Integrity Data Hub (IDH) Overview Presentation

July 26, 2018
1:00-2:00 PM EST

http://integrity.naswa.org
Webinar Administration/Ground Rules

• Session will be recorded and posted
• We will respond to questions on a preliminary basis
  • Please hold question until the end – raise hand function
  • Official responses will be posted to questions received via email
• Questions should be submitted electronically to DataHubRFP@naswa.org
• All questions and official responses will be posted on the RFP webpage at: http://www.itsc.org/Pages/datahubrfp.aspx
Agenda

• To provide background information
  • UI Integrity Center (the Center)
  • Integrity Data Hub (IDH) project
• To review the IDH Request for Proposal
  • Purpose
  • Response requirements
  • Solicit feedback/questions
• To review the Data Hub technical architecture
UI Integrity Center

• In 2012, USDOL selected New York State Dept. of Labor (NYSDOL) to lead the development of the UI Integrity Center

• NYSDOL subsequently passed leadership responsibilities for the Center to NASWA/CESER

• The Center is charged with developing:
  • “innovative UI program strategies to reduce improper payments, prevent and detect fraud and recover any improper payments made” (UIPL 28-12).

• Additional information about the Center
  • http://naswa.org/integrity/
Integrity Data Hub (IDH)
Expanding the Data Hub

• Future enhancements
  • Filtering function
  • Partial IP addresses

• Future capabilities
  • Suspicious green-dot card accounts
  • Fraud alerting
  • Multi-state database

• Participation
  • Engage all State Workforce Agencies
  • Expanded use benefits all participants
Background

• Many states collecting information on suspicious actors
  • IP addresses
  • Email addresses/domains
• Emergent need to focus on prevalence of ID theft
• Build a repository to allow states to submit and retrieve information on suspicious actors
  • Allow participation of states with varying levels of technology and volume
  • Minimize potential issues with data sharing
  • Develop a foundation on which to achieve larger IDH vision
1. Pilot states submit information on suspicious actors.
2. Pilot states provide similar information initial/weekly claims.
3. System compares claims data fields to suspicious actors to identify matches and sources.
5. System deletes states' claims data after matching.
Submit/Lookup Processes

- SAR accommodates different methods for collecting data
  - Manual
    - Data entry
    - Upload
  - System-to-System
    - Secure FTP (sFTP)
    - Web service
Data Hub Key Functionality

• Submit
• Lookup
• Modify
• Reporting
  • Administrative
  • Analytical
• Administration
• User access/roles
Data Hub Documentation

• Participation agreement
  • Not to be used to auto-block claims
  • Specify SAR state administrator
• Fact sheet
• User guide
• Technical info and products
  • SAR FTP and web service integration guides
  • Clients
    • Java client/.NET client
• Data transmission and security summary
Data Hub Status

• Data Hub Phase 1 is complete
  • SAR launched Fall 2017
    • SAR data has been loaded into database (over 13,000 suspicious actors)
    • Working with ~26 states to integrate in UI systems and processes

• Data Hub Phase 2 initiatives
  • SAR enhancements
  • Ongoing security assessment/testing
  • SAR application and database monitoring/optimization
  • Expanded capabilities
IDH RFP Purpose

• To identify industry partners to augment the Center project team in the ongoing development of Data Hub Phase 2 capabilities

• The Center is seeking assistance in the following areas:
  • Project Strategy/Management
  • Software Engineering/Development
  • IT Security
  • System Monitoring/Administration
  • Requirements Development
  • Data Architecture/Management
  • Database Architecture/Management
  • ETL Management
  • Program Analysis
  • System Architecture
RFP Structure/Process

Strategic Support
- Tech Vendor
  - 1 Award

Dev. Env. Support
- Tech Vendor
  - 1 Award

Rqmts. Mgmt. Support
- Tech Vendor
  - 1 Award

Development Support
- Tech Vendor
- Tech Vendor
- Tech Vendor
  - Multiple Award
Phase 2 RFP Elements

• Strategic Support
  • Weekly project status meetings
  • Periodic virtual/on-site meetings and working sessions
  • Quarterly strategic planning sessions
  • Preparing, reviewing and maintaining project plans and documents

