

Project Investment Justification

Data Lakehouse

HS23002

Department of Health Services

Contents

1. General Information	2
2. Meeting Pre-Work	2
3. Pre-PIJ/Assessment	3
4. Project	3
5. Schedule	5
6. Impact	6
7. Budget	7
8. Technology	7
9. Security	10
10. Areas of Impact	11
11. Financials	13
12. Project Success	14
13. Conditions	15
14. Oversight Summary	15
15. PIJ Review Checklist	17

1. GENERAL INFORMATION

PIJ ID: HS23002

PIJ Name: Data Lakehouse

Account: Department of Health Services

Business Unit Requesting: Business Intelligence Office

Sponsor: Susan Robinson

Sponsor Title: Chief Business Intelligence Officer

Sponsor Email: susan.robinson@azdhs.gov

Sponsor Phone: (480) 435-3929

2. MEETING PRE-WORK

2.1 What is the operational issue or business need that the Agency is trying to solve? (i.e....current process is manual, which increases resource time/costs to the State/Agency, and leads to errors...):

ADHS has many disparate data sources without a centrally governed data analytics and data set repository. This limits the Agency's agility to respond to health emergencies, slows down information sharing with partners, and lowers initial confidence in fact-based data analytics.

In the as-is strategy today are an excessive amount of point-to-point data feeds that solve individual business needs. This increases sustainability risks for both the Information Technology Services and Business Intelligence Office units as well as security concerns. This approach increases the probability of data errors and does not allow for a single version of the truth when determining or reporting on public health anomalies, epidemics, pandemics, and other public health needs.

Mortality data is often at the epicenter of ADHS analytics and is often difficult to curate. A centralized repository of curated mortality data is not currently available.

2.2 How will solving this issue or addressing this need benefit the State or the Agency?

The long term vision of the project is to set the foundational groundwork, infrastructure, governance, and sustainability processes in place for future enterprise use of curated data sets for analytics and data sharing for all ADHS divisions and external stakeholders as a single source of truth.

2.3 Describe the proposed solution to this business need.

The short term vision of this project is to:

1. Construct, install, and configure a sustainable ADHS Data Lakehouse environment.
2. Ingest, transform, and present ADHS mortality data in the new Data Lakehouse environment to the identified ADHS stakeholder groups.

2.4 Has the existing technology environment, into which the proposed solution will be implemented, been documented?

Yes

2.4a Please describe the existing technology environment into which the proposed solution will be implemented.

2.5 Have the business requirements been gathered, along with any technology requirements that have been identified?

Yes

2.5a Please explain below why the requirements are not available.

3. PRE-PIJ/ASSESSMENT

3.1 Are you submitting this as a Pre-PIJ in order to issue a Request for Proposal (RFP) to evaluate options and select a solution that meets the project requirements?

No

3.1a Is the final Statement of Work (SOW) for the RFP available for review?

3.2 Will you be completing an assessment/Pilot/RFP phase, i.e. an evaluation by a vendor, 3rd party or your agency, of the current state, needs, & desired future state, in order to determine the cost, effort, approach and/or feasibility of a project?

No

3.2a Describe the reason for completing the assessment/pilot/RFP and the expected deliverables.

3.2b Provide the estimated cost, if any, to conduct the assessment phase and/or Pilot and/or RFP/solicitation process.

3.2e Based on research to date, provide a high-level cost estimate to implement the final solution.

4. PROJECT

4.1 Does your agency have a formal project methodology in place?

Yes

4.2 Describe the high level makeup and roles/responsibilities of the Agency, Vendor(s) and other third parties (i.e. agency will do...vendor will do...third party will do).

The ADHS IT Team will perform

1. Schedule Management
2. Vendor Management
3. Budget Management
4. Scope Management
5. Cloud Infrastructure Support
6. Security Assessments
7. Tool Access Management
8. Data Access (rows, columns, tables, warehouse)
9. Enterprise System Architecture
10. Enterprise Application Architecture

The ADHS Business Intelligence Office will perform

1. User Requirements
2. User Acceptance Testing
3. Business Architecture

The Vendor Professional Services organization will perform

1. Tooling installation
2. User Requirements Facilitation
3. Development Lifecycle Management (Agile)
4. Quality Assurance Testing
5. User Acceptance Facilitation

The ADHS Enterprise Data Management Office will perform:

1. Data Governance Framework Coordination and Oversight
2. Data Management Tools Product Review and Selection
3. Data Architecture Review in Collaboration with IT

4.3 Will a PM be assigned to manage the project, regardless of whether internal or vendor provided?

Yes

4.3a If the PM is credentialed, e.g., PMP, CPM, State certification etc., please provide certification information.

