Statement of Work (SOW)

The Clinical Proteomic Technologies for Cancer (CPTC) Data Center (DC) and Portal

1 General Background

This statement of work outlines tasks for the CPTC DC (Data Center) and CPTC-Portal to coordinate the data and metadata that will be generated in the second phase of CPTC. Despite significant progress in understanding cancer at the molecular level, the sheer complexity of the over 200 diseases that comprise "cancer'" is a daunting barrier to developing the interventions needed to diagnose, treat, and prevent cancer. Vital to the progress in these areas is the discovery and understanding of cancer-specific aberrations at various molecular and cellular levels. Although proteins reflecting the genomic changes in cancer have the potential to become clinically meaningful biomarkers, their discovery and validation has proven to be challenging. As a result, few biomarker candidates have translated into clinical utility.

Two key barriers in the early stages of biomarker development are: 1) a limited understanding of the changes in cancer genomes that translate into functional differences at the proteomic level; and 2) insufficient technologies that could be widely applied to reproducibly detect and quantify these aberrant proteomic changes across samples from cancer and control populations. Significant barriers to the development of cancer protein biomarkers include insufficient inter-laboratory reproducibility, lack of standards for proper study design, various analytical barriers, biospecimen collection/handling, data acquisition/analysis, and a notable absence of standards and high quality reagents. The progress in the field has also been slowed by the lack of a coherent "pipeline" to connect biomarker discovery with well-established methods for qualification. Various cancer-related proteomic changes have been identified in numerous published studies. However, these studies mostly came from diverse research groups working independently. Consequently, the findings are typically based on an insufficiently low number of samples to have adequate statistical power needed for rigorous evaluation of the observed protein aberrations as specific, clinically relevant cancer biomarkers.

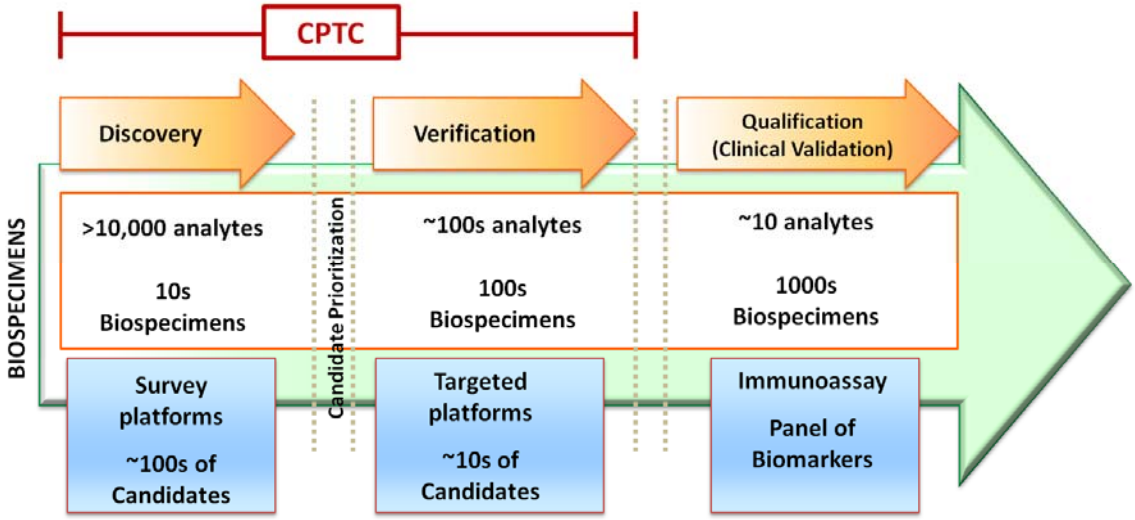
Recognizing this need for an evidence-based, proteomics pipeline, the NCI launched in 2006 the Clinical Proteomic Technologies for Cancer initiative (CPTC). At that stage (Phase I), the CPTC initiative was focused on removing technical barriers in order to enable the accurate, efficient, and reproducible identification and quantification of a meaningful number of proteins to drive clinically-relevant biomarker qualification studies. Phase I of the CPTC initiative has demonstrated the effectiveness of a multi-disciplinary, multi-institutional approach in addressing long-standing problems of analytical variability in proteomics and ways to overcome the inherent variability of specific analytical platforms in order to uncover and quantify real biological differences.

Although discovery efforts oriented on cancer protein biomarkers identify many hundreds of candidate biomarkers, CPTC investigators recognized that only a few would eventually prove clinically useful. Therefore, developmental strategies must allow for an efficient testing of many biomarker candidates to identify and verify those few that would be suitable for further clinical implementation. Addressing this need, researchers involved in the first phase of the CPTC initiative designed a two-step strategy (further referred to as the developmental "pipeline") for the efficient, timely, and cost-effective development of protein (and peptide) biomarkers for qualification studies. The two steps, referred to as "Biomarker Discovery" and "Biomarker Verification", are outlined below and in Figure 1.

