

RFP Functional Technical Requirements

January 13, 2014

ANSWERS Functional and Technical Requirements

The purpose of this document is to provide general functional and technical requirements to more fully explain the primary components of the ANSWERS system and the processes the contract technical resources will be required to create through their involvement in the project.

The system design and maintenance will focus on flexibility and scalability to be able to readily and efficiently respond to stakeholder needs. The functional and technical requirements identified are based on the best information available at the time of the RFP development but are evolving and subject to change.

The first part of this document addresses the functional requirements. Flow charts, which are attached, have been developed to give a graphical depiction of the SLDS process and the various databases and applications necessary to achieve the goals of the project. In order to better convey the process to the offeror, we have summarized the major components of the flow charts and included the high level requirements necessary for development. This summary is meant to serve as a high level informational overview and provide offerors a better understanding of the types of activities their teams will be asked to perform. Detailed requirements will be defined prior to the start of development.

The second part of this document addresses the technical requirements and the technology environment in which the databases and applications will be developed. Again, this is presented at a high level for the purposes of familiarizing the offeror with the environment in which the contract technical resources will develop the core components of the ANSWERS SLDS.

Functional Requirements and Flow Chart Overviews

There are seven flow charts provided to depict the various databases, applications and functionality that will need to be developed to support the ANSWERS SLDS processes. There are requirements implicit in each process that the flow charts depict; some of the requirements span across all flow chart processes, these are referred to as ‘global requirement’. Some requirements are specific to a particular process depicted in the flow charts. These requirements are identified in the requirements table below using the flow chart number as reference (ex: FCI-1 is a requirement associated with Flow Chart I).

The remainder of this section will highlight the requirements at a high level. Once the contract technical resources have been selected, ANSWERS staff will work with them to define more detailed requirements suitable to be used in devising a technical design.

| | |
|---|--|
| <p><i>Flow Charts I and II.</i> Represent the major components to be developed; Pre-Processing/Validation, ANSWERS Identity Management System (AIMS) for Matching and Identity Management, the Staging Environment, and the Data Warehouse Environment. The flow charts also depict the processes that will take place within each environment and the manner in which the data are initiated for load into the system. Below are the global functional requirements necessary for development of the four major components. Specific requirements to address the processes as depicted in flow charts IV-VII will follow the global requirements.</p> | |
| Global Requirement | G-1: The ability to provide a secure web service for data providers to use in submitting data to ANSWERS |
| | G-2: The ability to provide access to the secure web service from inside and outside the ANSWERS environment |
| | G-3: The ability to provide a secure login process for data providers and other users of the ANSWERS system <ul style="list-style-type: none"> • Data Providers • ANSWERS Staff |
| | G-4: The ability to run processes in non-upload mode to allow for data clean-up and possibly other processes prior to loading the data |
| | G-5: The ability to allow data to be submitted in variety of file formats |
| | G-6: The ability to customize the format in which the data should be submitted <ul style="list-style-type: none"> • Each data provider should have a specified format for their data to be submitted • Source documents have been provided by the data providers to assist with this requirement |
| | G-7: The ability to allow the data provider to submit multiple data files in succession during the same login session |
| | G-8: The ability to provide electronic feedback to the data provider about the status of the data submitted <ul style="list-style-type: none"> • Submission successful • Submission failed • Submission pending, etc. |
| <p><i>Flow Chart III – Pre-Processing/Validation</i> this flow chart depicts the process of validating the submitted data files through two different validations, reporting any ‘failures’ to the data provider, and allowing the iterative process to continue until all files pass validation and are ready to proceed to AIMS for matching. Below are the functional requirements to support this process.</p> | |
| Flow Chart III (FCIII) | FCIII-1: The ability to validate incoming data files against expected values and rules as outlined in the data dictionary (See attachment 8 – Data Source Summary/ANSWERS Codebook Sample) <ul style="list-style-type: none"> • The data dictionary will be an application that needs to be developed to provide the rules that will be called upon at the time of validation (see data dictionary database (TR-11) below for further technical requirements) |
| | FCIII-2: The ability to have multiple validation levels <ul style="list-style-type: none"> • First validation will validate that the data elements are being submitted in the expected format per rules in the data dictionary (Ex: data length, data type, etc.) • Second validation will validate that the data file is reasonable as |