4.4 Is the proposed procurement the result of an RFP solicitation process?

No

4.5 Is this project referenced in your agency's Strategic IT Plan?

Yes

5. SCHEDULE

5.1 Is a project plan available that reflects the estimated Start Date and End Date of the project, and the supporting Milestones of the project?

Yes

5.2 Provide an estimated start and finish date for implementing the proposed solution.

Est. Implementation Start Date

Est. Implementation End Date

1/20/2023 12:00:00 AM

9/4/2023 12:00:00 AM

5.3 How were the start and end dates determined?

Based on project plan

5.3a List the expected high level project tasks/milestones of the project, e.g., acquire new web server, develop software interfaces, deploy new application, production go live, and estimate start/finish dates for each, if known.

Milestone / Task	Estimated Start Date	Estimated Finish Date
Tooling Procurement	01/20/23	02/17/23
Project Kickoff	02/01/23	02/03/23
February 2023 Vendor Payment Estimate \$120,000 Data Ingestion	02/01/23	02/15/23
Platform & Data Pipeline Build	02/06/23	03/19/23
MVP Training Strategy/Plan	02/06/23	04/02/23
CI/CD Pipeline & Infrastructure-As-Code	02/06/23	04/16/23
Data Ingestion	02/06/23	03/05/23
March 2023 Vendor Payment Estimate \$190,000 Platform & Data Pipeline Build, Data Validation and Transformation, MVP Training Strategy/Plan	03/01/23	03/15/23
Data Validation and Transformation	03/06/23	04/02/23
Data Curation	03/20/23	04/30/23
Dashboard & Production Deployment	03/20/23	04/16/23

Snowflake SQL API	03/20/23	04/16/23
DAVE Data Set Extracts	03/20/23	04/16/23
April 2023 Vendor Payment estimate. \$210,000 Data Curation, DAVE Data Set Extracts, Consumption Views & Role-Based Security, Snowflake SQL API, CI/CD Pipeline & Infrastructure-As-Code, Data Pipeline Observability, Dashboard & Production Deployment,	04/01/23	04/15/23
Consumption Views & Role-Based Security	04/03/23	04/30/23
Training Development & Delivery	04/03/23	06/11/23
Data Pipeline Observability	04/03/23	04/30/23
May 2023 Vendor Payment Estimate \$190,000 UAT & Production Deployment	05/01/23	05/15/23
UAT & Production Deployment	05/01/23	05/28/23
Hypercare	05/29/23	06/11/23
June 2023 Final Vendor Payment Estimate \$131,976 Hypercare, Training Development & Delivery	06/01/23	06/15/23
Project Closeout	06/11/23	09/04/23

5.4 Have steps needed to roll-out to all impacted parties been incorporated, e.g. communications, planned outages, deployment plan?

Yes

5.5 Will any physical infrastructure improvements be required prior to the implementation of the proposed solution. e.g., building reconstruction, cabling, etc.?

No

5.5a Does the PIJ include the facilities costs associated with construction?

5.5b Does the project plan reflect the timeline associated with completing the construction?

6. IMPACT

6.1 Are there any known resource availability conflicts that could impact the project?

No

6.1a Have the identified conflicts been taken into account in the project plan?

6.2 Does your schedule have dependencies on any other projects or procurements?

No

6.2a Please identify the projects or procurements.

6.3 Will the implementation involve major end user view or functionality changes?

No

6.4 Will the proposed solution result in a change to a public-facing application or system?

No

7. BUDGET

7.1 Is a detailed project budget reflecting all of the up-front/startup costs to implement the project available, e.g, hardware, initial software licenses, training, taxes, P&OS, etc.?

Yes

7.2 Have the ongoing support costs for sustaining the proposed solution over a 5-year lifecycle, once the project is complete, been determined, e.g., ongoing vendor hosting costs, annual maintenance and support not acquired upfront, etc.?