Biomarker Discovery. In the first step of the CPTC-established process, cancer-specific biomarker candidates are discovered (identified) using metrics-driven protein survey technologies that interrogate appropriate biospecimens. The discovery platforms (based on mass spectrometry and immunochemistry) have proved to be robust and have revealed a large number of protein biomarker candidates. The biomarker candidates identified in the Discovery step must next be evaluated in biospecimen collections larger than those used initially.

Biomarker Verification. Some candidates can be analyzed using commercially available reagents (notably, antibodies for immunoassays). However, moving candidates from discovery to qualification typically requires overcoming various bottlenecks reflecting the lack of commercially available reagents (antibodies) in adequate numbers, their high costs, and/or lengthy production times. These limitations are addressed in a comprehensive manner by the Verification step of the CPTC pipeline. Verification involves the development of targeted, quantitative assays, which are commonly multiplexed and suitable for examination of a larger number of biospecimens to ensure appropriate statistical power. The Verification step and the established dedicated assays are meant to be cost effective and timely in terms of funneling those few biomarker candidates for further clinical qualification studies.

Figure 1: CPTC Pipeline

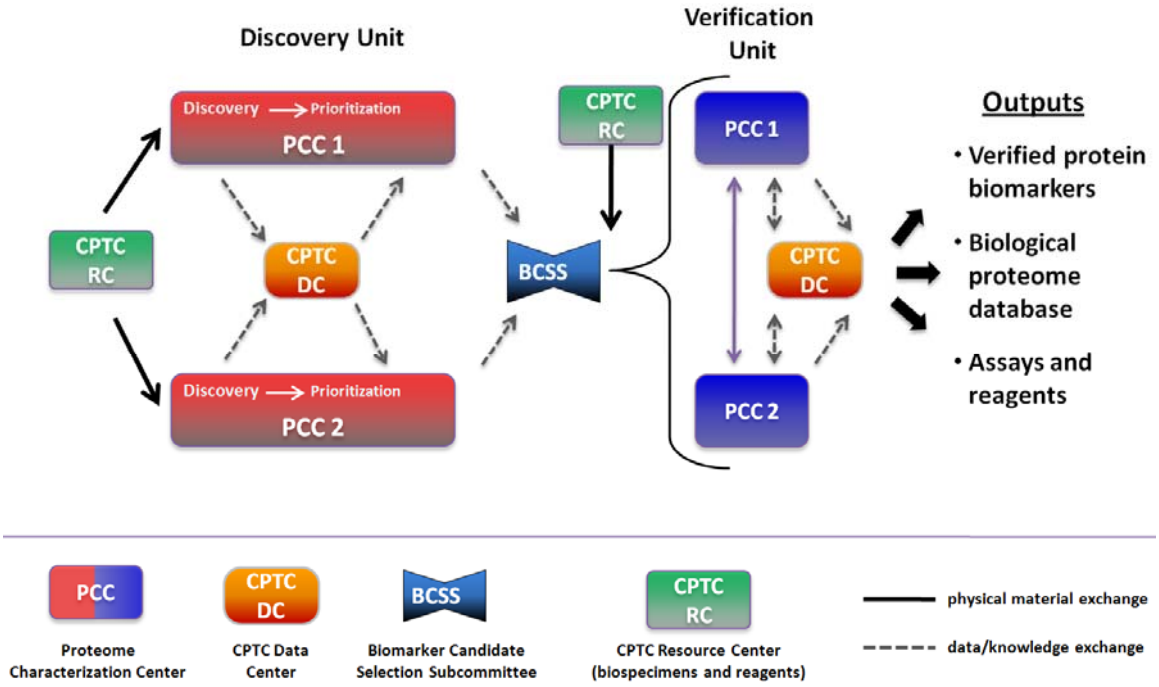


Recently, significant progress has been made in characterizing and sequencing the genomic alterations in statistically robust numbers of samples from several types of cancer. For example, The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov) and other similar efforts are identifying genomic alterations associated with specific cancers (e.g. copy number aberrations, rearrangements, point mutations, epigenomic changes, etc.). The availability of these multi-dimensional data to the scientific community sets the stage for the development of new molecularly targeted cancer interventions. Understanding the comprehensive functional changes in the proteome that arise from the genomic alterations or other factors is the next logical step in the development of high-value protein biomarkers that can drive the rational development of new diagnostics and therapies to achieve personalized cancer medicine.

In the first phase of CPTC, research centers shared data by depositing data files to a data repository entitled Tranche. Since most of these data files were generated in technology assessment and benchmarking experiments, experimental annotation focused on the technical aspects of sample analysis and processing. Data production in the second phase of CPTC is expected to exceed that of the first phase. Additionally, experimental annotation will include more refined descriptions of the samples involved, as well as de-identified clinical data accompanying each sample. Finally, since many of the samples will also have undergone genomic analysis, complete data management of CPTC data files will include connectivity between the proteomic data (produced under CPTC) and genomic data (produced under TCGA).

