

Attachment 4: Examples of Data Submission/Sharing Forms

Cancer Data Service (CDS) Data Submission Form – Requiring Revision

Proteomic Data Commons (PDC) Data Submission Form

Integrated Canine Data Commons (ICDC) Data Submission Form

Non-NCI Study Information Form

Cancer Data Service (CDS) Data Submission Form

(Word document sent upon request) – Requiring Revision

The following form has been revised to include edited questions, new questions, and be administered as an online form embedded in the Cancer Research Data Commons Data Submission Portal. The CRDC team aims to make this the standardized form for all data submissions to the Cancer Research Data Commons, ideally replacing the Proteomic Data Commons Data and Integrated Data Commons data submission forms once the CRDC data submission process is available to all.

OMB No.: 0925-7775

Expiration Date: 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?

Proteomic Data Commons (PDC) Data Submission Form (Word document sent upon request)

OMB No.: 0925-7775

Expiration Date: 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 online to complete this form so that NCI can consider your study for submission into the Proteomic Data Commons.

Public reporting burden for this collection of information is estimated to average 30 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.

Please complete the following document and send to: PDCHelpDesk@mail.nih.gov.

Please include a narrative describing your study and its scientific benefit for inclusion in the Proteomic Data Commons (PDC).

Please include the following information:

1. Name/Identifier of Study with a brief description
2. Grant ID and funding source (if applicable)
3. IRB approval numbers (if applicable)

4. Scientific Point of Contact (Name, Phone, Email)
5. Data Manager Point of Contact (Name, Phone, Email)
6. Data access policy (choose one):
 - a. Open-access – no-embargo
 - b. Open-access – [embargo](#)
7. Cancer type(s) included in study
8. Number of [cases](#) included in study (please indicate if demographic and diagnosis data are available)
9. Information on the Proteomic Data Analysis Protocol
 - a. Type of acquisition – DDA, DIA
 - b. Experiment type – Label Free, iTRAQ, TMT, etc.
 - c. Analytical fractions – Proteome, phosphoproteome, etc.
 - d. Instrument make and model
 - e. Additional proteomic data analysis protocol including experimental design
10. Additional data types included in study and experimental strategies used (list all that apply and indicate target repository for additional data types such as the National Cancer Institute's Genomic Data Commons):
 - a. Imaging

- b. Genomics
- c. Immunology
- d. Clinical
- e. Other (specify)

11. Amount of data (in TB, # of files)

12. Include description of treatment, relapse/recurrence, and/or outcome data available with this dataset (if applicable)

13. The overall scientific benefit of including this study in the PDC

14. Publications associated with this study, if any.

15. Time constraints on processing/loading/releasing the data to the public

16. Data standards used, if any.

Please attach (if available):

- 1. Data Dictionary
- 2. Biospecimen and experiment metadata
- 3. Data Model/Schema diagram indicating how collected data relates to subjects, visits, samples, etc.

**Integrated Canine Data Commons (ICDC) Data Submission
Request Template (Word document sent upon request)**

OMB No.: 0925-7775

Expiration Date: 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 online to complete this form so that NCI can consider your study for submission into the Integrated Canine Data Commons.

Public reporting burden for this collection of information is estimated to average 30 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.

Please complete the following document and send to: icdchelpdesk@mail.nih.gov. Please include a narrative describing your study and its scientific benefit for inclusion in the ICDC.

Please include the following information along with the narrative:

1. Name/Identifier of Study
2. Grant ID and funding source (if applicable)
3. IACUC/IRB approval numbers (if applicable)

4. Scientific Point of Contact (Name, Phone, Email)
5. Data Manager Point of Contact (Name, Phone, Email)
6. Data access policy (choose one): Open-access – no-embargo, Controlled-access – no embargo, Open-access – embargo, Controlled-access - embargo
7. Cancer type(s) included in study
8. Number of subjects included in study
9. Sample Source (e.g., CCOGC, other biospecimen repository, self-collected) - if other than self-collected, those identifiers will be required during submission
10. If self-collected, was a replicate sample also submitted to another biospecimen repository (e.g., CCOGC). If so, those identifiers will be required during submission.
11. Data types included in study (check all that apply): Imaging, genomics, proteomics, immunology, clinical, other (specify)
12. Amount of data (in TB)
13. The overall scientific benefit of including this study in the ICDC prototype
14. Any publications associated with this study, if any
15. Time constraints on processing/loading/releasing the data

16. Data standards used, if any (e.g., SEND)

17. Anticipated budget needed to prepare data set for submission

Please attach (if available):

1. Data Dictionary specific to study

2. Data Model/Schema diagram indicating how collected data relates to subjects, visits, samples, etc.

Non-NCI Study Information Form (Administered via MS Form)



Non-NCI-Funded Study Information Form

Please complete this form to the best of your ability. An NCI subcommittee will review this information to determine if the study will be accepted for storage in an NIH repository. The committee will consider the scientific value of the data, its usage of NIH resources, and the data use limitations (DULs). For questions about this process, please contact NCI's Office of Data Sharing.

OMB No.: 0925-7775

Expiration Date: 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 online to complete this form so that NCI can consider your study for submission into an NIH repository.

Public reporting burden for this collection of information is estimated to average 10 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.

* Required

1. What is the date? *

Format: M/d/yyyy

2. What is your first and last name? *

3. What is your email address? *

Please enter email address only, no extra characters or spaces

4. Please provide the title and description of your study *

Enter a brief description of the study along with the title and highlight the uniqueness of the study.

5. Please insert draft abstract from your manuscript. *

6. Other Relevant Information *

Enter any other relevant information that would assist reviewers in finalizing decisions. For example, how is this information different from other studies or will benefit the research community?

Additional Information

7. What are the data use limitations (DULs) for this study? *

Enter DULs. Please reference the Standard Data Use Limitations table (<https://sharing.nih.gov/genomic-data-sharing-policy/institutional-certifications/completing-an-institutional-certification-form#step-5>) (<https://sharing.nih.gov/genomic-data-sharing-policy/institutional-certifications/completing-an-institutional-certification-form#step-5>) for more information.

8. What is the sample size? *

Enter the number of samples.

9. How much storage is needed for the study? *

List specific units (GB, KB, MB, etc.)

10. What data types need to be deposited? *

List all data types for depositing into an NIH repository (i.e. WGS, WXS/WES)

11. What is the file format of the data? *

List the format of the data intended for submission (i.e. FASTQ, CRAM, BAM, etc.)

12. What is the average coverage range per base? *

If the coverage is 30X, the range would be 0-60X.

13. Please include information about the read length. *

For example, is it generated with paired-ends?

14. To which journal has a manuscript been submitted? *

15. What date do you anticipate submitting your manuscript to this journal? *

Format: M/d/yyyy