• Development Environment
  • Assessment
  • Implementation Support
    • Tools (emphasis on open-source)
    • Processes and Methods
    • Standards
    • Project Governance
Phase 2 RFP Elements

- Business Analysis/Requirements Management
  - Requirements gathering sessions
  - Interviewing state end users and other stakeholders
  - Use cases and design documentation
  - Test plans and procedures
  - Planning and coordinating testing

- Software Development
  - Software engineering/development
  - System architecture review/assessment
  - Data architecture/management
  - IT and data security
  - System monitoring/administration
  - Database architecture/management
  - Extract/Transform/Load (ETL) management
RFP Structure/Process

Strategic Support

Tech Vendor

Spec + RFQ

Development Support

Tech Vendor

Tech Vendor

Tech Vendor

Quote(s)

Tech Vendor

Task Order
## RFP Timeline

<table>
<thead>
<tr>
<th>Project Activity</th>
<th>Timeline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Hub RFP Webinar</td>
<td>26 Jul 18</td>
</tr>
<tr>
<td>Final Clarification Questions</td>
<td>2 Aug 18</td>
</tr>
<tr>
<td>Questions and Responses Posted</td>
<td>9 Aug 18</td>
</tr>
<tr>
<td>Proposals Due</td>
<td>7 Sep 18</td>
</tr>
<tr>
<td>Offeror Presentations</td>
<td>Week of 17 Sep 18</td>
</tr>
<tr>
<td>Best and Final Offer Pricing (optional)</td>
<td>28 Sep 18</td>
</tr>
<tr>
<td>Award (anticipated)</td>
<td>12 Oct 18</td>
</tr>
</tbody>
</table>
RFP Response (1 response per organization)

• Company Overview – 1 page
  • Brief description of the company, products/services, size and Point of Contact (PoC) for communication

• Project Summary Citations – Up to 3 citations – 3 pages total each citation
  • Experience in providing strategic/technical support for projects of similar content, size and scope to the Data Hub
  • Include: summary, size/scope, initial/final budget, agency, agency PoC
RFP Response (per SOW area)

• Technical/Management Approach – 10 pages per SOW area
  • How will personnel from your organization be selected to provide selected services?
  • How will you organize and manage providing strategic services?
  • How will you ensure availability of key staff?
  • What deliverables will you develop and/or maintain?

• Key Personnel Resumes – 3 resumes, 2 pages, per SOW area
  • Provide three resumes (two pages maximum per resume) for key personnel to be assigned to the project for each SOW area
  • The same resume may be provided for multiple SOW areas
  • Include: name, proposed labor category, percentage of time allocated to the Data Hub project, and relevant work experience
RFP Response

• Cost
  • Complete cost template sections as appropriate

**Part A: Program Strategic Services – Firm Fixed-Price**
Provide annual firm-fixed pricing for up to 30 hours per month for program strategic services.

<table>
<thead>
<tr>
<th>Base Year</th>
<th>Option Yr. 1</th>
<th>Option Yr. 2</th>
<th>Option Yr. 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Part B: Program Strategic Services – Firm Fixed-Price/Time & Materials**

Provide a firm fixed-price for conducting a comprehensive assessment and preparation of a detailed report providing recommendations for establishing an independent, stand-alone distributed development environment within the Center. Also, please provide T&M pricing for the labor categories listed to support implementation.

**Assessment Report and Recommendations**

<table>
<thead>
<tr>
<th>Deliverable</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comprehensive assessment report and recommendations</td>
<td></td>
</tr>
</tbody>
</table>

**Implementation Support**

<table>
<thead>
<tr>
<th>Labor Category</th>
<th>Education &amp; Experience</th>
<th>Base Yr. Labor Rate</th>
<th>Option Yr. 1 Labor Rate</th>
<th>Option Yr. 2 Labor Rate</th>
<th>Option Yr. 3 Labor Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sr. Implementation Specialist</td>
<td>MS + 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implementation Specialist</td>
<td>BS + 5, MS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jr. Implementation Specialist</td>
<td>BS + 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Part C: Requirements Development/Business Analysis Services – Firm Fixed-Price/Time & Materials**
Provide T&M hourly rates for the following labor categories for base period and all option periods. Please use a 3% escalation.