The key programmatic components of CPTC include: a resource center (RC); proteome characterization centers (PCCs); a CPTC Steering Committee (SC); a CPTC Biomarker Candidate Selection Subcommittee (BCSS); a data center (DC); and a data portal. Each PCC consists of a discovery unit, verification unit, and administrative core. A schematic representation of the CPTC project is shown in Figure 2. This SOW describes the requirements for the CPTC data center and data portal components of the CPTC program.

Figure 2: CPTC Workflow



2 Scope of Work

Independently, and not as an agent of the Government, the Contractor shall perform the services described below. The Contractor shall provide qualified personnel, material, equipment, and facilities not otherwise provided by the Government during performance of this contract.

The scope of this SOW is to develop and support the CPTC data center and data portal. Over the next five years, 4-6 tumor types will be characterized by the PCCs using a variety of platforms and experimental approaches to molecularly characterize cancer cells.

The Contractor shall furnish the necessary professional service experts to create and support the activities of the CPTC DC.

- The Contractor shall provide professional services in support of the CPTC DC. These services shall be performed with input from and in conjunction with NCI program management staff, the Contracting Officer's Technical Representative (COTR), and with representatives from other PCCs, as well as a designated representative or Point of Contact from CBIIT.
- The Contractor shall work with the COTR, the point of contact from CBIIT (appointed by the NCI COTR), and representatives from other CPTC program centers to establish and execute program-wide procedures for data transmission, quality assurance checks, compliance with requirements for adoption of data

standards, and other requirements related to the management, processing, aggregation, and distribution of CPTC data.

- The Contractor staff shall identify issues, conduct discussions, perform analyses, make recommendations, forge consensus, produce documentation, implement processes, and use their skills to achieve the Objectives of this SOW.

The Contractor shall provide experts with the following levels of expertise: PhD or equivalent-level knowledge and at least one person with five years past experience with life science data analysis and management, including data from patient demography and clinical annotations, pathology, proteomics, and genomics; at least one person with five years experience with high throughput methods for data processing, quality analysis and compliance with data standards; at least one person with five years experience in the management and analysis of tissue specimen, genomic, and proteomic data.

The Contractor shall provide software engineering and biomedical informatics support as described by Section 5 (Tasks to be Performed). The scope of this work includes the following types of activities:

- Professional services in support of the CPTC DC and data portal
- Requirements gathering and analyses
- High- and low-level systems architecture and design
- Software development services in accordance with caBIG® development standards as outlined in this SOW
- Database development services
- Project management
- Documentation of software, data model, and file structures
- Software testing
- Deployment, and
- Post-deployment on-going support

The Contractor shall aggregate all biomedical CPTC data into a public resource that is accessible through a web interface developed and maintained by the contractor.

3 Objectives

This statement of work (SOW) shall meet the following major objectives:

- Design, build, specify, enhance, test, deploy, and support the CPTC DC, targeted to centrally store and/or coordinate data produced by PCCs and coordinate CPTC data with that of TCGA on corresponding samples.
- Establish and execute standard operating procedures (SOP) for data transmission from PCCs to the DC. SOP will be developed in collaboration with representatives from PCCs and RC. DC will play a coordinating role among various PCCs generating similar data, and will be the final arbiter for definition of standard formats.

- Design and implement data analysis procedures that perform quality checks on incoming data and report anomalies and non-conforming data to the data source sites.
- Train a representative from each PCC and RC with appropriate SOPs for data transmission, and procedures for addressing non-conforming data.
- Implement a data management pipeline to process data and prepare it for public distribution in formats and systems that are consistent with compatibility guidelines defined by the caBIG® program as outlined in this SOW.
- Post the data to the publicly accessible distribution points after quality checks and processing are completed and that NCI's data release policy guidelines have been followed and cleared.
- Design, build, specify, enhance, test, deploy and support the CPTC Data Portal, targeted to basic and clinical researchers, for browsing and downloading subsets of data that are selected based upon various characteristics of data provenance, source, and content. The CPTC Portal shall display data from the CPTC project and other sources.
- Provide regular and detailed reports of the status of CPTC data, data quality, and data usage at all stages of the data management pipeline.

4 Technical Description

4.1 CPTC Workflow

The structure of CPTC is designed to facilitate the efficient implementation of the 2-step developmental pipeline. Briefly, the pipeline originates with the CPTC Resource Center, which shall provide a set of high quality, genomically characterized, clinical biospecimens (tumor and normal whenever available) from patients with a single type of cancer. De-identified clinical and genomic data from these biospecimens shall be available through the TCGA Data Access Portal.

