



National Research Council Canada  
Finance and Procurement Services Branch  
d'approvisionnement

Conseil national de recherches Canada  
Direction des services financiers et

**LETTER OF INTEREST  
LETTRE D'INTÉRÊT**

**RETURN BIDS TO :  
RETOURNER LES SOUMISSIONS  
À:**

**National Research Council  
Canada (NRC)  
Finance and Procurement  
Services**

**See herein for bid submission  
instructions/  
Voir la présente pour les  
instructions sur la presentation  
d'une soumission**

<b>Title/Sujet Omics Federated Learning Network</b>	
<b>Solicitation No./N. de l'invitation 22-58145RFI</b>	<b>Date 27 February, 2023</b>
<b>Solicitation Closes/L'invitation prend fin at/à 14 :00 on/le March 10, 2023</b>	<b>Time Zone/Fuseau Horaire EDT</b>
<b>Address Enquiries To/Adresser demandes de renseignements à :</b>  Elizabeth.Pitre@nrc-cnrc.gc.ca	

**Instructions: See Herein**

**Instructions: Voir aux présentes**

Proposal To:

We hereby offer to sell to Her Majesty the Queen in right of Canada, in accordance with the terms and conditions set out herein, referred to herein or attached hereto, the goods, services, and construction listed herein and on any attached sheets at the price(s) set out therefor.

Proposition aux:

Nous offrons par la présente de vendre à Sa Majesté la Reine du chef du Canada, aux conditions énoncées ou incluses par référence dans la présente et aux annexes ci-jointes, les biens, services et construction énumérés ici sur toute feuille ci-annexée, au(x) prix indiqué(s).



<p><b>Vendor/Firm Name and Address</b>  <b>Raison sociale et adresse du fournisseur/de l'entrepreneur</b></p>
<p><b>Telephone No./N. de telephone</b>  <b>Facsimile No./N. de télécopieur</b></p>
<p><b>Name and title of person authorized to sign on behalf of Vendor/Firm (type or print)</b>  <b>Nom et titre de la personne autorisé à signer au nom du fournisseur/de l'entrepreneur (taper ou écrire en caractères d'imprimerie)</b></p>
<p><b>Signature</b>  <b>Date</b></p>

## **Omics Federated Learning Network**

### **1.0 SCOPE**

The purpose of this request for information (RFI) is to assist the National Research Council of Canada (NRC) to gather information and to seek a cost estimate/model for the procurement and setup of an omics federated learning network (FLN) solution (requirements described in section 3.0).

A FLN is defined as a hub-node platform where users access data sets from nodes and perform analyses which are computed and stored on the hub. For this FLN, the term “analyses” is broadly defined and can pertain to a wide range of tools from genomic assembly to machine learning models. Thus, it is expected that analyses could be refined and augmented as tools are trained and experience learning. The goal of the FLN is to support industry in the use of omics (e.g. genomics, proteomics, metabolomics) as a tool; it is anticipated that the FLN would also provide access to datasets and models and facilitate the reuse of omic data investments.

Canada may use the information gathered in the RFI process to determine changes that should be included that will meet its needs and be consistent with Industry standard practices.

Following the RFI process, the next step will be for the NRC to assess the information and prepare for a potential request for proposal (RFP).

Once responses are received and reviewed, Respondents may be invited to discuss their responses at a Post RFI Response Submission. Industry Engagement Questions are included with this RFI (see Annex A) to provide Industry the opportunity to review and prepare written comments which may serve to facilitate this consultation process.

### **2.0 BACKGROUND**

Genomics is the study of the genetic information of an organism encoded in DNA and other molecules, which includes the structure, function, evolution, mapping, and editing of genomes. Currently, genomic data sets are largely underexploited; however, there is an opportunity in combining genomics with big data, artificial intelligence (AI), gene editing and synthetic biology to extract additional value from data, the outcome of which may be used to drive economic growth, and respond to global and national challenges, such as climate change, implementation of precision health, pandemics, food and energy security.

This FLN will be built with data shared by collaborators. Models created through this initiative will facilitate and enable data inputs, genomic and genetic predictions. This solution will also support the exchange of information between independent entities and participating groups (i.e., government, academic and industry organizations) who self-manage data sets.

