



ADOA - ASET

Arizona Strategic Enterprise Technology

Project Investment Justification

Version 01.01

A Statewide Standard Document for Information Technology Projects

Project Title:

AHCCCS Data Quality Tools PIJ

Agency Name:	AHCCCS
Date:	April 25, 2014
Agency Contact Name:	Joanne Obenour
Agency Contact Phone:	
Agency Contact Email:	

[Hover for Instructions](#)

I. Management Summary*

The agency's vision is to transform the existing Data Warehouse and Decision Support System (DW/DSS) into a mature Information Infrastructure (II) that seamlessly integrates disparate data to facilitate emerging requirements of current and future information needs such as those for the Health Information Technology for Economic and Clinical Health (HITECH) Act, Patient Protection and Affordable Care Act (PP&ACA), International Classification of Diseases Version 10 (ICD-10), Health Information Exchange (HIE), Health Insurance Exchange (HIX), and Quality Measures.

The DW/DSS, implemented in 2005, introduced the agency to a common repository of Medicaid data for analysis and reporting. This was a major step in addressing the need for usable, accessible, flexible, and timely information through a user-friendly interface. The data structure was consistent with that of the MMIS with few changes or transformations. While this solution allowed non-programmer access to the data, it also required expert understanding of the data and business rules to produce meaningful results. In the end, the business users were able to create and run simple queries leaving the more complex queries to the programmers as before.

By 2009 the DW/DSS had been transitioned from Business Objects to COGNOS and expanded to incorporate Hawaii data. This step strengthened the security of the data, provided geographic redundancy, and enabled remote access which allowed more staff to work from home and become part of the Virtual Office and Telework program. The data structure was not changed. With limited training in COGNOS, the business users continued to pass the complex queries to the programmers.

More recently in 2012, as part of the HIPAA 5010 implementation, the claims and encounter data was remodeled through the use of an Extract Transform and Load (ETL) tool made available through the MACBIS project. The other data in the DW/DSS is currently going through the same transition. Implementing an ETL tool provides for a common set of business data rules, formats, and values to be established that can be applied across all of the enterprise data, and new agency-level data fields can be defined for consistent application of the data; this eliminates redundant data definition and adds consistency to commonly used terms.

While all of these changes and enhancements were being implemented, the DW/DSS was increasingly becoming a more critical tool and source of the essential program data used to manage the Medicaid programs. Business and IT personnel increasingly rely on these tools to provide ad-hoc, day-to-day, periodic as well as state/federally mandated reporting.

During 2013, members of the Data Warehouse development and operational team, Data Warehouse super users, and several consultants with expertise in enterprise data analytics and systems including

representatives from the existing system vendor met with business areas across the agency including clinical areas such as health plan operations, quality management, medical management, and fee for service management to determine reporting needs and identify any issues in using the data warehouse to meet those needs. The team also met with data warehouse experts to analyze the existing infrastructure both from a technical view and from the reporting needs of the business areas, knowing that the desire was to leverage the existing hardware and software.

The following table shows our findings and how we plan to resolve the problem areas.

Findings	Problem	Requirements	IAPD Project
Difficult to effectively access data	Process of developing reports requires detailed understanding of operational source systems	Develop an integrated data infrastructure to enable seamless cross-content analysis	Metadata Management
Lack of understanding the data, need for clear definitions	Little or no written definitions; data names vary across systems	Deploy Metadata Management tools to create an enterprise Metadata Library including lineage, transformations and definitions	Metadata Management
Need to create and define new data fields for commonly used information	Each report must combine data to obtain value for commonly used information	Develop normalized, codified data elements to enable consistent reporting	Metadata Management
Too dependent on people knowledge	Only data experts know the data and how it fits together	Create a common vocabulary to describe, document and identify Enterprise data	Metadata Management
Limited confidence in data; no standard defined data quality control processes	Source data is not validated for completeness, consistency or accuracy	Deploy data quality control solution to proactively analyze, cleanse and codify incoming legacy system data sources	Metadata Management

Findings	Problem	Requirements	IAPD Project
Lack of confidence in the validity of data	Little or no methods or processes to confirm/test the validity of source and target data	Deploy an integrated Data Quality solution compatible with existing tools and infrastructure	Environmental and Infrastructure
No way to display data geographically	Lack of coding and mapping tools that can relate data to its geographic location	Deploy an integrated Geocoding/Geomapping solution compatible with existing tools and infrastructure	Environmental and Infrastructure
Limited baseline reports are run for comparison and trending over time; each report run requires validation	Need set of standard reports that are run and reviewed on a regular basis	Develop standard set of intra-agency reports and dashboards	Best Practice in Design, Development and Reporting
Reporting process is complex; requires significant understanding of the data and the reporting tools; no simple examples to follow	Reports created without solid knowledge of the data may produce unpredictable results and require significant validation	Define effective report development methodologies and practices using normalized, codified data elements	Best Practice in Design, Development and Reporting
Need to combine unlike sets of data	Difficult to combine disparate data sources into single query	Develop integrated dimensionally-modeled data architecture to enable cross-content reporting and analysis	Conformed Business Data Marts
Need timely access to data	Data is only updated periodically creating a data latency issue for users	Develop new ETL processes to acquire data in timeframes more consistent with business needs	Conformed Business Data Marts

