IN STEP: Intramural NCI STRIDES-based Transition and Exploration Program

Program Rationale: While the extensive computational and data management capabilities available in the cloud hold great promise for accelerating the pace of biomedical research, Investigators within NCI's Intramural Research Program (IRP) currently have no formal mechanism for seeking cloud-based research support. In contrast, on-premises compute, storage and software are readily available and centrally funded through capital expenditures. This difference in availability biases Investigators to use on-premises resources even in cases in which the most efficient approach to address a specific research question may be cloud-based and has uncoupled IRP needs from the development of NCI's cloud-based research infrastructure.

IN STEP Goals and Objectives: By launching the Intramural NCI STRIDES-based Transition and Exploration Program (IN STEP), NCI aims to **1) initiate a process for Intramural Investigators to seek cloud-based research support** and **2) accelerate development of NCI's cloud-based research infrastructure**, hosted by NCI's Center for Biomedical Informatics and Information Technology (CBIIT). IN STEP can be viewed in part as an implementation vehicle for the recently launched NIH STRIDES initiative, which provides ICs with cost-effective access to advanced cloud-based computational and data management capabilities, professional consultations, technical support, and cloud training for scientists, developers, engineers, and system administrators. Through IN STEP, CBIIT will work with the IRP to explore strategic deployment of a centrally-funded cloud-based research infrastructure tailored to IRP needs, leverage STRIDES discounted pricing on cloud compute, storage and services, and provide multiple sources of cloud engineering support to transition IRP projects to the cloud.

IN STEP objectives:
- o Identify intramural research needs that can be most effectively met in the cloud
- o Evaluate the utility of existing NCI Cloud Resources to support IRP needs, where appropriate
- o Fund immediate deployment of selected high priority NCI IRP projects to the cloud
- o Develop a 3-5 year projection of IRP cloud resource requirements
- o Plan for sustained funding and support of cloud resources for the IRP

FY20 IN STEP Call for Letters of Interest

Overview: Identifying appropriate cloud-based projects and building the necessary infrastructure to execute and sustain them is inherently a collaborative effort that depends on IRP scientists at all levels to define the needs, IT professionals from CBIIT, CIT, FNL and cloud service providers to develop technical solutions, and NCI leadership to prioritize the needs and support implementation a successful long-term cloud strategy.

The process will begin with a **Call for Letters of Interest** (LOIs) as a means to identify how current IT infrastructure limitations are impeding specific IRP research objectives and initiate exploration of possible cloud-based solutions. LOIs will inform needs assessment, resource projection, and future planning activities, and will be further developed via consultation into full proposals for consideration for immediate cloud deployment. Note that it is **not necessary for IRP scientific staff to describe how final cloud-based solutions would be structured**, but rather to define the scientific/technical objectives and describe current limitations.

Eligibility: To carry out the IN STEP objectives, CBIIT invites IRP Investigators, Staff Scientists, and Staff Clinicians to submit LOIs describing potential cloud projects and/or cloud resources they wish to see developed over the next 3-5 years to support intramural research.

Examples of possible types of projects include but are not limited to:
- o Full-stack computationally-based workflows requiring cloud services or workflow engines that are unavailable or difficult to replicate on premises or would benefit from elastic cloud compute (e.g. image processing and quantitative analysis workflows). This could include support for small-scale development and testing of new workflows or large-scale deployments of established workflows.

- o Cloud platforms used by numerous NCI groups that could be brought together under a single umbrella contract to achieve lower cost, greater flexibility, hybrid on-premises/cloud architecture, and/or enhanced collaboration opportunities.

- o Intramural/extramural collaborative development of computationally-based applications that would be enabled by simplified cloud-based federated log-in procedures.

- o Hosting of large investigator-developed resources that exceed the capacity of externally-facing on premises resources to support heavy use by extramural users (e.g. pathology imaging repositories, *in silico* screening applications, molecular modeling tools etc.)
- o Projects that could leverage any of the existing NCI Cloud Resources, supported by Seven Bridges, the Institute for Systems Biology, and/or the Broad Institute.

