and contextually relevant data.

V. Definitions

Data: factual information that can be digitally transmitted or processed

Party: one (1) of the agency signatories of this document

Originating Party: the party that maintains primary responsibility for the data in question at the time of transmission, access, or reporting

Receiving Party: the party that accesses or is the recipient of data transferred pursuant to the terms of the Agreement

Sensitive Data: data containing identifiable, confidential, or otherwise protected information such as Personally Identifiable Information (PII)

Non-Sensitive Data: de-identified data, aggregated or summary data, or data that is otherwise a matter of public record

VI. Data Overview

This Agreement provides for the exchange of identifiable, individual-level data pertaining to informational domains which includes, but is not limited to, the following:

- Identifiers
  - Full Name
  - Date of Birth
  - Agency Identifiers
  - State-Issued Identifiers
  - Social Security Number

- Criminal History
  - Arrests
  - Incarcerations
  - Sentencing and Disposition
  - Criminal Associations

- Demographics
  - Gender
  - Race
  - Ethnicity
  - Relatives/Associates
  - Residential History

- Education Records

- Risk/Needs Assessments

- Behavioral Records (e.g., misconduct, disciplinary violations)

- Mental Health Information

VII. Public Safety Data Committee

A Public Safety Data Committee (Committee) shall be responsible for the implementation and execution of the provisions of this Agreement and shall consist of the following members:

1. One (1) representative, from each of the parties, who has direct supervisory responsibility for the data and research activities of his/her respective agency; and

2. One (1) Data Administrator (DA), from each of the parties, who will administer the data for appropriate data transfer and sharing.

The Committee shall address all matters pertaining to this Agreement requiring resolution by decision, which matter shall be resolved by a majority vote of Committee members. The DIS representative and DA shall abstain from voting in such matters unless a majority vote cannot be reached, in which case the DIS representative may cast one (1) vote to resolve a tie vote.

VIII. Access

An access request to transfer or share data shall be initiated by the DA of the Receiving Party to the DA of the Originating Party.

The DA of the Originating Party shall compile and maintain a comprehensive list of individuals with access to the data encompassed within this Agreement. The list shall include, at a minimum, the following information:

- Name
- Title
- Agency/Organization
- Primary Place of Work
- Contact Information
- Effective Date(s) of Access
- Access Purpose

Access to sensitive data identified in this Agreement shall be limited to individuals authorized for such access by each agency's DA and only for purposes aligned with the intent of this Agreement. DAs shall not have the authority to approve access to sensitive data to individuals external to their agency.

Access to, or reporting of data covered under the terms of this Agreement, shall be governed in a manner commensurate with each originating party's existing data governance policy and practice.

IX. Confidentiality

All parties understand and agree that confidential information may be used only as permitted by applicable state and federal laws, including without limitation Ark. Code Ann. § 9-27-309; Ark. Code Ann. § 12-12-1008; and 5 U.S.C. §552(a). Receiving parties agree to establish appropriate administrative, technical, and physical safeguards to protect the confidentiality and security of the data and to prevent unauthorized use or access to the data. The safeguards shall provide a level and scope of security that is not less than the level and scope of security established by the originating party.

X. Responsibilities

The information provided pursuant to this Agreement shall not be sold or distributed to any entity not a party to this Agreement without written permission of the originating party and shall not to be used for any purpose(s) other than the purpose(s) specified in this Agreement. If a request is made to a receiving party to make available information covered under the provisions of this Agreement for which it is not the originating party, said party will refer the request to the originating party.

XI. Representations, Warranties, and Liabilities

All parties shall, to the best of their ability, ensure the accuracy and completeness of the information and data transferred or shared under the provisions of this Agreement. However, transfer or sharing of data under the provisions of this Agreement shall not constitute warranty or representation to the other parties of this Agreement regarding the data. It shall be the responsibility of the receiving agency to verify the accuracy of the data, and all data transferred pursuant to this Agreement should be independently verified by the receiving agency. An originating party accepts no liability that may arise from the use of the data by a receiving party.

XII. Severability

Nothing in this Agreement is intended to conflict with or violate state or federal laws, regulations, and policies. If a term or provision of this agreement is inconsistent with a law or authority, then that term or provision shall be invalid, but the remaining terms and provisions shall remain in full force and effect. If any provision of this agreement is unenforceable, unlawful, or void, the provision shall be deemed severable from the Agreement and shall not affect the validity and enforceability of the remaining provisions.

XIII. Amendments

Any party may request an amendment to this Agreement. The amendment shall be submitted in writing and shall take effect once it has been executed by all parties. No waiver of any provision of this Agreement shall be valid unless in writing and signed by all parties.

XIV. Administrative Provisions

Nothing in this Agreement shall be construed to obligate any of the parties to any current or future expenditure of resources in advance of the availability of funds. Nor does this Agreement obligate the parties to spend funds on any particular project or purpose when funds are available. This Agreement in no way restricts the parties from participating in similar activities or arrangements with other public or private agencies, organizations, or individuals.

XV. Termination

Termination of this Agreement may be requested by any party at any time by submitting a written request to the Committee stating the reason(s) for the request. Termination requests shall be signed, dated, and submitted by a party's Agency Director or designee. Upon receipt, the Committee shall convene within ten (10) business days to consider remediation strategies. A requesting party agrees to submit to, and in good faith engage in, remediation proceedings held by the Committee. Successful remediation shall be confirmed by the requesting party's written request to the Committee to withdraw the termination request. If the termination request is not withdrawn upon the conclusion of the remediation process, the requesting party consents to a 90-day period from the date of the conclusion of the remediation proceedings before the cessation of data transfer or sharing.

The parties shall establish procedures for timely deletion of the terminating party's data. Confirmation of the deletion of records shall be supplied by the Data Administrator of the receiving party to the terminating party according to a timeline set by the parties.

If a party violates the provisions of this Agreement, then the other parties shall reserve the right to cease data sharing activities immediately following a decision by the Committee that a violation has occurred. Any party that becomes aware of a use or disclosure of data that violates the provisions of this Agreement shall report such use or disclosure to the Committee for investigation and resolution.

XVI. Execution

In witness whereof, the parties' authorized representatives attest to and execute this agreement effective on November 10, 2017.
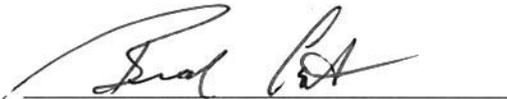
Sheila Sharp

Director
Arkansas Community Correction

Wendy Kelley by Rhonda Patterson

Wendy Kelley
Director
Arkansas Department of Correction

John Felts by

John Felts
Chairman
Arkansas Parole Board

Brad Cazort
Director
Arkansas Crime Information Center

Yessica Jones
Director, Chief Technology Officer
Arkansas Department of Information Systems

Exhibit C

Stricken language would be deleted from and underlined language would be added to present law.
Act 936 of the Regular Session

1	State of Arkansas	*As Engrossed: S3/26/19*
2	92nd General Assembly	A Bill
3	Regular Session, 2019	SENATE BILL 589
4		
5	By: Senator J. English	

6

For An Act To Be Entitled

8 AN ACT TO AMEND THE LAW CONCERNING THE DUTIES OF THE
9 DATA AND TRANSPARENCY PANEL; AND FOR OTHER PURPOSES.

10

11

Subtitle

13 TO AMEND THE LAW CONCERNING THE DUTIES OF
14 THE DATA AND TRANSPARENCY PANEL.

15

16

17 BE IT ENACTED BY THE GENERAL ASSEMBLY OF THE STATE OF ARKANSAS:

18

19   *SECTION 1. Arkansas Code § 25-4-127(b)(3), concerning the creation and*
20 *duties of the Data and Transparency Panel, is amended to read as follows:*
21   *(3) The directors, or their designees, of the following*
22 *departments* *and offices:*
23   *(A) The Arkansas Crime Information Center;*
24   *(B) The Department of Arkansas State Police;*
25   *(C) The Department of Career Education;*
26   *(D) The Department of Community Correction;*
27   *(E) The Department of Correction;*
28   *(F) The Department of Education;*
29   *(G) The Department of Finance and Administration;*
30   *(H) The Department of Health;*
31   *(I) The Department of Higher Education;*
32   *(J) The Department of Human Services;*
33   *(K) The Department of Information Systems;* ~~*and*~~
34   *(L) The Department of Labor;*
35   *(M) The Department of Workforce Services; and*
36   *(N) The Office of Skills Development;*

1
2   *SECTION 2. Arkansas Code § 25-4-127(c), concerning the creation and*
3   *duties of the Data and Transparency Panel, is amended to read as follows:*
4   (c) The panel shall:
5   (1) Perform a feasibility and cost study on the development of a
6   statewide data warehouse program;
7   (2) Evaluate and identify data to be included in the statewide
8   data warehouse program;
9   (3) Determine and recommend procedures necessary for the
10 implementation of a statewide data warehouse program;
11   (4) Oversee a statewide data warehouse program implemented in
12 this state;
13   (5) Evaluate and identify data that may be provided to the
14 public in accordance with data standards and specifications developed by the
15 Department of Information Systems; ~~and~~
16   (6) Provide annual reports to the Joint Committee on Advanced
17 Communications and Information Technology; and
18   (7) Develop a unified longitudinal system that links existing
19 siloed agency information for education and workforce outcomes to
20 continuously conduct a business systems assessment to:
21   (A) Help the leaders of this state and service providers
22 develop an improved understanding of individual outcomes resulting from
23 education and workforce pipelines in Arkansas;
24   (B) Identify opportunities for improvement by using real-
25 time information; and
26   (C) Continuously align programs and resources to the
27 evolving economy of this state.
28
29
30   */s/J. English*
31
32
33   **APPROVED: 4/12/19**
34
35
36

Feasibility and Cost Study on the Development of a Statewide Data Warehouse Program

Findings and Recommendations Report

Prepared by the Arkansas Data and Transparency Panel

Table of Contents

Executive Summary:

The purpose of this study is to evaluate the feasibility and costs of the development of a statewide data warehouse program, its ability to address identified gaps and objectives, and to make recommendations for an optimal implementation approach based on existing public sector programs and industry best practices.

In Act 1282 of 2015, the general assembly identified needs for increased data access, use, and management and called for an Open Data and Transparency Task force to study these findings and make recommendations. An Arkansas Legislative Audit special report published in 2015 identified the potential benefits of a centralized data warehousing approach in addressing some of the findings identified. In January 2017, the Open Data and Transparency Task force recommended a cost and feasibility study for a statewide data warehouse program, which was codified by Act 912. A gap analysis with executive agency leadership reinforced data integration, data sharing, and data access as top needs, with a focus on a citizen master record, and subsequent 2019 legislation emphasized the need for real-time data access for continuous program alignment with evolving needs.

The identified needs and recommendations were synthesized into six program objectives aimed at establishing and supporting the accessibility, sharing, integration, quality, timeliness, and analysis of state data assets. The objectives identified are:

- **Enable openness, transparency, and pervasive, self-service data access and delivery**
- **Share data to enhance its value while enforcing privacy and security**
- **Support data-driven decision making and analytic maturity through development and support of analytic skills and shared services**
- **Integrate data for improved cross-agency analysis and reduced duplication of data and efforts**
- **Enable real-time assessment, support, alignment, and automation of decisions, programs, and resources**
- **Manage enterprise data as a state asset.**

As part of this study, research was conducted of existing data warehouse, analytics, and information management programs in multiple US states, 1 Australian state, and a Canadian federal agency. Industry analyst interactions and a study of published research were leveraged to assess current best practices from both public and private sector. The study found many successful examples using a variety of different approaches that resulted in cost savings, operational efficiencies, increased program performance, and effective policy insights.