Discovery Units of each PCC shall use a set of proteomic technologies to comprehensively analyze the protein composition of this set of biospecimens. To create a comprehensive data set for advancing proteomic research consistent with the goals of this program, all data from the Discovery Units, as well as corresponding data from TCGA, shall be shared within the network via the CPTC DC. Based on this aggregate dataset, each Discovery Unit shall select and prioritize cancer-related proteins for potential use in the verification phase.

The selected lists of prioritized proteins by the Discovery Units shall be consolidated and collectively prioritized by the Biomarker Candidate Selection Subcommittee. This network-wide subcommittee shall consider the clinical, biological, and technical aspects of each protein and develop a verification plan for the proteins, laboratories, and resources involved in the verification step. These verification plans shall be subject to final approval by the CPTC Steering Committee, which may include assigning to a given PCC the verification of proteins discovered by another PCC. Both of these

committees shall consist of representatives from each PCC as well as NCI Program Staff. While the contractor may be asked to attend meetings of these committees, the contractor shall not have a vote on either of these committees. Additionally, these committees shall not have authority over the contractor. While committee decisions may have bearing on the contractor's tasks, all modifications to the contractor's tasks must be approved by the NCI Contracting Officer.

The Verification Unit of each PCC shall develop targeted assays against corresponding protein targets as assigned by the verification plan. The CPTC Resource Center shall facilitate assay development by providing key reagents (reference materials) as required. The Verification Units shall then verify the biomarkers by performing these assays in clinically relevant cancer biospecimens (provided by the CPTC Resource Center). The number of biospecimens shall be sufficiently large so as to ensure appropriate statistical power. Data from these verification studies shall be shared via the CPTC DC. These verified proteins shall serve as high value targets to other initiatives involved in clinical qualification studies.

Achievement of the goals of this reissuance for CPTC shall address the growing gap between multi-dimensional data at the genomics and proteomics level needed to support the cancer research community in their efforts to develop new, clinically useful protein biomarkers.

All data from both the discovery units and verification units shall ultimately be shared with the general research community via the CPTC Portal.

4.2 CPTC Data Center

Two categories of software are anticipated to be needed to carry out the CPTC DC tasks in this SOW. The first category includes scripts and programs to support data receipt, processing, analysis, and transformation, quality checks, and reporting. The Contractor shall re-use existing public scripts, programs, and software tools to carry out these functions, and develop new scripts, programs, and tools as needed. Since this category of software is for internal processing purposes only, it is not expected to conform to caBIG® compliance standards, though the data that is disseminated by the DC shall be compliant (see "caBIG® Compliance", Section 5 Task 11 below).

Any CPTC DC related software developed with funding under this contract shall be released under an open-source license that complies with caBIG® licensing guidelines https://cabig.nci.nih.gov/working_groups/DSIC_SLWG/index_html/document_view. caBIG® compliance of data is not required when the DC is required to deposit certain data types in non-caBIG® public repositories, such as those hosted by the National Center for Biotechnology Information (NCBI). Decisions about deposition in such non-caBIG® repositories shall be made in conjunction with the COTR, who will obtain input from the point of contact at CBIIT and with NCI CPTC program management.

4.3 CPTC Data Portal

The second category of CPTC DC related software includes data management systems to support public access to the data after it has been fully processed by the DC. The contractor is encouraged to build on existing systems as use of entirely new software systems in this category are not in scope for this SOW.
The Contractor shall design, develop, build, and maintain the CPTC Portal. The Contractor shall maintain a detailed specification of the CPTC Portal functionality as outlined below. The following is a high-level description of an example of the data types and functions to be supported by the Portal.

4.3.1 Data Types to be Supported by the Portal

CPTC shall generate several kinds of data including but not limited to the following: global protein identification, targeted protein quantification, analysis of specific types of post-translationally modified proteins, and protein-protein interaction. The below table enumerates the types of platforms that have been used to date by the CPTC Centers.

Platform
ABI 4000 Q Trap
ABI 4700
ABI 4800
ABI 5500 Q Trap
ABI 5800
ABI DE-STR Voyager
ABI Q Trap
ABI QSTAR Elite
ABI QSTAR Pulsar
ABI QSTAR XL
ABI QTRAP 4000
Agilent 6410 Chip Cube
Agilent 6460 Chip Cube
Agilent XCT+
Firefly 3000 immunoassay system
Invitrogen Protoarray NCV version 2
Luminex 100/200
NanoPro 1000 immunoassay system
Thermo Fisher TSQ Vantage Triple Stage Quadrupole
ThermoFisher LTQ
ThermoFisher LTQ-Orbitrap
ThermoFisher LTQ-XL
ThermoFisher Orbitrap Velos
ThermoFisher Orbitrap XL
ThermoFisher Orbitrap XL ETD
ThermoFisher TSQ Quantum Ultra Triple Stage Quadrupole
Waters QTOF Premier
Waters Synapt G2
Waters Xevo TQ

For each of the platforms used by CPTC investigators, the RC and PCCs shall produce several levels of data, including but not limited to the following:

- Raw data: primary-level data directly from the instrument for a single sample, not compiled across samples, and not interpreted for the presence or absence of specific proteins or features. Examples: raw mass spectrum file; Thermo *.raw, or AB Sciex *.wiff file.
- Processed data: data for single samples that has been processed using software to identify and examine:
    - Total ion current chromatograms
    - Peptide peak lists and their associated intensities
    - Protein identifications from single samples or fractions (Example: Proteins identified from a single HPLC fraction)
    - Compiled protein identifications from multiple samples or fractions (Example: A list of the total proteins compiled from proteins identified from multiple HPLC fractions, or a comparison of the proteins identified from different biospecimens)
- Analyzed data: Processed data which has been further analyzed in order to identify and examine:
    - Post-translational modification analysis of proteins (Examples: phosphorylation or glycosylation)
    - Analysis of protein-protein interactions (Example: antibody/antigen binding)
    - Quantitative/Semi-quantitative analyses of proteins, post-translational modifications, and protein-protein interactions
    - A comparison of all of the above analyses across different biospecimens of the same and differing types
- Summary finding: a quantified association, across classes of samples, among two or more specific proteins, features, sample characteristics, or clinical variables. Example: a finding that a particular protein or post-translational modifications of a protein is found to be amplified in 80% of a specific tumor type.

All data shall be linked to other data sources such as the clinical and genomic data that is available through the TCGA Data Access Portal and other genomic, epigenomic, and expression analysis sources.

4.3.2 Functions to be Supported by the Portal

The Portal shall support at least the following four generic usage scenarios:

1. Provide a high level summary of the data that exists in the Portal (tumor types, data types, etc.), the progress of data release for a given tumor type, portal functions accessible to the public, caveats about data that may have been modified and/or redacted, and a user friendly means to assist in data navigation.
2. Support generation of reports on the status of data within the DC. Reporting should allow for multiple views and inform different audiences:
    - Multiple views may include:
        - Time series of data transfer into and out of the DC

    - Modifications of data due to incomplete information or inability to meet quality metrics
    - Workflow tracking of data associated with a single specimen
    - Progress toward program goals aligned with tumor types
  - Support for multiple audiences may include:
    - CPTC Management
    - Quality Management Requirements (to be specified)
    - Individual Centers (PCCs)
    - Basic researchers
    - Clinical researchers
3. Provide a means to generate multiple views of the data to users, based upon a variety of possible characteristics of selection, using variables that are intrinsic to the data – absent of calculation and biological or clinical correlation.
   - View raw and processed data for a given assay for a single sample.
   - View raw and processed data for a given assay for a set of samples.
   - View proteomic and genomic data for a single sample.
   - View proteomic and genomic data for a set of samples.
   - Download all raw data (by disease, data type, platform, generating PCC center)
4. Facilitate data quality assessment and review of raw and processed data
   - Develop a workflow or process that integrates consistent and regular data quality checks.
     - Example: As data is sent to the DC, an example workflow might articulate steps for data acceptance, data integration into the DC database, database value checks, database comparison to original data, and regular data quality assurance checks over time to minimize data corruption.

Regular reporting of these tasks and their results, periodicity, and format to be decided by the Contractor and the NCI COTR, shall be necessary to effectively manage the DC data systems.

Some CPTC Data will be restricted to registered users. As part of this SOW, the Contractor shall implement appropriate access controls for the DC and the Portal.

5 Tasks to be Performed

The Contractor shall be responsible for ensuring the accuracy, timeliness, and completeness of all tasks performed under this effort.

TASK 1 - PROJECT MANAGEMENT

TASK 1.1 - Project Management Plan

The Contractor shall develop a Project Management Plan (PMP) describing the means to accomplish project objectives. The PMP shall describe the technical approach, deliverables, methods for producing deliverables and the following management plans:

- Scope Management – scope definition; work breakdown structure (WBS) - encompassing planning, tracking, execution, and completion activities for all tasks; and change control.
- Data source management – activities to manage interactions with and between the various data source PCCs and the potentially un-harmonized formats of the data those PCCs shall deliver to the DC.
- Time Management – Outline activities, methods for producing deliverables, resource requirements, project durations and schedules, and project schedule controls. If an agile development process is to be employed, the activity sequencing of development sprints may be outlined here.
- Quality Management – quality assurance processes and controls; deliverable validation.
- Communications Management – stakeholder management and performance reporting. Given the large number of stakeholders and participants in CPTC, the NCI COTR and the Contractor shall agree upon an appropriate communications burden and process for the DC Contractor.
- Risk Management – risk identification and management processes – to include a risk register to facilitate, monitor, and manage risks, as well as the communication of the status of risk prevention, mitigation, and response.
- Procurement Management – process for major purchases and acquisitions, such as the purchase of licenses and support for large information management systems.
- Change Management – address the inclusion of new data types from new and improved platforms put into use as CPTC progresses, and to address how data will be forward-migrated over the course of the project.