The FLN platform is expected to have an initial scope of approximately 10 nodes but should be scalable. It will: 1) be built with very large data sets from multi-sectoral federated data networks for omics (genomics, transcriptomics, metabolomics, lipidomics, etc.) and phenotyping data; 2) facilitate interoperability; 3) provide participants with advanced analytics; and 4) be compatible with other federated networks. The FLN will enable the reuse, sharing and standardization of

data management, leveraging the significant investment of federal funding and data being created in federal government, academic and industry laboratories.

### 3.0 REQUIREMENTS

The following are the requirements for the FLN.

#### 3.1 TECHNICAL REQUIREMENTS:

##### 3.1. *Infrastructure and platform:*

- 3.1.1. the hub must be deployed in a public cloud service provider (CSP) which provides Canadian data residency.
  - 3.1.1.1. The CSP must have a supplier arrangement with Shared Service Canada (SSC) cloud brokerage.
- 3.1.2. must implement and comply with the GA4GH standards [<https://www.ga4gh.org>] when applicable.
- 0.1.3. must conform to the Canadian Government IT policies.
- 0.1.4. must respect all cyber security and data privacy compliance as required.
- 0.1.5. must provide infrastructure (e.g. infrastructure as code) and applications for the hub.
- 0.1.6. must provide a toolkit for node onboarding. The toolkit will facilitate dataset exchange between nodes and provide a framework for dataset standardization in the FLN.
- 0.1.7. must provide a toolset for the management of datasets visibility and accessibility across the FLN.
- 0.1.8. must implement the necessary controls and management system to manage users accessing the FLN.
- 0.1.9. must provide interfaces:
  - 0.1.9.1. have a graphical interface front end.
  - 0.1.9.2. have an Application Programmer Interface back end which allows the access to the FLN and all its tools.
- 0.1.10. must have the capacity to simultaneously analyze;
  - 0.1.10.1. a minimum of 11 Pb of data in a stable level of production, based on current estimations
  - 0.1.10.2. more than the minimum amount of data as the network grows for genomics and associated phenotyping workflows;
- 0.1.11. must be able to use the different classes of cloud storage for efficiency and cost saving purposes.
- 0.1.12. must be available from the Internet and support multiple end user device operating systems, including Windows, Mac and Linux.
- 0.1.13. have accounting capabilities for the FLN administrators.
- 0.1.14. must either provide:
  - 0.1.14.1. a complete open-source system where dedicated personnel from the host organization maintains and updates the FLN.
  - 0.1.14.2. a commercial system where source code will be shared confidentially with the NRC and retain usage rights in case of abandonment by the supplier.
- 0.1.15. must provide ongoing after sale support to provide security system updates and feature enhancement requested by the NRC.
- 0.1.16. support English and French interfaces for end users and administrators and have proper documentation and user's manuals.

## 0.2. *Scope:*

- 0.2.1. have the capacity to;
  - 0.2.1.1. onboard a minimum of 10 nodes comprised of government departments, academia and industry.
  - 0.2.1.2. maintain a stable level of production based on current estimates.
  - 0.2.1.3. scale as needed as more organizations register to the network.
- 0.2.2. focus on agricultural and environmental genomics.

## 0.3. *Data processing:*

- 0.3.1. must have the capacity to build, retain and run models or analysis pipelines against individual or aggregations of existing and/or future data sets.
  - 0.3.1.1. models: including, but not limited to: 3D image reconstruction, neural network for image analysis.
  - 0.3.1.2. analysis pipelines for datasets, including, but not limited to: structural genomics, metagenomics, transcriptomics, metagenomics, variant calling, metabolomics, phenotyping, computer vision.
- 0.3.2. must have the ability to increase power to previously run and saved analysis pipeline or models by providing to them supplemental datasets.
- 0.3.3. must provide a model/analysis importing/developing toolkit that would allow node users to run a novel pipeline/analysis that is not included in the FLN.
- 0.3.4. must have the ability to support the sharing of private datasets among participating partners (node-to-node).
- 0.3.5. must be able to pull data into the hub from publicly available data repositories (e.g. NCBI SRA).
- 0.3.6. must have the capability to store metadata for datasets, interoperable with other federated network models and databases.

