Integrity Data Hub (IDH) Request for Information

<u>Questions and Responses</u>
As of 09/02/16 (updates highlighted in Red)

1. It is our understanding that prior to the issuance of this RFI a UI fraud detection system had been built in collaboration with Elder Research. Will the proposed solution be a replacement for this system or would it work in conjunction with this system? If in conjunction, how does NASWA envision them integrating or co-existing (at a high level)?

> **The Center has an ongoing pilot project with Elder Research to develop a Data Analytics/Predictive Modeling (DAPM) solution for State UI agencies. The DAPM project is still underway. The resulting product will be evaluated for potential applicability to the IDH, similar to the other tools for which we are requesting information.**

2. When the current system was developed it was funded by a grant from the US DOL. Would any replacements of this system (resulting from this RFI and eventual RFP) also be funded by a similar Federal grant, or is there an alternate funding model envisioned? If alternate model, would it be entirely Federally funded (i.e. single revenue stream) or be funded by each individual state utilizing the system (i.e. multiple revenue streams)?

> **The Center is evaluating alternatives for funding IDH enhancements as well as ongoing IDH operations and maintenance costs. Rough Order Magnitude (ROM) cost information collected as part of the RFI is expected to inform this process.**

3. Is Cassandra the only database platform that will be considered for this project?

> **No. Vendors are welcomed and encouraged to recommend alternative database platforms as appropriate.**

4. What is the Center's rationale for using Apache Cassandra in early versions of the IDH? What factors would drive a change in the data storage platform for the IDH in future versions? Are there other data storage options which the Center has in mind for future versions?

> **The rationale for using Apache Cassandra is based on the Center's current pilot project to develop a Suspicious Actor Repository (SAR). The SAR will utilize Apache Cassandra as its database platform. Several factors could potentially drive a change in future IDH versions including volume and security requirements. The Center has not selected storage options for these future IDH versions and intends to use information gather from the RFI to inform these decisions.**

5. Does the Center have in place or does it plan to put in place any data governance framework? For example, will data addition and updates be undertaken in a federated approach or will the Center oversee this process centrally?

> **No formal decision on data governance has been made, but the Center expects to oversee the addition of data sources centrally working closely with USDOL OUI.**

6. Will States push data to the IDH and if so will ETL activities be performed by the States to meet the specifications of the IDH or will IDH resources perform the ETL work to align raw data with IDH data structures?

States will push data to the IDH in a specified format. Extract, Transform and Load (ETL) activities are expected to be performed by IDH resources, potentially through subcontracting or staff augmentation.

7. What is the envisioned frequency for adding new data sources to the IDH? What expectations exist in terms of responsiveness to new data?

The Center is interested in adding new data sources as quickly as possible; however, each data source is expected to have unique technical and policy issues that do not allow us to accurately predict the frequency and responsiveness.

8. How does the Center envision the distribution of effort among Parsed Claims, Cross-Match, and Analytics teams if these are different vendors? For example, are Parsed Claims and Cross-Matching answerable to Analytics to generate features and aggregations, or are these wholly separate? Outputs could also display a combination of underlying claims data, cross-matched data, and analytics. What is the Center's vision for how the different functional areas will interoperate, both within the IDH (among Parsed Claims, Interface/Cross-Match, and Analytics/Scoring) and between the IDH functional areas and the Output?

We expect parsed claims data to be analyzed and scored based on matching data.

9. The Center has emphasized their interest in products to support the development of the IDH. With the comprehensive scope of the IDH, solutions incorporating both products and tailored services may better serve the Center's goals. Will the Center accept solutions-based responses?

The Center is interested in both specific products as well as product/service solutions that can support development of the IDH.

10. How is the Center planning to engage states and / or incentivize state participation?

The Center plans to engage and incentivize states based on the ability to effectively prevent and identify improper payments. In addition, the Center has additional resources to assist states with implementation and support costs.

11. What states does the Center anticipate will participate?

The Center hopes to ultimately have participation by all state and territories.

12. At what level of data granularity does the Center envision being included into the Hub?

The Center would prefer the most data granularity achievable.

