

# Collated Q&A: NLx Research Hub RFP

## General Clarifications

- The Proposal Submission Elements section of the RFP asks for offerors to complete “Attachment (2) Pricing Summary information for each SOW area as appropriate.” Please disregard this instruction. You may use any template or format you deem appropriate for a Firm Fixed-Price calculation, broken down by SOW sub-bullets if possible.
- There may be communication with offerors to clarify elements of their proposals.
- All participants will be invited to propose a Best & Final offer.
- The table below shows the original RFP timeline and a modified timeline, to accommodate requests for additional response time following this Q&A release.

RFP Activity	Original Timeline	New Timeline
Q&A Period	October 19-22	October 19-22
Collated Q&A Responses Released	October 28	<b>October 29</b>
Proposals Due	November 2	<b>November 6</b>
Best and Final Offer Pricing (optional)	November 6	<b>November 13</b>
Award (anticipated)	November 13	<b>November 20</b>

## Team

**1) Can you describe the NASWA team makeup – how many people will be working with the selected vendor?**

The NASWA team will include: Charlie Terrell, Director, NLx; Emma Northcott, NLx Program Manager; Lou Ansaldi, Technical Director; and Kevin Knubel, IT Specialist. In addition, NASWA hired a consultant Project Manager & Strategic Advisor for the Phase I Project (Adriana Lopez) who will support the team for this phase as well. These team members will contribute expertise in data architecture and cloud technologies.

**2) Who are the stakeholders that would define the requirements?**

Staff from NASWA and DirectEmployers will be the primary contacts for requirements definition. Data test users from Phase I, including representatives from state workforce agencies, think tanks, and universities, may be consulted as needed.

## Contract

**3) Is there an anticipation that the FFP bid will be for the full solution or will there be a possible split between Parts A and B?**

As stated in the RFP (p. 5), NASWA anticipates making a combined award for SOW Areas A and B.

**4) Given the need for a FFP for both parts A and B, are you anticipating that the Part B FFP may be dependent on the outcomes of Part A or do both parts need to have a set FFP up front?**

It is acceptable to propose an estimated cost and solution for Part B based on current knowledge and the understanding that the final scope of work may be modified based on the findings of Part A. Specific aspects of the FFP may be negotiated in the contracting stage.

**5) Is the targeted completion date of February 2021 a firm requirement?**

It is not a firm requirement; however, if an offeror expects a later completion date they should name the targeted date and the rationale for a longer project period.

## Requirements

**6) Do you have any existing technical documentation, database design, or data field requirements that you can share?**

In the requirements gathering phase, we plan to share the following:

- Summary of Phase I governance activities and deliverables (Data Trust Agreement and exhibits, Data Sharing Agreement, etc.)
- User discovery interviews and documentation
- Use case/user profile notes
- Analysis of source data systems
- AWS S3 bucket components and screenshots
- Documentation for JobsPosting Format Conversion ETL
- Data validator tool documentation
- Data dictionary
- Data download instructions, descriptors, and analysis tips
- GitHub repository
- Prototype API testing notes
- Phase I technical infrastructure notes and learnings
- Phase I data test user process notes and learnings

**7) Can the existing code become available to the bidders to understand the salvageable vs. level of updates?**

The “Old Archive Data” and other materials may be referenced by the selected vendor throughout the project as points of comparison to the real-time NLx feed from DirectEmployers. However, the expectation is that this scope of work covers the creation of entirely new data stores to capture NLx data from 2021 onward.

**8) Is the use of MySQL required, preferred, or mandatory for the new environment?**

The NASWA project team is open to other processes, so long as they seamlessly integrate within the AWS environment. NASWA prefers open source tools as feasible.

## Database

**9) What is the current DirectEmployers data field structure? Is it a relational database?**

Yes – NASWA receives and archives job postings from DirectEmployers in a relational database. There are approximately 50 data fields (not all fields are required for every job posting).

