NLx
National Labor Exchange
Research Hub

Request for Proposal

**Technical Support for Development of a
Data Warehouse**

October 2020

Contents

Introduction

The National Labor Exchange (NLx) Research Hub was established to provide workforce development professionals, academic researchers, and other organizations that rely on labor market information with high-quality, transparent, real-time and historical data that represents the diversity of jobs available in the labor market. The efforts of the NLx Research Hub are managed by the National Association of State Workforce Agencies (NASWA) and its Center for Employment Security Education and Research (CESER), through National Science Foundation (NSF) funding.

The NLx is a partnership between NASWA and DirectEmployers Association. The NLx collects and distributes job openings exclusively found on corporate career websites and state job banks. All openings are unduplicated, currently available, and from vetted employers. In 2019, the NLx made available over 15 million unduplicated job openings, from employers of all sizes.

One of NASWA's key tasks during our Phase I project is to refresh the data warehouse infrastructure that will store the "historical" (2007-2020) and "real-time" (2021-onward) NLx job posting data to power the NLx Research Hub. NASWA currently utilizes a historical SQL database, developed in 2010, that no longer serves the needs of the NLx Research Hub. A copy of the historical data (through January 2020) has been migrated to an AWS S3 bucket, with various data versions converted to redacted and un-redacted CSV flat files. As of January 2020, there were approximately 130 million observations (individual job postings) in the historical database. Each observation contains approximately 60 data fields situated across 8 tables. As outlined in Figure 1 below, NASWA's preference is to triangulate on redundant databases within an AWS envelope; NASWA is also open to other architectures that enable rollback to a backup pristine data store that is frequently updated or synchronized with the primary data store. In essence, the proposed architecture includes two of the same databases into which real-time NLx data will be piped simultaneously from DirectEmployers's servers. For this architecture, there must be guarantees that the two data stores are in sync. Only one of these will provide data for external user access, which will likely be supported through APIs and sFTPs. (Note: The development of APIs and sFTPs are not part of this scope of work).

While the current SQL Server database represents 1.2 terabytes of data, it is not expected that this data will be migrated into the new database structure under this scope of work. Nor is it expected that an external/public-facing interface will be developed under this scope of work.

NASWA is working closely with other partners to develop plans for expanding the capabilities of the NLx Research Hub. These expanded capabilities may include: interfacing with additional public and private, external data sources for expanded cross-matching – such as BLS or Census data – providing advanced data analysis and reporting, and developing researcher- and worker-centered labor market information tools and data visualizations that leverage the dataset.

If you intend to respond, and are interested in receiving additional information on the NLx Research Hub project and current status, please submit a request to Emma Northcott, Program Manager, at enorthcott@naswa.org.
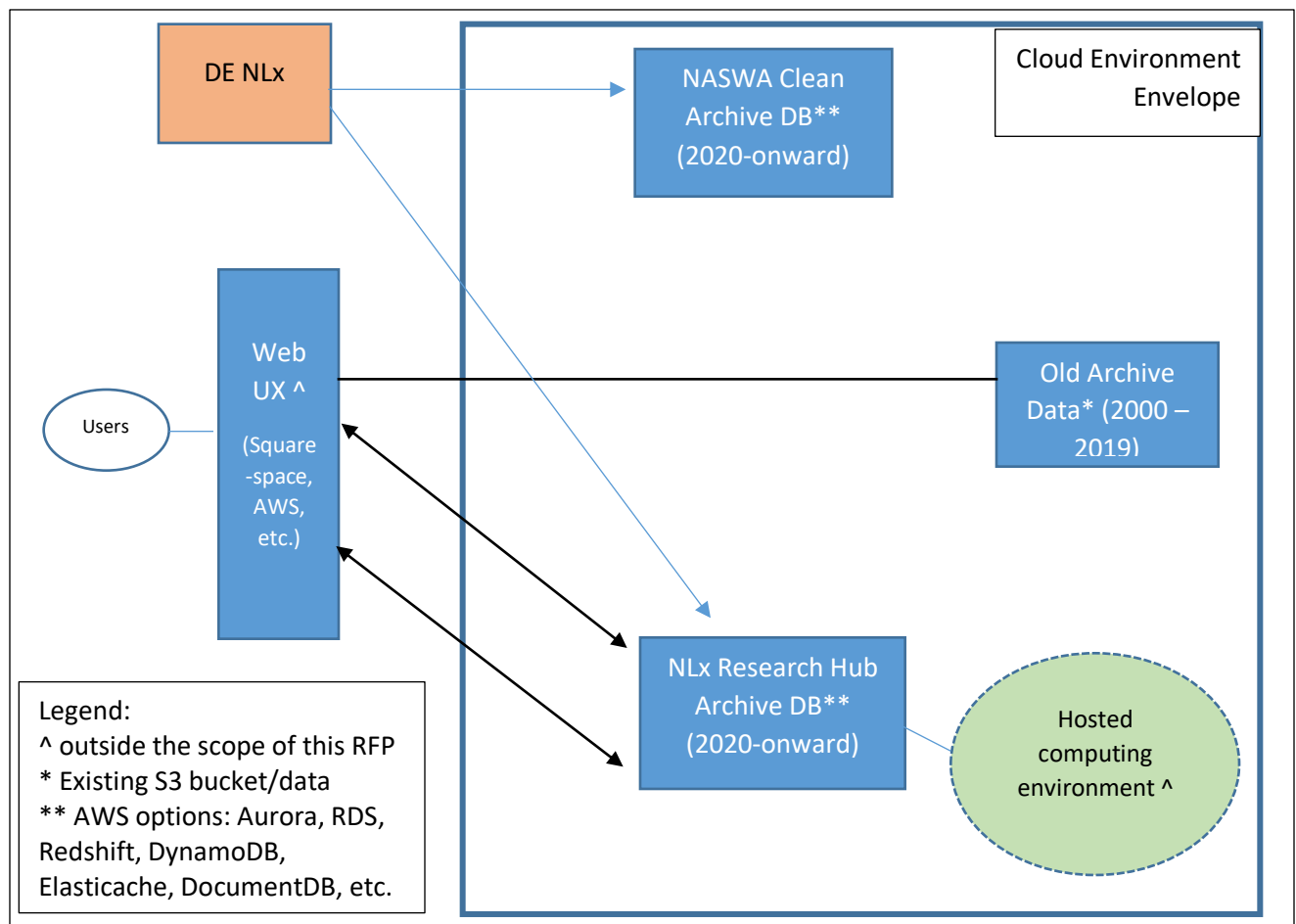
Purpose