<table>
<thead>
<tr>
<th>Labor Category</th>
<th>Education &amp; Experience</th>
<th>Base Yr. Labor Rate</th>
<th>Option Yr. 1 Labor Rate</th>
<th>Option Yr. 2 Labor Rate</th>
<th>Option Yr. 3 Labor Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sr. Business Analyst</td>
<td>MS + 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business Analyst</td>
<td>BS + 5, MS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jr. Business Analyst</td>
<td>BS + 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Part D: Technical Support – Firm Fixed-Price/Time & Materials

Provide T&M hourly rates for the following labor categories for base period and all option periods.

<table>
<thead>
<tr>
<th>Labor Category</th>
<th>Education &amp; Experience</th>
<th>Base Year Labor Rate</th>
<th>Option Yr. 1 Labor Rate</th>
<th>Option Yr. 2 Labor Rate</th>
<th>Option Yr. 3 Labor Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sr. Project Manager</td>
<td>BS +10, MS +5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Project Manager</td>
<td>BS +5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sr. Software Developer</td>
<td>MS +5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software Developer</td>
<td>BS +5, MS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jr. Software Developer</td>
<td>BS +2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cybersecurity Analyst</td>
<td>BS + 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sr. Database Administrator</td>
<td>MS +5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Database Administrator</td>
<td>BS +5, MS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jr. Database Administrator</td>
<td>BS +2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sr. Solution Architect</td>
<td>MS +5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solution Architect</td>
<td>BS +5, MS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jr. Solution Architect</td>
<td>BS +2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Architect</td>
<td>BS +10, MS +5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sr. Data Scientist</td>
<td>BS +10, MS +5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Scientist</td>
<td>MS +5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ETL Specialist</td>
<td>MS +3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sr. System Administrator</td>
<td>MS +5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System Administrator</td>
<td>BS +5, MS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jr. System Administrator</td>
<td>BS +2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
RFP Administrative Items and Guidance

• Confidentiality
  • Mark all confidential information as such

• Format
  • MS Word or Adobe Acrobat
  • 8.5” x 11”, 12-point font, ½ inch margins, page numbers

• Questions
  • Submit to: DataHubRFP@naswa.org
  • Please put “RFP QUESTION” in the subject line

• RFP Responses
  • Submit to: DataHubRFP@naswa.org
  • Due: 5:00pm ET September 7, 2018
RFP Administrative Items and Guidance

- RFP Information
- Webinar recording
- Webinar presentation
- Questions and responses
- Updates (as needed)
Integrity Data Hub (IDH)

Applications, Architecture & Technology Stack
IDH Applications

• SAR Portal
  • Lookup & Submit SAR Records
  • User Administration
  • Reporting

• SAR Web Services
  • Real-time
  • Lookup and Submit SAR Records

• SAR FTP Processing
  • Batch
  • Lookup and Submit SAR Records

• SSO Portal

Submit - putting suspicious actor data into the system. Lookup - searching suspicious actor data.
IDH Applications

- Java Programming Language
- Java Server Faces (JSF)
- Spring Framework
- CXF Web Services
- Web Service Clients
IDH Architecture Key Features

- Open Source
- AWS Cloud Based Environment
  - Flexibility, Availability, Scalability
  - Redundancy - separate servers in different availability zones
  - Platform tools allow small team to manage many servers
- Security - Developed with NIST Based Best Practices
- Interface Channel Flexibility
  - State Integration SFTP, Lookup File, Web Services
Open Source Ecosystem

cassandra
FORGEROCK
Linux
Java
APACHE
HTTP SERVER PROJECT
DATASTAX
OpenAM
OpenDJ
IDH AWS Cloud Environment

https://sar.naswa.org
Apache Httpd – Web Server

- Handles Web Requests from the Internet
- SAR Portal Application
  - [https://sar.naswa.org](https://sar.naswa.org)
    - Requests on urls protected by OpenAM policy agent
    - Validates if the OpenAM security token exists with OpenAM and is valid
- SAR SSO Application
  - [https://sarsso.naswa.org](https://sarsso.naswa.org)
- Forward To Internal Load Balancers for Application Servers
Apache Tomcat - Application

