Mini Supporting Statement A

NCI Cancer Research Data Commons (CRDC)
Data Submission Forms

Sub-study under,

"Generic Clearance for National Cancer Institute (NCI)

Resources, Software and Data Sharing Forms"

OMB# 0925-0775,

Expiration Date: 06/30/2025

Date: August 2, 2022

Erika Kim

Biomedical Informatics Program Director
Center for Biomedical Informatics and Information Technology
National Cancer Institute
9609 Medical Center Drive, MSC 9746 Rm 6W236, Rockville, MD 20892
(240) 376-5026
Erika.kim@nih.gov

List of Attachments

Attachment 1: Data Submission Form - Proteomics Data Commons Data (PDC)

Attachment 2: Data Submission Upload Template - Proteomic Data (PDC)

Attachment 3: Data Submission Form - Cancer Data Service (CDS)

Attachment 4: Data Submission Request - ICDC

Attachment 5: Instructions – Cancer Data Service (CDS)

Attachment 6: Instructions – Data Submit PDC

Attachment 7: Online Portal - Proteomic Data Commons (PDC)

Attachment 8: Online Portal - Integrated Canine Data Commons

Attachment 9: Email - Cancer Data Service

Attachment 10: Privacy Impact Assessment - Proteomic Data Commons (PDC)

Attachment 11: Privacy Impact Assessment - Cancer Data Service (CDS)

Attachment 12: Privacy Act Memo

<div align="center">**Mini Supporting Statement A**</div>

A.1 Circumstance Making the Collection of Information Necessary

The Health Omnibus Programs Extension of 1988 (Public Law 100-607, Nov. 4, 1988, 102 Stat. 3048) and its amendments require the National Cancer Institute (NCI) to establish an information and education program to collect, identify, analyze, and disseminate on a timely basis, through publications and other appropriate means, information on cancer research, diagnosis, prevention, and treatment (Sections 410 and 412 of the Public Health Service Act (42 USC § 285 and 285a-1)). To disseminate information and data, the National Institutes of Health (NIH) created the NIH Data Sharing Policy and Implementation Guidance (Final NIH Policy for Data Management and Sharing), which will require investigators to submit a data sharing plan beginning January 25, 2023 (https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm). All data, particularly those generated through public funds, should be considered for data sharing. Furthermore, when the data are shared, they should be made as widely and freely available as possible while safeguarding participants' privacy and protecting confidential and proprietary data.

NCI's Cancer Research Data Commons (CRDC) is a data science infrastructure to empower researchers with the data, tools, and computational capacity to perform cross-domain analysis of large cancer data sets. As of 2022, the CRDC encompasses and connects multiple cloud-based data repositories to serve as a central location to support public data sharing for NCI-funded programs. Though the data submission and access process may vary based on the type of data, its sharing restrictions, and where the repository is in its development, these components require the ability to ask questions that help intake, correctly index, and make available a variety of cancer data sets to the community.

A.2 Purpose and Use of the Information Collection

The following forms represent the current repositories accepting new data submissions through their online portals. The information collected serves an operational purpose of indexing the data in the system and helping to prioritize which data sets should be made available through the repositories. Though the CRDC's goal is to make large data sets available, the time, resources, and cost to clean, quality-check, standardize, upload, and maintain these data can be high. To help prioritize data that have the maximum scientific benefit to the community, the CRDC repositories leverage scientific governance committees to review and consult on the data submission request forms.

The information in these forms is used to help the review teams assess the operational cost of uploading, maintaining, and making that data set available, taking into consideration:
- The scientific benefit of making this data set available to the larger community, and
- The guidelines govern access and secondary use of that data set.

The specific uses and processes for each of the current data commons intaking data sets are outlined below:

- **Attachments 1 and 2: Proteomics Data Commons Data (PDC) Submission Form and Data Upload Template**

  Researchers interested in making proteomic data available to the larger community can apply to submit their data set to the PDC (Attachment 1). PDC's data submission form assesses whether this study should be stored and maintained through this data commons. Once submitted, the PDC governance committee reviews these forms to determine if the data set would be available through the PDC. If accepted, the information in the form will also be used to create a private and secured workspace for the investigator to upload the raw data and metadata files. Researchers will be asked to use (Attachment 2), which gives templated fields for researchers to add the raw data and metadata so that it can be uploaded directly and accurately to the repository. Unlike attachment 1, attachment 2 is a tool to assist researchers in correctly updating the data according to the standards of the repository.

