Common Fund STRIDES Decision Support Tool

August 2019

Purpose

STRIDES (Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability) is an initiative that involves NIH-funded other transactions agreements with Cloud Service Providers (CSPs) through their billing partners for discounted cloud costs for NIH and its awardees. The Common Fund provided $8 million to the Google/Carahsoft STRIDES other transaction award and $8 million to the AWS/Four Points award in FY 2018. **Because the Common Fund made this payment through the STRIDES Other Transaction awards, all cloud costs for Common Fund Data Coordinating Centers and other large data awards are expected to be paid through STRIDES as in kind contributions to the Data Coordinating Centers/large data awards.** This document describes how to access in kind value for cloud expenses for CF programs via STRIDES.

Definitions

- Cloud Services Provider (CSP) - Currently, STRIDES cloud service providers include only Google Cloud and Amazon Web Services (AWS). These CSPs provide specific capabilities, such as storage, computation/analytics, and data transfer services, in an on-demand, pay-as-you-go manner.
- Cloud Services Billing Partner - An intermediary who handles accounting and billing for the CSPs. Google's Billing Partner is Carahsoft; AWS's Billing Partner is Four Points Technology.
- Common Fund Program - Defined to include any current or planned program or project funded by the NIH Common Fund. See https://commonfund.nih.gov/initiativeslist for a complete list of current programs.
- Daily Operation of Cloud Environments - a set of activities and costs that trigger the inclusion of terms and conditions in award documents for the purposes of accessing in kind contribution of cloud services from the STRIDES agreements. Such activities include, but are not limited to, data storage, provision of access, other account management, account auditing, charge/usage tracking, fulfilling utilization notification responsibilities, and the like.
- Data Egress Fees - Cloud providers charge for all outgoing traffic from their cloud environments. This includes everything from sending out web pages, returning results of API calls and sending out data files when a user requests a download. Downloading data for local use has been a common practice that was appropriate when datasets were smaller and local compute was the only option; however, datasets are now much larger and cloud compute resources far more capable. As a result, NIH plans to foster a more cloud-centric model that minimizes data movement out of the cloud and in turn minimizes any associated data egress fees.

- STRIDES team - federal employees and contractors of the NIH's Center for Information Technology (CIT) that liaise with the NIH Institutes, Centers, and Offices on the STRIDES vehicle.

About STRIDES

STRIDES is an initiative that aims to address the challenges associated with establishing a cloud-based, FAIR data ecosystem by establishing partnerships with commercial CSPs and harnessing the power of the commercial cloud in support of biomedical research. See https://datascience.nih.gov/strides for additional information. Google Cloud and AWS currently partner with NIH on the STRIDES initiative.

What STRIDES Offers

STRIDES provides access to the most advanced, cost-effective computational infrastructure, tools, and services to help accelerate advances in biomedical research. STRIDES provides the following:
- **Discounts on Cloud Services** – Favorable pricing on computing, storage, and related cloud services.
- **Professional Services** – Access to professional service consultations and technical support from cloud service provider partners. Technical support may include engineering labor.
- **Training** – Access to training for researchers, data owners, and others to help ensure optimal use of available tools and technologies.
- **Potential Collaborative Engagements** – Opportunities to explore methods and approaches that may advance NIH's biomedical research objectives (with scope and milestones of engagements agreed upon separately).

STRIDES agreements ease access to and use of cloud services on supported cloud services platforms (currently Google and AWS).