- Custom applications written in Java perform application level processing
  1. SAR Portal
  2. SAR SSO Login
  3. SAR Web Service
  4. SAR FTP Service
- Open AM - SSO software runs on Tomcat
Apache Cassandra - Database

- NOSQL Database
- DataStax Community Edition 3.0
- Distributed Key-Value Store, Low Latency, Lookup By Key Use Cases
- 4-Node Cluster (m4.2xlarge 8 vCPU 32GB RAM)
  - High Availability - Cluster can still function if one node is not performing
  - 2TB disk space per node (SSD, gp2)
  - Currently 184 MB of data allocation per node
- Predictable Linear Scalability with Added Nodes
Open AM - Single Sign On

- Open Source Configurable Application By Forge Rock
- Administration Console for Single Sign On (Tomcat)
- Configuration Store is Configured to Use Open DJ
- APIs Validate Users and Manage Passwords
- Policy Agent API for Request Validation

- LDAP Server for Open AM to store user configuration
dedicated replicated Open DJ LDAP Servers
- User Store Replicated
IDH AWS Cloud Environment

- Production
  - 16 Servers (Web, App, DB, LDAP)

- Staging
  - Same Topology as Production
  - Facilitates Load Testing
  - Available during core hours 7AM-7PM EST*

- Development
  - SAR server and SAR SSO server
IDH AWS Cloud Environment

• Activities By Environment

**DEV**
- NASWA
  - Development
  - Testing
  - No PII Data

**STAGING**
- NASWA
  - Application Testing
  - Load Testing
- STATES
  - System Training
  - FTP & Web Service Integration
  - No PII Data

**PRODUCTION**
- NASWA
  - Sys Admin
  - App Admin
- STATES
  - App Users
  - FTP
  - Web Service
  - PII Data
IDH Redundancy - Web Tier

• Load Balancing
  • SSO Web Servers
  • SSO Login Application Servers
  • SAR Web Servers
  • SAR Core Application Servers
• LDAP Servers - Replication
IDH Redundancy - Data Tier

- Cassandra cluster
- Replication factor of 2
- Each piece of data is automatically replicated
IDH Redundancy - Data Tier

• One Node Is Down
• Data Still Available
• Tradeoff for Disk Space
• Replication Factor of 3 Common to Increase probability of Availability
• Currently in one Region but across multiple availability zones
IDH Scalability - Application

• Horizontal Scaling
  • Architectural foundation for expandability with load balancing
    • Web servers
    • Application servers
    • Web services
  • FTP processors can also be added

• Vertical Scaling
  • Increase processing power and capability of servers
IDH Scalability - Auto Scaling

- UI demand changes with changes in economics
- Nationwide peak times will vary
- Monitor server usage and automatically deploy more servers to handle the load within minutes
- Servers can then automatically be shutdown when demand is lower saving costs which saves capacity
- Scheduled or On Demand

Reference:
https://docs.aws.amazon.com/autoscaling/ec2/userguide/as-using-sqs-queue.html
IDH Scalability - Data

Cassandra

- Ability to add nodes as throughput or storage needs increase
- Distributed key value storage allows linear scalability

4 Nodes
1000 Transaction/Sec

8 Nodes
2000 Transaction/Sec
Security

- Data is encrypted in motion and at rest
- Extensive use of https internally and public facing traffic
- Encrypted AWS drives for the database cluster
- Captcha two factor SAR portal authentication
- Secure Data Channels
  - SFTP using asymmetric encryption keys
  - Soap 1.1 Secure Web Service using asymmetric encryption
- Security IV&V on going
- Continuous activity
IDH Architecture Summary

• AWS Cloud Based Scalable Architecture
• Secured through NIST best practices
• Extensive use of open source software
• Flexible state integration channels to accommodate state level of technical resources