- **Attachment 3: Cancer Data Service (CDS) Data Submission Form**

  Researchers with data (that cannot be submitted to other NIH repositories) can apply to submit their data set to the CDS to satisfy data sharing requirements mandated by journals or funding agencies. The form serves as a first touchpoint to gather basic information so that a member of the CDS infrastructure team can schedule a consultation call to discuss further data storage needs. The information may also be used to identify other repositories better suited to managing the data set.

- **Attachment 4: Integrated Canine Data Commons (ICDC) Data Submission Form**

  Researchers wanting to share naturally occurring canine cancer data (clinical, genomics, etc.) can apply to upload their data to the ICDC. Once submitted, the ICDC governance committee reviews these forms to determine if the data set should be made available through the ICDC. If accepted, the ICDC infrastructure team will reach out to coordinate the data transfer, standardization, and upload.

A.3 Use of Information Technology to Reduce Burden

If the data request form is approved, there are online instructions to upload the data (**Attachments 5 and 6**). All data submission forms are available on the corresponding repositories' online portal (see **Attachments 7, 8, and 9** for examples of the webpages/email outreach referencing these forms). The form is a downloadable word document that can be completed and emailed to the repository's help desk team. This downloadable form allows investigators to circulate it internally for any review or approvals before sharing it with the repository. Through this process, investigators submit the required information directly to CRDC, thereby minimizing the burden for investigators, institutions, and NCI staff. The online system uses time-saving features, such as the use of pull-down and scrolling menus to fill data fields, "find as you type" (or "type-ahead") functionality, and text fields that allow investigators and requesters to cut and paste information from other sources.

The NCI Privacy Act Coordinator was consulted and determined that a Privacy Impact Assessment (PIA) is required for PDC and CDS **(Attachments 10 and 11)**. ICDC only hosts canine data.

A.4 Efforts to Identify Duplication

The CRDC has created a data submission working group to streamline and unify the data submission processes for the research community. The working group has reviewed these forms to find opportunities to avoid duplication and potentially create a standard data submission form template.

A.5 Impact on Small Businesses or Other Small Entities

No small businesses or other small entities will be impacted.

A.6 Consequences of Collecting the Information Less Frequently

Data submission forms are collected as research data is generated. Constraining or limiting these data submission forms can slow down and limit the amount of data accessible via these repositories, significantly impacting the CRDC's goals to provide access to critical data sets to the community. Additionally, limiting the submission can negatively impact the data submitter, who may be required by a journal, funding agency, or their institution to deposit their data into a public repository by the deadline indicated in their data management and sharing plan. If researchers miss this deadline, they may lose the ability to publish the data or no longer comply with the terms of their grant award.

A.7 Special Circumstances Relating to the Guidelines of 5 CFR 1320.5

There are no special circumstances for this information collection request relating to guidelines of 5 CFR 1320.5.

A.8 Comments in Response to the Federal Register Notice and Efforts to Consult Outside Agency

N/A

A.9 Explanation of Any Payment of Gift to Respondents

No payment or gift will be made to respondents.

A.10 Assurance of Confidentiality Provided to Respondents

Researchers who seek access to individual-level data are typically required to enter into a data-sharing agreement. Data-sharing agreements, which come by many terms, including "license agreements" and "data distribution agreements," generally include requirements to protect participants' privacy and data confidentiality. They may prohibit the recipient from transferring the data to other users or require that the data be used for research purposes only, among other provisions, and they may stipulate penalties for violations.

For access to and submission of data, researchers are both NIH-funded and non-NIH-funded investigators. Making these researcher's names available is an important ethical underpinning of the NIH GDS Policy as it allows NIH to be transparent in informing research participants, the scientific community, and the public on how data are being shared, with whom, and for what research purpose in addition to fostering future research collaborations.