STRIDES Management

The NIH Center for Information Technology (CIT) programmatically manages STRIDES and assists NIH Program Officers and PD/PIs with such efforts as:
- Helping to determine whether cloud services will benefit research program objectives
- Facilitating discussions with Google and AWS to assess which cloud platform(s) might be best suited for the research needs
- Understanding accounting and billing structures, especially as they pertain to Common Fund use of STRIDES, and structuring the creation of billing accounts (how money flows to pay for cloud services)
- Connecting NIH or PD/PIs with Google and AWS account managers, as well as engineering and technical support staff (as needed)

- Creating administrative and/or user accounts on cloud platforms, when needed
- **Providing billing and usage reports to PD/PIs, Program Officers, and other specified delegates**

Eligibility to Use STRIDES

All Common Fund-supported programs/projects are eligible to establish a STRIDES account so that data storage and compute costs are paid via the STRIDES Other Transactions awards. Common Fund programs that involve data resources are intended to provide public data users with free access to data (unless specified otherwise for individual programs), so the STRIDES Other Transactions awards are intended to cover costs associated with cloud-based activities on the data (up to specific per-project limits as approved by OSC leadership; see below).

Common Fund Program Staff Step-by-Step Guidance

A. Project Eligibility and Readiness

Common Fund program staff should first determine whether the project (e.g., dataset, analytical tools, research platforms, and the like) is ready for launch in a cloud environment or, for existing data sets, transition to the cloud environment. If data are already in the Google or AWS cloud environment, the program is still eligible for the cost-savings offered by STRIDES. To determine eligibility:

1. **Is the project already in the cloud environment? If yes, proceed to Question 3**. If no, proceed to Question 2.

2. **Is the project "ready" for the cloud environment? If yes, proceed to Question 3**. If no, you'll want to prepare the project for the cloud environment before proceeding further. If program staff are unsure about cloud readiness, they should consult the Cloud Guidebook, Common Fund Data Ecosystem team, or STRIDES team.

3. **Is the project generating a public data resource (i.e., data that are expected to be accessed by many users as a research resource) or does it otherwise involve large data sets?** If yes, proceed to Question 4. If not, proceed to Section B.

4. **Have you engaged with the Common Fund Data Ecosystem Project?** The Common Fund is supporting the development of a common set of resources for all Common Fund data resources through the Common Fund Data Ecosystem Project. It is expected that all Common Fund Data Resource Programs leveraging STRIDES agreements contribute to/participate in the Common Fund Data Ecosystem Project. **If yes, proceed to B.**

Setting up a STRIDES account. If no, please contact the Common Fund Data Ecosystem (CFDE) Program Team (cfde@od.nih.gov).

B. Setting up a STRIDES account

Submit a Recommendation

A requestor (an NIH Program Official), in collaboration with the Common Fund Program Leader, Common Fund Program Officer, STRIDES onboarding representative must:

1. Prepare a recommendation to use STRIDES for the project of interest. The recommendation includes:
   - The type and amount of resources (financial or other) being requested
   - A description of what the resources will be used for
   - If relevant, a description of the need to support egress fees incurred by any large scale data download functionalities that are deemed to be a core function of the cloud environment. Include an estimate of anticipated costs.
   - A summary of the current FAIR status of the data and the PI's commitment to making the data FAIR and/or working with the Common Fund Data Ecosystem as appropriate. This section must include the following components, but each can be brief (2-3 sentences):
     - An initial review of the resource with regards to FAIR. If an initial review has not been completed, this should be stated with details about the current status of the resource and reasons why a FAIRness review has not previously been completed.
     - Plans for improving the resource's FAIRness over time.
     - Plans for more formal, periodic reassessments (e.g. using FAIRshake) to track the improvement in the FAIR score over time.
   - Which cloud service provider(s) will be utilized, if known
   - A budget for in kind value to be provided
   - A description of funds already provided for compute and storage within the active award and how those funds will be repurposed.
   - A description of cost requirements for the awardee to adopt community standards for FAIR data
   - A proposed timeline for establishing the STRIDES account
   - Proposed terms and conditions of award (see "Sample Award Terms and Conditions" section below).
   - Description of touchpoint(s), if any, with the Common Fund Data Ecosystem Project Leader
2. Email the recommendation to the Common Fund Program Leader. S/he will then convey the recommendation to the OSC Director and the STRIDES Program Lead, with a cc to the OSC Associate Director and the OSC Operations Team Leader.