Feasibility of addressing the identified objectives was found to be high with key recommended approaches including:

- **Pervasive Business Intelligence, web, mobile, and Application Programming Interfaces for data access**
- **Compliant, secure data sharing, via a layered approach of authorized role and organization-based access to identified data and broader access to deidentified aggregate data**
- **Analytic tools and training for increased value from existing data and more data-driven decision making**
- **Integration of cross-agency data via a Data Hub approach leveraging centralized Master Data Management**
- **Real-time data access supported as needed via data federation**
- **Self-service, agency-level Data Quality Management, Master Data Management, and Stewardship at the source**
- **Comprehensive Data Governance for standards, security, privacy, compliance, and change management**

The annual cost to meet Arkansas' identified needs is estimated at $4M. The cost of other statewide data programs is over $8-9M annually with larger efforts having up to 30 support staff. This is a significant investment, but other states report returns from statewide data warehousing programs ranging from $25M per year to $1M per day. Other programs use a combination of funding strategies including federal and private grants, general revenue appropriations, and usage-based chargeback.

The key recommended next steps include:

- **Develop a program charter to formalize the scope of the program**
- **Formalize a multi-department agreement to allow departments for compliant, secure and efficient data sharing**
- **Determination of initial and sustaining program funding approaches**
- **Enablement of a data hub (value-driven broker for cross-agency data sharing and analytics) to act as an agent of individual agencies in integrating and providing secure, compliant access to and analysis of inter-agency data**

Background

Act 1282 of 2015 outlined findings related to state data and formed the Open Data and Transparency Task Force (ODTF) to determine the best practices for maintaining and delivering state data. The task force met throughout calendar year 2016 and published their findings and recommendations in January 2017. Act 912 of 2017 included many of the task force's recommendations. Two key provisions of this act were the creation of the position of Chief Data Officer (CDO) and the completion of a feasibility and cost study on the development of a statewide data warehouse program. The purpose of this study is to **evaluate the feasibility and cost** of implementing a statewide data warehouse program and to **recommend best practices and approaches** for addressing the identified state data needs.

Methodology

The approach taken to conduct this study included:

- Analysis of identified Arkansas data needs and opportunities as outlined by the general assembly, Open Data and Transparency Task Force, and executive agency leadership
- Synthesis of related identified needs into program objectives
- Study of current data management and analytic best practices from research and advisory firm Gartner and data management professional organizations via analyst inquiries and analysis of published research.
- Study of other public sector entities[1] who have implemented centralized data warehouse programs or other approaches to the delivery and support of integrated data management and analytics
  - Analysis of published literature and presentations
  - Interviews with public sector chief data officers and analytic program leadership
  - Analysis of budget data from state transparency portals
- Application of successful best practice approaches to program objectives and feasibility assessment
- Cost and benefit estimation of implementing the identified approaches
- Recommendation of next steps towards program implementation

Data Needs → Program Objectives → Feasibility Study → Approaches → Costs → Next Steps

Figure 1 - Study Methodology

---

[1]Alabama, Connecticut, Florida, Indiana, Iowa, Kentucky, Massachusetts, Michigan, Minnesota, Mississippi, New Jersey, New South Wales Australia, North Carolina, Tennessee, Texas, Treasury Board of Canada

Identified Data Needs

Act 1282 of 2015

The General Assembly identified the following data-related findings in section 1.b.1 of Act 1282 of 2015 (To create the Open Data and Transparency Task Force to determine the best practices for the state to achieve the most efficient system for maintaining and delivering the state's public records and data.):

(A) State agencies contain great amounts of valuable information and reports on all aspects of life for the citizens of this state, including without limitation health, business, public safety, labor, and transportation data;

(B) The tremendous amount of data maintained by state agencies can result in **duplication of efforts, data, records, and parts of data and records** that may result in the maintenance of **inconsistent data and records concerning the same citizen**;

(C) The **lack of a quick and efficient delivery system** to respond to legislative and executive branch inquiries is harmful to the policy-making process and ultimately costs taxpayers money;

(D) Progressive states have evolved to **become data-driven governments that use data as a strategic asset** to improve the delivery of services to the state's citizens and to **become more efficient stewards of citizens' data**;

(E) **Ensuring the quality and consistency of public data** is essential to maintaining the data's value and utility;

(F) New information technology has fundamentally changed the way people search for and expect to find information and can aggregate large quantities of data to allow the state to **provide better information to citizens** with increasing efficiency and thoroughness; and

(G) The state should:

    (i) Evaluate ways to appropriately, efficiently, and securely **share data between and within state agencies** to **allow for quicker, more impactful cross-agency analysis** to allow policymakers to make quicker, more informed decisions; and

    (ii) Use the innovations in information technology to **enhance public access to public data** to make the state more transparent and to promote public trust while **eliminating waste, fraud, and abuse in the execution and delivery of government services**.

Arkansas Legislative Audit Special Report on Centralized Data Warehousing

A special report published by Arkansas Legislative Audit (ALA) in November 2015 on "Potential Benefits of a Centralized Data Warehouse for the State of Arkansas" studied other states, counties, and municipalities that are leveraging centralized data warehouses to reduce tax and program fraud, conserve state dollars, enhance public safety efforts, and improve health care. The identified benefits of centralized data management and access include:

- **Appropriate and authorized access** to large data sets for reporting and analytics.
- **Improved quality and accuracy of data**.
- **Sharing of data** among state and local entities.
- Greater efficiency through **reduction of duplicate efforts**.

The study made the following recommendations for steps towards implementation of a statewide data warehouse:

- Regarding a statewide centralized data warehouse, ALA staff recommend that the General Assembly consider:
  - **Authorizing a feasibility study** identifying the IT requirements and costs associated with centralized data warehousing.
  - Creating legislation **authorizing a Chief Data Officer** to lead the State's research into and potential development and implementation of a centralized data warehouse project.
- Should the feasibility study conclude that a centralized data warehouse would be beneficial to the State, ALA staff recommend that, during the development and implementation process:
  - Access to the centralized data warehouse is controlled
  - Secure transmission and storage of data are ensured.
  - Current facilities and other resources available are used.

Open Data and Transparency Task Force and Act 912 of 2017

Per Act 1282 of 2015, the Open Data and Transparency Task Force (ODTF) met throughout calendar year 2016 to study best practices related to the identified findings. Based on recommendations from the ALA special report and other interviews, testimony, and analysis, the ODTF determined that:

"A feasibility and cost study should be performed to **determine the specific requirements needed for a statewide data warehouse solution**. At a minimum, the hardware, software, physical location, staffing and communication networking requirements should be considered and evaluated. A cost/benefit analysis should also be performed."

"The results of the feasibility and cost study should be provided to the General Assembly so that the objectives, magnitude, and scope of the data warehouse program may be revised as appropriate. At this time the General Assembly may take action on appropriating any necessary funds for the remainder of the program."

Along with the creation of the positions of Chief Data Officer and Chief Privacy Officer and the formation of the Data and Transparency Panel (DTP), this requirement was codified as A.C.A. § 25-4-127.c.1 by Act 912 of 2017.

CDO Gap Analysis of Agency Data Needs

Upon formation in September 2017, the office of the Chief Data Officer (CDO) conducted a gap analysis of data and data management capabilities needing improvement. Interviews were held with the director and executive staff of each executive agency resulting in an inventory of identified gaps/needs. Analysis of the identified gaps identified the following core data needs:

Data Need	Percentage of Responses
Data Sharing	48.9%
Data Integration	26.7%
Data Access	6.7%
Data Quality	6.7%
Data Standardization	6.7%
Data Governance	4.4%

Figure 2 - Gap Analysis Identified Data Needs

The CDO team recommended undertaking three major initiatives to address the gaps identified in the analysis and to increase the overall data management capability and maturity across all agencies. These initiatives include:

- Creation of a Comprehensive **Multi-Agency Data Sharing Agreement**
- Implementation of a **Master Data Management** (MDM) System and **Master Citizen Record**
  - A central MDM hub ensures consistent identification
  - A central MDM hub affords more accurate identification
  - A central MDM Hub saves time and effort in data integration
- Adoption of a **Data Governance Model** and **Data Quality Management Standards**
  - Creating a data catalog
  - Creating a data asset inventory
  - Establish data governance framework
  - Establish data quality management program

Act 936 of 2019

During the 2019 legislative session, Act 936 amended the law concerning the duties of the data and transparency panel to add additional member agencies and to add the following duties:

(7) Develop a unified longitudinal system that **links existing siloed agency information** for education and workforce outcomes to **continuously conduct a business systems assessment** to:
   (A) Help the leaders of this state and service providers **develop an improved understanding of individual outcomes** resulting from education and workforce pipelines in Arkansas;
   (B) **Identify opportunities for improvement by using real-time information**; and
   (C) **Continuously align programs and resources** to the evolving economy of this state.

Act 943 of 2019

Act 943 of 2019 (To create the data sharing and data-driven decision-making task force) reiterated the findings listed in Act 1282 of 2015, noted that the ODTF began to address these problems through creation of the CDO, CPO, and DTP, and found that:

(6) The state should continue those efforts by **evaluating ways to implement a shared services model for statewide data sharing** in order to drive innovation and facilitate efficiency across state agencies, improve the delivery of services, and to better serve the citizens of this state.

To this end, the act created the **Data-Sharing and Data-Driven Decision-Making Task Force.** The task force shall meet between July 1, 2019 and December 31, 2019 and file a written report of its activities, findings, and recommendations.

The task force shall:

(B) Recommend specific solutions and legislation necessary to **create a statewide data sharing system** for maintaining and sharing public data that is owned, controlled, collected, or maintained by a state agency; and
(H) Recommend funding mechanisms to support the use of statewide data sharing, including without limitation **data analytics, machine learning**, and **innovative technologies to link data between agencies, to support data driven decision making** for all state agencies.
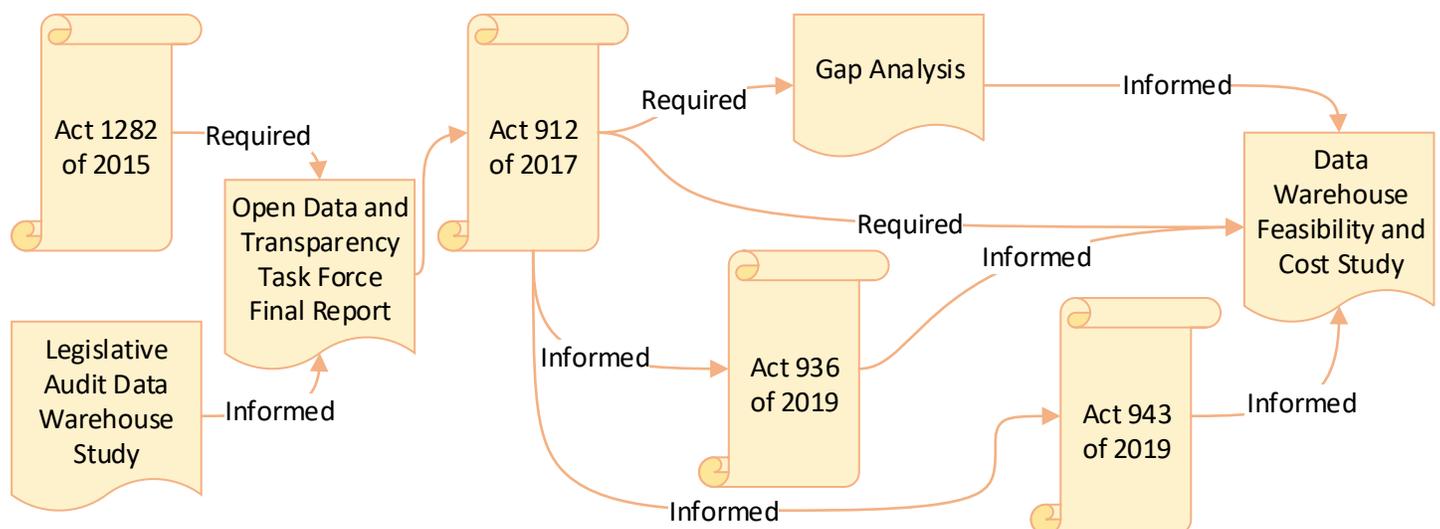

Figure 3 - Legislation and Studies Requiring or Informing the Data Warehouse Study

Program Objectives

The identified needs were synthesized into six program objectives aimed at establishing and supporting the accessibility, sharing, integration, quality, timeliness, and analysis of state data assets.