Findings	Problem	Requirements	IAPD Project
Analytical tool very complex to use and does not satisfy business needs	Current tool is a home-grown solution that is not well supported, not easily modified or extensible	Deploy an integrated, standard Health Quality solution	Quality Initiative Reporting

This PIJ proposes to implement Data Quality Tools to address the Environmental and Infrastructure findings, and specifically address the quality of address information as a first step in improving the quality of the data.

II. Project Investment Justification (PIJ) Type*

Yes No Is this document being provided for a Pre-PIJ / Assessment phase?

If Yes,

Identify any cost to be incurred during the Assessment phase.	\$
Based on research done to date, provide a high-level estimate or range of development costs anticipated for the full PIJ.	\$

Explain:

[Click here to enter text.](#)

Yes No Will a Request for Proposal (RFP) be issued as part of the Pre-PIJ or PIJ?

III. Business Case

A. **Business Problem***

As AHCCCS identifies current and future objectives to improve program performance, providing appropriate, cost-effective care and improving outcomes, it is critical that source data is correct, complete, accurate, and trusted before it is loaded in the Data Warehouse, as the Data Warehouse is the source for agency analysis and reporting. The AHCCCS Data Warehouse is a repository of data from many other systems both internal and external, and each system has its own unique set of data edits. We have found that the entry of the data in these systems introduces lots of inconsistencies and errors especially with addresses of members and providers. These inconsistencies make it virtually impossible to adequately map members across the state and determine whether available providers are located within reasonable distance of where their patients reside. AHCCCS must be able to monitor health plan compliance of this new contract requirement, and proposes

implement tools to identify discrepant address data and facilitate the integrity of future data across the AHCCCS Data Warehouse.

B. Proposed Business Solution*

AHCCCS proposes to implement data quality tools to the Data Warehouse that will identify discrepancies in address data as it is being loaded in the Data Warehouse, and will add a geographic locator for each address. Any data discrepancies will be reported to the data owners for correction. The solution consists of the following:

- Informatica Data Quality (IDQ) – provides the capabilities needed to discover, search, profile, standardize, build rules, deploy rules, create scorecards, manage exceptions and more.
- IDQ Accelerator – pre-built functionality to address common data quality issues.
- Data Quality Identity Match – component of geo coding
- Lab – interface to IDQ
- Standard Year 1 Maintenance of software value
- United States Premium Address Cleansing — component of geo coding
- United States Point Premium Geo Coding – component of geo coding
- Server – an additional server is needed for the data quality tools
- Consultant Services – consultant expertise will be needed to plan and implement the Data Quality tools

C. Quantified Benefits*

- | | |
|-------------------------------------|---------------------|
| <input checked="" type="checkbox"/> | Service enhancement |
| <input type="checkbox"/> | Increased revenue |
| <input type="checkbox"/> | Cost reduction |
| <input type="checkbox"/> | Problem avoidance |
| <input type="checkbox"/> | Risk avoidance |

Explain:

The Data Quality tools will provide the ability to proactively monitor and cleanse source data across the enterprise and promote collaboration between business and IT. The solution enables Geo-coding/Geo-mapping capabilities that can provide comprehensive reports and maps to represent network sufficiency and accessibility. Other uses of these tools include the enhanced ability to identify provider and member activity by geo-location, disaster response capability, and fraud detection capabilities. The agency’s acute care contracts require that health plans have contracted providers within a certain geographic distance from each of their members, but without valid addresses in consistent format the agency cannot easily monitor contract compliance with this regard. These tools will enable the agency to more adequately monitor compliance and improve confidence in the data.

AHCCCS will leverage these new capabilities to monitor quality care and target best practices to areas/groups of people who may be underserved. Specifically, existing capabilities for examining and tracking quality-of-care measures for Medicaid can be expanded to focus on vulnerable populations to determine if performance for different groups varies by enrollee race/ethnicity, socioeconomic status (SES) and/or geo-location, to name just a few. Additionally, Geo-coding/Geo-mapping provides the ability to calculate distances, or time-to-travel that may be used to highlight potential barriers to care.