The PMP shall include an MS Project formatted document including the WBS, and MS Word documents describing the other plans listed above. The draft WBS shall encompass any milestone reviews described in this SOW, as well as any requested Contractor supplied dates for Deliverables. Based on review of the PMP, the NCI COTR shall provide approval to move forward on planned activities. The Contractor shall request prior approval for all activities not included in the PMP or any modifications to the plan after approval has been given. The final PMP, including all approved changes, shall be submitted as a deliverable at the end of the project.

TASK 1.2 - Monthly Status Reports

The Contractor shall document technical and financial status in detailed Monthly Status Reports. All monthly technical and financial reports shall be delivered in the template format provided by the COTR during the project kickoff meeting. An acceptable Monthly Status Report (MSR) format will be decided upon by the NCI COTR and the Contractor. An anticipated MSR shall include the following:

- Program status overview, including status of deliverables/milestones.
- Description of activities completed during the month.
- Description of activities planned for the next month.
- Description of resolution(s) to previously – identified issues/obstacles.
- Delineation of risk factors that may delay completion of planned items and recommended solution(s).
- Budgeted total and budgeted monthly hours vs. actual monthly and total hours.
- Planned Value, Earned Value, and Actual Costs for the month and project.
- Estimated Cost to Completion.
- Task/Cost variance (for >10% variance include explanation/analysis).
- Other direct cost (ODC) progress/costs.

In addition, the Contractor shall provide status teleconferences with the NCI COTR and the point of contact from CBIIT. These teleconferences will generally be held at a weekly frequency, but focused status teleconferences to address specific issues may also be requested by the NCI COTR.

TASK 1.3 - Final Project Summary Report and Project Close Out

The Contractor shall submit a Final Project Summary Report at the conclusion of the contract. This report shall describe the work accomplished, issues encountered and resolutions to them, recommendations for future enhancements, potential implementation strategies, inventory of the data files, and lessons learned. In combination with the final software code, documentation, and data files, this report shall provide the necessary information to operate the DC and portal. The final version of the project summary report, including any approved changes made during review, shall be delivered to the COTR in electronic form (Microsoft Word® format) at the completion of the project.

TASK 2 - SYSTEM REQUIREMENTS GATHERING, ARCHITECTURE, AND DESIGN

TASK 2.1 - System and Software Requirements

The Contractor shall meet with individuals and groups representing major stakeholders and potential Portal users as determined by the Contractor and NCI COTR, including the CPTC PCC representatives, CPTC Program Staff representatives, and additional experts to be identified in order to develop the System Requirements. The Contractor shall summarize the requirements in a Software Requirements Specification (SRS) document. The SRS is a complete description of the behavior of the systems to be developed and shall explicitly state the functions and capabilities of the CPTC Data Center and Portal. The SRS shall contain the PCC, NCI, and CPTC Requirements, Functional User Requirements, Non-functional Requirements, and Performance and Design Requirements.

Of significance to this task is the Functional User Requirements, which are largely comprised of the Use Cases. The Contractor shall develop a set of Use Cases specific to the intended use and operation of the systems, subsystems, and/or features to be developed. The set of Use Cases shall clearly identify how the system is intended to be used in its functional context. The System Use Cases shall make specific reference to any enterprise Use Cases described in the Technical Description, Objectives, and Scope of Work sections of this RFP document. In the event the Contractor employs an agile programming development process, User Stories and Epics may be substituted for Use Cases.

TASK 2.2 - System and Software Design

The Contractor shall produce system and software architecture and design artifacts. Specifically, the Contractor shall submit a System and Software Design (SSD) document. The SSD document shall, at a minimum, describe the hardware and software architecture, and external interfaces.

The Contractor shall conduct and submit at least one (1) formal design review to the NCI COTR. For iterative developments, after the initial design review, updates to the design documentation shall be completed and approved by the NCI COTR before each planned increment.

The Contractor shall develop the Portal and shall develop the object model, data model, and data access methods to be used by the Portal. For any new objects, the Contractor shall annotate with Common Data Elements already in NCI's Data Standards Repository (caDSR) or create new Common Data Elements according to CBIIT's established process.

TASK 2.3 - User interface design

The Contractor shall prepare a Portal site map and wire-frames for Portal pages. In preparing the user interface design, the Contractor shall consult with a usability-testing organization, which may be provided by the NCI.

TASK 3 - DEPLOYMENT PLAN

TASK 3.1 - Deployment Plan

The Contractor shall develop a Deployment Plan that outlines:

- The assumptions/dependencies/constraints of the system and the operational environment
- Operational readiness of the system. This shall include such items as a deployment diagram, site and environmental requirements, assessment of system readiness, product content, etc.
- The roll-out process – especially if it is a phased process
- Training requirements and expectations
- System documentation such as user manuals, data primers, system administration manual, etc.
- Transition to support
- Operations and maintenance planning

The Contractor shall execute the Deployment Plan and include the deployment activities within the testing protocols. The Deployment Plan shall include a successful deployment package for download by the public. A successful deployment for each release shall include a single, downloadable package, with a comprehensive System Administration guideline, and result in successful execution of the system executable files that function as prescribed in the system documentation. In the event the DC software requires outside, third-party, open-source executable files, system requisites shall be clearly documented in the System Administration document.