Identified Requirement / Need	Category	Program Objective
Improve data access and delivery (Act 1282 of 2015, Gap Analysis)	**Accessibility**	**Enable openness, transparency, and pervasive, self-service data access and delivery**
Enhance accessibility and quality of public data (Act 1282 of 2015)		
Improve data sharing between and within state agencies (Act 1282 of 2015)(Gap Analysis)	**Sharing**	**Share data to enhance its value while enforcing privacy and security**
The state should continue those efforts by evaluating ways to implement a shared services model for statewide data sharing (Act 943 of 2019)		
Create a statewide data sharing system for maintaining and sharing public data that is owned, controlled, collected, or maintained by a state agency (Act 943 of 2019)		
Become more data driven (Act 1282 of 2015)	**Analysis**	**Support data-driven decision making and analytic maturity through development and support of analytic skills and shared services**
Data analytics, machine learning, and innovative technologies to link data between agencies, to support data driven decision making (Act 943 of 2019)		
Improved capacity for cross-agency analysis (Act 1282 of 2015)	**Integration**	**Integrate data for improved cross-agency analysis and reduced duplication of data and efforts**
Reduce duplication of efforts and data (Act 1282 of 2015)		
Inconsistent data and records concerning the same citizen (Act 1282 of 2015)		
Eliminate waste, fraud, and abuse in the execution and delivery of government services (Act 1282 of 2015)		
Links existing siloed agency information (Act 936 of 2019)		
Implementation of a Master Data Management System and Master Citizen Record (Gap Analysis)		
Continuously conduct a business systems assessment (Act 936 of 2019)	**Timeliness**	**Enable real-time assessment, support, alignment, and automation of decisions, programs, and resources**
Identify opportunities for improvement by using real-time information (Act 936 of 2019)		
Continuously align programs and resources (Act 936 of 2019)		
Improve data stewardship (Act 1282 of 2015, Gap Analysis)	**Quality**	**Manage enterprise data as a state asset**
Ensure data quality (Act 1282 of 2015)		
Adoption of a Data Governance Model and Data Quality Management Standards (Gap Analysis)		

These objectives align with the top three business expectations identified by respondents in the Gartner 2017 Third Annual Chief Data Officer survey which included "**enhance data quality, reliability and access**", "**enhance analytical decision making**" and "**create internal and/or operational efficiencies.**"

Data Warehouse Program Approaches

Act 912 of 2017 called for a feasibility and cost study on the development of a **statewide data warehouse program**.

In defining a data warehouse program, it is important to note that:

A) The **definition of and approaches** used to deliver the intended benefits of data warehousing **have evolved over time** and **continue to mature** rapidly in response to evolving needs such as new data types, new processing requirements, data science, and faster development cycles as well as the advent of new enabling technologies.

B) As noted in the Open Data and Transparency Task Force Final Report, **data warehousing alone can only partially address the findings** identified in Act 1282. In order to fully address all findings, **additional complementary solution components** such as business intelligence, data quality management, and master data management **are required**. These are common components of mature data warehousing or enterprise information management programs and were **considered as part of the scope of a statewide data warehouse program**.

The following is a brief summary of the common styles of data warehousing currently in use.

Traditional Data Warehouse
A traditional repository-style data warehouse, or enterprise data warehouse (EDW), is a **collection of data** in which **disparate data sources** can be brought together in an **integrated, time-variant** information management strategy.

This style of data warehouse generally:

*   Houses **well-known and structured data** in a **persistent repository** of data that is copied from source systems
*   Supports high performance access for **well-known, predefined and repeatable analytics needs**
*   Contains data that is **highly modeled** during the development process
*   Transforms data during **batch extract, transform, and load (ETL) processes**, typically nightly
*   Supports **retention of historical data** that may not be retained by the source systems
*   May support smaller, subject-oriented datamarts for specific analytic needs

There are numerous successful examples of traditional repository-style data warehouses that support program or agency-level needs currently deployed across Arkansas state government, some since the 1990's.
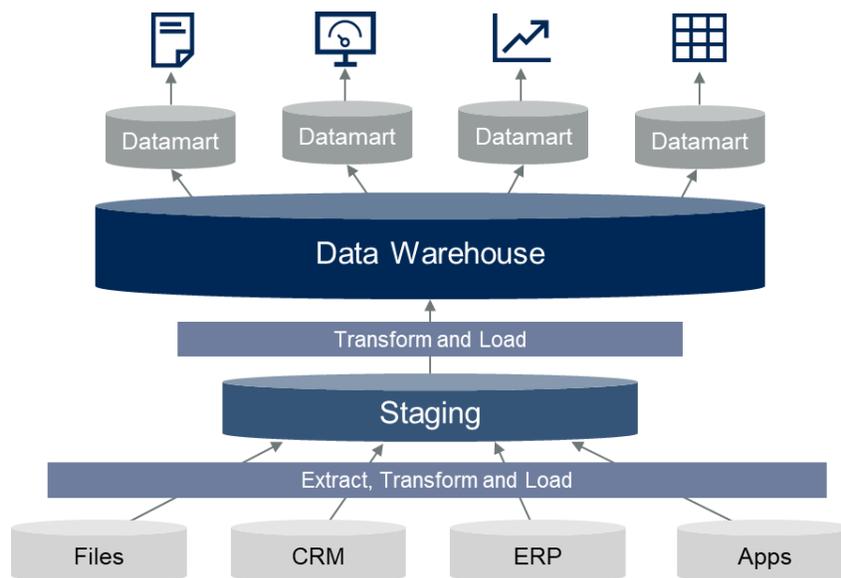
Figure 4 - Traditional Repository-Style Data Warehouse (Source: Gartner)

Data Lake

The concept of the Data Lake emerged in the mid-2000's to **supplement the Traditional Data Warehouse** in order to **support new data types, new processing requirements**, and the **need for development agility**.

A data lake is a **collection of generally raw data and events** that have been captured and ingested for use with **limited structure, transformation** to a specific schema, **or quality assurance**. Data lakes usually support **exploratory analysis, rapid prototyping, and data science activities**, often with data of **unknown quality and utility**. The data lake generally supports **minimal controls for data governance** and is typically intended more for use by data analytics professionals than by business users.

The data lake differs from the data warehouse primarily in that the data isn't modeled prior to use. This has the benefit of reducing the time to transform and ingest the data, which increases agility and reduces time wasted on data that is found to be of limited value or that was required for one-time use. Because the data isn't pre-modeled for specific known uses, it has more flexibility. However, this requires the data to be modeled by the user upon retrieval, requiring more time and technical capability.

Data lakes have gotten a poor reputation as "data swamps" because of the lack of quality controls, but this stems from a frequent misconception of the intended purpose of the data lake. The data lake's goal is to exploit new, unknown data sources. The Arkansas Department of Information Systems (DIS) leverages the data lake as a means for rapid ingestion and use of new data in conjunction with the stable, governed data warehouse environment. If the value, utility, and quality of data are established, they are promoted into the more governed environment where they are more intentionally modeled, placed under stewardship, and made available for widespread use.

Data lakes are sometimes misunderstood as a replacement for data warehouses, but **modern data and analytics initiatives require the capabilities of both types**. The data warehouse is optimized for performance-oriented, scalable, repeatable delivery on consistent data. The data lake is optimized for rapid access, free-form discovery, and flexibility. These two optimization goals are at odds, so these environments are complementary, not interchangeable. The evolving form of the logical data warehouse encompasses both concepts.
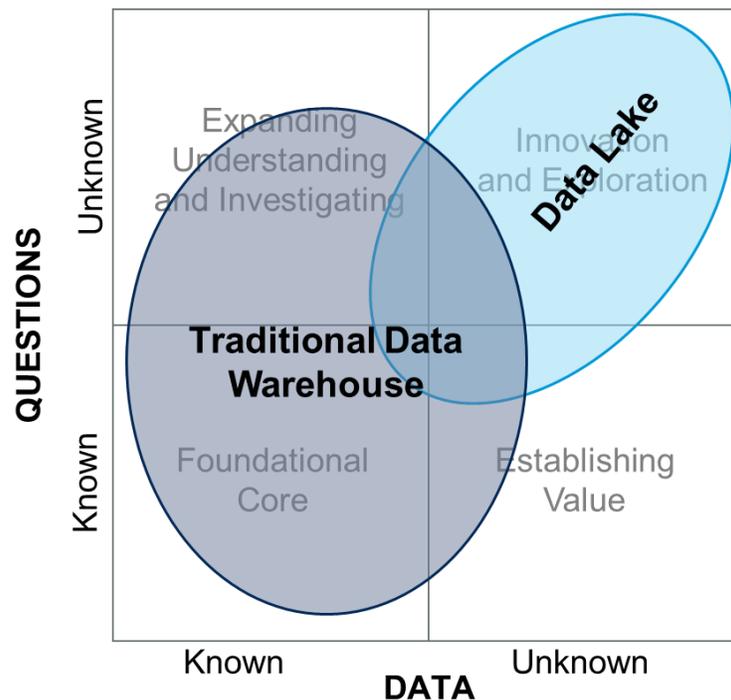

Figure 5 - Data Lake Relationship to Data Warehouse (Source: Gartner)

Logical Data Warehouse

In the early 2010's, the Logical Data Warehouse (LDW) brought the data warehouse and data lake together into a **logically consolidated view across all types of data**. It also introduced **Data Federation as the primary interface** for data access and analysis over multiple data sources.

Data federation is where the data stored in multiple sources (of the same or different type) are made accessible to data consumers by using **on-demand data integration**, rather than executing data movement and physically storing integrated data. Data federation acts as a **real time virtual repository** that hides the underlying complexity of the data landscape from consumers so that changes can be made in a nondisruptive way behind the scenes of the data federation tier.

The logical data warehouse **prioritizes data connection over data collection** but **still includes the data warehouse** as a system of compromise for meeting performance or historical retention needs, data lakes for agility, and possibly other forms of data sources and stores such as event processing or streaming data.

The Arkansas Department of Information Systems adopted the Logical Data Warehouse architecture in 2013 and has experienced many benefits including faster development cycles (and lower accompanying development costs), support for real-time reporting, lower storage costs, and reduced impact from changes to the underlying data sources. Some critical components required to enable maturation to logical data warehousing included the establishment of agency-level master data management and the accompanying stewardship and governance over naming and coding of data, the inclusion of a data federation layer in the agency's business intelligence platform, and support for business intelligence semantic layers that can span multiple distributed data sources of different types. These topics are covered in more depth later in this study. While the agency uses a federate-first strategy to decision support, traditional data warehouse and data lake approaches are still key components as needed for performance, historical retention, and agility.
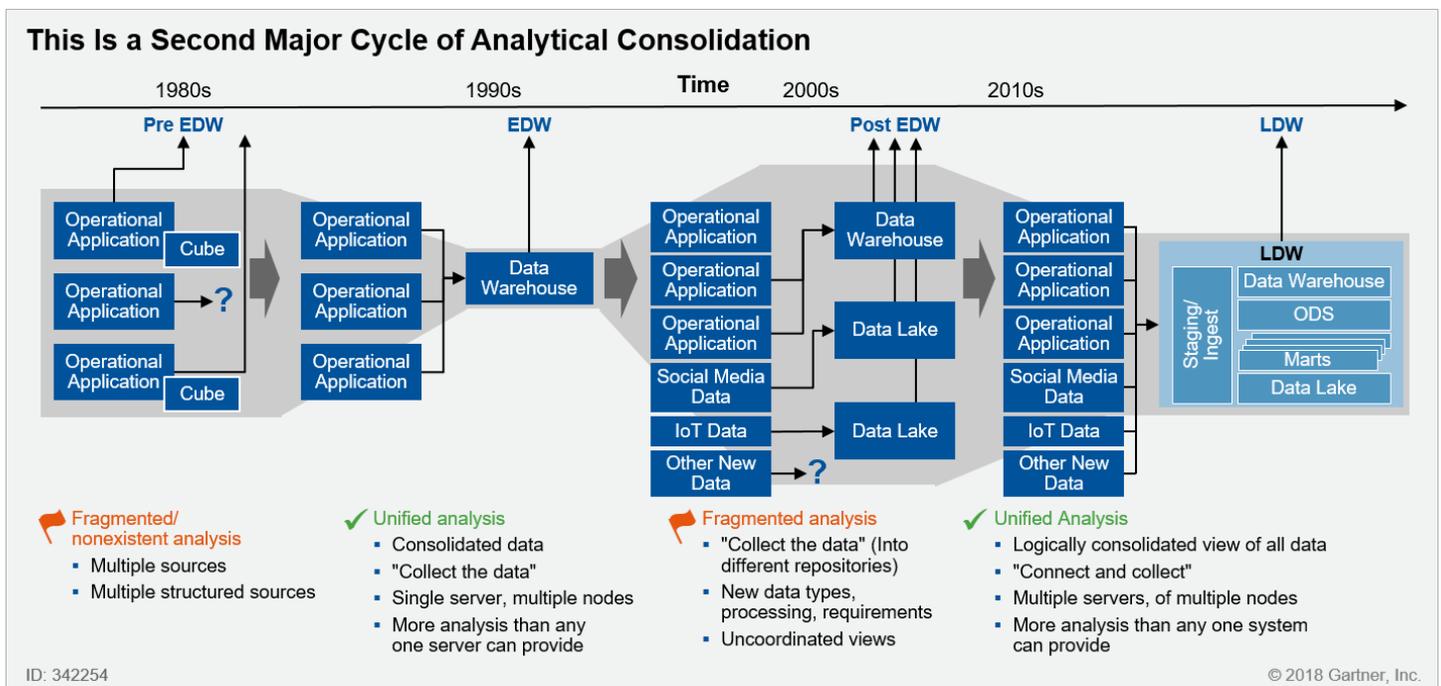