**10) Is there an anticipation of establishing a data model that will support the exchange between NLx and NASWA repository?**

Yes, a data model is anticipated.

**11) In the proposed architecture, there are two redundant databases. What is the reason for this? Relatedly, please clarify “architectures that enable rollback to a backup pristine data store that is frequently updated or synchronized with the primary data store.” Is there a current challenge this would help resolve? Are you open to other architectures/solutions?**

The proposed architecture includes one redundant database OR a frequently-updated backup data store to ensure data integrity and emergency recovery. However, NASWA is open to other architectures.

**12) In the proposed architecture, how “in sync” do the two databases need to be? Can they eventually be consistent and if so, would it need to be within a few seconds, minutes, or hours?**

They could eventually be consistent, with a maximum sync time of 24 hours.

**13) Is the NLx Research Hub specifically required to be FedRAMP Moderate or is that chosen as a general baseline target?**

NASWA has chosen to be FedRAMP Moderate for this and other projects.

**14) There was mention of Excel archives of data – is there anticipation that all of that information becomes part of the new repository and/or segmented into a reporting structure?**

The Excel archives are associated with the “Old Archive Data” from 2019 and earlier. The goal is to include the CSV files within the same cloud environment envelope, but stored separately from the “real-time” NLx dataset.

**15) Does the data structure of the historical data match the data structure of the DirectEmployers data in AWS?**

An expectation of the requirements gathering phase is to continue reconciling the “Old Archive Data” fields with the DirectEmployers data fields (this work began in Phase I of the NSF grant). There are some generated fields in the “Old Archive Data” store that do not match the current DirectEmployers feed, and a smaller number of DirectEmployers data fields that the SQL database does not capture.

**16) Does NASWA have a data warehouse data model in mind such as Star? Or is that a task for the selected vendor during the requirements phase?**

The NASWA project team is open to exploring any options for data models proposed by the selected vendor, but prospective vendors must provide supporting rationales for their proposed approach.

**17) Will the entire data universe in the warehouse including historical data be available for querying by users of the system? Or will NASWA and the selected vendor create subsets of the data based on the requirements for different user groups?**

The goal is to make subsets of the data available based on the needs of different users. Ideally the data could be partitioned by year and data field. This will enable ease of transfer to end-users and ensure proper data governance (for example, some users will not be approved to see employer names).

## Security

**18) Will the database be utilizing AWS security controls for access or will we need to consider adding a third party access or database driven authentication system?**

AWS security controls should be sufficient for access.

**19) While we understand that APIs/sFTP development is not part of this effort, to properly design/architect we should consider at least collecting during requirements what the potential use cases will be for end user data access via APIs/sFTP. Should the documentation/design/architecture be focused with this end goal in mind?**

Yes. Some of those use cases were already collected in Phase I, and can be shared with the selected vendor during the requirements gathering phase.

## ETL

**20) The ETL pipeline mentions data verification and remediation; do you have any tools/integrations in mind like USPS Address verification or similar?**

NASWA expects offerors to propose their recommended approach along with rationale.

**21) Should we assume that the ETL process is implemented before data is added to Archive DB and account for the ETL server as standalone data processor vs. processing post data pull from DE NLx?**

NASWA expects offerors to propose their recommended approach along with rationale.

**22) Does the web UX need to read directly from the archive, or can that be incorporated into the data warehousing solution selected?**

NASWA expects offerors to propose their recommended approach along with rationale.

**23) Do you have an existing or preferred ETL/ELT tool such as Informatica, Talend, Fivetran, etc.?**

NASWA expects offerors to propose their recommended approach along with rationale, and costs must be considered.

## Data

**24) Is there a need to keep the raw data as well or just copies of the cleaned data after ETL process?**

Copies of the cleaned data after ETL process should be sufficient.

**25) Please clarify the requirements for vendor data migration as a task of Part B.**

There are no requirements for vendor data migration in Part B.