3. Expect the review process to take 5-10 business days, with a decision transmitted by the Office of Strategic Coordination Leadership via email to the Common Fund Program Leader and Program Officer for the relevant Common Fund program.

After Approval to Draw-down on Direct Funds

1. The OSC Leadership will communicate the official decision to the OSC Program Leader, the OSC Trio for the program, and the OSC Operations Team Leader.
2. The OSC Program Leader will convey the approval to the Program Officer for the relevant award; the OSC Operations Team Leader will convey the approval to the Chief GMO of the IC where the award is administered. The Program Officer, with assistance from the OSC Operations Team Leader as needed, will resolve questions and concerns about the Proposed Terms with the Chief GMO; the Chief GMO will finalize the award terms.
3. The Grants Specialist for the award will incorporate the award terms at the next award renewal or when funds are otherwise added to the award.
4. The Program Officer will work with the PI to establish a phone call consultation between the STRIDES team, the OSC Program Leader, the Grants Specialist, and the designated institutional official for the award. This consultation will explain the in-kind services to be provided and the roles and responsibilities as outlined in the Terms of Award. (Note: this call may occur before the award is revised with the new terms)
5. The consultative phone call will confirm the preferred cloud service provider(s) that the program wishes to use.
6. The Program Team will work with the STRIDES Team, the PI (or their designee) and the relevant cloud services billing partners to set up the account structure.

Sample Award Terms and Conditions for Grants, Cooperative Agreements, and Other Transactions that will receive in kind services through STRIDES

- Through the Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) initiative agreements, NIH provides in-kind value for cloud storage of data, costs associated with all users who compute on the data in a cloud environment, and costs required to engage engineers or other staff of Cloud Service Provider companies if needed. NIH will not award funds to the institution for this purpose. Allowable in-kind value associated with data egress are not covered unless specifically noted below:
  - Egress charges to be covered (if any) [specify]
  - The value of the coverage of all costs to be provided is [specify].
- As a condition of receipt of the in-kind cloud service coverage, the PI and relevant designees will participate in the Common Fund Data Ecosystem and adopt community standards developed through that effort.
- The PD/PI(s) and Authorized Organizational Representative (AOR), must work with the NIH Program Official to: 1) establish a STRIDES Billing Account for this award; 2)

establish a process for establishing sub-accounts for users associated with the STRIDES Billing Account ; 3) set spending limits for each sub-recipient/sub-awardee/consortia member/public user, as appropriate; and 4) monitor account use of STRIDES funds including limitations on data egress. It is ultimately the responsibility of the PD/PI(s) to oversee work to be billed against the STRIDES Billing Account and to ensure that costs do not exceed the amount specified above.

- The NIH, through the Program Official named in the Notice of Award, will provide regular billing information to the AOR with a copy to the PD/PI(s)to track costs and to help prevent usage overruns. The NIH will provide automated alert notifications to the PD/PI(s) and AOR when 75% of the usage reserved for this purpose has been spent. The PD/PI and AOR must provide written acknowledgement of receipt of this alert and a plan for how the recipient will monitor the remaining usage for the 25%. If usage exceeds the amount reserved by 25% or more, the NIH may elect to adjust or offset the award by the amount exceeded. Please contact your Grants Management Official if there are questions surrounding this specific term.
- At closeout of the award, or if NIH decides to provide support for cloud services via alternate means, the PI or designee will transfer administrative activities associated with the STRIDES account to an NIH designee, including relevant documentation concerning user groups, permissions, etc.
- [This bullet should be included for existing awards] Funds previously provided through this award for the purposes of paying for cloud services may be used to support the PD/PI(s) in efforts to ensure that data meet community standards to become Findable, Accessible, Interoperable, and Reusable (FAIR). These costs include the following [list the specific costs].