NASWA is interested in identifying potential industry partners to augment its project team in the development of the NLx Research Hub cloud environment, as part of the NSF Convergence Accelerator Phase 1 project period. Interested parties are asked to provide information on their personnel, corporate experience, and related labor categories and pricing information in the following areas:

- Requirements development
- Database architecture/management
- System architecture/monitoring
- IT security
- Software engineering/development
- ETL development
- AI, machine learning, and data science

Technical Architecture

The proposed system diagram for the NLx Research Hub is summarized in the diagram below.

Figure 1: NLx Research Hub Concept Diagram

The items enclosed in the rectangle labeled "Cloud Environment Envelope" are within the scope of this RFP – including the ETL pipelines that transfer data from the external DE database to the cloud environment but excluding the hosted computing environment – and should utilize an open source software stack in an AWS cloud based environment, including considerations for an AWS FedRAMP Moderate environment. Depending on the outcomes of the requirements definition phase and future capabilities planning, this stack may include Aurora, RDS, Snowflake, Redshift, Glue, DynamoDB, Elasticache, Memcached, DocumentDB, and/or other tools. The AWS cloud based environment provides scalability, security, and flexibility, and was the preferred system for most Phase I data test users. 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).

"DE NLx" refers to the set of job posting files collected by DirectEmployers every 24 hours, which should be added into the two databases ("NASWA Clean Archive DB" and "NLx Research Hub Archive DB") through ETL pipelines.

"Old Archive Data" refers to the "historical" data that was migrated to an AWS S3 bucket in Phase I.

"Hosted Computing Environment" refers to a tentative plan to build computing capabilities within the envelope itself.

Scope of Work

The scope of work for this RFP includes two separate task areas as described below.

Part A: Technical Requirements Development/Business Analysis Resources – Firm Fixed-Price

The selected contractor shall provide qualified resources to assist the project team in developing and documenting business and technical requirements, and supporting business analysis activities in support of developing expanded NLx Research Hub capabilities. Support shall include the following:

1) Reviewing Phase 1 technical documentation, data dictionaries, and test user process outputs;

2) Coordinating and facilitating requirements gathering sessions;

3) Preparing requirements documentation;

4) Developing design documentation;

5) Preparing test plans and procedures; and

6) Planning and coordinating testing (functionality, usability, user acceptance, etc.).

Part B: Database Development, Implementation, and Technical Support – Firm Fixed-Price

The selected contractor shall provide qualified resources to assist the project team in database development and associated technical activities. Development/Technical support shall include the following:

1) System architecture review/assessment;

2) Modernization and alignment with the current DirectEmployers data field structure;

3) Data architecture/master data management;

4) Database architecture/management;

5) Data model development, data mapping, and data cleansing;

6) IT and data security;

7) Extract/Transform/Load (ETL) pipeline development, including data verification and remediation;

8) ETL protocol creation to process data for outputting;

9) Index maintenance plan development; and

10) Database configuration.

Contract Structure

SOW Area A and B:

NASWA anticipates making a combined award for SOW Areas A and B. Based on quotes from offerors, NASWA will award a Task Order (TO) using the labor rates established in the base agreement. Each TO awarded shall include a Statement of Work, estimated labor hours and pricing, total TO funding authorized, and a timeline for completion/deliverables as appropriate; as well as any additional information needed. Each TO will be awarded using Firm Fixed-Price (FFP) pricing.

If you intend to respond, and are interested in receiving additional information on NASWA's standard contractual Terms and Conditions, please submit a request to enorthcott@naswa.org.