Yes

7.3 Have all required funding sources for the project and ongoing support costs been identified?

Yes

7.4 Will the funding for this project expire on a specific date, regardless of project timelines?

Yes

7.5 Will the funding allocated for this project include any contingency, in the event of cost over-runs or potential changes in scope?

Yes

8. TECHNOLOGY

8.1 Please indicate whether a statewide enterprise solution will be used or select the primary reason for not choosing an enterprise solution.

The project is using a statewide enterprise solution

8.2 Will the technology and all required services be acquired off existing State contract(s)?

Yes

8.3 Will any software be acquired through the current State value-added reseller contract?

Yes

8.3a Describe how the software was selected below:

A discovery of products was conducted based on requirements that were elicited.

8.4 Does the project involve technology that is new and/or unfamiliar to your agency, e.g., software tool never used before, virtualized server environment?

Yes

8.5 Does your agency have experience with the vendor (if known)?

Yes

8.6 Does the vendor (if known) have professional experience with similar projects?

Yes

8.7 Does the project involve any coordination across multiple vendors?

Yes

8.8 Does this project require multiple system interfaces, e.g., APIs, data exchange with other external application systems/agencies or other internal systems/divisions?

No

8.9 Have any compatibility issues been identified between the proposed solution and the existing environment, e.g., upgrade to server needed before new COTS solution can be installed?

No

8.9a Describe below the issues that were identified and how they have been/will be resolved, or whether an ADOA-ASET representative should contact you.

8.10 Will a migration/conversion step be required, i.e., data extract, transformation and load?

No

8.11 Is this replacing an existing solution?

Yes

8.11a Indicate below when the solution being replaced was originally acquired.

Roughly 12 years ago or in approximately 2010.

8.11b Describe the planned disposition of the existing technology below, e.g., surplus, retired, used as backup, used for another purpose:

The SAS code will be retired.

8.12 Describe how the agency determined the quantities reflected in the PIJ, e.g., number of hours of P&OS, disk capacity required, number of licenses, etc. for the proposed solution?

This was performed via a discovery where requirements were gathered. Based on the requirements, capacity and licensing were proposed.

8.13 Does the proposed solution and associated costs reflect any assumptions regarding projected growth, e.g., more users over time, increases in the amount of data to be stored over 5 years?

Yes

8.14 Does the proposed solution and associated costs include failover and disaster recovery contingencies?

Yes

8.14a Please select why failover and disaster recovery is not included in the proposed solution.

8.15 Will the vendor need to configure the proposed solution for use by your agency?

Yes

8.15a Are the costs associated with that configuration included in the PIJ financials?

Yes

8.16 Will any app dev or customization of the proposed solution be required for the agency to use the project in the current/planned tech environment, e.g. a COTS app that will req custom programming, an agency app that will be entirely custom developed?

No

8.16a Will the customizations inhibit the ability to implement regular product updates, or to move to future versions?

No

8.16b Describe who will be customizing the solution below:

The professional services vendor.

8.16c Do the resources that will be customizing the application have experience with the technology platform being used, e.g., .NET, Java, Drupal?

Yes

8.16d Please select the application development methodology that will be used:

Agile/Scrum

8.16e Provide an estimate of the amount of customized development required, e.g., 25% for a COTS application, 100% for pure custom development, and describe how that estimate was determined below:

5%

8.16f Are any/all Professional & Outside Services costs associated with the customized development included in the PIJ financials?

Yes

8.17 Have you determined that this project is in compliance with all applicable statutes, regulations, policies, standards & procedures, incl. those for network, security, platform, software/application &/or data/info found at aset.az.gov/resources/psp?

Yes

8.17a Describe below the compliance issues that were identified and how they have been/will be resolved, or whether an ADOA-ASET representative should contact you:

8.18 Are there other high risk project issues that have not been identified as part of this PIJ?

No

8.18a Please explain all unidentified high risk project issues below:

9. SECURITY

9.1 Will the proposed solution be vendor-hosted?

No

9.1a Please select from the following vendor-hosted options:

9.1b Describe the rationale for selecting the vendor-hosted option below:

9.1c Has the agency been able to confirm the long-term viability of the vendor hosted environment?