Figure 6 - Evolution of Logical Data Warehouse

Data Hub

The most recent evolution of approaches for serving data sharing and access needs is the Data Hub, an architectural pattern for enabling the **seamless flow and governance** of data across a range of business systems**, including applications, data warehouses and data lakes**.

Producers and consumers of data connect with each other via the data hub, with governance controls and standard models applied to enable effective data sharing. Predicated on the concept of common models, data hubs are mainly focused on driving consistent semantics (consistent naming and meaning) but can support a range of use cases (both operational and analytical) and processing strategies.

Data hubs are about sharing of data with effective governance. By having endpoints (systems, processes and organizations) that connect to the hub as either providers or consumers of data, the hub becomes a point of mediation and a place to apply governance controls. The data hub enables better scalability and manageability of data flows, because it provides transparency and independence between data producers and consumers.

Many organizations have identified a need to share data faster, but they also want to maintain governance guardrails. This is where a hub-centric approach comes in. In this context, the data hub acts as a consistent arbiter between producers and consumers of data, providing consistent semantics and governance policy across operational use cases.
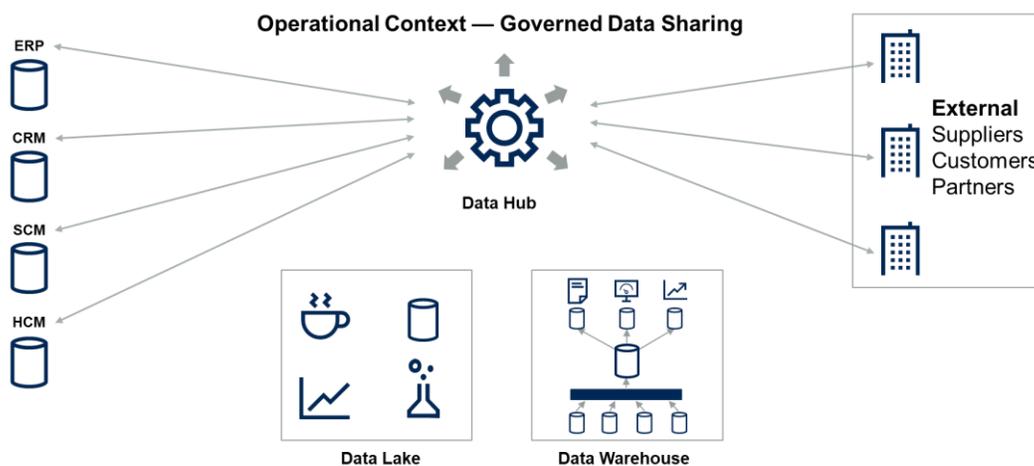

Figure 7 - Data Hub (Source: Gartner)

Recommended Approach

In order to support all identified program objectives, it is recommended to use the current "data hub" approach, which includes the use of more traditional approaches where applicable.

- Use the data hub's capabilities for enabling seamless sharing of all types of data.
- Use the data hub's capabilities around consistent semantics to feed both data warehouses and the data lake.
- Use a data warehouse as needed to collect and organize data for known questions and data.
- Use a data lake for unknown, less-structured data, unclear questions and discovery-oriented analysis.

Including a combination of interconnected data warehouse, data lake and data hub capabilities will enable:

- A wider range of use cases (both analytical and operational)
- A greater reach in connecting data assets across the state
- A more flexible approach to governing and provisioning data

12

Key Case Studies

State of Indiana

One of the most well publicized and successful statewide data warehousing, sharing, and analytics programs is Indiana's Management Performance Hub (MPH).

History

The MPH is an advanced data analytics management system for the State of Indiana government that performs business analytics and in-depth case studies on specific problems. MPH also houses a public transparency site enabling citizens to access a variety of real-time information about state government.

The MPH initiative began with the kick-off of an infant mortality study in August 2013 and was followed with an executive order by Governor Mike Pence in March 2014 requiring interagency data sharing with the MPH. The MPH website was launched in June 2014, and a physical MPH center opened in September 2014. In 2017, House bill 1470 codified and funded MPH as a state agency.

Approach

Over 5 trillion rows of data are received from agencies via flat file and loaded into a 3TB in-memory data platform. Access to detailed data is highly controlled in the most secure system in the state. This system is used for data integration and for data science tasks across the full breadth of state data. Analysts who work directly with the privacy data have all undergone background checks and work in a secure room. There are also strict agreements in place on how the data can be used in the form of MOUs. A separate instance houses anonymized data for broader agency and public consumption. Much of the anonymized data is available for public access on the Indiana Data Hub.

The implementation of in memory database technologies has greatly improved query speeds (1000x faster than SQL for simple queries, 5000x faster than SQL for complex queries) and data compression (90+% compression).

Results

MPH has conducted studies to guide policy and increase transparency in areas including recidivism, income tax fraud, highway crash prediction, Medicaid optimization, education and workforce, and the opioid epidemic.

MPH automatically calculates real time performance measures for all executive agencies.

Now that data is accessible in one location and can be pulled quickly, analysts can spend more time examining data.

Cost

MPH currently has 20 full time employees and an annual budget of $9M consisting of $6M in general funds, $1.7M in database management funds, and $1.3M from the Department of Insurance fund.

Lessons Learned

Don't try to include all state data at once. Take a more phased approach. Start with use cases that make a difference.

Treasury Board of Canada

The Treasury Board of Canada Secretariat has had great results by starting with widespread access to data.

History

The Treasury Board of Canada Secretariat (TBS) is a Federal agency that manages 250K employees across 86 departments and an annual budget of $250B CAD. The agency's strategic outcome is that government is well managed and accountable, and that resources are allocated to achieve results. Faced with 100's of silos, little data sharing, and inconsistent reporting, TBS developed and began implementing a phased, multi-year data sharing and access plan.

Approach

TBS started by conducting analysis and research on leading public sector examples and showing "Art of the Possible" examples to executive stakeholders. They formed a strategy in 2013 to consolidate data silos, encourage and enable data sharing, and establish common data elements. In early 2014, TBS visited Washington, DC to visit with the Recovery Accountability and Transparency Board for a deep dive into the Recovery.Gov implementation to learn from a successful example.

Software was purchased in March 2014 and an enterprise data warehouse was developed, starting with mature, standardized data sets. Data was provisioned from multiple sources and formats and included direct connection to source data for operational reports. Each business unit has data asset owners with stewardship over their own data in the data warehouse. In Fall 2014, the Central Online Reporting System (CORS) was launched to early adopter clients at first and then rolled out to 250K users starting with ad-hoc and predefined Business Intelligence reporting.

New capabilities have been rolled out yearly including visual data discovery, dashboards, planning and budgeting, and in-memory analytics. Their focus is on master data management and providing tools, training, and hosting. Their goal is not to provide reporting but to provide departments the ability to provide their own reporting.

Results

CORS has had an excellent adoption rate and resulted in gains in operating efficiencies. Stakeholders who used to wait 2 weeks for someone else to provide data are now accessing data on their own instantly during meetings. As the analytic maturity of all users evolved, they began using more advanced tools and the users are now demanding predictive and prescriptive analytics and driving the business case for growing the program. Agencies are even contributing staff to the development and admin team.

Cost

The cost of this program was not available. It started with centralized funding and is moving to a chargeback model for recovery of consumption-based expenditures as adoption and use grows. The starting team size was 12.

Lessons Learned

- Find "real world" public sector examples
- Strategic "business-focused" IT model
- Leverage a multi-year project plan
- Communicate project status updates and outcomes clearly and often
- They trained staff proactively before project
- User groups
- Super end users from business embedded in technology team during development sprints
- Core active users drive future direction

State of Michigan

Michigan's statewide data warehouse is the longest running and has the highest reported return on investment.

History

Michigan's statewide data warehouse started out in 1993 as the first Medicaid data warehouse in the nation and had an early focus on Temporary Assistance for Needy Families (TANF) and Medicaid fraud. It has now grown to span 21 agencies and more than 10K users from across Michigan state government. It is operated by the Data Center Operations group within the Department of Technology, Management, and Budget (DTMB).
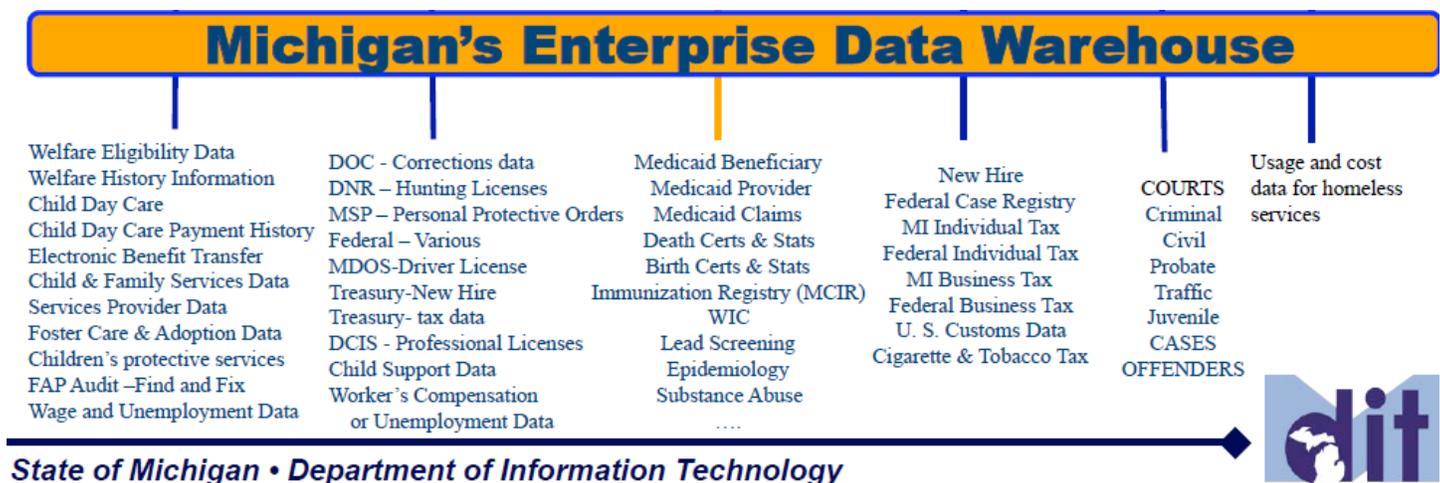
Approach

Michigan uses a traditional repository-style data warehouse. A unique client identifier (UCI) has been in place for many years to assist with the unique identification of individuals. The UCI is now being replaced by the Master Person Index (MPI) from the state's Master Data Management (MDM) platform, which was established to improve the quality of data held by each of the departmental systems.