## **3.2 OPERATIONAL SUPPORT REQUIREMENTS:**

- The solution must be managed and provided as a service which includes service level agreements for operational support.
- The service model will include operational support; including system and information security, data management, identity management, patching, upgrades and general system management.

## **3.3 COSTING REQUIREMENTS**

- For the purposes of this Request for Information, NRC is also seeking a cost model for a solution that responds to the above requirements. This model should provide cost estimates that include;
  - Data computation and storage estimates,
  - Licensing or a licensing cost model with costs (named user based, concurrent user based, enterprise licensing, etc...)
  - Annual operational support costs, including all software licensing
- The costing model should be of the following form:
  - A base cost for the implementation of the FLN
  - A cost in relation to the quantity of expandable features of the FLN (e.g. storage, number of nodes, number of pipelines and models to maintain)

#### **4.0 OBJECTIVES OF THE REQUEST FOR INFORMATION**

The purpose of this Request for Information (RFI) is to engage Respondents to:

- a) Provide industry with an early opportunity to assess, comment and suggest changes to the requirements and approach to building this federated network;
- b) Determine the capability of Respondents to provide for the requirements and provide associated cost estimates to assist in the establishment of the scope of the FLN;
- c) Obtain Respondent feedback on any issues that would impact their ability to bid on the potential RFP to deliver on Canada's requirements;
- d) Gather industry knowledge, expertise and recommendations with regard to best practices that would increase the success of the project and/or identify any risks that would impact the project.
- e) Enhance competition, access and fairness.

#### **5.0 NATURE OF THE REQUEST FOR INFORMATION**

This RFI is not a bid solicitation and will not result in the award of any Contract or Standing Offer nor will this RFI result in the creation of a list of potential bidders. Therefore, potential suppliers of any goods and services described in this RFI should not reserve stock or facilities, nor allocate resources, as a result of any information contained in this RFI.

Whether or not a submission is received in response to this RFI will not preclude any organization from participating in any future procurement process. Also, this RFI will not necessarily result in the procurement of any of the services that it describes. It is simply intended to solicit feedback with respect to its content.

Nothing in this RFI shall be construed as a commitment from the National Research Council Canada (NRC). NRC may use any non-proprietary information obtained as part of this review in the preparation of future requirements.

#### **6.0 INSTRUCTIONS FOR RESPONDING TO THE REQUEST FOR INFORMATION**

##### **6.1 NATURE AND FORMAT OF RESPONSES REQUESTED**

The department is seeking input and responses to specific questions (refer to Annex A) from industry stakeholders covering important elements of the requirement prior to proceeding with finalizing its procurement strategy.

Respondents are invited to provide comments regarding the technical requirements by completing and returning Annex B. Respondents should explain any assumptions they make in their interpretation of the requirements.

## **6.2 RESPONSE COSTS**

NRC will not reimburse any respondent for expenses incurred in responding to this RFI.

## **7.0 TREATMENT OF RESPONSES**

### **7.1 USE OF RESPONSES**

Responses will not be formally evaluated. However, the responses received may be used by NRC to develop or modify procurement strategies or any draft documents contained in this RFI. NRC will review all responses received by the RFI closing date. NRC may, in its discretion, review responses received after the RFI closing date.

### **7.2 REVIEW TEAM**

A review team composed of representatives of NRC will review the responses. NRC reserves the right to hire any independent consultant, or use any Government resources that it considers necessary to review any response. Not all members of the review team will necessarily review all responses.

### **7.3 CONFIDENTIALITY**

Respondents should indicate and mark any portions of their response that they consider proprietary or confidential. NRC will handle these portions in a confidential manner in accordance with the Access to Information Act of Canada.

### **7.4 FOLLOW-UP ACTIVITY**

NRC may, at its discretion, contact any respondents to follow up with additional questions or for clarification of any aspect of a response. NRC may meet with any respondent without having the obligation to meet with all respondents. NRC may, at its discretion agree to meet with respondents to provide respondents with the opportunity to present and/or demonstrate their capabilities in relation to this RFI. Respondents' presentations are at no obligation to NRC and respondents will be responsible for all costs associated with NRC's invitation to make a presentation.