| | |
|--|---|
| | compared to other data elements in the files per the rules in the data dictionary (Ex: if a student has a graduation date they must have a degree type, etc.) |
| | FCIII-3: The ability to easily update and enhance validation rules |
| | FCIII-4: The ability to require that data pass the expected validation prior to moving to the next process |
| | FCIII-5: The ability to provide real time feedback about the quality of the data <ul style="list-style-type: none"> • Identify which data do not pass the validation tests and why • Identify how many records do pass the validation tests |
| | FCIII-6: The ability for a data provider to resubmit data files after corrections have been applied to data that failed validation tests |
| | FCIII-7: The ability for a data file that passes validation to move to the next process <ul style="list-style-type: none"> • For most data sources the next process will be matching |
| <i>Flow Chart IV Matching Detail.</i> This flow chart represents the processes necessary to accept data from many different sources and match it against previously received data, matching data about the same person or entity and then assigning the same ID to those records so they will be linked in the system or creating a new ID if there are no matching data that already exists in the system. A master index will be created representing each unique individual or entity in the system. | |
| Flow Chart IV (FCIV) | FCIV-1: The ability to apply various matching algorithms to compare an incoming record against existing records in the AIMS system <ul style="list-style-type: none"> • Examples of matching algorithms data elements are: Agency ID, Name, Date of Birth, Social Security Number • Detailed algorithms will be supplied prior to development |
| | FCIV-2: The ability to require exact matches of data elements in the matching process |
| | FCIV-3: The ability to match records based on a combination of element matches <ul style="list-style-type: none"> • Based on the incoming data source |
| | FCIV-4: The ability to perform ‘fuzzy’ matches of data and accommodate common typographical errors <ul style="list-style-type: none"> • Ex: - The names match to 90% accuracy or the first and last names are transposed |
| | FCIV-5: The ability to easily configure and change the matching algorithms |
| | FCIV-6: The ability to create an extracted file of personally identifiable information which can be provide to another data source to add in additional elements for the individuals included in the file |
| | FCIV-7: The ability to identify and correct matching <ul style="list-style-type: none"> • Disconnect or split records apart that were matched in error • Run the records back through the match process to identify potential correct matches • Connect or link records that were not matched previously but should have been |
| | FCIV-8: The ability to create an audit log of all activity associated with matching and un-matching (in the case of errors) records |
| | FCIV-9: The ability to create reports about the matching results |

| | |
|---|--|
| | <ul style="list-style-type: none"> • How many records matched at each algorithm level • How many records did not match and a new record was created • Total records read • Total records updated • Etc. |
| | FCIV-10: The ability to append Personally Identifiable Information (PII) of matched records to the AIMS Person Record and pass the record to the ID Management process |
| | <p>FCIV-11: The ability to create a new person record when no matching record exists in the AIMS system</p> <ul style="list-style-type: none"> • Assign a unique AIMS ID • Create a unique ANSWERS ID (See AIMS to ANSWERS ID Database (TR-12) below for further technical requirements) • Add the PII to the Person Record • Move the new Person Record to the ID Management process |
| | <p>FCIV-12: The ability to search the AIMS system for records that have missing SSNs and create an extract to send to the Alaska Permanent Fund Dividend program and /or the Institutional Student Information Record (ISIR) to attempt to locate the SSN</p> <ul style="list-style-type: none"> • Once the SSN is identified it would be loaded to the AIMS system |
| <p><i>Flow Chart V AIMS Identity Management.</i> This flow chart depicts the process for de-identifying records by removing the PII fields and replacing the AIMS ID (which is only used within the identity management system and is not shared outside the AIMS system) with an ANSWERS ID, which is the key used within the ANSWERS data warehouse to connect data about the same person together. Additionally each file is assigned a series of batch identifiers that will be needed in the future to identify a particular batch for an adjustment or a correction. The ID Management process occurs in the web service up until the point the records are transmitted to the staging area.</p> | |
| Flow Chart V (FCV) | <p>FCV-1: The ability to create and assign batch identifiers to each record and batch that enters the AIMS ID system</p> <ul style="list-style-type: none"> • To be used to easily remove entire batches of data from the database due systemic issues • To allow individual records that were matched in error to be easily found and corrected • Batch identifiers should include: <ul style="list-style-type: none"> ○ Reporting year the data applies to (2011, 2012, etc.) ○ Specific reporting period within the year if applicable due to some data providers reporting at a time period other than annual, i.e., monthly ○ Date of the data upload ○ Batch ID of the specific data load ○ Flag to indicate that the data are either in a preliminary state (not ready for review) or final state (ready for review) ○ As of date which will indicate the data provider's internal approval or release of data deeming it good to be used by researchers ○ Moved date to indicate when the data entered ANSWERS ○ Other identifiers as needed |

