SUBMISSION PROCESS OVERVIEW

Submitting to CDS is rather straightforward.

First, we ask submitters to start by filling out a Submission Questionnaire and sending it to us (cdshelpdesk@mail.nih.gov). We'll schedule a short interview to discuss the requirements for submitting to CDS, to make sure we understand the data, and that we are a good fit for you.

Prior to submitting the data, you'll need to submit a manifest of your submissions metadata. This is like a summary of what is in your submission and includes things such as filenames, MD5 checksums for the files, and file size. We use the metadata manifest to index the data so it can be made available via NCI Cloud Resources. Once the metadata manifest has been approved you will receive credentials to start your data submission in your bucket on Cloud One. Please note that our indexing process is very time consuming and can take anywhere from six to eight weeks.

Then, once you are ready to start your submission, we'll send you and instructions on how to upload your data to a cloud S3 bucket. We have open office hours and can walk you through the technical details to help you get started. Once you are done uploading, reach out to us to notify us and we will lock the bucket (you will continue to have read-only access to the bucket).

After the indexing is complete, the data will be available through NCI's Cancer Research Data Commons (CRDC). NCI requires authentication through eRA Commons and dbGaP authorization to access controlled data.

FAQ

Q: How is the data made available?
- Through NCI Cloud resources: access analysis tools, workflows, & workspaces
- Through NCBI (dbGaP/SRA)
- Both controlled access and open access data
- Access is controlled based on dbGaP whitelists
- General rule: direct data download not supported

Q: What data is eligible for CDS?
- Open to any NCI Program; Data from *NCI funded programs,* available to research community for secondary data sharing
- Accepts various data types (e.g. genomic, proteomic, imaging)

Q: How long does submission process take?

- At the beginning: it can take about a week to create a bucket and credentials for your submission
- Indexing: after you send us the metadata manifest and the uploading is complete,

SUBMISSION PROCESS OUTLINE

1. Contact CDS
   a. CDS Helpdesk email (cdshelpdesk@mail.nih.gov)
2. Initial Interview
   a. Discuss data storage, sharing requirements, and requirements such as the metadata manifest.
   b. Please fill out questionnaire in advance (it will be sent by CDS) as it helps guide the discussion
3. Pre-requisites:
   a. Register study with dbGaP (if controlled access)
      i. Select Data Commons Framework (DCF) as the external database
      ii. Ask the GPA to provide KanakaDurga Addepalli with streamlined access
      iii. Add **Note for Data in CDS:** The **data** files for this study for this study are available via the NCI Cancer Data Service (CDS). More information about the NCI CDS is available here: https://datacommons.cancer.gov/repository/cancer-data-service. Other files can still be downloaded from dbGaP.
   b. Metadata manifest
   c. Other requirements (e.g. NCI funded)
   d. Note: when working with dbGaP curators submitter will need to add DCF as external database
4. (Submission "Begins")
   a. Submitter provides list of submitters
5. CDS creates buckets
   a. CDS Sends credentials (for submitter, PI)
6. Submitter uploads data to buckets using credentials
7. Submitter notifies CDS that submission is complete
   a. CDS then locks the bucket – making it read-only
8. CDS begins indexing studies: time required: ~ 6 to 8 weeks
   a. Indexing and A&A via Data Commons Framework (DCF)
9. Study is released and available
   a. Searchable via CGC at SevenBridgesGenomics (SBG)