TASK 3.2 - Product Demonstration

The Contractor shall provide a Product demonstration as prescribed by the NCI COTR . This shall be an end-user oriented demonstration of the product. In addition, if grid services are developed, a demonstration of all developed services, both data and analytical, shall be performed.

TASK 4 - DOMAIN MODEL, DATA ELEMENTS, AND TERMINOLOGY SELECTION AND SPECIFICATION

In conjunction with the CPTC RC, PCCs, and TCGA components, the Contractor shall come to agreement on the standard vocabularies, data elements, and data models needed to represent data flowing to the DC.

Working in concert with the PCCs, standard terminologies for annotations and representations shall be selected. Such terminology standards are expected to be drawn from the NCI Thesaurus, available from the NCI Enterprise Vocabulary Services (EVS, http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/vocabulary). In cases where the NCI Thesaurus does not contain suitable terminology, sources from other redistributable terminologies hosted by NCI EVS or approved by the NCI COTR and the Contractor, shall be used. Terminologies that are not hosted by NCI EVS or which have not been approved by the NCI COTR shall not be selected for use in CPTC. The DC shall be responsible for executing these processes, and will receive support from caDSR and EVS staff as needed. The final selection of information models and corresponding registered data elements shall be documented and reported to all CPTC Centers.

The following reports shall be submitted upon completion of each implementation phase:

- Report detailing information model for each category of data, accompanied with the models themselves in suitable file formats.
- Report detailing specific caDSR-registered data elements, to be used as the basis for specification of data processed and distributed by the DC.
- Report detailing the terminology sources to be used for controlled vocabulary.

TASK 5 - SYSTEM IMPLEMENTATION

The Contractor shall develop software code to implement Portal functions and shall conduct unit and integration testing. In order to manage the development activities, the Contractor shall maintain a Product Backlog detailing the functions to be developed,

priority of each function, and estimated development time per function. The Contractor shall review and assess the system's testing and quality assurance procedures. Where existing open source software is available to perform a function (particularly existing caBIG®-compatible software), the Contractor shall adapt and use existing software rather than developing new software, unless there are compelling reasons to the contrary which must be approved by the NCI COTR.

TASK 6 - QUALITY MANAGEMENT SYSTEM (QMS)

The Contractor shall conduct basic Quality Assurance (QA) testing on data that it receives from the RC and PCCs. The DC shall prepare a QA Test Plan for approval by the NCI COTR and the point of contact from CBIIT. This plan must address:

- Transmission integrity
- Data completeness
- Data validation
- Reporting back to data generation centers
- Reporting out to the CPTC program staff

TASK 6.1 - Implementation:

The Contractor shall implement the QA plan within the DC. The QA plan will be vetted by the NCI COTR and the point of contact from CBIIT and shall be in compliance with the, as yet to be established, CPTC Quality Management System.

TASK 6.2 - Operation:

The Contractor shall conduct Quality Assurance as part of the DC activities during the period of performance and maintain an up to date QA log.

TASK 6.3 - Data Quality Non-conformities:

If the Contractor detects a systematic QA problem with one or more of the PCCs or the RC, it shall communicate this problem to the appropriate Center and maintain a record of the issue(s) within the QA log. The Contractor shall work with the identified Center to attempt to ascertain the cause of the problem and shall engage the Center to identify solutions. If, after three months, this activity proves ineffective, the Contractor shall notify the Government in writing that a Center is non-compliant.

TASK 7 - REPORTING

The QMS is meant to serve primarily the management of the CPTC Program. The Contractor shall develop reports designed to meet the throughput and quality metrics for the CPTC Program. A number of metrics regarding proteomic measurement have already been identified, will be shared with the Contractor after contract award, and shall serve as a basis for a portion of the quality management reports. The COTR will work with the Contractor to approve all report formats, content, and frequency. These reports will need

to be an additional component of the QMS. The reports will be tabulated information with graphical summaries and accessible via the internet.

TASK 8 - STAFF TRAINING

The Contractor shall ensure that all staff, including subcontractors and consultants, are properly trained in handling secure data and IT security. A staff training plan shall be developed and must be approved by the NCI COTR and CPTC Program staff before implementation of any staff training.

TASK 9 - DATA COLLECTION AND PROCESSING

TASK 9.1 - Standard Operating Procedures (SOP).

The Contractor shall conduct a series of discussions with the RC and PCCs to define and establish appropriate Standard Operating Procedures for transmission and processing of data from the various data generation centers to the DC. These SOPs shall address the following Tasks. It is expected that the current DC SOP shall serve as the standard, and should be modified as appropriate.