Results

- The return on investment (ROI) for the program is **one million dollars per day**.
- Provides one place for multiple agencies and organizations to get information about large segments of Michigan's citizens and the government efforts to serve them.
- Used to reduce health care fraud, support health care analysis and outreach, and analyze mental health care
- Statewide Homeless Assistance Data online Warehouse (SHADoW) provides de-identified longitudinal database for studying and improving homeless assistance
- Used for rate setting for hospitals and managed and long-term care providers
- Used to determine parent location for child support
- Supports tax audits and validate tax filings
- Used to manage partner performance
- Supports fraud detection
- Used for supporting policy improvements
- Used for epidemiology studies

Cost

The annual operating cost is $8.1M and is recovered via chargeback to agencies served.

Michigan's Enterprise Data Warehouse

Welfare Eligibility Data	DOC - Corrections data	Medicaid Beneficiary	New Hire	COURTS	Usage and cost
Welfare History Information	DNR – Hunting Licenses	Medicaid Provider	Federal Case Registry	Criminal	data for homeless
Child Day Care	MSP – Personal Protective Orders	Medicaid Claims	MI Individual Tax	Civil	services
Child Day Care Payment History	Federal – Various	Death Certs & Stats	Federal Individual Tax	Probate	
Electronic Benefit Transfer	MDOS-Driver License	Birth Certs & Stats	MI Business Tax	Traffic	
Child & Family Services Data	Treasury-New Hire	Immunization Registry (MCIR)	Federal Business Tax	Juvenile	
Services Provider Data	Treasury- tax data	WIC	U. S. Customs Data	CASES	
Foster Care & Adoption Data	DCIS - Professional Licenses	Lead Screening	Cigarette & Tobacco Tax	OFFENDERS	
Children's protective services	Child Support Data	Epidemiology			
FAP Audit –Find and Fix	Worker's Compensation	Substance Abuse			
Wage and Unemployment Data	or Unemployment Data	….			

State of Michigan • Department of Information Technology

Recommended Approach and Feasibility

The methodology used for identifying and assessing the feasibility of optimal approaches for addressing the program objectives included:

- Evaluating each needs-based program objective to identify successful approaches in use by other states, successful approaches in use at a smaller scale in Arkansas, and industry best practice.
- Recommending one or more approaches to address each program objective.
- Evaluating the feasibility of each recommended approach through assessment of successful Arkansas implementations, prototyping with representative tools and data, and studying successful applications in other states.
- Mapping approaches into a cohesive framework to mitigate known barriers to compliant sharing, access to, and analysis of timely, integrated, high quality data between data providers and consumers in Arkansas.

Accessibility

Objective: *Enable openness, transparency, and pervasive, self-service data access and delivery*

According to the Data Management Body of Knowledge (DMBOK), all data has associated costs and risks, but **data has value only when it is actually used** or can be useful in the future. To achieve the maximum value and utility from state data assets and any data warehousing initiative, enablement of widespread access to the data should be a key component. This should include the ability to easily find, understand, search, download, analyze, and retrieve data through self-service user interfaces in both human-readable and machine-readable formats.

Recommended Approach: *Business Intelligence*

All data warehousing approaches put data into technologies that require special programming dialects to access. The earliest decision support systems required programming staff to handle any data requests and no self-service access to data by business users. A key enabler to self-service data access emerged in 1991 with the invention of the "semantic layer", which translates technical data structures into user-friendly business terms and allows for easy, drag-and drop query and analysis without technical knowledge of the data stores or query languages. The semantic layer provides the ease of use and flexibility to answer any question on the fly. Because the data relationships, calculations, and business rules are centrally defined, it also ensures consistency between analysis across users and time.

Self-service access to and analysis of data through the semantic layer are provided through Business Intelligence (BI) tools and platforms. BI platforms are used to query and analyze data retrieved via semantic layers, format the results, make the analysis repeatable and parameter-driven for flexibility, and schedule and distribute the results to the intended audience in a wide variety of formats.

Many users do not even need direct access to BI tools to receive benefits. A common use of BI platforms is to schedule distribution of data and reporting to a wider audience via email, file server, web server, etc. This combination of automation and distribution is used widely to gain operational efficiencies in state business processes.

Publication of data from BI tools can also include public audiences. During the hurricane Katrina relief effort, family members who were put on different buses in New Orleans were evacuated to relief areas in different cities or states. BI tools were used to facilitate overnight development of searchable public-facing reporting of evacuee locations to help reunite families. Public-facing transparency dashboards for the American Reinvestment and Recovery Act (ARRA) including interactive maps of recovery project spending and benefits were also easily provided via BI tools.

As noted in the case study, the Treasury Board of Canada has seen tremendous value from widespread BI access.

While data warehousing has little utility without business intelligence, business intelligence can provide early value without requiring a data warehouse. Arkansas Community Corrections (ACC) leveraged a BI semantic layer over the electronic Offender Management Information System (eOMIS) without a specialized data warehouse schema starting in 2009 and have realized many benefits to operational efficiency and decision support. They have since developed and provided thousands of reports and automatically distribute many reports to the field daily. ACC has recently built an eOMIS data warehouse for increased performance, historical retention, and incorporation of external data sources.

The feasibility of this approach is considered to be high because thousands of Arkansas state and school district employees currently leverage BI tools to access and analyze data from hundreds of state data sources, it is an extremely mature and established technology that has become a basic commodity asset for most organizations. Many Arkansas organizations are already using one or more business intelligence tools, and the prevailing industry trend is to use multiple business intelligence tools as needed for specific business needs over standardizing on a single tool. To maximize access to the largest audience of consumers while minimizing disruption, it is recommended to provide access to a state shared business intelligence platform but also support data access to other state business intelligence tools through secure, compliant database connectivity.

It is recommended that Business Intelligence tools be a key component of the statewide data warehouse strategy to ensure easy, self-service access to and detailed, flexible analysis of state data for all consumers.

Recommended Approach: *Web Portals and Mobile Access*

Not all users or use cases require detailed analysis of data. There is also value and utility in publishing raw structured data sets for consumption. Other uses cases may be better served by purpose-built web or mobile device interfaces to provide specific information via a guided experience and customized interfaces. This is a common starting point for government transparency initiatives.

Arkansas already has financial data sets available for exploration or download via the transparency.arkansas.gov website provided by the Arkansas Department of Finance and Administration. The myschoolinfo.arkansas.gov website provided by the Arkansas Department of Education supports online search, download, comparison, and analysis of public school data. Arkansas also has several successful web applications for use by citizens including the Gov2Go app developed by the Information Network of Arkansas. Given the prevalence and maturity of both open and closed web and mobile access portals, the feasibility of this approach is considered high.

In order to support new and existing web and mobile access to Arkansas data assets, it is recommended that support for web and mobile application access to the statewide data warehouse data be a key consideration.

Recommended Approach: *Application Programming Interfaces*

While Business Intelligence tools and web/mobile access can provide people with easy, self-service access to data, it is also becoming increasingly important to also provide applications with access to data via machine-readable formats to support automated processing. This need is most commonly met through the provisioning, documentation, and support of Application Programming Interfaces (APIs) or reasonably structured, non-proprietary data sets, without licensing restrictions, and easily downloadable in bulk. When applied to open data, these same interfaces could provide benefits to both authorized systems and the general public.

Application Programming Interfaces are widely used for internal information exchange in Arkansas government and are a mature and widespread technology. The Arkansas GIS Office makes Arkansas location data and services available via open APIs, and the City of Little Rock's Open Data portal supports API access for all published data sets.

To serve the broadest range of purposes and users, it is recommended to include support for access to data via Application Programming Interfaces as a component of the statewide data warehouse program.

Sharing

Objective: *Share data to enhance its value while enforcing privacy and security*

Increased data sharing was identified as the top need by the CDO Gap Analysis of Agency Data Needs, and the first recommendation from that study is the **creation of a comprehensive multi-agency data sharing agreement**. This would address issues with the lack of consistency between agreements and the lack of agility in creating or expanding agreements. The Public Safety Interagency Data Exchange Agreement signed in November 2017 was a vital first step to enable a culture of data sharing within the state, but **standardized data sharing agreements will need to be in place across all agencies** in order for a statewide data warehouse program to be effective.

The State of Texas has addressed this challenge via a **statewide data sharing compact** with standard terms in place of ad hoc peer to peer data sharing agreements. The State of Indiana **required all agencies to share data** by Governor's executive order (14-06 of 2014), which was later codified by Act 1470 of 2017. Interagency data sharing is facilitated by the Indiana Management Performance Hub acting as a centralized broker, which is discussed in further detail below.

From an operational perspective, the primary concerns that can impede or prevent more widespread data sharing are related to the need to ensure data security, privacy, and compliance, which necessarily complement and overlap each other. Compliance is ensuring that an organization meets its obligations under applicable state or federal laws, regulations, contractual obligations, and institutional policies. Privacy is related to ensuring appropriate collection, sharing, and use of personal data and personally identifiable information. Both privacy and compliance are typically the responsibility of the Chief Privacy Officer (CPO). Privacy and compliance risks are often barriers to sharing personally identifiable, regulated, or sensitive information.

Security is important for both compliance and privacy. Security is the primary responsibility of the Chief Information Security Officer (CISO) and involves protecting data from impermissible access.

The key challenge lies in establishing methods for sharing personal, sensitive, or regulated data in a manner that complies with all applicable policies but still retains value and utility.

Recommended Approach: *Authentication and Authorization*

Two means for providing secure, compliant access to data include the use of Authentication (verifying a user's identity) and Authorization (granting permissible access based on organization and role) are very common and mature.

The most basic form of authorization is restricting access to entire data sources, reports, dashboards, semantic layers, or collections of content based on a user's organization and role. However, most modern BI and data platforms have robust functionality allowing for more nuanced controls that support restriction to specific rows, columns, or other elements of the same data set. This supports compliant access to the greatest number of users via the smallest number of standardized delivery mechanisms to develop and maintain. Row-level security has been successfully used since 1999 to provide human resource managers at all agencies appropriate access to their agency's sensitive historical human resource data in the Arkansas Human Resources Management System (AHRMS) data warehouse via standard reports used by all agencies while the stewards at the Office of Personnel Management (OPM) have unrestricted access.

Authenticated use also allows for complete auditing of user activity, which is required by some regulations and can provide additional benefits for assessing and driving adoption and use of tools and data.

Recent Business Intelligence development efforts by the Department of Finance and Administration (DFA), such as the Performance Goals and Compensation calibration dashboards used by all agencies, are accessed via the statewide employee self-service portal (EASE) using the same authentication mechanism. Because the user account is tied to the employee record in the Arkansas Administrative Statewide Information System (AASIS), access is revoked automatically when an employee's status, role, or organization changes. This pre-existing authentication source could be effectively

leveraged by the statewide data warehouse for a large percentage of the potential user population. Additional authentication sources would likely be necessary for non-AASIS users or for data access requiring more stringent multi-factor authentication.

Unauthenticated access is an option for public access to data determined to be open and sharable without requirement for authenticated, authorized access.

The feasibility of this approach is considered to be high because it has widespread successful use across many Arkansas agencies and systems.

It is recommended that the statewide data warehouse include support for providing both authenticated, authorized access to restricted or closed data and unauthenticated access to open data.

Recommended Approach: *Compliance Views (*Deidentification, Aggregation, Masking)

While authorization and authentication are established mechanisms for ensuring permissible organization and role-based access, their primary purpose is to restrict the sharing of sensitive data publicly or between organizations. However, there are closed data sets that are beneficial for public, research, or inter-agency use that are unavailable due to security, privacy, or compliance. The most common privacy and compliance barrier is related to the sharing of personally identifiable information.

Much of the value of the closed data than cannot be shared in its raw form is not lost when personal identities are omitted. There are many valuable use cases that require only aggregate statistics and de-identified data.