Due Date: LOIs should be submitted by October 31, 2019 to instep@mail.nih.gov

LOI Content: Narratives of 2-3 pages should address the following points; additional pages for figures and references may be used if needed.

1. **Scientific and technical objectives** the Investigator seeks to achieve, discussion of current approach(es), why this work should be considered a high priority, and how it will uniquely enable high-impact research.

2. **Technical or infrastructure-based limitations and/or resource gaps** currently hindering development of more effective approaches to achieve the scientific objectives.

   - If Investigators have attempted to explore or implement cloud solutions previously but encountered difficulties, those efforts should be described here.

3. Briefly describe **how cloud resources are envisioned to enable the scientific or technical objectives and/or overcome current barriers**.

   - This statement is intended to stimulate thinking about cloud solutions and **does not need to be fully developed for the LOI**.
   - Investigators may find it useful to consider the key cloud characteristics of scalability, accessibility, cost-effectiveness, flexibility, and security when thinking about using cloud services for research.
     - o For scientists and developers, cloud environments are especially well-suited for rapid, iterative, development and testing of computationally-based workflows because of the availability of pre-existing services, minimal upfront investment, accessibility for collaborators, and the ability to rapidly scale up and deploy workflows deemed to be successful.
     - o Cloud is especially attractive when rapid access to the fastest nodes or newest software is desired, because the cloud procurement cycle is exceptionally fast by comparison to capital expenditures, and users pay only for what they use.
     - o For large, collaborative efforts, cloud-based federated log-in procedures can be faster and easier than setting up accounts for outside users to access externally-facing resources and are equally secure.
     - o Using cloud-based compute is frequently a faster and less expensive approach for projects that depend on data that is stored in the cloud, compared to downloading large datasets to store and operate on locally.
     - o Rapidly evolving compute-intensive software that is required by numerous users can be much easier to maintain from a cloud installation, because updates to both software and hardware can be done centrally rather than on individual workstations.

4. **Collaborators** or other potential beneficiaries of the proposed cloud infrastructure.

5. **Data flow** from beginning to end, e.g., from the data-generating instruments through the processing workflows, to the end user. Diagrams are encouraged but not required.

6. **Key references** highlighting the science and/or methods that will be advanced.

Proposal Development, Review and Selection Process:

- LOIs will result in consultations with CBIIT staff or other consultants to explore technical feasibility and the potential for broad utility/applicability. Note that LOIs proposing similar infrastructure may be grouped together for further development. Some LOIs may be routed to existing cloud or on premises resources, as appropriate.

- LOIs that are deemed to be technically feasible, cloud appropriate, and have the potential for broad utility will be routed for Divisional approval and coordination for full proposal development.

- Cloud development project plans and cost estimates will be developed by CBIIT for inclusion with full proposals.

- Full proposals will be reviewed for scientific merit and ranked for funding priority by the IN STEP Review Panel, which will include scientific experts from within and beyond the IRP appropriate to specific proposals under review.

- Final funding decisions for projects to be executed in FY20 will be made in consultation with CBIIT and NCI leadership in **January 2020**. Selection of projects will depend on availability of funds and cost of the highest-priority projects. Meritorious projects that cannot be funded in the first round will remain in consideration for development as additional funds are identified.

- Information gained about overall IRP cloud needs will be used by CBIIT and NCI leadership to develop projections for future IRP cloud development needs, identify funds and resources for cloud development, and further refine the cloud project development process.

- Follow-up activities will include a workshop for IN STEP awardees and other stakeholders focused on discussing how Cloud has advanced NCI's data-intensive projects and evaluating IN STEP objectives, outcomes, and lessons learned.

Questions on IN STEP or requests for assistance with developing or submitting LOIs may be directed to the In STEP mailbox at instep@mail.nih.gov, where they will be answered by either Janelle Cortner or Carl McCabe. Interested Investigators may also contact either directly (cortnerj@mail.nih.gov or Carl.McCabe@nih.gov).