Leidos Biomedical Research, Inc.

Frederick National Laboratory for Cancer Research

October 16, 2018

Dr. Toby Hecht
Contracting Officer's Representative
MSC 9726
9609 Medical Center Drive
Bethesda, MD 20892

Reference: Contract HHSN261201500003I

Subject: Task Order HHSN26100076

NCI Action: Review and Acceptance of Task Order Deliverable

Dear Dr. Hecht:

In accordance with the above referenced contract and task order, the deliverable summarized below is provided for your review and acceptance.

Table 1: Deliverable Summary

Task Order Number:	HHSN26100076	**Project Title:**	Development of an Integrated Canine Data Commons (ICDC)
Deliverable Item Number:	1	**Deliverable Description:**	Project Implementation Plan
Reporting Period:	Kick-off Meeting	**Quantity:**	1
Primary Program Manager (PPM):	John Otridge	**Contracting Officer's Representative (COR):**	Toby Hecht
PPM Email:	John.Otridge@nih.gov	**COR Email:**	Toby.Hecht@nih.gov
PPM Phone:	240.276.5653	**COR Phone:**	301.435.9162

Respectfully,

Connie Suders
Contract Administrator

Project Implementation Plan
Development of an Integrated Canine Data Commons (ICDC)

Acceptance of the Project Implementation Plan signifies the start of the project and agreement by all stakeholders that the project scope can be accomplished within the time and budget described. The Risks, assumptions and constraints are accepted.

Primary Program Manager: John Otridge, Ph.D.

Scope:

The objective of this project is to leverage CBIIT CRDC experience and knowledge, and its development of Data Commons Framework Services (DCFS), to create a new, dynamic data commons for canine cancer data, including not only clinical outcomes and genomics findings from canine clinical trials being conducted by the Comparative Oncology Program (COP) in collaboration with DCTD, but also the trials' molecular, pharmacological, microenvironment, medical imaging and other study data.

The ICDC is envisaged to provide the research community with access to multiple types of data, for example clinical outcomes and diagnostic information (including genomic data), medical images, clinical assessments and laboratory correlates such as pharmacokinetics and pharmacodynamic biomarker responses including immunofluorescence microscopy images. Users will be able to query the data via a web-based interface as well as an API. The production system will also have a more comprehensive set of web pages providing users with more details about the projects and other canine specific information. It is possible that the ICDC can serve as a major hub of canine and canine:human relationship information that the NCI wishes to share with the research community, and the infrastructure we propose to create will enable such an environment. The project will involve collecting the perspectives of the comparative/canine oncology extramural community (eventually a major user group) regarding data models/data structure, use cases/user queries, and comparative oncology applications via input and advice from a newly-formed ICDC Steering Committee. This Steering Committee will also be an important source of data from the canine clinical trials.

Period of Performance:

- The period of performance for this task order begins on September 24, 2018. The Base Requirement for this task order has a duration of 24 months. The completion date for the Base Requirement is September 23, 2020.

- If the Government exercises its option pursuant to the Option Provision of this task order, the period of performance for Option 1 will begin immediately following the completion of the Base Requirement and have a duration of 12 months. The completion date for Option 1 is September 23, 2021.

- If the Government exercises its option pursuant to the Option Provision of this task order, the period of performance for Option 2 will begin immediately following the completion of Option 1 and have a duration of 24 months. The completion date for Option 2 is September 23, 2023.

Cost:

- Base Requirement = $1,991,470
- Option 1 = $1,486,358
- Option 2 = $1,916,810
- Total (Base Requirement + Options) = $5,394,638

- Subcontractor Ceiling
  - Base Requirement = $535,377
  - Option 1 = $389,938
  - Option 2 = $517,417
  - Prior written consent of the Contracting Officer is required to exceed the ceilings set forth above. Prior written consent of the Contracting Officer is also required for any international subcontracts or consultants, as well as for any legal services subcontracts or consultants.

- Equipment Ceiling = $0 – Prior written consent of the Contracting Officer is required to acquire Capitalized Equipment.