One option for mitigating sharing barriers is to provide multiple views to the same data for different audiences and purposes:

- **Personal and Identifiable Individual-Level Data View** – Authenticated and authorized access to personal and identifiable individual-level data can be limited to users with an authorized role and purpose.
- **De-Identified Individual-Level Data View** - Some data can be more broadly shared at the individual record level if personal identifiers are sufficiently obscured or removed so that the record cannot be attributed back to the person to which it belongs. These views would require authenticated and authorized access.
- **De-Identified Aggregate Data View** - Some regulations prohibit the disclosure of even deidentified data at the individual record level because data could potentially be re-identified in the case of very small populations. In these cases, data can only be shared as aggregate counts. Some regulations require that aggregate data not be shared below a minimum statistical cell size or n-size in order to avoid re-identification risk. This requirement can be met through the use of suppression (excluding data), perturbation (adding noise), or blurring (reducing the precision of data) as data is accessed through the view depending on the data set being accessed and the allowed minimum cell size. These views could be openly available if sufficiently aggregated.

A similar approach could be applied to other data such as location that might only be shareable if sufficiently obscured to ensure privacy, compliance, and security.

It is not known if this exact approach is currently in use by Arkansas stakeholders, but similar approaches are currently in use to deliver data to users from different tables, views, or data sources in the same report based on user profile. Anonymization and aggregation views have been tested successfully on a representative data platform and are generally considered to be a feasible approach that requires additional testing and validation.

It is recommended that the statewide data warehouse include support for various compliance views in order to increase the compliant sharing of sensitive data with additional authorized or public users.

Analysis

Objective: *Support data-driven decision making and analytic maturity through development and support of analytic skills and shared services*

While not currently available to every consumer or for every relevant data source, access to historical Arkansas data via business intelligence outputs such as reports, dashboards, scorecards, and ad hoc query and analysis are currently available to over 10K state and school employees. These tools and techniques are categorized as Descriptive Analytics because they describe what happened in the past, or in the case of real time access, what is currently happening.

The Accessibility recommendations in this study focus on increasing the availability of access to data consumers, but there are also important considerations for increasing the maturity of analytic capabilities available to existing and future data consumers. Increasing awareness of, skills in, and technical capabilities for more mature types of analytics can enable more efficient and effective decisions and processes by reducing time-to-decision, recommending optimal actions and interventions, predicting future events or performance (which can support early intervention), and automating some decisions or processes where appropriate (with human oversight).

Three broad categories of business analytics or advanced analytics beyond traditional business intelligence include:

- Diagnostic Analytics – Used for diagnosing root causes and drivers of events based on relationship and patterns in the data. While descriptive analytics tools can support some of the same capabilities for known questions, tools and techniques for diagnostic analytics tend to offer more support for automated detection of unknown relationships, anomalies, or patterns without always knowing the question in advance.
- Predictive Analytics – Leverages predictive modeling and machine learning techniques to predict likely future outcomes based on patterns and relationships in current and historical data.
- Prescriptive Analytics – Extends predictive analytics by recommending the decisions and actions that are most likely to lead to the desired outcome based on current and historical data.
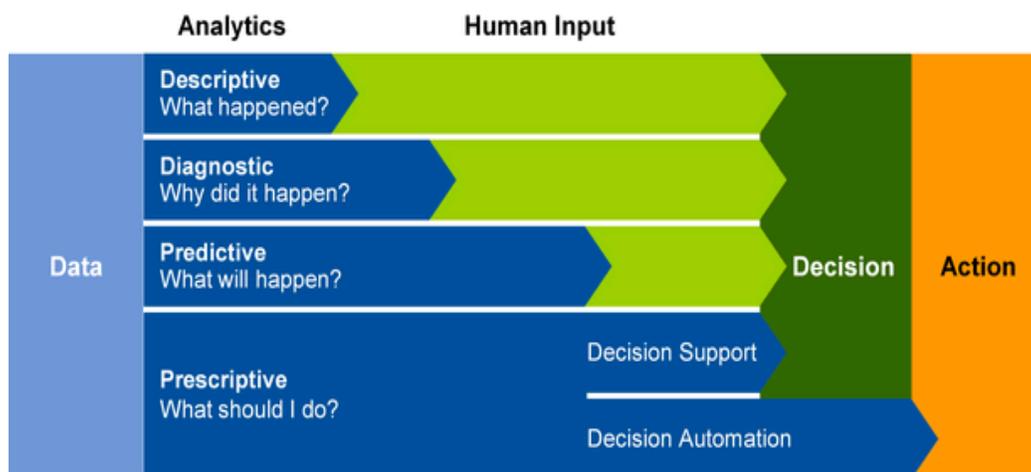

Figure 8 - Analytic Maturity (Source: Gartner)

While these are the three most commonly known and addressed categories of more advanced analytics, it is a broad and rapidly evolving field with new tools and approaches constantly emerging. Some other prevalent types of analytics include spatial (location-oriented), time series (patterns and forecasts over time), graph (focuses on relationships), text (sentiment and other derived meaning from unstructured text), and process mining (mapping and optimizing flow of events). This is not an exhaustive list and merely seeks to illustrate the need for versatile analytics capabilities to address evolving needs and approaches.

Recommended Approach: *Analytics Tools (Predictive/Prescriptive, Machine Learning, Spatial, Data Science Automation)*

As noted above, analytics technologies, approaches, and requirements evolve constantly and rapidly. An exhaustive recommendation for analytics capabilities is beyond the scope of this study and would become quickly dated. It is instead recommended to include in the data warehouse approach certain key, established capabilities that have either broad adoption and versatility or known applicability to Arkansas use cases. These key, foundational analytics capabilities include:

- **Predictive and Prescriptive Analytics** – The capability to predict outcomes and prescribe effective actions are broadly applicable to strategic and operational needs and can improve the efficiency and effectiveness of many decisions, processes, and services.
- **Machine Learning** – There are a core set of mature, foundational machine learning algorithms that can be applied and combined to perform a wide variety of analytic tasks. These should include Decision Trees, Support Vector Machines, Nearest Neighbor, Linear Regression, Logistic Regression, Neural Networks, K-Means, Association Rules, Random Forest, and Naïve Bayes.
- **Spatial Analytics** – Location is one of the few common elements and intersection points across many data sets and enables incorporation of valuable location-based external data such as that from the Census Bureau and the Bureau of Labor and Statistics.
- **Data Science Automation** - Up to 80% of the data science process can be spent on data preparation, feature evaluation and selection, preparation of training and testing sets, and model evaluation. To maximize the productivity of a limited number of data science resources, the capability should exist for at least semi-automated support for model preparation and evaluation tasks as well as that for ongoing model performance evaluation and retraining after model deployment. Deployed models should not require ongoing human interaction for operational use but should support automated scheduling or dynamic use.

Because needs and capabilities expand and evolve, the data warehouse approach should include the flexibility to easily leverage additional capabilities and techniques as needed to address emerging and evolving needs such as incorporation of additional libraries, functions, and services including, but not limited to, R (statistical programming language), Python (general open source programming language with), and TensorFlow (open source machine learning library).

Availability of analytics tools alone will provide little value unless paired with availability and development of analytics skills. The Indiana Management Performance Hub and other programs pair centralized data science resources with evangelization and training efforts to raise analytic maturity statewide. The State of Tennessee provides a shared pool of four data scientists with different specialties for use on agency and inter-agency analytics efforts. It is recommended to include at least one data scientist on the data warehouse program staff to support analytics use cases.

There are a few successful examples of the use of advanced analytics in Arkansas state government, but adoption is not widespread. Predictive and prescriptive analytics were developed and tested successfully during a proof of concept (POC) project with the Arkansas Department of Corrections, Arkansas Department of Community Corrections, Arkansas Parole Board, and Arkansas Crime Information Center to leverage data and analytics to reduce recidivism. A predictive recidivism model was developed in less than 90 days with good accuracy compared to current instruments and peer models in other states. The project also produced a prescriptive model that recommends optimal interventions for successful community re-entry and made extensive use of spatial analytics that brought new insights from existing data. Based on the results of this effort, success in other states, and the success of smaller-scale efforts in Arkansas, this approach is considered to be feasible. While difficult to implement, analytics have the potential for tremendous value.

It is recommended that the statewide data warehouse include staffing and support for varied and extendable analytics capabilities including at a minimum support for predictive and prescriptive analytics, common machine learning algorithms, and spatial analytics.

Integration

Objective: *Integrate data for improved cross-agency analysis and reduced duplication of data and efforts*

Once data sharing concerns are addressed through controls on the privacy, security, and compliance of data access, interagency data sets must still be integrated with each other to gain a broader perspective than is possible using any individual system alone.

The two biggest challenges to data integration are:

- The lack of semantic consistency (using the same names and identifiers for the same people, places, events, things, etc.) increases the difficulty of matching data across systems.
- Security, privacy, and compliance concerns when personally identifiable or sensitive information must be disclosed by one or more parties in order to establish the required data relationship, even if the ultimate data product will only be disclosed in deidentified or aggregate form.

Recommended Approach: *Statewide Master Data Management*

Integrating data from multiple systems together is frequently challenging due to a lack of standard, consistent identifiers for the same entity (person, place, location, thing) across systems. Data regarding organizations frequently suffer from lack of a standard identifier. This can result in many records for the same organization in even the same source and impede analysis such as spend on a single vendor or payments to a single organization.

One effective method of addressing this issue is establishment and use of common identifiers across systems. In some cases, multiple agencies are served by the same information system such as the Arkansas Administrative Statewide Information System (AASIS), allowing for centralized control over the consistency of names and identifiers. When multiple information systems are used to describe the same entity, there must be more planning and coordination to establish and maintain consistent naming and globally unique identification of entities.

For entities with a limited, relatively static range of values, the decision can be made for multiple systems to use the same authoritative source for encoding data. One common example used in many Arkansas information systems is the Federal Information Processing System (FIPS) code for US states. One benefit of this approach is that national standards may be used by other states, facilitating the possibility of interstate integration and comparison of data or required for Federal reporting.

One successful example of standards-based interagency data integration and reporting in Arkansas is the effort to comply with Federal reporting requirements for the American Recovery and Reinvestment Act of 2009 (ARRA). The act resulted in the award of 218 Federal grants with award amounts totaling over $3.7B to Arkansas government organizations, but it also made continued funding contingent upon compliance with quarterly Federal reporting requirements. A key aspect of these requirements was that very little submission of free form text was allowed. Reporting of grant programs, recipient organizations, vendors, grant activity types, and locations were all supported only through specified coding standards such as FIPS codes for counties and state, DUNS numbers for organizations, or NAICS codes for industries. Upon report submission, the validity of codes was checked, and invalid or conflicting codes resulted in a failed submission. Arkansas was a centralized reporting state, meaning that the Department of Finance and Administration Office of Intergovernmental Services (IGS) was responsible for all Federal reporting to the White House Office of Management and Budget for all state agencies. By coordinating on standards for encoding and reporting data, quarterly reports for 28 state agencies using 3 different information systems (plus AASIS) were successfully integrated and submitted without issue over a 4-year span. This semantic consistency also allowed for consistent national aggregation and reporting and was beneficial for supporting Arkansas' own ARRA transparency portal.

Common identifiers can enable data from across sources to be effectively integrated but doesn't help with duplication of data and effort when the same data is being created and maintained in multiple systems. It also introduces the challenge of maintaining and sharing a consistent list of valid identifiers across all sources when dealing with large, dynamic lists such as people, organizations, and addresses.

The commonly used and recommended approach for addressing the coordination of consistent identifiers across systems is Master Data Management (MDM). Master data management is a technology-enabled discipline comprised of the processes, governance, and tools required to create, maintain, integrate, monitor, and share master and reference data. Master data are the key business entities or "nouns" of an organization that are widely used across departments, processes, and systems (people and roles, organizations, locations). Reference data are a set of standard permissible values that are not specific to the enterprise but are used by multiple systems to facilitate data sharing (industry codes).