## Future Work

**26) Will there be ongoing database administration work for the selected vendor to include index management, backups, etc. or will this be done by NASWA once the Part B work is complete?**

Parts A and B should result in deliverables based on the assumption that NASWA will steward any ongoing database administration work. However, there may be opportunities for the selected vendor to engage in that work depending on future funding.

**27) Regarding “the architecture must support future connections to a web interface allowing external access, as well as hosted computing environments (not part of this scope of work)” – what other clients/uses do you expect to see?**

Future work may include developing standardized queries to deliver data subsets to end users, creating a hosted computing environment within the cloud envelope for users who have restrictions on downloading NLx data into their own systems (primarily state users), mapping NLx data to other data standards and schema, integrating other datasets, etc.

**28) Do you have an idea of anticipated future traffic to the Web UX or alternative uses?**

Several test users indicated they would likely download data once per week or per month, while others expected to ping the UX more frequently. NASWA plans to make the data available to a small number of early users and scale up on a pilot basis.

**29) Please share what projects, if any, you anticipate arising that will require support for an AI or machine learning approach?**

As one example, partners and/or users may apply AI and natural language processing techniques to extract job requirements from job descriptions and map them to the specifications of O\*NET occupation codes, which can then be correlated with Classification of Instructional Programs (CIP) codes. The resulting categories can be used to map candidate courses an applicant could take to acquire specific job skills, and vice versa, they can map candidate jobs that require the skill taught by a particular course.

**30) How is the archived data anticipated to get used as part of the proposed solution?**

One goal is for the NLx Research Hub to house simple dashboards/visualizations with descriptive statistics and search functionality. End users may also analyze the archived data to categorize job postings by requirements, veterans preference, likely job setting, and estimated annual pay; match resumes to job postings based on skills/credentials; build tools to inform jobseekers

about training programs for local in-demand jobs; predict local economic trends; link to other datasets for further analysis; etc.

**31) Does NASWA anticipate future expansion of the data warehouse by adding UI claims data from states or from the IDH to the warehouse in an effort to compare employer hiring against employer layoffs?**

The project team does not expect to integrate the NLx Research Hub with other NASWA products and services at this time.

**32) Is there a business continuity plan for the data warehouse? Is there a disaster recovery plan for the data warehouse? Are these plans part of the scope of work of the selected vendor?**

Data backup and recovery plans are included in the scope of work for the selected vendor (see page 2 of RFP). The NLx Research Hub is intrinsically linked to NLx operations, and future business continuity plans must align with NLx operational needs. NASWA welcomes input from the selected vendor on business continuity plans, particularly around cloud investments and managed services opportunities. Rationales and cost considerations should be included.

## Business

**33) Other than building more scalability and flexibility for the future as well as creating a basis for APIs, is there anything in particular that the current system does not do that you would need it to do?**

Harmonizing the data fields between the current system and the real-time NLx feed is a critical task. Additionally, data users leverage a wide variety of systems for accessing, loading, and analyzing the data. In the Phase I test user process, this required significant technical assistance at every stage of the process from initial credential setup, data download, and analysis – NASWA would like to provide less direct technical assistance in the future. We determined that both sFTP and APIs would be necessary delivery alternatives to help achieve this.

**34) Who are potential users of the data warehouse?**

Economists, analysts, and program staff working in state workforce agencies and WIOA partner agencies; academic researchers; nonprofit researchers and program analysts; and community college staff are all considered intermediary users – those who might access the “raw” data. Jobseekers and employers are the primary audiences for derived data products, tools, and visualizations.

**35) Will NASWA acquire business intelligence tools for users to query the data warehouse or will users be allowed to use their own tools? If so how will that be managed?**

Future projects may result in NASWA acquiring business intelligence tools, and formal governance protocols are already in place to guide the allowable uses of users’ own tools and their research outputs.