The names and institutional affiliations of the researchers (both data submitters and data requesters) may be posted publicly on a website. Thus there is no assurance of confidentiality afforded to the researchers. However, it is essential to emphasize that no personal information is requested from researchers submitting or accessing data beyond their name and institutional affiliation. Data submitters are largely NIH-funded investigators whose

names and institutional affiliations are already a matter of public record. All information will be kept secure to the extent allowable by law.

The Privacy Act is applicable as determined by the NIH Privacy Officer in the Privacy Act Memo (**Attachment 12**). This data collection is covered by the following Privacy Act System of Records:

- 09-25-0200, "Clinical, Basic and Population-based Research Studies of the National Institutes of Health (NIH), HHS/NIH/OD."
- 09-25-0036, "Extramural Awards and Chartered Advisory Committees (IMPAC 2)."
- 09-90-1401, "Records About Restricted Dataset Requesters."

There is no assurance of confidentiality provided to the applicants. However, their information will be kept private to the extent provided by law.

A.11 Justification for Sensitive Questions

No sensitive questions are asked in these forms.

A.12.1 Estimates of Hour Burden Including Annualized Hourly Costs

The PDC and ICDC forms are anticipated to take 30 minutes, and the CDS form will take 60 minutes to complete. The total annual burden hours are expected to be 180, with an estimated cost to the respondents of $9,888.

A.12-1 Estimated Annualized Burden Hours

Form Name	Type of Respondent	Number of Respondents	Number of Responses per Respondent	Average Time Per Response (in hours)	Total Annual Burden Hours
PDC – Data Submission – Attachment 1	Individuals	60	1	30/60	30
CDS – Data Submission – Attachment 3	Individuals	60	1	60/60	60
ICDC – Data Submission – Attachment 4	Individuals	60	1	30/60	30
Totals			**180**		**120**

A.12-2 ANNUALIZED COST TO RESPONDENTS

A.12-2 Annualized Cost to the Respondents

Type of Respondents	Total Annual Burden Hours	Hourly Wage Rate*	Respondent Cost
Individuals	120	49.44	$ 5,932.80
Total			**$ 5,932.80**

*Source of the Hourly Wage Rate is provided by the May 2021 National Occupational Employment and Wage Estimates, Bureau of Labor Statistics, Occupation title "Medical Scientists" 19-1040 https://www.bls.gov/oes/current/oes_nat.htm

A.13 Estimate of Other Total Annual Cost Burden to Respondents or Record Keepers

There are no annual costs to the respondents or record keepers.

A.14 Annualized Cost to the Federal Government

Provide estimates of the annualized cost to the Federal government. This includes FTE and contractor costs.

A.14-1 Annualized Cost to the Federal Government

Staff	Grade/Step	Salary**	% of Effort	Fringe (if applicable)	Total Cost to Gov't
Federal Oversight					
PDC Data Submission	GS 14/6	$147,272	2%		$2,945.44
CDS Data Submission	GS 14/1	$126,233	3%		$3,786.99
ICDC Data Submission	GS 14/6	$147,272	2%		$2,945.44
Contractor Cost					$0.00
Travel					$0.00
Other Cost					$0.00
Total					**$9,677.87**

**The salary in the table above is cited from https://www.opm.gov/policy-data-oversight/pay-leave/salaries-wages/salary-tables/pdf/2022/DCB.pdf

A.15 Explanation for Program Changes or Adjustments

N/A.

A.16 Plans for Tabulation and Publication and Project Time Schedule

Results collected from these Data Submission Request Forms will not be published on the portals. This information may be shared with a repository that will follow its own operational timelines for uploading study data and publishing the study information on its portal.

A.17 Reason(s) Display of OMB Expiration Date is Inappropriate

There is no request for exemption from displaying the expiration date of OMB approval.

A.18 Exceptions to Certification for Paperwork Reduction Act Submissions

There are no exceptions to the Certification for Paperwork Reduction Act Submissions.