Timeline

The RFP timeline of events is shown below:

RFP Activity	Timeline
Q&A Period (Via email submission)	October 19 - 22
Collated Q&A Responses Released	October 28
Proposals Due	November 2
Best and Final Offer Pricing (optional)	November 6
Award (anticipated)	November 13

NASWA reserves the right to invite offerors to participate in detailed discussions, clarifications to responses, and presentations/demonstrations subsequent to the proposal due date.

The preliminary project timeline of events is shown below:

Project Activity	Deliverable(s)	Timeline
Vendor Onboarding/Requirements Gathering	Written requirements documentation	November/December 2020
Database and ETL buildout, testing, and iteration	Technical documentation and virtual demo with client	January 2021
Final Delivery of SOW Areas A and B	Fully developed and implemented data warehouse	February 2021

Proposal Submission Elements

For a proposal to be considered complete, respondents are requested to address each of the items elaborated upon below.

1) Company Overview (all respondents)

Provide a brief description of your company, services, business size (revenue, employees, customers), subcontractors, and points of contact, including name, address, phone numbers, and email addresses.

> *Limit response to 1 page.*

2) Project Summary Citations (all respondents)

Include up to two (2) project summary citations that outline your organization's experience in providing strategic and/or technical support for projects of similar content, size and scope to the NLx Research Hub SOW areas you propose to support. For each project summary citation, please include the following: project summary, project size/scope, project budget, agency/organization, and agency/organization point of contact.

> *Limit response to 3 pages for each citation, up to 2 citations.*

3) Technical/Management Approach (for each SOW area)

For each SOW area proposed please provide information on your planned approach to include addressing the questions listed below.

Part A: Technical Requirements Development/Business Analysis Resources
- What process and tools will you use to collect and document business and technical requirements?
- How will you ensure participation from all stakeholders in developing requirements?
- How will personnel from your organization be selected to provide business analysis services?
- How will you organize and manage providing business analysis services?
- What deliverables will you develop and/or maintain?
- Have you provided similar services to other projects? What were the results/benefits provided?
    - Please provide project examples where the bidder has worked requirements with multiple entities.
- Outline your planned process for ongoing communications with the NASWA project manager and project team.

Part B: Database Development and Technical Support
- What tools and standards for database development will you use (please be specific and enumerate all)?
- How will personnel from your organization be selected to support the project (both initial assessment and database development support).
- How will you ensure availability of key staff?
- Have you provided similar services to other projects? What were the results/benefits provided?

- Outline your planned process for ongoing communications with the NASWA project manager and project team.
- Has the bidder architected, designed, and implemented a model whereby virtual/remote staff are actively engaged in software development processes, including:
  - o Database design;
  - o Associated tools;
  - o Standards; and
  - o Governance.

  Please provide details of at least two projects, and address all items listed in Scope of Work, Part B (addressing how these projects are similar to the current NLx Research Hub as described herein, including use of AWS).

Limit response to 5 pages per SOW area.

4) Key Personnel Resumes (for each SOW area)

Please provide three resumes (two pages maximum per resume) for key personnel to be assigned to the project for each SOW area proposed. The same resume may be provided for multiple SOW areas. Resumes should include: name, proposed labor category, percentage of time allocated to the NLx Research Hub project, and relevant work experience.

5) Pricing

Complete Attachment (2) Pricing Summary information for each SOW area as appropriate.

Evaluation Criteria

The NASWA project team will evaluate all proposals using the following evaluation criteria and award base contracts to the contractor(s) that represents the best value for NASWA.

Criteria	Weight
Corporate Experience	30%
Technical/Management Approach	30%
Key Personnel	20%
Pricing	20%

Proposal Description and Process

Participation in this RFP process is voluntary. All costs incurred in responding to, or in participating in this RFP, will be the responsibility of the vendors, or other third-party organizations participating in the RFP, and not that of NASWA.

Confidentiality

Any document submitted in response to this RFP that contains confidential information must be marked by a watermark on the appropriate pages as "Confidential." The confidential information must be clearly identifiable to the reader as confidential. All other information will not be treated as confidential. Note all confidential information is for NASWA's use evaluating proposals in response to this RFP.

Instruction and Response Guidelines

Responses to this RFP shall adhere to the page limits specified, and must be in narrative form and provide details on vendor product capabilities. Responses must be viewable with Microsoft Word or Adobe Acrobat and printable on 8.5" x 11" paper, must use 12-point font, the margins of each page should be at least ½ inch, and each page should contain a page number in the footer.

Reponses must be received electronically by 5:00 p.m. Eastern Daylight Time on November 2, 2020. Reponses will be sent to the email address of the sender along with any additional email addresses included in the submittal.

Please ensure that the submittal is in Microsoft Word or PDF format. All responses must be submitted electronically to the following email address: enorthcott@naswa.org.

Telephone calls regarding this RFP will not be accepted. Questions may be submitted by email up to 5:00 p.m. Eastern Daylight Time, October 22, 2020.