## **8.0 SUBMISSION OF RESPONSES**

### **8.1 IDENTIFICATION OF RESPONSE**

The Respondent must provide a contact name, email address and telephone number when submitting their response. In the event that a response is not sufficiently clear, NRC reserves the right to seek additional information at their sole discretion.

### **8.2 TIME AND PLACE FOR SUBMISSION OF RESPONSES**

Respondents are requested to submit responses by March 10, 2023, **2 p.m. Eastern Daylight Time**. The review of responses will be ongoing as they are received. Respondents may submit their responses after the due date but these may or may not be reviewed.

Respondents are requested to submit responses to this RFI to Elizabeth Pitre, [Elizabeth.pitre@nrc-cnrc.gc.ca](mailto:Elizabeth.pitre@nrc-cnrc.gc.ca)

### **8.3 ENQUIRIES**

Only enquiries which clarify the questions asked or feedback requested may be answered with respect to this RFI. All enquiries must be submitted via email to Elizabeth Pitre, [Elizabeth.pitre@nrc-cnrc.gc.ca](mailto:Elizabeth.pitre@nrc-cnrc.gc.ca)

### **8.4 OFFICIAL LANGUAGES**

Responses to this RFI are requested to be presented in either of the Official Languages of Canada (English or French).

## **9.0 INDUSTRY ASSISTANCE - FEEDBACK AND DIRECTION**

Outlined below are 6 aspects the Project would like to get feedback and direction on. They are unordered and if possible, we ask you to provide your insight on all of them, however the 3 questions regarding the cost, the risks and the opportunity for improvement are the most critical ones.

### **1) Approach**

Please provide your opinion and thoughts on the approach detailed.

### **2) Schedule**

Please provide an estimate and overview on the schedule and schedule-related aspects such as timelines, milestones, etc. as they relate to the SI SOW and the overall Project (the



latter implies there are aspects you see that are out of scope for the SI SOW, if so please highlight these).

**3) Cost**

Please provide an estimate and overview on the cost and cost related aspects such as resource levels, travel, etc. as they relate to the SI SOW and the overall Project (the latter implies there are aspects you see that are out of scope for the SI SOW, if so, please highlight these).

**4) Risks/Concerns**

Please provide your opinion and thoughts on the risks and concerns as they relate to the SI SOW and the overall Project.

**5) Opportunity for Improvement**

Please provide your opinion and thoughts on any aspect that can be improved upon as they relate to the SI SOW and the overall Project.

**6) Feedback**

Please share your thoughts and feedback on the RFI, the documents, the process, or any other relevant aspect.

## ANNEX A. Industry Engagement Questions

Briefly introduce your company:

### 1. General Technology specifications for each objective:

Q. Does your company have experience in implementing a FLN that can enable agricultural data to be leveraged for research, analysis and prediction in support of the one or more of the desired requirements listed above?

- If yes, describe how your solution breaks down traditional barriers to data research, including how data access and consent is managed?
- If no, can you expand on how your solutions could be adapted for agricultural data?

Q. Agricultural data in Canada resides in numerous databases and institutions including, academia, industry and government.

- What restrictions or limitations would a FLN have in accessing different types of agricultural data across industry, academia and government?
- How do you envision a role for the FLN to facilitate data sharing and reuse of data to enable federated learning and analysis across the various datacenter?

Q. Do you have experience accessing data from systems held by Canadian-based and US-based datacenters and owners?

Q. Do you have experience working with data repositories collated from across Canada and organizations?

Q. Do you have experience working in environments where data generated from two different organizations or database, centers are used for analysis?

Q. Please describe how you would link data from two disparate IT systems for analysis using different IT vendors?

Q. Of the listed requirements in Section 3.0, are there any for which you would not recommend using a federated data solution? Alternatively, should Canada consider procurement based on subset of requirements? If so, how would this approach impact the FLN?

Q. Please describe the storage requirements needed for active and cold datasets based on the requirements listed in Section 3.0?

## **2. Operational steps implicated to set up:**

Q. Please outline the phases of work from start to finish in setting up and using a FLN in order to enable federated analysis, sharing, and federated learning across multiple existing digital systems and centers.

Q. What process is undertaken to achieve interoperability of data? For example, what protocol do you use to establish consensus on standards, including data content (e.g. semantic, terminology) and technical (exchange, integration) standards.