9.1d Has the agency addressed contract termination contingencies, e.g., solution ownership, data ownership, application portability, migration plans upon contract/support termination?

9.1e Has a Conceptual Design/Network Diagram been provided and reviewed by ASET-SPR?

9.1f Has the spreadsheet located at <https://aset.az.gov/arizona-baseline-security-controls-excel> already been completed by the vendor and approved by ASET-SPR?

9.2 Will the proposed solution be hosted on-premise in a state agency?

Yes

9.2a Where will the on-premise solution be located:

Other

9.2b Were vendor-hosted options available and reviewed?

9.2c Describe the rationale for selecting an on-premise option below:

9.2d Will any data be transmitted into or out of the agency's on-premise environment or the State Data Center?

9.3 Will any PII, PHI, CGIS, or other Protected Information as defined in the 8110 Statewide Data Classification Policy be transmitted, stored, or processed with this project?

Yes

9.3a Describe below what security infrastructure/controls are/will be put in place to safeguard this data:

Project security compliance measures are developed by the Arizona Department of Homeland Security, the Arizona Department of Administration, and the Department of Health Services to align with StateRAMP cybersecurity and other information security best practices. This includes but is not limited to:

1. NIST 800-53, Revision 4, and 800-18, Revision
2. ASET 8120
3. FedRamp
4. HIPAA
5. FERPA standards.

These are outlined in the Arizona Strategic Enterprise Technology Division's P8000 Information Security Standards. These standards relate to physical and logical security, network and system security, application security, and privacy and compliance.

10. AREAS OF IMPACT

Application Systems

Database Systems

Data Warehouse/Mart

Software

Hardware

Hosted Solution (Cloud Implementation)

Security

Telecommunications

Enterprise Solutions

Business Intelligence System

Contract Services/Procurements

11. FINANCIALS

Description	PIJ Category	Cost Type	Fiscal Year Spend	Quantity	Unit Cost	Extended Cost	Tax Rate	Tax	Total Cost
AWS Workspaces	License & Maintenance Fees	Development	1	12	\$562	\$6,744	0.00 %	\$0	\$6,744
Azure Dev Ops Cloud Seats	License & Maintenance Fees	Development	1	16	\$79	\$1,264	860.00 %	\$109	\$1,373
Snowflake	License & Maintenance Fees	Development	1	1	\$11,806	\$11,806	0.00 %	\$0	\$11,806
Matillion AWS Infrastructure (m5.large)	License & Maintenance Fees	Development	1	1	\$2,064	\$2,064	0.00 %	\$0	\$2,064
Ancillary Tool AWS Infrastructure	License & Maintenance Fees	Development	1	1	\$4,500	\$4,500	0.00 %	\$0	\$4,500
Splunk (workload credits)	License & Maintenance Fees	Development	1	1	\$2,500	\$2,500	0.00 %	\$0	\$2,500
Carahsoft Professional Services	Professional & Outside Services	Development	1	1	\$841,976	\$841,976	0.00 %	\$0	\$841,976
Visual Studio	Software	Development	1	5	\$299	\$1,495	860.00 %	\$129	\$1,624
Matillion (ETL Tool)	Software	Development	1	1	\$78,000	\$78,000	0.00 %	\$0	\$78,000
Snowflake	License & Maintenance Fees	Operational	2	1	\$11,806	\$11,806	0.00 %	\$0	\$11,806
AWS Hosting etc.	License & Maintenance Fees	Operational	2	1	\$9,705	\$9,705	0.00 %	\$0	\$9,705
Matillion (ETL Tool)	Software	Operational	2	1	\$78,000	\$78,000	0.00 %	\$0	\$78,000
Snowflake	License & Maintenance Fees	Operational	3	1	\$11,806	\$11,806	0.00 %	\$0	\$11,806
AWS Hosting etc..	License & Maintenance Fees	Operational	3	1	\$9,705	\$9,705	0.00 %	\$0	\$9,705
Matillion (ETL Tool)	Software	Operational	3	1	\$78,000	\$78,000	0.00 %	\$0	\$78,000
Snowflake	License & Maintenance Fees	Operational	4	1	\$11,806	\$11,806	0.00 %	\$0	\$11,806
AWS Hosting etc..	License & Maintenance Fees	Operational	4	1	\$9,705	\$9,705	0.00 %	\$0	\$9,705