IV. Technology Approach

A. *Proposed Technology Solution**

The proposed solution adds Data Quality tools so that source data will flow from the ETL tool through IDQ into the Data Warehouse creating new tables for the geographic locator information. Any data discrepancies found will be captured and reported to the data owners for correction. The corrected data will be extracted during the next ETL cycle and flow through the IDQ into the Data Warehouse. Over time, new rules may be added to the ETL process for automatic correction of common data discrepancies.

-

B. *Technology Environment*

The current Data Warehouse environment contains multiple servers:

- The Cognos reporting tool utilizes 3 servers (2 dispatches and 1 load balancer), and an additional development server
- Informatica ETL (extract, transform, and load) utilizes 2 servers (development and production)

Oracle (v 11.2) has 3 databases:

- New Production (prd11)
- Old Production (azdssp1) – being transitioned into a new structure
- Development (azdssa1)

Informatica ETL tools:

- Power Center (v9.5.1) runs in a Linux environment
- Power Exchange (v9.5.1) is mainframe connection
- Informatica Power Center (v5.2.1)

This project adds:

- 1 new virtual server to run the Data Quality tool (standard network configured two-processor virtual server host with VM vSphere Enterprise software, Backup software, Microsoft Data Center for each processor, and Oracle Linux license)
- Informatica Data Quality (IDQ) which includes Address Doctor for Geo-Coding

Solution is fully compatible with the existing Data Warehouse

C. Selection Process

Approach was to work with members of the Data Warehouse development and operational team, Data Warehouse super users, and several consultants with expertise in enterprise data analytics and systems including representatives from the existing system vendor to identify the needs and issues and recommend solutions that would leverage the existing hardware and software.

V. Project Approach

A. Project Schedule*

Project Start Date: 4/1/2014 **Project End Date:** 9/30/2014

B. Project Milestones

Major Milestones	Start Date	Finish Date
Acquisition	4/1/2014	5/1/2014
Installation for Testing	5/1/2014	6/1/2014
Component Testing	5/1/2014	6/1/2014
System Testing	5/1/2014	6/1/2014
Implementation	6/1/2014	6/30/2014
Post Implementation Support	7/1/2014	9/30/2014

VI. Roles and Responsibilities

A. Project Roles and Responsibilities

The Business Intelligence Team will work together to test and implement these products.

B. Project Manager Certification

- Project Management Professional (PMP) Certified
- State of Arizona Certified
- Project Management Certification not required

C. Full-Time Employee (FTE) Project Hours

Total Full-Time Employee Hours	500
Total Full-Time Employee Cost	\$25,000

VII. Risk Matrix, Areas of Impact, Itemized List, PIJ Financials

VIII. Project Approvals

A. Agency CIO Review*

Key Management Information	Yes	No
1. Is this project for a mission-critical application system?	X	
2. Is this project referenced in your agency's Strategic IT Plan?	X	
3. Is this project in compliance with all agency and State standards and policies for network, security, platform, software/application, and/or data/information as defined in http://aset.azdoa.gov/security/policies-standards-and-procedures , and applicable to this project? If NO , explain in detail in the "XI. Additional Information" section below.	X	
4. Will this project transmit, store, or process sensitive, confidential or Personally Identifiable Information (PII) data? If YES , in the "XI. Additional Information" section below, describe what security controls are being put in place to protect the data.	X	
5. Is this project in compliance with the Arizona Revised Statutes (A.R.S.) and GRRC rules?	X	
6. Is this project in compliance with the statewide policy regarding the accessibility to equipment and information technology for citizens with disabilities?	X	

B. Project Values*

The following table should be populated with summary information from other sections of the PIJ.

Description	Section	Number or Cost
Assessment Cost (if applicable for Pre-PIJ)	II. PIJ Type - Pre-PIJ Assessment Cost	\$
Total Development Cost	VII. PIJ Financials tab	\$477,660.93
Total Project Cost	VII. PIJ Financials tab	\$870,168.93
FTE Hours	VI. Roles and Responsibilities	1000

C. Agency Approvals*

Contact	Printed Name	Signature	Email and Phone
Project Manager:	Craig Srsen		
Agency Information Security Officer:	Jim Wang		
Agency CIO:	Jim Wang		
Project Sponsor:	Jim Wang		
Agency Director:	Tom Betlach		

IX. Optional Attachments

A. *Vendor Quotes*

X. Glossary

XI. Additional Information

The AHCCCS Data Warehouse uses role-based security so that participants can only access what they need to perform their job functions.

Links:

[ADOA-ASET Website](#)

[ADOA-ASET Project Investment Justification Information Templates and Contacts](#)

Email Addresses:

[Strategic Oversight](#)

ADOA-ASET_Webmaster@azdoa.gov