- Please provide details on the technical process and on how data custodians and stakeholders can achieve consensus on standards.

Q. Do you offer support and customization services for your solution?

Q. Please describe in detail the steps required for the implementation of the solution in a live environment, including onboarding, testing, training, change management, and any other activities.

Q. Does your organization have any experience using federated data solutions for agricultural omics data types listed in requirements section? Please describe the type of data.

Q. Is your solution currently being used by other clients? Could you briefly describe who they are and what was deployed?

Q. How much training and testing commitments for end-users/clients is required before launch and ongoing?

Q. What resource implications (to your company and to the client) are there for the set up and delivery of a federated data solution for both SaaS and a hosted solution? What timeframe would be expected across the various phases of work to establish both?

Q. What baseline human resource requirements are associated with delivery of a SaaS or hosted solution outlined above?

## **3. Privacy and security requirements:**

Q. If the FLN requires the collection and storage of sensitive and/or protected information, what protocols are in place to ensure protection and safeguarding of this protected information?

Q. What frameworks and agreements need to be developed in order for the solution to be deployed, such as Data Sharing Agreements (DSAs)?

Q. Is there standard policies in place for privacy, security, retention and management of data? What are the user requirements? What IT security compliance frameworks is your solution compliant with?

Q. Does your system allow for dataset security level tagging?

**4. Costs (software, development, ongoing maintenance):**

Q. What baseline costs are associated with the delivery of the solution including human resources? Are there any ongoing service or subscription costs?

Q. Please describe any service level agreements in place, including service availability, troubleshooting, time to repair, etc.

Q. How does the volume and quality of agricultural data influence price? What other factors will influence prices? Does the number of implicated stakeholders, data custodians and users impact the price and service costs?

Q. What is the delivery timeline for the identified solution from start to finish?

Q. How are accessibility standards addressed in your solution?

**5. Back-end technology requirements:**

Q. After a solution is deployed can new queries or data sources be amended? If yes, what steps are required to re-deploy?

Q. How are the results delivered to the client? Can clients access the solution directly?

Q. What type of service maintenance might be required for software and hardware?

Q. What human resource requirements are needed to maintain either a hosted or SaaS system outlined above?

Q. Can your solution integrate with existing federated networks and federated analysis and learning platforms? Please describe the process, including activities associated with engaging users of these systems.

Q. Is there a technology solution in place to capture and record data sharing, learning and analysis?

Q. Does your company foresee any challenges with deploying federated data solutions in areas of agriculture where data is considered highly unstandardized?

Q. Is your solution cloud-based?

## 6. Functionality:

Q. Briefly describe how your solution is able to conduct data analytics and research, including when data is developed using different standards and/or held in different systems for the data types described in Section 3.0.

Please outline how your service/solution would meet the following minimum expectations:

- Provide a streamlined and efficient mechanism to access, gather, analyze, and share (I.e., interoperability), using a common look and feel platform, specific data from a number and variety of different sources, platforms, and digital systems;
- Must adhere to GC Protected B security and personal information protections and protocols;
- Provide scalable service tiered options with associated pricing that Canada could select from;
- Be a Canadian owned and operated company, ensuring that all personnel meet on a need to know and role specific basis Canada's security clearance required obligations.
- Provide easy to use interface for efficient user experience; Please describe how this is accessed? E.g. web-browser, mobile application, etc.

Q. Are there any other factors that the Government of Canada should take into consideration as it assesses the setup of a procurement mechanism for the use of federated data solution for agriculture data?