13. Are data elements considered for inclusion in the Hub already identified?

The RFI requests vendor inputs on data fields needed to provide effective analysis and cross-matching. We anticipate that a greater number of fields will improve performance, but will impact elements such as volume and storage.

14. Does the Center have a preference for pricing / fee structure?

No.

15. Who would be the Project Steering Team? From the Center specifically? What level of involvement could be expected?

The task project team includes the Center IDH Project Manager as well as other project resources including: business analysts, IT professionals, system architects, programmers, and database analysts.

16. Assuming states retain ownership of the data after committed to the hub, if a state elected to de-participate in the Hub in the future, should there be functionality that performs that action?

Yes.

17. Does the Center have a preference for how state data is integrated into the Hub?

No, however data is expected to be collected using a web-based cloud implementation.

18. Does the Center have a preference for any specific technologies for the Hub? Is there technology / software already owned by the Center that should be considered in pricing?

No. The Center is interested in obtaining information on all products and solutions with potential applicability to the IDH.

19. How does the Center define "product" in the "Product Overview" response requirement section of the RFI powerpoint? Is the Center's use of "per product" synonymous with per functional area or intended to mean per product within each functional area?

Product is defined as separate products or solutions offered by the vendor with applicability to the IDH. We understand that some vendors may have more than one product/solution and we want those vendors to have the ability to define each and how it would apply to the IDH.

20. Given that state statutes vary across the country, what data security standard does the Center want to adopt for this solution?

The Center intends to use a web-based FEDRamp/NIST-compliant infrastructure.

21. What resources does the Center anticipate allocating to the project?

The task project team includes the Center IDH Project Manager as well as other project resources including: business analysts, IT professionals, system architects, programmers, and database analysts.

22. Is the center open to standing up this project in phases? If so, is there a particular order or preference?

The Center is open to developing the IDH in phases, no order of preference has been established.

23. What is the target date in which the Center would like this Hub to be operational?

No specific date has been determined. Information collected by the RFI is expected to inform this process.

24. What is the anticipated timeline of this project (i.e. RFP release, start date, etc)?

No specific timeline has been determined. Information collected by the RFI is expected to inform this process.

25. What is the Center's budget for this project?

No specific budget has been determined. Information collected by the RFI is expected to inform this process.

26. What does the Center perceive maintenance and support will entail?

Vendors should respond to the RFI with their software maintenance policy; i.e. yearly maintenance and support costs, upgrades policies, etc.

27. What is the total number of Initial UI claimants per week, month, and year? Could be an average or a range.

        See CY 2015 claims data included at the end of this document.

28. What is the total number of weekly certification claims per week, month, and year? Could be an average or a range.

        See CY 2015 claims data included at the end of this document.

29. The RFI states the IDH is envisioned to operate as a cloud Infrastructure as a Service (IaaS) FedRAMP compliant system. Is the vendor responsible for hosting the system or will ITSC host? To what degree will ITSC be involved if offsite?

        The Center/ITSC will be responsible for hosting the system.

30. Is it anticipated that this solution will only apply to UI claimants? Please confirm our assumption that the IDH is for beneficiaries only as references throughout the document are for UI claims.

        The IDH is currently planned to address UI claimants. However, the possibility of interfacing with other areas such as UI Tax and other federal benefits programs is possible as the system evolves.

31. Does NASWA currently use a cloud infrastructure in your IT enterprise?
If yes, who is your cloud provider?

        We are currently establishing the cloud infrastructure environment; a specific provider will be selected in the coming weeks.

32. What are the data quality challenges NASWA faces in working with the states?
E.g., what is the variability in the format, content, and complexity of these data sets?

With the potential of 53 entities participating in the IDH using 45+ different benefit systems, there will be many issues with the variety and quality of the incoming data.

33. What is the data volume transacted with each state (monthly and annually)?

        See claims data included at the end of this document.

34. How many other external data sources are you using now, and what is the total monthly and annual data volume for these data sources?

        The IDH is under development and does not currently interface with any external data sources.