Matillion (ETL Tool)	Software	Operational	4	1	\$78,000	\$78,000	0.00 %	\$0	\$78,000
Snowflake	License & Maintenance Fees	Operational	5	1	\$11,806	\$11,806	0.00 %	\$0	\$11,806
Matillion (ETL Tool)	Software	Operational	5	1	\$78,000	\$78,000	0.00 %	\$0	\$78,000
AWS Hosting etc..	License & Maintenance Fees	Operational	5	1	\$9,705	\$9,705	0.00 %	\$0	\$9,705

Base Budget (Available)	Base Budget (To Be Req)	Base Budget % of Project
\$386,056	\$0	29%
APF (Available)	APF (To Be Req)	APF % of Project
\$0	\$0	0%
Other Appropriated (Available)	Other Appropriated (To Be Req)	Other Appropriated % of Project
\$0	\$0	0%
Federal (Available)	Federal (To Be Req)	Federal % of Project
\$957,294	\$0	71%
Other Non-Appropriated (Available)	Other Non-Appropriated (To Be Req)	Other Non-Appropriated % of Project
\$0	\$0	0%

Total Budget Available	Total Development Cost
\$1,343,350	\$950,586
Total Budget To Be Req	Total Operational Cost
\$0	\$398,044
Total Budget	Total Cost
\$1,343,350	\$1,348,630

12. PROJECT SUCCESS

Please specify what performance indicator(s) will be referenced in determining the success of the proposed project (e.g. increased productivity, improved customer service, etc.)? (A minimum of one performance indicator must be specified)

Please provide the performance objective as a quantifiable metric for each performance indicator specified.

Note: The performance objective should provide the current performance level, the performance goal, and the time period within which that performance goal is intended to be achieved. You should have an auditable means to measure and take corrective action to address any deviations.

Example: Within 6 months of project completion, the agency would hope to increase "Neighborhood Beautification" program registration by 20% (3,986 registrants) from the current registration count of 19,930 active participants.

Performance Indicators

This project will result in numerous strategic, tactical and operational successes for ADHS. The following measurable success metrics have been established by the project's steering and leadership team to more clearly define the necessary quality of the project's execution.

1. The project's approved budget, schedule, and scope was not exceeded by 10%.
2. 100% of the defined tooling for the operation and support of the Data Lakehouse to support mortality data were operational within 30 days of the final project release.
3. Data load, transform, and data curation performance is a key expectation of a data lakehouse. Therefore, it is the expectation that all developed data loads be accomplished within the currently allotted time frame 99.9% of the time. It is understood that more real time strategies may be employed that make this key success factor not applicable. The expectation of presentation performance is that all newly developed analytical feeds be as performant as the pre-existing feeds (e.g., connection through analytic tools, including Tableau) 100% of the time.
4. 90% of current ADHS end users of the data are able to access the data within the Data Lakehouse, according to their specified access permissions.
5. 90% of ADHS end users have been provided training, or have access to training, to successfully perform their pre-existing activities using Lakehouse data.
5. All current analytics and reports produced for the Bureau of Vital Records have been successfully migrated unless deemed unnecessary within 90 days of the final release.

13. CONDITIONS

Conditions for Approval

Should development costs exceed the approved estimates by 10% or more, or should there be significant changes to the proposed technology scope of work or implementation schedule, the Agency must amend the PIJ to reflect the changes and submit it to ADOA-ASET, and ITAC if required, for review and approval prior to further expenditure of funds.

Monthly reporting on the project status is due to ADOA-ASET no later than the 15th of the month following the start of the project. Failure to comply with timely project status reporting will affect the overall project health. The first status report for this project is due on March 15, 2023.

14. OVERSIGHT SUMMARY

Project Background

The Arizona Department of Health Services (ADHS) is tasked to promote, protect, and improve the health and wellness of individuals and communities in Arizona. The Arizona Department of Health Services promotes and protects the health of Arizona's children and adults. We strive to set the standard for personal and community health through direct care, science, public policy, and leadership. The department operates programs in the following areas; disease prevention, community public health, environmental health, maternal and child health, emergency preparedness, the regulation of childcare centers, assisted living centers, nursing homes, hospitals, other health care providers, and emergency services, ADHS also operates the Arizona State Hospital (ASH), which provides psychiatric services to some of our state's residents in need of heightened care.