## ANNEX B. Technical Requirements Table

### TECHNICAL REQUIREMENTS:

Section	Description	Yes/No	Comments
<b>3.1 Infrastructure and platform</b>	<p>3.1.1. the hub must be deployed in a public cloud service provider (CSP) which provides Canadian data residency. 3.1.1.1. The CSP must have a supplier arrangement with Shared Service Canada (SSC) cloud brokerage.</p> <p>3.1.2. must implement and comply with the GA4GH standards [<a href="https://www.ga4gh.org">https://www.ga4gh.org</a>] when applicable.</p> <p>3.1.3. must conform to the Canadian Government IT policies.</p> <p>3.1.4. must respect all cyber security and data privacy compliance as required.</p> <p>3.1.5. must provide infrastructure (e.g. infrastructure as code) and applications for the hub.</p> <p>3.1.6. must provide a toolkit for node onboarding. The toolkit will facilitate dataset exchange between nodes and provide a framework for dataset standardization in the FLN.</p> <p>3.1.7. must provide a toolset for the management of datasets visibility and accessibility across the FLN.</p> <p>3.1.8. must implement the necessary controls and management system to manage users accessing the FLN.</p> <p>3.1.9. must provide interfaces: 3.1.9.1. have a graphical interface front end. 3.1.9.2. have an Application Programmer Interface back end which allows the access to the FLN and all its tools.</p> <p>3.1.10. must have the capacity to simultaneously analyze;  3.1.10.1.a minimum of 11 Pb of data in a stable level of production, based on current estimations 3.1.10.2. more than the minimum amount of data as the network grows for genomics and associated phenotyping workflows;</p>		

Section	Description	Yes/No	Comments
<p><b>3.1 Infrastructure and platform</b></p>	<p>3.1.11. must be able to use the different classes of cloud storage for efficiency and cost saving purposes.</p> <p>3.1.12. must be available from the Internet and support multiple end user device operating systems, including Windows, Mac and Linux.</p> <p>3.1.13. have accounting capabilities for the FLN administrators.</p> <p>3.1.14. must either provide:  3.1.14.1. a complete open-source system where dedicated personnel from the host organization maintains and updates the FLN.  3.1.14.2. a commercial system where source code will be shared confidentially with the NRC and retain usage rights in case of abandonment by the supplier.</p> <p>3.1.15. must provide ongoing after sale support to provide security system updates and feature enhancement requested by the NRC.</p> <p>3.1.16. support English and French interfaces for end users and administrators and have proper documentation and user's manuals.</p>		
<p><b>3.2 Scope:</b></p>	<p>3.2.1. have the capacity to;  3.2.1.1. onboard a minimum of 10 nodes comprised of government departments, academia and industry.  3.2.1.2. maintain a stable level of production based on current estimates.  3.2.1.3. scale as needed as more organizations register to the network.</p> <p>3.2.2. focus on agricultural and environmental genomics.</p>		
<p><b>3.3 Data processing:</b></p>	<p>3.3.1. must have the capacity to build, retain and run models or analysis pipelines against individual or aggregations of existing and/or future data sets.</p> <p>3.3.1.1. models: including, but not limited to: 3D image reconstruction, neural network for image analysis.</p> <p>3.3.1.2. analysis pipelines for datasets, including, but not limited to: structural genomics, metagenomics, transcriptomics, metagenomics, variant calling, metabolomics, phenotyping, computer vision.</p> <p>3.3.2. must have the ability to increase power to previously run and saved analysis pipeline or models by providing to them supplemental datasets.</p>		

Section	Description	Yes/No	Comments
<b>3.3 Data processing</b>	<p>3.3.3. must provide a model/analysis importing/developing toolkit that would allow node users to run a novel pipeline/analysis that is not included in the FLN.</p> <p>3.3.4. must have the ability to support the sharing of private datasets among participating partners (node-to-node).</p> <p>3.3.5. must be able to pull data into the hub from publicly available data repositories (e.g. NCBI SRA).</p> <p>3.3.6. must have the capability to store metadata for datasets, interoperable with other federated network models and databases.</p>		

#### **OPERATIONAL SUPPORT REQUIREMENTS:**

- The solution must be managed and provided as a service which includes service level agreements for operational support.
- The service model will include operational support; including system and information security, data management, identity management, patching, upgrades and general system management.

#### **COSTING REQUIREMENTS**

- For the purposes of this Request for Information, NRC is also seeking a cost model for a solution that responds to the above requirements. This model should provide cost estimates that include;
  - Data computation and storage estimates,
  - Licensing or a licensing cost model with costs (named user based, concurrent user based, enterprise licensing, etc...)
  - Annual operational support costs, including all software licensing
- The costing model should be of the following form:
  - A base cost for the implementation of the FLN
  - A cost in relation to the quantity of expandable features of the FLN (e.g. storage, number of nodes, number of pipelines and models to maintain)