35. In section 4 of the RFI response instructions, you ask "How soon could your product be implemented in a cloud based, multi-state format?" Would you please explain what you mean by "multi-state format"?

        The IDH is envisioned to collect claimant data from multiple states, to perform cross-matching/analysis and return data to multiple states concurrently.

36. The IDH Concept displayed on page 5 of the RFI references 'initial/weekly' ingestion of claims data. Can NASWA specify whether this refers to claimant data, benefit claim payments, or both? Specifically, does NASWA envision this being utilized at the point of benefit enrollment, at the point of benefit payment, or both?

> **The Center expects to gather claimant data, initial claims data and weekly claims data that may or may not include payment data.**

37. Does NASWA envision piloting the IDH with a single state? If so, which states have expressed interest?

> **The Center does anticipate piloting the IDH in multiple states; however, specific states have not been determined. The Center is currently engaged in a pilot project to develop a Suspicious Actor Repository (SAR) that includes participation from: MT, NE, UT, ID, MD, TX, FL, and NY.**

38. Does NASWA have an estimated timeline for implementing an IDH solution?

> **No specific timeline has been determined. Information collected by the RFI is expected to inform this process.**

39. How does NASWA envision state UI agencies accessing the multi-state solution (E.g. Task orders, separate contracts, etc.)

> **The Center envisions participating states will sign an IDH participation agreement that outlines system access, utilization, roles and responsibilities, etc. This agreement has not been developed.**

40. What data sets/sources do you envision interfacing with?

> **Specific data sources have not been selected, but candidate sources may include items such as: death records, suspicious physical addresses, suspicious IP addresses, incarceration records, and others that may provide effective cross-matching.**

41. What specific data elements from initial and continued claims will be collected?

> **The RFI requests vendor inputs on data fields needed to provide effective analysis and cross-matching.**

42. Will data be collected on all initial and continued claims or only on those deemed worthy of investigation after being cross matched?

> **The IDH is envisioned to collect and perform cross-matching and analysis of all initial and continued claims.**

43. What data elements from the states data are expected to be made available for analysis through the hub. And will these data elements be restricted to the sample populations or the entire universe. Will they be suspect claims only, or will they include initial/continued claims?

> **States will provide a subset of data collected on initial and continued claims. Specific data fields have not been finalized. The RFI requests vendor inputs on data fields needed to provide effective analysis and cross-matching.**

44. If we have a prebuilt Dept of Labor Data model built in SQL server, can we use that as Database platform instead of Cassandra?

Vendors are welcomed and encouraged to recommend alternative database platforms as appropriate.