- Travel Ceiling = $0 – Prior written consent of the Contracting Officer is required for travel expenses.

General Approach:

An FNLCR led team will be created to drive development of the ICDC. The team will consist of members who have experience with the specific canine data sets and with the CRDC infrastructure. The team will work closely with the DCTD and CBIIT leadership and will use an Agile approach to develop and implement the prototype and production ICDC versions of the system. This approach will be characterized by frequent engagement points with the client to ensure that all stakeholders are engaged, decisions are made/tracked collaboratively, progress is made in the agreed upon direction, and everyone is aware of the next steps in the project. There will be three milestones for the project. Milestone 1, which will meet the Base Requirement Task, will be complete upon instantiation of a prototype and an updated estimate for the cost and time to develop the production system. The updated production costs will be shared with the NCI for consideration on how to move forward for Milestone 2, which will meet Option Task 1, which will be completed with instantiation of a production version of the ICDC. Milestone 3, which will meet Option Task 3, will be complete after we have updated and maintained the Production system through the end of the period of performance.

A prototype ICDC based on the existing technology from the CBIIT Data Commons Framework project will be established by end of year two (end of the Base Requirement period) with the first version of production by the end of year three (completion of Option Task 1). During years four and five, it is expected there will be ongoing maintenance and incremental improvements to the system (more data of the same type and format from the same projects, modifications to the supported features, etc.) but not large changes (large as defined by levels of effort).

The prototype will work with existing data from existing canine clinical trials projects. The production system will manage more data from the original projects and add data derived from DCTD-COP collaborative projects, with the following studies as high priority sources of Use Cases:

- P30 Cancer Center Grant Supplement-funded genomic analysis of 25 cases each of 6 types of canine cancer, including newly developed immuno-oncology analysis pipelines.
- UO1/U24 grant award-funded canine immunotherapy trials and companion studies of laboratory correlatives.
- ID/IQ TO53-associated canine clinical trials of FNLCR-generated therapeutic antibodies targeting canine CTLA4, an immune checkpoint that is a therapeutic target for treating human cancer.
- Two planned canine clinical trials of NExT agents (deoxycytidine analogs, protease inhibitor).

The ICDC system, both in prototype and production forms, will enable users to query the relevant data via a web-based interface as well as an API. The production system will also have a more comprehensive set of web pages providing users with more details about the projects and other canine specific information. It is possible that the Integrated Canine Data Commons can serve as a major hub of canine and canine:human relationship information that the NCI wishes to share with the research community and the infrastructure we propose will enable such an environment.

Governance Approach:

- The PPM from FNLCR will chair a Steering Committee (SC) formed around extramural Principal Investigators/leaders in the fields of canine oncology, tumor genomics analysis, tumor immunity and immune-oncology and laboratory correlates as key members, including the study PI's of the clinical trials and grant funded projects listed above. The Steering Committee is expected to:
    - provide FNLCR with key input and opinions regarding ICDC governance regarding data models, data queries, and desired applications in the ICDC to integrate with NCI direction.
    - provide guidance on what data/projects will be included in the ICDC
    - define data policy items such as data lifecycle (will data live forever) and how will updated data be handled

    - extramural members of the SC will be expected to provide curated, structured datasets for populating the prototype ICDC that illustrate their opinions about data models and queries.
- FNLCR will create an ICDC Implementation Team (i.e., End Users Group) that will include members from the PI laboratories, who will need to do the actual work of curating, structuring and transferring the data sets according to the data models.
- FNLCR will act as an advocate/resource for the ICDC within the CRDC and other groups as relevant/requested (such as Imaging Data Commons or NIH Commons). It will also explore ICDC outreach activities, for example, an ICDC booth at conferences that will also be a source of Use Cases.