The technology component of master data management is a master data "hub" that is used to coordinate the exchange or "harmonization" of codes and values between subscribing systems. If one system can be determined to be the authoritative system of record, its values can be made available for reference from, or pushed to, subscribing systems that require the same data, reducing duplicative efforts to maintain the same data. When multiple systems need to maintain different, potentially overlapping lists of the same type of entity (location), MDM can be used to synchronize all subscribing systems into a single master list with a combined "golden record" consisting of the most complete and timely information from the combination of all systems and establishing a common unique identifier.
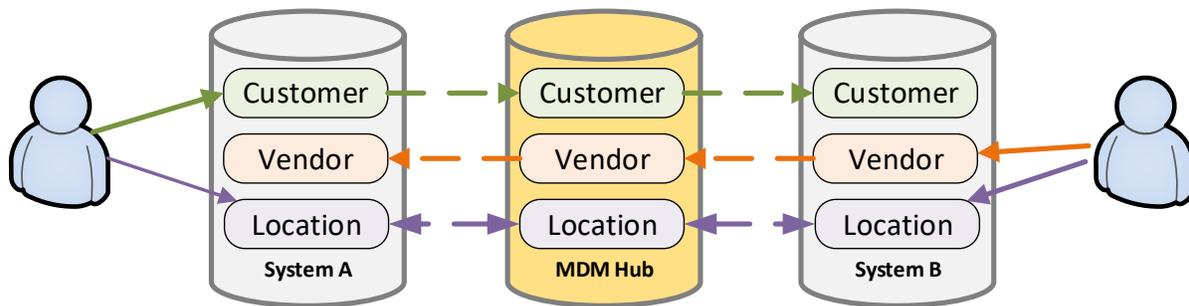

Figure 9 - Master Data Management

The key benefits of this approach include:

- Each record is only created and maintained in one place and reused globally, reducing duplicate effort
- All systems use the same codes and names, making reporting, integration, and reconciliation easier
- Consolidated "golden records" can be more complete, timely, and accurate than in any individual system
- Provides a single location for stewardship of master data, typically with support for data quality rules
- Increases operational efficiency through easier, more comprehensive access to data
- Minimizes the impact of changes by serving as an intermediate broker between systems
- Serves as a key enabler for logical data warehousing by storing the relationships between systems and entities

Another key component of MDM is entity resolution, which is the process of linking records for the same entity.

The Arkansas Department of Information Systems and Arkansas Department of Parks and Tourism both have successful implementations of MDM systems providing valuable business benefits. Due to the maturity and prevalence of this technology, existing successes in Arkansas, specific requirements to provide MDM in Act 912 of 2017, and additional recommendations from the CDO Gap Analysis, the feasibility and value of this approach is considered to be very high.

It is recommended that the statewide data warehouse include capabilities for statewide master data management to support the coordination of consistent identifiers across agency systems.

Recommended Approach: *Centralized Broker*

Master data management can help address the problem of semantic consistency, but implementing statewide master data management, particularly for citizen data, requires access to personally identifiable information from multiple agencies in order to perform the matching process required to maintain a unique list of records with a consistent identifier. Similarly, the integration of data related to various citizen services requires access to private and sensitive data from multiple agencies, even if the data distributed to consumers is deidentified and aggregated enough to be considered open. Agencies can share their own deidentified, aggregate data, but for interagency data integration efforts, the requirement for one party to have access to two or more sets of personally identifiable information has been a frequent and persistent barrier.

The State of Indiana addressed this issue by establishing in the Indiana Management Performance Hub (MPH) a group which can serve as a trusted broker between and on the behalf of all agencies. MPH acts as an agent of the agencies they serve to integrate, analyze, and distribute data on behalf of and subject to governance defined by the providing agencies. MPH was established by Governor's executive order (14-06 of 2014) and later codified by Act 1470 of 2017:

> *Sec. 13. The MPH is considered to be an agent of the executive state agency sharing government information and is an authorized receiver of government information under the statutory or administrative law that governs the government information. Interagency data sharing under this chapter does not constitute a disclosure or release under any statutory or administrative law that governs the government information.*
> http://iga.in.gov/legislative/2017/bills/house/1470#document-af23f3bf

This successful approach could be applied in Arkansas by leveraging a centralized group to act as agents on behalf of providing agencies in integrating and sharing data in compliance with established governance policies. The data should be stored in a secure zone in the state data center that meets the physical, network, multi-factor authentication, and other requirements of the most sensitive data hosted. Support staff should be required to have the same background checks, training, and certifications (CJIS, HIPAA, etc.) as the agencies they represent when integrating, analyzing, and sharing data on their behalf. The security, governance, architecture, and operating procedures for the data integration layer should be designed and overseen by the state Chief Data Officer, Chief Privacy Officer, and Chief Information Security Officer to ensure stringent security, privacy, and compliance. The group should be located within the Department of Information Systems along with the CDO, CPO, and CISO.

The feasibility of this approach is demonstrated by success using this approach in the State of Indiana.

It is recommended that the integration layer of the statewide data be managed by a centralized group housed at the Department of Information Systems and acting as agents of providing agencies.

Recommended Approach: *Data Integration Layer*

Data should be integrated using current best practices for balancing timeliness, performance, agility, and ease of use. In order to meet the widest array of operational and analytical data integration needs, the data integration layer should:

- Include support for a data warehouse as needed for performance or historical retention
- Include support for a data lake for agility and exploratory analysis
- Use a data hub approach to logically integrate data via virtualized integration views over master data hubs, data warehouse, data lake, and federated local data sources

This pattern has been used successfully at DIS and is considered to be a feasible option for statewide integration.

It is recommended that the statewide data warehouse leverage the data hub approach by logically integrating data from a centralized data warehouse, data lake, master data hub, and local data sources.

Timeliness

Objective: *Enable real-time assessment, support, alignment, and automation of decisions, programs, and resources*

Some of the initial barriers that led to the development of separate reporting databases and data warehouses were:

- Source systems were designed for efficiently processing individual transactions (inserts, updates, deletes) and not for mass query and analysis. They were frequently either too slow for efficient analysis or using them for analysis was detrimental to the availability and performance of the transaction system.
- Early business intelligence tools typically only supported connection to a single data source, but decision support and analysis needs required the combination of data from multiple data sources.

These factors, along with the need to retain historical data, required a second copy of the source data to be maintained that could be optimized for analysis, integrated into a single consolidated data store, and isolated from impacting the source systems. Since the batch extract, transform, and load (ETL) technology available at the time was also detrimental to the performance of the source systems, reporting databases or data warehouses were typically loaded overnight when they would have the least impact to business users.

This approach enabled for consolidated, performant data access that met many business needs, but resulted in some decision latency due to the nightly, weekly, or monthly batch processes required. It also resulted in a redundant volume of data to store, maintain, secure, and keep consistent with the source.

The value of some data decreases with time, and some business processes such as operational support benefit greatly from applying decision support and analytics to real-time data. Advancements in the technologies and approaches for hosting source system data and connecting source data with consumers have now matured to the point that real-time access is a feasible and integral component of a modern data and analytics infrastructure.

Recommended Approach: *Federation*

The primary approach used for enabling real-time access to distributed data sources is Data Federation. As previously defined in the logical data warehouse section, Data Federation is a form of data virtualization where the data stored in multiple sources (of the same or different type) are made accessible to data consumers by using on-demand data integration, rather than executing data movement and physically storing integrated data. This is essentially a reversal of the extract, transform, and load process which distributes queries to the data and applies necessary integration and transformation on demand as opposed to moving and transforming the data to a separate data store in a batch process.

In order to minimize impact on the source system, several approaches are used by the federation layer to avoid sending computationally expensive queries to transactional data source. Leveraging advances in database and hardware technology, some source systems are also becoming robust enough to handle greater analytic workloads and are even being designed for hybrid transactional and analytical use.

The Arkansas Longitudinal Data System developed by the Arkansas Department of Education is a successful example of batch federation for analysis of integrated inter-agency data sets. The Arkansas Department of Information Systems successfully uses federation for providing distributed real-time access to operational data. Because this is a mature, common technology, the feasibility of this approach is considered to be high.

It is recommended that the statewide data warehouse include support for federated, real-time access to disparate data sources when appropriate.

Quality

Objective: *Manage enterprise data as a state asset*

In the 90s, data quality was often corrected into the data downstream in the data warehouse, which is the equivalent of inspecting quality into the product in manufacturing by checking for and discarding or repairing defective products. Ensuring quality data products, much like manufacturing, is more efficiently and effectively done as close to the source of creation as possible. Only the data creator, at the moment of data creation, truly knows what real-world person, item, or event they are trying to represent.

The most effective place to address data quality is at the source and through business rules and restricting entry to valid domain at point of entry. This is not always practical for some systems but is a best practice that can be adopted through attrition over time. A less invasive and more immediate way to start measuring and managing data at the local agency-level is through the use of data quality management tools for profiling, measuring, and monitoring the quality of local data assets.

When age is first profiled in many Arkansas data sources, it typically results in individuals who are negative or hundreds of years old. Self-service data quality management tools can enable data stewards to profile the data to identify and measure any issues with completeness, obvious inaccuracies (people with negative ages), problems with consistency of representation, etc. These stewards can then establish data quality rules and dashboards for measured improvements on data quality problems and potentially identify controls to mitigate future issues at the source.

Data quality management and stewardship portals are also frequently the source for collecting and sharing metadata with consumers and establishing data lineage that can be used to assess the impact of potential data change.

While data quality management tools can help identify and monitor local data quality issues to speed and ensure correction, a more proactive approach is the use of agency-level master data management. The benefits of master data management have been discussed in the integration section in relation to statewide master data management. However, there are many agency-specific data elements that may never need to be included in the statewide master data hub. Local, agency-level master data management can yield substantial gains in improving the quality of agency data, increasing the ease of access to key agency data.

Finally, Data Governance is needed over all aspects of the data warehouse program, including a comprehensive data catalog and shared metadata repository.

It is recommended that the statewide data warehouse include support for local data quality management tools and stewardship portals, agency-level master data management, and comprehensive data governance.

Summary of Recommendations

The following summarizes how the identified program objectives map to each group or layer of recommended approaches:

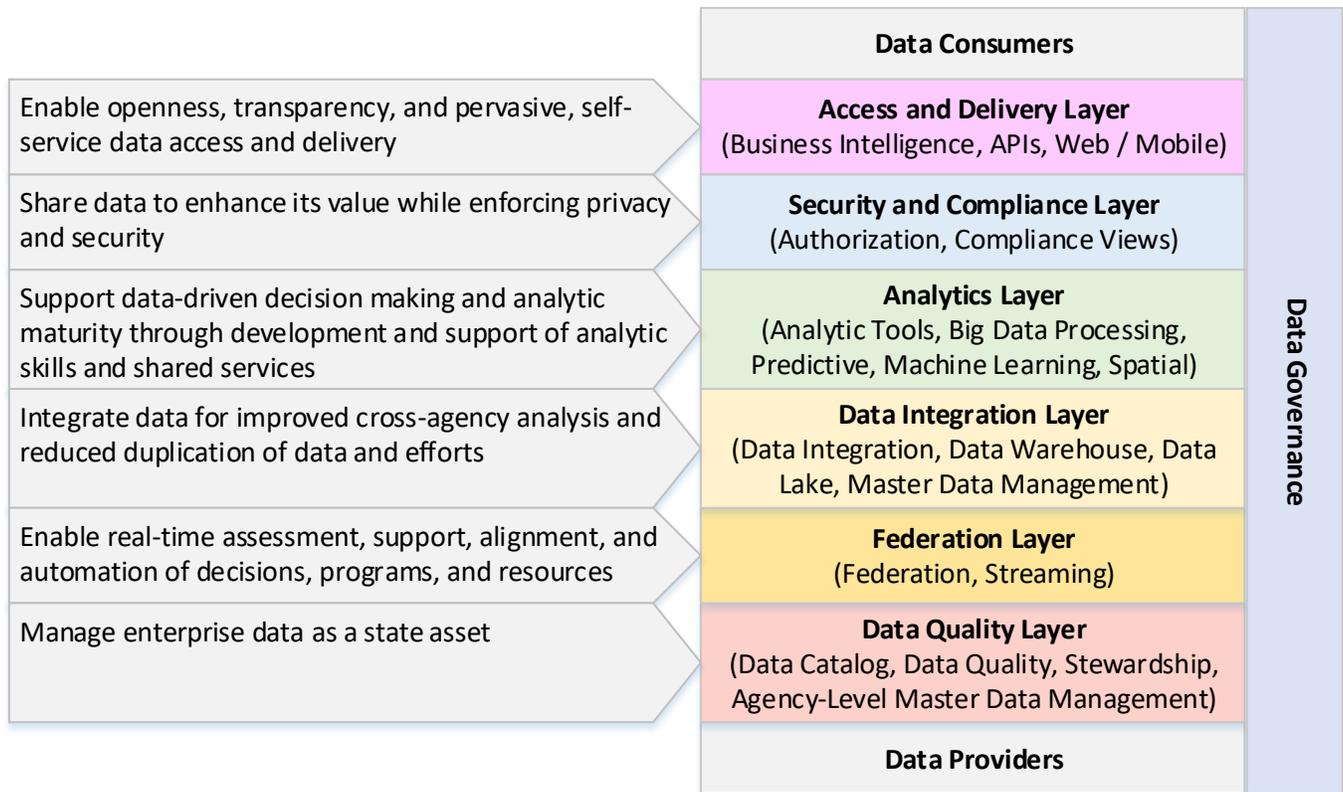| Objectives	Layers	Data Governance
Enable openness, transparency, and pervasive, self-service data access and delivery	**Data Consumers**	
	Access and Delivery Layer (Business Intelligence, APIs, Web / Mobile)	
Share data to enhance its value while enforcing privacy and security	**Security and Compliance Layer** (Authorization, Compliance Views)	
Support data-driven decision making and analytic maturity through development and support of analytic skills and shared services	**Analytics Layer** (Analytic Tools, Big Data Processing, Predictive, Machine Learning, Spatial)	
Integrate data for improved cross-agency analysis and reduced duplication of data and efforts	**Data Integration Layer** (Data Integration, Data Warehouse, Data Lake, Master Data Management)	
Enable real-time assessment, support, alignment, and automation of decisions, programs, and resources	**Federation Layer** (Federation, Streaming)	
Manage enterprise data as a state asset	**Data Quality Layer** (Data Catalog, Data Quality, Stewardship, Agency-Level Master Data Management)	
	Data Providers	

Figure 10 - Program Objectives Mapped to Recommended Approaches

Some dependencies and rationale considered when ordering the layers from provider to consumer include:

- Widespread access via Business Intelligence tools, database connectivity, Application Programming Interfaces, mobile and web applications, etc. should be on top of the security and compliance layer to allow for decentralized development and delivery via a wide variety of tools but leveraging centralized management of data access in compliance with security, privacy, regulatory, and other policies and applicable laws.
- Analytics should have the capability to be performed on identified and integrated individual-level data if needed and permissible and then presented with applicable anonymization, aggregation, filtering. Therefore, the analytics layer should be located between the data integration layer and the security and compliance layer.
- Integration should have the capability to be performed dynamically across both federated or streaming real-time data and persisted data from the centralized master data management hub, data warehouse, and data lake. Therefore, the data integration layer should be located between the analytics and federation layers.
- Data quality should be managed as close to the source as possible and with self-service access by the actual agency-level data stewards. The data quality layer should be provided directly to data providers with the capability of being used with local data sources.
- Data governance should encompass all aspects of the statewide data warehouse. This includes the collection, integration, and provisioning of metadata from providers to consumers as well as centralized change management for shared data and services.

The Recommended Arkansas Data Hub (Statewide Data Warehouse) Approach

The following diagram illustrates in more detail how the recommended approaches detailed in this study fit together to comprise a holistic approach to delivering upon the identified needs. The full "data hub" architecture expands on the traditional data warehouse to enable compliant, governed data sharing for both operational and analytical needs.
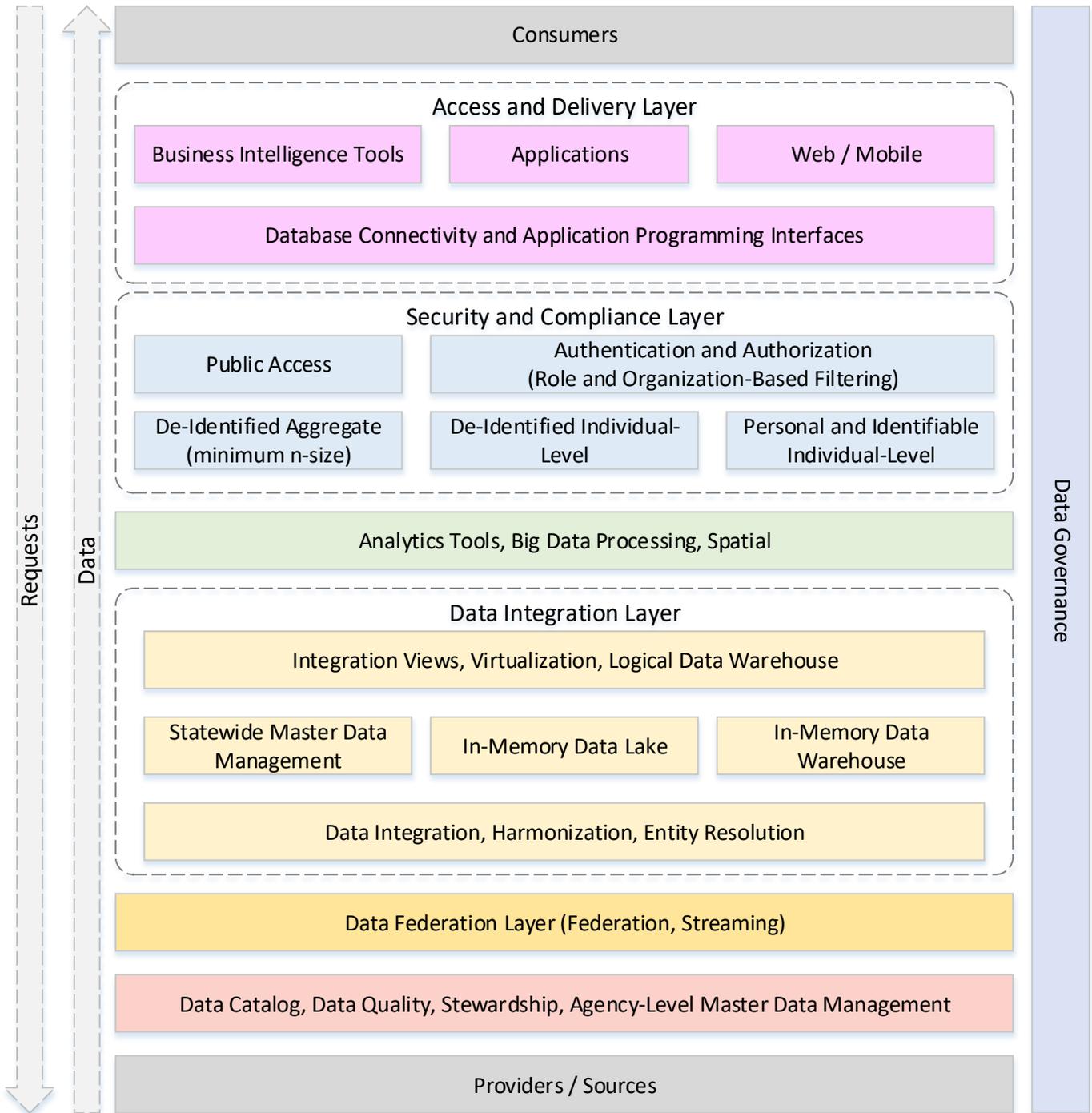

Figure 11 - Recommended Arkansas Data Hub Approach

Costs and Benefits

Costs and Funding Models

Costs

The annual costs for establishing and supporting a statewide data warehouse program are estimated at **$3.9M per year for years 1-5 and $2.8M per year for years 6 and beyond**. The lower annual cost after year 6 is due to the capital expenditures for software purchases being fully amortized (assumes a 5-year useful life). Hardware cost is also based on amortization over a 5-year useful life but assumes replacement at end of useful life.

These costs were based on conservative estimates for scaling established state data platforms, staffing the program based on the current Arkansas classification and compensation plan, and adding recommended additional capabilities and capacity based on costs for representative interoperable technologies.

Software costs were estimated based on purchase and annual software support costs for representative software platforms and products to address the recommended approaches to be included in the program scope. Hardware costs were based on platform sizing equivalent to that used by the Indiana Management Performance Hub.

Staffing costs were estimated based on full or partial resource allocations for program staff including the CDO, CPO, Data Warehouse Lead, Database Administration Lead, Data Scientist, Data Engineers, Data Warehouse Specialists, Platform System Administrator, Database Administrator and Project Manager.

Total annual costs for **other statewide data warehouse programs studied ranged from $8M to $9M per year**. The Indiana Management Performance Hub is the closest analogue to the proposed program. It has similar hardware and software sizing, but a larger staff size (20 full time employees plus contract staff for data science).

Funding Models

Other statewide programs use **a combination of funding strategies** including federal and private **grants, general revenue** appropriations, and usage-based **chargeback**. The Indiana Management Performance Hub's $9M annual cost is covered by a $6M appropriation of general funds, $1.7M in database management funds, and $1.3M from the Department of Insurance fund. The $8M annual operating cost for Michigan's statewide data warehouse is recovery by chargeback to agencies served.

There are Federal grants for inter-agency data efforts being leveraged by other states to start and support statewide data warehouse efforts. Relevant grant funding opportunities should be monitored and leveraged as appropriate for helping to start Arkansas inter-agency data sharing and analysis efforts.

Benefits

The financial benefits of this initiative are expected to exceed the costs. Some examples of financial impact of enhanced data sharing and integration from other states include:

- The 2018 Annual Report published by the Indiana Management Performance Hub estimated a **$40M return on investment over 18 months** ($4.50 return for every $1 invested taxpayer dollar) as quantified by serviced agency project owners.
- The State of Michigan is **saving $1M per day** by linking data across programs in 21 different agencies and extending access to 10K users.
- The Texas Workforce Commission implemented an early detection program for unemployment insurance fraud leveraging interagency data integration and **avoided $71.84M in costs** in FY15-FY17.
- Additional financial ROI examples are cited in the 2015 Arkansas Legislative Audit (ALA) Special Report on the "Potential Benefits of a Centralized Data Warehouse for the State of Arkansas" including:
    - The State of Washington recoups more than **$10M per year** in fraudulent tax refunds.
    - The State of Georgia detected **$25M over 2 years** in fraudulent tax returns.
    - The State of New York **increased collections by $100M** and **reduced fraudulent refunds by $1.2B** in 2010.

An Arkansas proof of concept project exploring the use of advanced analytics to reduce the risk of recidivism resulted in a prototype with a projected cost savings of **$8.1M over 3 years** with a 1% reduction in recidivism.

Beyond financial benefits, there are many notable examples of how integrating, sharing, and analyzing state data can yield improvements to citizen safety, health, and quality of life. The Indiana MPH 2018 Annual Report includes details on MPH projects that improved policy and operational support in combatting the Opioid epidemic, optimizing Medicaid, increasing government transparency, projecting recidivism risk, enabling agency performance management, and aligning education and workforce development.

Next Steps

The recommended next steps towards implementation of a statewide data warehouse program include the following:

- Formalize a **multi-department data sharing agreement** for compliant, secure and efficient data sharing between departments
- Determination of initial and sustaining program **funding approaches**
    - Identify and leverage **potential grants** for program startup
- Develop a **program charter** to formalize the scope of the program
    - **Discussion and agreement** on data warehouse program components and architectural approaches
    - Determine **prioritization mechanism** for implementing and addressing use cases
- Implementation of a **data hub** (value-driven broker for cross-agency data sharing and analytics) to act as an agent of individual agencies in integrating and providing secure, compliant access to and analysis of inter-agency data