ADHS has multiple data sources that are essentially different in kind, which impacts the ability to collect, communicate and verify the data. The current approach is utilizing point-to-point data feeds that solve the individual business needs they are intended to have and increase security risks, potential data collection errors. This slows the agency's ability to effectively and efficiently collect, analyze and share the data. ADOA-ASET believes

the project is inline with the Arizona It Strategic plan and has the ability to offer additional improvements to the agency and state in both the short and long term.

Business Justification

This project will focus on ingesting, transforming, and presenting ADHS mortality data in the new Data Lakehouse environment, improving the agency's response time and accuracy in the data during statewide health emergencies. The project will also set the foundational groundwork, infrastructure, governance, and sustainability processes for future data to be incorporated. The future potential use of the curated data sets for analytics and data sharing for all ADHS divisions as a single source of data.

ADOA-ASET sees this project as following the Arizona Statewide IT Strategy, the solution will provide data sharing improvements, a cloud-first approach and will improve the enterprise security for ADHS data.

Agency Performance Indicators

1. The project's approved budget, schedule, and scope was not exceeded by 10%.
2. 100% of the defined tooling for the operation and support of the Data Lakehouse to support mortality data were operational within 30 days of the final project release.
3. Data load, transform, and data curation performance is a key expectation of a data lakehouse. Therefore, it is the expectation that all developed data loads be accomplished within the currently allotted time frame 99.9% of the time. It is understood that more real time strategies may be employed that make this key success factor not applicable. The expectation of presentation performance is that all newly developed analytical feeds be as performant as the pre-existing feeds (e.g., connection through analytic tools, including Tableau) 100% of the time.
4. 90% of current ADHS end users of the data are able to access the data within the Data Lakehouse, according to their specified access permissions.
5. 90% of ADHS end users have been provided training, or have access to training, to successfully perform their pre-existing activities using Lakehouse data.
5. All current analytics and reports produced for the Bureau of Vital Records have been successfully migrated unless deemed unnecessary within 90 days of the final release.

Implementation Plan

The solution will be hosted within the State network in the via DHS controlled AWS GovCloud environment, and the vendor will only be providing professional services. No AZRAMP or SSP required.

The ADHS Team will provide

Executive Sponsor - Susan Robinson, ADHS Chief Business Intelligence Officer
Product Owner - Laura Erhart, ADHS Informatics Section Lead
Business and Technical SME's
Business SME - Laura Fox, Healthcare Informatics Manager
Enterprise Architect - Nita Surathu, Business Architect
Mortality Data SME - Shobha Vaddireddy, Senior Vital Records Programmer/Analyst
Data Architect - Ravi Pitti, ADHS Chief Data Officer
Data Engineer - Carahsoft
AWS system administrators
Mike Shaw, Senior System Administrator
Alex Warren, System Administrator
BI Engineers - Sridivya Mahankali - Vital Records Analyst
UAT Approvers
Laura Fox, Healthcare Informatics Manager

The Vendor will provide

Accountable Executive- part time
Engagement Lead- part time

Solution Owner
Solution Architect
Cloud Engineer
Data Engineer
Learning and Development Lead- part time
Learning and Development Consultant- part time
Visual Analytics Architect- part time
Quality Assurance Analyst

The role descriptions for the vendor team are available on pg 10-12 of the Task Order Request.

Vendor Selection

The vendor was selected off of state contract -CTR046098. The agency reviewed four professional services' response quotes.

Budget or Funding Considerations

71% of the project development costs will be covered by Federal funding. The remaining 29% will be from the base budget. There are no Taxes added to certain qualifying License & Maintenance Fees due to the tax exemption.

15. PIJ REVIEW CHECKLIST

Agency Project Sponsor

Susan Robinson

Agency CIO (or Designee)

Paula Mattingly

Agency ISO (or designee)

John Stark

OSPB Representative

ASET Engagement Manager

ASET SPR Representative

Emily Gross

Agency SPO Representative

Gina Caster-Corwin

Agency CFO

Marianne Morrow