Technical Approach:

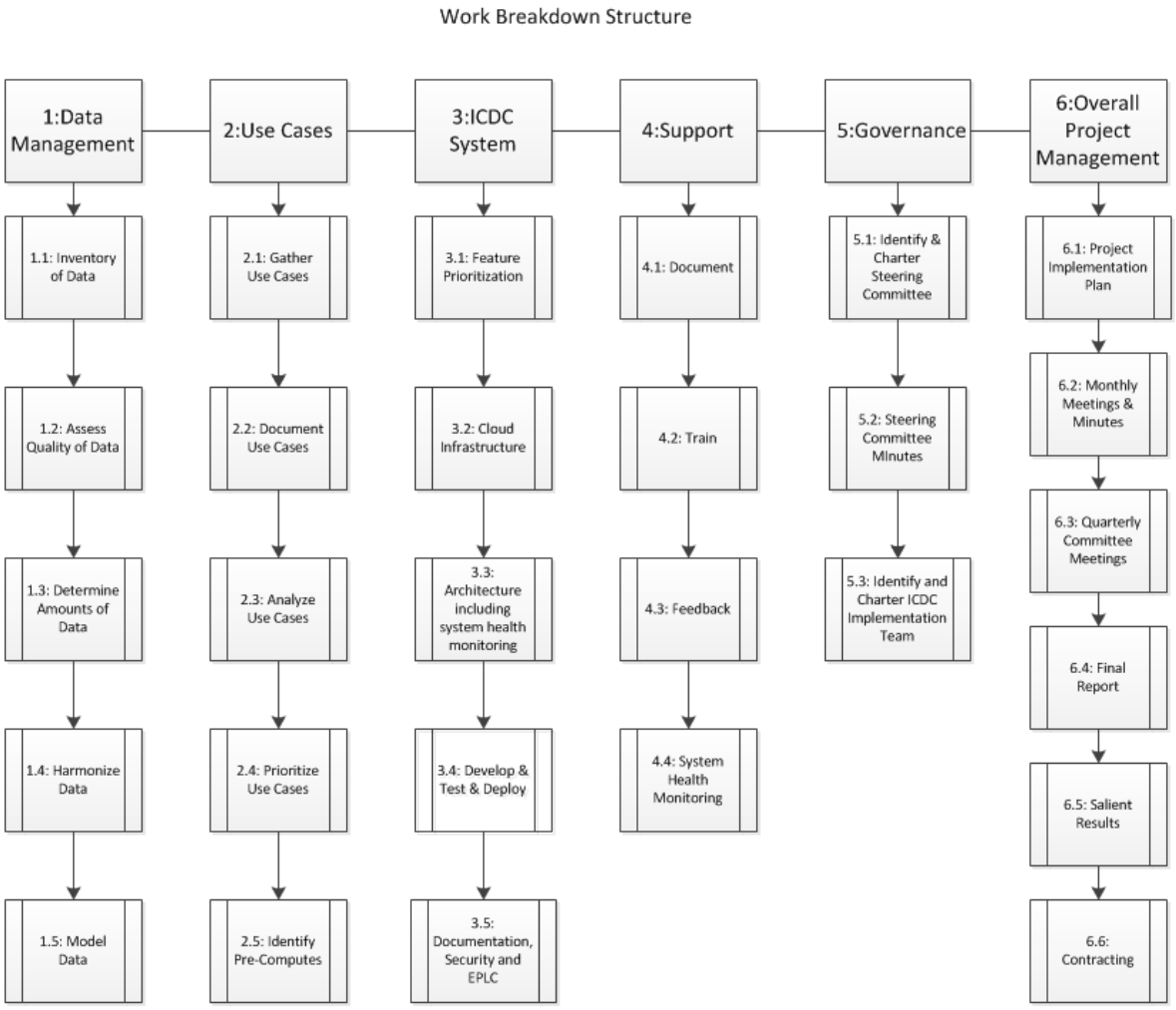
- The ICDC infrastructure will use the NCI Data Commons Framework/Gen3 open-source stack of services, developed by University of Chicago, and will be instantiated on Amazon Web Services. This includes access controls, authorization and authentication, file indexing, data submission support and metadata indexing and search. We will subcontract with University of Chicago where they will provide Subject Matter Experts (SME) to support us if we have any questions during planning and implementation.
- Experimental data will be stored on either or both of Google or Amazon cloud systems depending upon the needs of the user community. For example, if an analysis requires all the input data to be available then all the data must be sources from the same cloud provider, i.e. Amazon or Google. If it is necessary to store the same data on multiple cloud providers, we will utilize services provided by the NCI Cloud Resources where possible.
- The image files will be stored at FNLCR in Frederick while the other data and metadata of ICDC will be stored on the cloud.
- The prototype is envisaged to be “bare-bones” enabling querying of data, downloading of image data only, enabling users to use “cloud” tools to analyze selected data.
- The prototype will be limited in the number of users and we will gather feedback from those users.
- Updated cost and schedule estimates for the production system will be generated during milestone one. Feedback from use of the prototype will be important factors in developing the updated estimates. As will revisiting the underlying architecture/technoalogy stack to ensure that the production system design uses the most upto date information and relevant technology.
- Milestone one ends after the prototype is instantiated and an updated cost and schedule for the production system is provided to the NCI.
- The production system will have open availability to the canine oncology research community.
- Production ICDC will include prototype functions data but extend the features (for example a feature for users to submit data) and add additional informational content. A Drupal based content management system will be used to enable NCI to edit much of

