

CDS New Request Questionnaire

OMB

Expiration

No.:

Date:

0925-7775

06/30/2025

Collection of this information is authorized by The Public Health Service Act, Section 411 (42 USC 285a). Rights of participants are protected by The Privacy Act of 1974. Participation is voluntary, and there are no penalties for not participating or withdrawing at any time. Refusal to participate will not affect your benefits in any way. The information collected will be kept private to the extent provided by law. Names and other identifiers will not appear in any report. Information provided will be combined for all participants and reported as summaries. You are being contacted by email to complete this form so that NCI can consider your study for submission into the Cancer Data Service.

Public reporting burden for this collection of information is estimated to average 60 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to: NIH, Project Clearance Branch, 6705 Rockledge Drive, MSC 7974, Bethesda, MD 20892-7974, ATTN: PRA (0925-7775). Do not return the completed form to this address.

The following sets of high-level questions are intended to provide an insight to CDS, into the data storage, access and secondary sharing needs and requirements of data submitters. It is requested that the submitters answer as many questions as they can. It is not required to answer all questions.

Data Characteristics

1. What are the principal types of data the program will be submitting (e.g., genomic, clinical, imaging)?
2. Will there be additional data types associated with the principal data types, not being submitted to CDS? For ex: Proteomics, Imaging etc.

3. Do you anticipate other additional data types to be submitted to CDS in future? For example, data type that does not fit the submission criteria to any of the present CRDC nodes.
4. Is the data from Humans?
NOTE: CDS accepts only Human data at this point.
5. What additional associated data would you be providing? For ex: Clinical/Phenomics data from study subjects (participants) and/or any other study associated metadata/searchable variables. Describe the format for each.
NOTE: CDS at this point will accept all metadata submitted.
6. What is the total number of samples and cases per study, being submitted?
7. For Genomics datasets, though CDS takes BAM files, it is preferred to submit CRAM files. Would you be able to provide CRAM files instead of BAM files?

Data Storage and Management

1. Who is the PI on the study?
2. How much data are you planning to submit to CDS?
 - a. By data type (if known)?
3. What is the reason you are looking for storage with CDS? What are your challenges related to the storage of data?
4. Do you have a preference of AWS versus Google cloud for storage? CDS provides AWS storage as of now and plans to provide Google storage in near future.

Data Submission

1. Who will submit the data, the PI (or the PI's team) or a collaborator?
2. Would there be multiple uploaders (ex: by data type or working groups)?
3. Is there a program timeline associated with the data Submission?
4. When do you plan to start submitting data to CDS?
5. Who is the primary point of contact for data submission?
6. Do you plan one or multiple submissions to CDS? For example, multiple studies or newer versions of the data for the same study.
 - a. If yes, do you have a timeline for the successive submissions?

- b. If this submission has data from a newer version of the study already submitted to CDS, do you want to retain data from the older version/s at CDS?
- 7. Do you have an Amazon/Google account for data submission? CDS submissions presently require that the data uploaders have an Amazon account.

Data Sharing

1. Is your data being released to broader research community for secondary sharing?
 - a. When is the data planned to be released for secondary sharing ?
2. Is your data sensitive, i.e., require controlled access?
3. Has the data been registered with any public sharing repository such as dbGaP?
 - a. If not, is there a reason?
 - b. If yes, please share the associated study ID, for ex: dbGaP PHS number.
 - c. Is the study RELEASED by dbGaP?
 - a. Is data currently shared through NCBI /dbGaP or other means? What is a plausible timeline?
4. Has the data already been submitted to any data repository? For Ex: SRA
5. CDS does not allow downloads. Given this, does CDS meet your data sharing needs?
6. Are there any data access limitations?
 - a. Is the data embargoed? If yes, would the data reside in CDS during that time? How would it effect user access?
7. Is any part of this data “open-access” ? for ex: VCFs from Genomics studies.
8. How can you assure the data does not contain PII and PHI and/or identifiable data elements?

Data Users/Access

1. How do the users access the data presently?
 - a. How well are these methods working today?
2. Will your data be made accessible through any other repository?

Data Analysis

1. For what purpose(s) do the approved users access the data?

- a. Conduct analyses / computations?
 - b. Cite in a publication?
2. Do they need to link this data to other data types in other repositories/CRDC nodes for analysis?

Data Post CDS Destination

1. Do you know the post-CDS destination for this data? For ex: to other CRDC nodes such as GDC, PDC, IDC etc.
 - a. Is there any plan to move data out of CDS buckets, before sharing publicly?

Is there any other information you would like to share about your data?