

Cancer Genomics Cloud Pilots Sources Sought

This is a Research and Development (R & D) Sources Sought notice. This is NOT a solicitation for proposals, proposal abstracts, or quotations. The purpose of this notice is to obtain information regarding the availability and capability of all qualified sources, including small businesses, to perform a potential R & D requirement. As such, this notice is also seeking Small Business information regarding: (1) the availability and capability of qualified small business sources; (2) whether they are small businesses; HUBZone small businesses; service-disabled, veteran-owned small businesses; 8(a) small businesses; veteran-owned small businesses; woman-owned small businesses; or small disadvantaged businesses; and (3) their size classification relative to the North American Industry Classification System (NAICS) code for the proposed acquisition. Responses to the information requested will assist the Government in determining the appropriate acquisition method, including whether a set-aside is possible. This requirement is assigned a code of 518210 in the North American Industry Classification System (NAICS), and the size standard for such requirements is \$30 million. All qualified organizations are encouraged to submit capability statements in response to the subject Sources Sought Notice.

THERE IS NO SOLICITATION CURRENTLY AVAILABLE. THIS SOURCES SOUGHT NOTICE IS STRICTLY FOR MARKET RESEARCH PURPOSES.

Background

The National Cancer Institute (NCI), Center for Biomedical Informatics and Information Technology (CBIT) is anticipating releasing a new Broad Agency Announcement for multiple cost-reimbursement contracts structured as phased awards for Cancer Genomics Cloud Pilots.

The current standard model for analyzing genomic data involves individual researchers downloading data stored at a variety of locations, adding their own data and then computing over these data on local hardware. This model has been successful for many years, but is becoming untenable given the enormous growth of biomedical data since the advent of large-scale scientific programs such as the Cancer Genome Atlas (TCGA) and Therapeutically Applicable Research to Generate Effective Treatments (TARGET) that use next generation sequencing technology. At its projected completion in 2014, it is expected that TCGA will generate approximately 2.5 Petabytes (PB) of data. The difficulties that are caused by such scale are three-fold:

1. Maintaining local copies of data at this scale is not cost effective and will continue to increase in cost as the data is growing faster than Kryder's Law¹
2. Providing computational capacity to analyze the data is similarly constrained
3. Electronic transfer of data at this scale is not practical.

With regard to the final point, the 2.5 PB expected TCGA core data set would require weeks to months to download even if the organization that is requesting the data has access to point-to-point 10 Gigabit networks². Anecdotal evidence, as well as the results of a request for information³ sent

¹ Kryder's law refers to the growth in storage capacity per unit cost and is an analog of the more commonly known Moore's Law.

² 2.5 PB is 20,000,000 Gigabits (Gb). At 10 Gb/sec, this equates to a 23 day download time. As a practical matter, transfers over such lines are unlikely to exceed 1-3 Gb/sec given the necessity of sharing such links with other users and the latency associated with file transfer protocols.

to all NCI grantees and intramural researchers indicates that with less than 0.5 PB of TCGA data available, much of the NCI-supported research community is already computationally limited by financial constraints and IT (network, storage, and computing) issues that exceed the capacity of individual investigators and most institutions.⁴

Purpose and Objectives

The purpose of this requirement is to support the development of a new model of computational analysis of biological data that has the potential to address the problems described above. This model (which has been articulated recently in a series of NCI Precision Medicine Workshops⁵ and publications by experts in the field of biomedical informatics⁶) involves the creation of a set of data repositories with co-located computational capacity and an Application Programming Interface (API) that provides security, data and resource access for developers of analytic tools. In this model, applications are brought to the data, rather than bringing the data to the applications. Such a “Cancer Genomics Cloud” has the capability to democratize access to NCI-generated genomic data and provide a more cost-effective way to provide computational support to the cancer research community.

Towards this end, the NCI is contemplating the development of multiple prototype Cancer Genomics Cloud pilots that will support computation on a core dataset of approximately 2.5PB and that will be available for use and testing by the scientific community. Key anticipated outcomes and benefits include the ability to:

1. Enhance access to community-generated cancer data and tools;
2. Take advantage of increases in computational efficiency and scalability;
3. Accelerate bioinformatics tool development;
4. Serve the individual researcher possessing limited or complete lack of high-performance computing access;
5. Better assess and optimize hardware need and costs for future production system(s).

The pilots are envisioned to be a community platform for cancer genome integration and analysis. If successful, the NCI will consult with the community and relevant oversight committee(s) to define, and potentially construct, production version(s) of one or more of the pilot clouds, or a successor design based on the results of the pilot evaluations. In spirit, and by design, this new initiative is envisioned to be the first step toward a full Cancer Knowledge Commons, along the lines referenced in the National Academy report on Precision Medicine, and would be coordinated with The National Institutes of Health’s Big Data to Knowledge (BD2K) initiative and other biomedical cloud activities.