informational content themselves as needed. We intend to work with Ontario Institute for Cancer Research (OICR) for development of key User Experience (UX) elements. OICR developed many of the Data Portal section of the Genomic Data Commons (GDC). We would like to provide something similarly engaging for the ICDC.

- The system infrastructure will be maintained through the project, updating necessary components for functionality and security needs.
- The three NCI Cloud resources (ISB, SBG and Broad) will be made aware of the ICDC and any information they need to interact with the ICDC data will be provided.
- The system will be monitored for security, performance and log management.
- A User Support desk will be created to support IT, how-do-I, and scientific questions from the user community. Any issues will be tracked and managed using NCI's Jira system.
- Documentation will be generated to support use of the system (user guides, FAQs, release notes, etc.). The format of the documentation will be written but could also include other media.
- The code will be open source and GitHub will be the repository.
- Development and maintenance of the system will be informed by best practices of NCI's CRDC. For example, being interoperable with the to-be-developed Cancer Data Aggregator.
- Enterprise Performance Lifecy Cycle (EPLC) will be performed using CBIIT's EPLC process. In collaboration with CBIIT we will "right-size" the processes the project will follow and this will be captured in the Project Process Agrement (PPA) document we will generate as part of EPLC's Planning Phase.
- We will use Jira to manage development of the project, enabling documentation, storage and traceability of use cases/requirements, test cases, issue tracking, test reports, and release version details. Jira supports export of information which will enable meeting EPLC documentation requirements for the following phases; Requirements, Development, Test and Implementation. We will create and manage a Design Document (including architecture and data) for the Design phase and an Implementation Plan for the Implementation phase of EPLC. A security risk assessment will be performed and documented and a training plan created part of the Implementation Phase documents.
- We will use an Agile methodology to develop the prototype and production versions of the ICDC. This is characterized by defining/refining/clarifying a subset of requirements, implementing and testing all in a few weeks. Then we repeat the cycle with a new subset of requirements. This repeated cyles are called sprints. Multiple tasks occur in a sprint; agreement on which requirements will be targeted for development, a more detailed definition of those requirements, implementation and testing of those requirements, planning for the next sprint and reviewing what was completed in a sprint. This methodology involves the NCI to clarify requirements both in terms of the clarification, prioritization and definition of what is complete.