UI Claims Data for CY 2015

state	Total 5159 report New ICs	Total 5159 report Additional claims	Total 5159 report Transitional claims	Total 5159 Report Liable Claims	Est # liable ICs	Est # ICs =New + transitionals +liable	total Weeks claimed	total weeks compensated
US Total	9,156,990	4,823,573	539,003	584,809	398,325	10,094,318	118,664,641	102,299,605
AK Total	37,762	25,496	2,026	12,033	7,161	46,949	558,551	471,709
AL Total	126,968	45,391	4,130	5,071	3,741	134,839	1,184,417	955,527
AR Total	90,163	52,936	1,739	4,495	2,843	94,745	1,097,391	828,205
AZ Total	174,488	29,987	2,677	6,736	5,753	182,918	1,540,413	1,315,617
CA Total	1,358,483	1,055,333	56,999	45,953	25,830	1,441,312	20,160,298	18,061,071
CO Total	114,530	27,558	4,089	11,514	9,272	127,891	1,659,836	1,387,948
CT Total	121,693	78,087	15,518	8,516	5,182	142,393	2,216,592	2,141,991
DC Total	15,155	1,438	277	18,564	16,961	32,393	447,596	400,391
DE Total	25,009	16,104	2,104	4,039	2,472	29,585	371,515	317,148
FL Total	366,243	72,302	4,681	9,937	8,305	379,229	2,950,666	2,233,348
GA Total	280,709	116,768	1,045	9,562	6,767	288,521	1,987,335	1,569,151
HI Total	30,092	36,570	754	3,100	1,400	32,246	430,261	363,805
IA Total	104,418	57,165	12,743	19,714	12,760	129,921	1,249,117	1,188,533
ID Total	42,929	24,820	1,854	1,814	1,168	45,951	436,733	363,003
IL Total	373,086	225,500	52,642	14,717	9,170	434,898	6,232,704	5,499,984
IN Total	161,243	46,360	7,226	6,666	5,179	173,648	1,495,059	1,363,894
KS Total	75,579	51,280	6,142	22,744	13,660	95,381	913,031	725,860
KY Total	122,835	67,263	3,626	4,844	3,123	129,584	1,252,136	1,128,598
LA Total	100,205	20,931	1,055	9,843	8,142	109,402	1,080,914	957,745
MA Total	218,859	94,658	12,430	21,567	15,074	246,363	3,967,268	3,447,298
MD Total	138,984	67,172	10,531	6,338	4,280	153,795	2,087,790	1,799,441
ME Total	36,038	17,861	2,670	1,709	1,144	39,852	466,704	401,332
MI Total	343,807	125,680	6,814	5,880	4,304	354,925	3,465,000	3,018,097
MN Total	141,415	87,979	25,267	14,521	8,918	175,600	2,484,938	1,948,800
MO Total	175,178	95,317	12,959	6,984	4,527	192,664	1,686,215	1,388,177
MS Total	63,480	22,466	5,592	3,231	2,419	71,491	636,340	515,192
MT Total	33,437	21,461	3,533	2,296	1,398	38,368	428,155	370,174
NC Total	201,507	29,196	3,896	11,571	10,092	215,495	1,767,421	1,209,283
ND Total	25,773	7,569	126	19,422	15,134	41,033	410,780	344,859
NE Total	38,481	16,998	1,218	3,643	2,524	42,223	380,599	294,044
NH Total	24,227	12,624	2,312	1,708	1,129	27,668	250,643	231,067
NJ Total	372,532	142,309	20,635	16,286	11,735	404,902	5,928,939	5,351,155
NM Total	46,465	12,454	1,849	1,804	1,426	49,740	668,364	589,304
NV Total	99,036	45,395	3,681	4,296	2,952	105,669	1,314,342	1,164,336

NY Total	581,625	422,686	56,745	22,361	12,930	651,300	9,283,984	8,022,736
OH Total	276,697	139,081	19,324	28,291	18,813	314,834	3,430,618	2,887,948
OK Total	88,513	18,051	1,692	9,000	7,501	97,706	1,130,938	954,137
OR Total	141,721	116,838	7,370	9,283	5,087	154,178	1,744,032	1,507,337
PA Total	538,572	491,708	43,704	50,157	26,216	608,492	7,653,762	6,370,512
PR Total	77,065	24,819	10,616	1,316	1,005	88,686	1,373,244	1,329,584
RI Total	38,273	27,321	3,139	1,461	853	42,265	543,172	487,340
SC Total	132,924	38,868	3,840	6,063	4,701	141,465	884,397	629,662
SD Total	10,592	2,669	1,090	1,314	1,051	12,733	101,790	85,151
TN Total	151,878	69,543	2,731	5,560	3,823	158,432	1,380,548	1,251,632
TX Total	715,038	156,505	43,099	38,014	31,136	789,273	8,736,398	7,604,398
UT Total	56,337	15,435	2,502	3,592	2,811	61,650	576,225	476,039
VA Total	152,131	37,083	603	24,215	19,488	172,222	1,687,391	1,377,723
VI Total	2,270	331	54	136	119	2,443	43,545	38,662
VT Total	18,406	15,341	2,569	2,333	1,277	22,252	272,706	236,448
WA Total	226,679	142,495	8,321	17,639	10,810	245,810	2,828,732	2,504,169
WI Total	193,221	227,363	31,425	7,357	3,379	228,025	2,650,601	2,173,736
WV Total	54,676	19,572	4,358	8,042	5,913	64,947	873,163	779,609
WY Total	19,563	7,436	981	7,557	5,467	26,011	261,332	236,695