| | |
|--|---|
| | <p>FCV-2: The ability to remove PII elements from each record prior to load into the staging area</p> <ul style="list-style-type: none"> • Name • Social Security Number • Etc. |
| | <p>FCV-3: The ability to replace the AIMS ID with an ANSWERS ID for each record entering the staging area (See AIMS to ANSWERS ID Database (TR-12) below for further technical requirements)</p> |
| | <p>FCV-4: The ability to create detail and summary reports</p> <ul style="list-style-type: none"> • Number of records read • Number of records updated • Number of records failed • Total number of records processed • Etc. |
| | <p>FCV-5: The ability to load batches of data to the staging area</p> <ul style="list-style-type: none"> • Data with batch identifiers and ANSWERS ID – no PII or AIMS ID |
| <p><i>Flow Chart VI Staging Detail.</i> This flow chart details the processes necessary to stage the data for loading into the data warehouse. Rational validation occurs to compare and identify discrepancies in person level characteristics for existing records such as the same individual with multiple graduation years for high school or different demographic data. Tolerance rules will be applied to determine if the discrepancies warrant stopping the process for further action to reconcile the differences prior to loading to the data warehouse. Reports are generated about the discrepancies at both a detail and summary level.</p> | |
| Flow Chart VI (FCVI) | <p>FCVI-1: The ability to use the data provider staging tables to compare records with existing records in the database using the ANSWERS ID as the link to associate records and identify discrepancies about specific fields</p> <ul style="list-style-type: none"> • Compare person level characteristics such as demographics, high school graduation date, school attended, etc. |
| | <p>FCVI-2: The ability to apply rules to the checking process as described in FCVI-1 and enforce an action to continue or stop processing depending on the discrepancy tolerances defined</p> <ul style="list-style-type: none"> • EX: if 50% of the records have a discrepancy in the high school graduation year stop processing |
| | <p>FCVI-3: The ability to process corrections for any discrepancies or errors found in the data</p> |
| | <p>FCVI-4: The ability to create summary and detail reports about the discrepancies found in the staging rational validation process</p> <ul style="list-style-type: none"> • List of data elements in question • Source of the existing data element • Number of records read • Number of records processed • Number of records that failed validation • Number of records updated |
| | <p>FCVI-5: The ability to close the staging environment for file acceptance while moving data to the data warehouse</p> |
| | <p>FCVI-6: The ability to extract, transform, load(ETL) the data to the existing data warehouse source tables as tables are validated</p> |
| | <p>FCVI-7: The ability to determine if the data were successfully loaded to</p> |