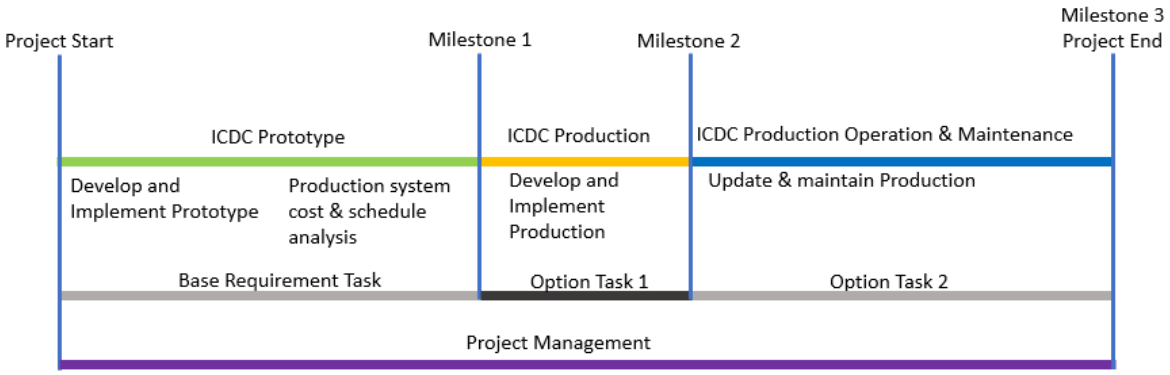
Use Case and Data Approach:

- Gathering, documenting, verifying and prioritizing use cases and workflows from the research community will be ongoing during the life of the project and will inform incremental feature development.
- Creating a data inventory that will be used for estimating storage costs.
- Modelling the existing/known data to understand how best to organize it to meet the user defined use cases, also including the needs of the larger NCI Cancer Research Data Commons community.
- Curate and harmonize data from selected projects for use in the ICDC.
- Human comparison data. What is the best way of representing this data, where is it sourced, how is it updated and what "links" should be provided to other non-ICDC resources. These and similar questions will be asked of the Steering Committee during the project with the answers documented and shared for use in driving future ICDC features.
- Working with the Cancer Data Aggregator team to ensure optimal outcomes with the ICDC data.
- Explore pre-computes. A community can benefit from a single "gold-standard" way of analyzing data. For example, mapping genomic reads onto a reference genome as is done by the Genomic Data Commons. The mapping is done once and shared with the community, it removes the need for the same mapping to be carried out by each user of the data. It is possible that we would avail of systems such as RedShift or BigQuery (AWS and Google respectively) that allow cost effective sharing of large sets of pre-computed data. However, at this point we do not know what pre-computes would be valuable so nor do we know how they could be implemented both from a User Experience and a technical perspective. We would explore pre-computes during the project and on a case-by-case basis work with the NCI to determine if we would implement any pre-compute.
- We will work with the Imaging Data Commons (IDC) to understand if and how canine images would be appropriate for storage there. Canine medical images will be processed into The Cancer Imaging Archive (TCIA) using the best available image processing pipeline in a separate project (see Technical Response to HNC17V-8 - The Cancer Imaging Archive (TCIA) Support).



Milestones and Schedule:

Milestone	Description	Completion Date
1	Base POP, Completed Prototype ICDC, reporting and review of cost/schedule	09/28/20
2	Option one, Completed Production ICDC System	09/28/21
3	Option two, Verified production version performance	07/28/23



Assumptions:

- NCI SME(s) will be provided, if needed, to the ICDC implementation team to interpret the data.
- NCI will provide access to researchers so use cases can be collected and documented.
- Cloud data storage will be handled under CBIITs allotment of space from Amazon and/or Google.
- There will be a strategic and tactical governance board that will advise the FNLCR with policy for maximizing success with the ICDC implementation team.
- System will be FISMA low since there is no human subject data. Any human data is level of gene to gene mapping that is independent of any human subject/patient.
- We will follow EPLC as defined by NCI CBIIT. It is assumed that the Concept gate is already complete and that a CBIIT owner, supported by the LBR Project manager, will lead taking the project through CBIIT's EPLC stage gates.
- EPLC deliverables consist of Project Process Agrement (PPA), Design Document (data and architecture), Implementation Plan, Test Plan, Requirements Definition and Traceability, Security Plan and Training Plan. If it is determined with CBIIT that other artifacts are required we will work with CBIIT to understand what, if any impact, that will have.
- Azure has not been considered but in theory could be encompassed but the effort to do so is unknown at this time.
- Presenting the user community with precomputed data (for example, mapping canine genomic data onto latest version of canine or human genome) would be very efficient but at present we have no use cases for precomputes.
- The project is not developing any analysis tools but if there are analysis tools that are cloud enable, for example via Docker, then the ICDC system will enable use of those tools.
- Users will need to pay for computes themselves via their own cloud accounts.
- Users will be able to leverage Jupyter (or other) notebooks that are cloud compatible.
- The Steering Committee to be formed under this Project will serve an advisory role by providing guidance and insight, but will not direct the project.
- GOVERNMENT PROVIDED MATERIALS will be available during the project:

- NCI/COP will provide canine cancer data sets from its existing projects in companion dogs with spontaneous cancers, and other datasets it may have access to and are permitted to use for this project in developing and testing the ICDC.
- Via its FNLCR contractor support, NCI/CBIIT will provide access to the CRDC/DCFS system and strategy as it is developed and finalized, so that ICDC development can utilize and leverage the same CBIIT approach.

Risks:

- The known Use Cases may only be a small fraction of the Use Cases the community requires. As such, our level of efforts estimates may not be enough to cover the effort required to meet the new use cases.
  - Impact: Minor
  - Probability: High
  - Mitigation Strategy: Frequent communication with the NCI program leadership to prioritize Use Cases to use in the Prototyping and Production stages.

- Amount of data to be stored is larger than the free-storage can handle, so could exceed our estimated costs.
  - Impact: Moderate
  - Probability: Low
  - Mitigation Strategy: Work with the NCI programs to identify this issue if it arises and evaluate options before implementing a solution.

- The level of detail in the SOW is low and the Data Commons concept is new. So there are a lot of unknowns that will only be encountered during implementation. So this adds a lot of uncertainty to the timelines and the effort estimates.
  - Impact: Moderate
  - Probability: Medium
  - Mitigation Strategy: Focus on uncovering those unknowns during the Prototyping stage so they do not arise late in the project at Production. At the completion of the Prototyping phase we will conduct an assessement of costs and schedule for the development of the Production system.

- Unable to staff the project in a timely fashion with either/or FNL or subcontractor staff. This could delay progress towards meeting milestones.
  - Impact: Moderate
  - Probability: Low
  - Mitigation Strategy: The initial phases will focus on activities such as data inventory, harmonization and use case definition that utilize existing or soon to be hired staff (anticipated to be onboard before project starts). This will allow time to find any additional staff or subcontractors to staff up.

- The Gen3 architecture is still new and not extensively documented or field tested. It is possible there are missing elements needed to fully support the ICDC or there are performance gaps in functionality or stability.
    - Impact: Moderate
    - Probability: Low
    - Mitigation Strategy: Gen3 was assembled by UChicago so by using UChicago as SME(s) during development we can be guided by them with respect to what features are incomplete and if there are roadmaps to complete those features. This will enable us to determine prioritization of ICDC system development to avoid any known issues and to plan around Gen3 development releases.

Constraints: None.

Technical Reports: All reports required shall be submitted in electronic format and shall be compliant with Section 508 of the Rehabilitation Act of 1973.

Other Reports/Deliverables:

Description	**Delivery Schedule**
Project Implementation Plan	Due at the Kick-Off Meeting (within 30 calendar days of the Period of Performance start date)
Quarterly Cost, Schedule, Performance Report	Due Quarterly, 15 calendar days after system updates
Monthly Meeting Minutes	Delivered within 5 business days of a monthly meeting, detailing project science updates with DCTD leadership, including Action Items and Decisions
Steering Committee Minutes	Captured and delivered within one week following the occurrence of a Steering Committee Meeting
Final Report	Due on or before the completion date of the Task Order
Summary of Salient Results	Due on or before the completion date of the Task Order

Change Management:

The Project Implementation Plan serves to define the known state of the requirements, milestones, risks, schedule and budgets. Change Management defines a means to manage communication about the elements in this plan and the impact of these changes on the project schedule, cost or success. Any requested changes to the proposed plan as captured in this document will be formally managed. Changes to project scope may impact schedule or cost of the project and we will develop an impact analysis memo to convey the impact of the requested change. In the event that the change request impacts the cost or schedule, we will request formal review by the Contracting Officer and LBR Project Management Operations office.