Project Requirements

Your capability statement shall identify and support the design, implementation, and documentation of an innovative Cancer Genomics Cloud pilot while identifying potential obstacles. The pilot would provide access to key NCI data sets as well as computing resources (i.e. storage, servers, high capacity networks, etc.) and users must be able to compute over their own data in

³ <http://ncip.nci.nih.gov/nci-cloud-initiative>

⁴ <https://wiki.nci.nih.gov/x/CLbBw>

⁵ Precision Medicine Workshops were held on 4/24/2012 and 9/10/2012

⁶ See, for example, Grossman, RL and White, KP, 2012, Journal of Internal Medicine 271:2, pp 122-130, Haussler, D, et al., EECS Department, University of California, Berkeley, Tech. Report UCB/EECS-2012-211.

conjunction with the NCI data sets. It is important that data be represented using standard formats, data elements, and vocabularies. Although the government is using the term “cloud” for convenience, the NCI is not anticipating requiring that the computing environment utilize commercial commodity cloud offerings. The government is interested in all possible solutions (including those that utilize commodity cloud offerings) so long as they are capable of meeting the government’s needs to support computing over peta-scale datasets.

The NCI further anticipates the Cancer Genomics Cloud pilots will serve diverse communities, including application developers, bioinformaticians, computational biologists, molecular biologists, and ultimately clinicians and patients. As such, the computational interface must support query and analysis tools that will be used by a broad range of stakeholders. NCI envisions any potential awardee will be expected to work in concert with other awardees and the supplier of the TCGA data set to identify, implement and demonstrate key interoperability features among the pilot clouds while complying with federal IT security requirements.

Anticipated Period of Performance

It is anticipated that the period of performance for this effort will be divided into 3 phases envisioned as: a 3 month design phase, a 12 month pilot creation phase, and a 6 month testing and evaluation phase; a total of 21 months. The NCI contemplates the third phase of the project will include testing by the community of biomedical researchers; as such, any potential offeror must have the ability to provide pilot cloud access to individuals and institutions as directed by the NCI.

Capability Statement

Any interested sources are requested to submit a written capability statement demonstrating their ability and experience in performing the requirements reflected in this notice. Submissions will be reviewed for their experience and expertise relevant to the design, implementation, and documentation of innovative Cancer Genomics Cloud pilots. Interested sources should demonstrate past performance and expertise in:

1. **Personnel/Experience:**
 - a) Ability to hire staff possessing pertinent expertise including availability, experience, and related, formal training; such staff must include experts in both information technology and cancer genomics;
 - b) In house capability and capacity to perform the work;
 - c) Prior completed projects of similar nature; including examples of prior completed Government contracts, references, and other related information
 - d) Organizational experience and management capability;

2. **Facilities:** Interested sources shall have the necessary dedicated office space and accommodations to fulfill the requirements; access to pre-existing information technology support space (server facilities with adequate power and cooling, network connectivity, etc.) either within the source’s facilities or through arrangements with third parties. It is not the government’s intention to finance the construction of new or expansion of existing information technology hosting facilities. The facilities shall have secure computer services that are in full compliance with NIH computer security requirements (<https://ocio.nih.gov/InfoSecurity/Policy/Pages/default.aspx>, <http://ais.nci.nih.gov/>). The source’s facility must have the ability to host online meetings between the Contracting Officer’s Representative (COR) at NCI and the contractor facility.

Submission Information

Any responsible source that can perform the requirements outlined above may submit a capability statement for consideration that clearly details the ability to perform the aspects of this notice. Submissions shall be provided as PDF formatted documents to the NCI Contracting Officer by the response date and time, Wednesday, July 17, 2013 at 4:00PM at scott.keasey@nih.gov. There is a 15 page limit, (excluding resumes) and submissions must be presented in English, in a single-spaced document, using a 12-point font size. No collect calls or facsimile transmissions will be accepted. All responses must be received by the specified due date and time in order to be considered. CAPABILITY STATEMENTS RECEIVED AFTER THIS DATE AND TIME MAY NOT BE CONSIDERED. Capability statements will not be returned.

Respondents must include the following information:

- A. Capability to meet FAR 52.219-14, Limitation in Subcontracting requirements (as applicable)
- B. Business Information and Identification:
 1. DUNS number;
 2. Organization name and address;
 3. Organizational Point of contact (POC) and email address;
 4. Size and type of business (e.g., Large, 8(a), HUBZONE, etc.) pursuant to the applicable NAICS code; as validated via the System Award Management (SAM). Additional information on NAICS codes can be found at www.sba.gov and/or www.census.gov. All potential offerors must be registered on the System Award Management, located at <https://www.sam.gov/portal/public/SAM/>;
 5. Organizational structure (corporation, LLC, sole proprietorship, partnership, limited liability partnership, professional corporate, non-profit, academic institution, etc.);
 6. References - POC, names, phone numbers, and emails of references who can verify the capabilities/experience identified in the responses;
 7. Potential partners for Teaming Arrangements, including name, address and POC information.

Disclaimer and Important Notes: This notice does not obligate the Government to award a contract or otherwise pay for the information provided in response. The Government reserves the right to use information provided by respondents for any purpose deemed necessary and legally appropriate. Any organization responding to this notice should ensure that its response is complete and sufficiently detailed to allow the Government to determine the organization's qualifications to perform the work. Respondents are advised that the Government is under no obligation to acknowledge receipt of the information received or provide feedback to respondents with respect to any information submitted. After a review of the responses received, a pre-solicitation synopsis and solicitation may be published in Federal Business Opportunities. However, responses to this notice will not be considered adequate responses to a solicitation(s). The Government reserves the right to use any non-proprietary technical information in any resultant solicitation.

Confidentiality: No proprietary, classified, confidential, or sensitive information should be included in your response.