| | |
|--|--|
| | the data warehouse source tables |
| | FCVI-8: The ability to wipe the staging environment or individual staging tables and ready them for the next data load when verification is received that the data loaded successfully |
| <i>Flow Chart VII Data Warehouse Cohort (Person) Table Details.</i> This flow chart depicts the process that loads data from the staging tables to the data warehouse tables. | |
| Flow Chart VII (FCVII) | FCVII-1: The ability to perform the ETLs on a pre-determined basis upon agency approval from the staging area by table |
| | FCVII-2: The ability to update existing warehouse data with incoming data based on pre-determined rules <ul style="list-style-type: none"> Rules would determine which data provider's data elements would take precedence over data from other sources for data represented in the person table, e.g., the most recent record gets the highest ranking because it is the most current |
| | FCVII-3: The ability to record all historical changes to the Person Table |
| | FCVII-4: The ability to record historical snapshots of the Person Table on a periodic basis to allow analysis into what the record reflected during a given timeframe |
| | FCVII-5: The ability to rebuild the Person Table when corrections are made in the underlying data provider historical data. Once the Person Table has been rebuilt need the ability to compare against existing record and document corrections in an audit record. |
| | FCVII-6: The ability to transform common data elements to a standardized version using the ANSWERS ID as the record identifier <ul style="list-style-type: none"> Dates for enrollment, exit, school year, reporting period, fiscal period, graduation, etc. Birth date may be blurred to ensure the confidentiality of the data Statuses such as enrollment, marital, graduation General location such as city, state, country, zip code Citizenship Residency Demographics Grade level Pre-identified groupings such as urban/rural, legislative districts, regions, etc. Etc. |
| | FCVII-7: The ability to accommodate look-up tables for codes stored in the data warehouse <ul style="list-style-type: none"> Locations such as cities, states, etc. Course identifies for postsecondary data (CIP Codes) Etc. |
| | FCVII-8: The ability to update the Person Table |
| | FCVII-9: The ability to create detailed transform reports of discrepancies and errors |
| | FCVII-10: The ability to keep historical records of all updates to the data source tables |
| <i>Technical Requirements.</i> Describe the environment in which the ANSWERS SLDS should be developed and operate. | |
| Technical Requirements (TR) | TR-1: The ability for the ANSWERS SLDS to be developed in and |

| | |
|-----------------------------|--|
| | <p>operate in the Microsoft SQL Server 2012 environment using additional development tools and software:</p> <ul style="list-style-type: none"> • Basic RDBMS • OLTP Database • OLAP Tools • Advanced Encryption • High Performance OLAP • .Net Framework • Data Extract Transformation and Loading Tools • Business Intelligence Tools • Reporting Tools |
| | <p>TR-2: The ability to apply security to all modules of the SLDS system and ensure that all data are protected</p> <ul style="list-style-type: none"> • Must adhere to the State of Alaska’s security policies • Secure log-on • Encrypted databases • Encrypted data as it moves through the various SLDS components • Encrypted data quality feedback reports |
| | <p>TR-3: The ability of the system to be easily scaled to accommodate additional years of data and additional data sources which may be added over time</p> |
| | <p>TR-4: The ability to implement a backup and recovery plan</p> |
| | <p>TR-5: The ability to accommodate multiple concurrent users</p> |
| | <p>TR-6: The ability to establish different security levels for users in defined roles</p> <ul style="list-style-type: none"> • Passwords must adhere to the State of Alaska’s policies for difficulty and expiration • Must contain a self-provisioning module that allows for retrieval of lost passwords or for users to easily reset their passwords • Must support encryption of passwords during data transmission and storage • Provide a mechanism to automatically locate and retire inactive accounts • Provide provisioning and de-provisioning of user accounts that can be scheduled ahead of time |
| | <p>TR-7: The ability to generate an audit trail of all activity taking place at any point in the SLDS system</p> |
| | <p>TR-8: The ability to support access to the data using a reporting tool</p> |
| | <p>TR-9: The ability to provide user help tools</p> <ul style="list-style-type: none"> • Data layout templates to be presented at the time of data load • Etc. |
| | <p>TR-10 The ability to provide technical documentation</p> |
| Other Database Descriptions | <p>TR-11: Data Dictionary Database- The ability to be used for multiple purposes throughout the data lifecycle, including</p> <ul style="list-style-type: none"> • Validation of incoming data feeds • Validation of lookup codes • Update and maintenance of the table structure in ANSWERS to reflect the data feed |

| | |
|--|---|
| | <ul style="list-style-type: none">• Creation of mappings for standard versions of fields and lookup codes• Creation of common demographic and residence tables and mappings• Identify Common Education Data Standards (CEDS) mappings for fields and lookup codes• Provide data definitions and meta data needed by system users to develop reports and information products |
| | <p>TR-12: AIMS to ANSWERS ID Database- The ability to have a one to one record containing an AIMS ID and the associated ANSWERS ID</p> <ul style="list-style-type: none">• Whenever a new AIMS ID is needed the corresponding ANSWERS ID will also need to be created.• Secure database not open to agency staff or researchers. |