TASK 9.1.1 - Sample identification:

The Contractor shall obtain initial sample information from the RC and maintain an index of sample identifications for the purpose of aggregating information from the various PCCs as well as linking CPTC data to TCGA data of the same sample. This sample index shall be used to communicate CPTC sample identifiers to all participating groups (RC and PCCs), and shall serve as the primary means to track sample and derivative data as it is generated and released through the CPTC Portal and DC.

TASK 9.1.2 - Data Transfer Schedules:

The Contractor shall work with the various data generation centers to determine the appropriate schedules for transfer of information to the DC, in accordance with the CPTC data release policies.

TASK 9.1.3 - Data Transfer Format:

Wherever possible, the Contractor shall receive data from the PCCs in a format that meets "caBIG® Compliance" as outlined in Task 11. However, in certain circumstances the Government recognizes that data transmission in some formats is not feasible. In those cases the Contractor may utilize alternate transfer mechanisms subject to the approval by the NCI COTR .

TASK 9.1.4 - Physical Mechanism of Transfer:

The Contractor shall define the procedures for the physical transmission of data (FTP, magnetic or optical media, etc.). As the CPTC program undertakes production, transmission of data may become a real obstacle for adequate release and access of information. In some cases, institutions will need to access terabytes of information on a consistent basis. Understanding the information transfer needs of CPTC and implementing an appropriate solution shall be a high priority of the Contractor.

TASK 9.1.5 - Security:

The Contractor, in cooperation with the RC and PCCs shall determine the appropriate level of security for the persistent data and during data transmission. This shall include cryptographic requirements (if any), secure transfer, etc. The Contractor shall create a process that is compatible with and in accordance with the terms of this contract, applicable Federal Information Security requirements (FISMA, HIPAA, 21CFR11, etc.) and the boundaries of the informed consent attached to the samples.

TASK 9.1.6 - Quality Assurance/Validation:

The Contractor shall provide appropriate Quality Assurance (QA) checks and data validation. This requirement is described in more detail in Task 6.

TASK 9.1.7 - Transformation:

The Contractor shall take the data feeds from the RC and PCCs and provide them to repositories as directed by the NCI COTR and the point of contact from CBIIT. The Contractor shall provide a procedure for supplying this data to data-persistence systems and repositories.

TASK 9.1.8 - ETL

The Contractor shall develop and release Extract/Transform/Load (ETL) processes to load CPTC data into data systems external to the DC. In addition, the Contractor shall transform the data into convenient bulk downloads that are consistent with caBIG® Compliance as outlined in Task 11. Additionally, the Contractor shall provide an interface to the ETL processes for TCGA data of corresponding CPTC samples.

TASK 9.1.9 - Reporting on Quality and Productivity Metrics:

The Contractor shall provide metrics to the CPTC program regarding data received from the RC and PCCs. The report shall include identification of all data that fails QA checks and data validation. The report shall also include data completeness metrics, measured against cases, samples, and reporting Center. These reports shall be delivered at least monthly, not later than the last day of the month following the activity.

TASK 9.2 - SOP Implementation:

TASK 9.2.1 - Data processing script and software program adoption and development.

The Contractor shall identify all data processing and transformation software scripts and programs that are needed to support the SOPs defined in Task 9.1. This software is for internal use in the DC, and is itself not directly subject to caBIG®-compatibility requirements. The Contractor shall first attempt to obtain data processing scripts and programs from public and open source bioinformatics resources. In cases where a need is not fulfilled by a public or open source software resource, after approved by the NCI Contracting Officer and COTR the Contractor shall develop and test the scripts, using the most efficient programming technology suited to the task. All scripts and programs, whether adopted from external resources or developed by the Contractor, shall be tested and validated prior to use in production.

TASK 9.2.2 - Implementation:

The Contractor shall implement the SOPs described in Task 9.1 and the associated processes for quality assurance and data persistence and distribution.

TASK 9.2.3 - SOP Validation:

The Contractor shall provide a test plan for each SOP and conduct several tests of the data transfer mechanism. If problems are discovered, the Contractor shall modify the SOP and/or the underlying process to correct the problems.

TASK 9.2.4 - Production:

The Contractor shall collect data from the RC and PCCs during the contract period as described by the SOPs created and implemented in Tasks 9.1 and 9.2. The Contractor shall perform all required activities described by the SOPs and this SOW.

TASK 10 - DATA PERSISTENCE AND DISTRIBUTION

TASK 10.1 - Plan for bulk data distributions on public FTP site.

As part of the SOPs, the Contractor shall develop plans for the uploading of bulk aggregated data files to the DC anonymous FTP site. The Contractor shall design bulk data download formats in accordance with caBIG® Compliance as outlined in Task 11. The plan shall provide for the immediate posting of all data that passes quality checks and shall not include data that fails quality checks, in accordance with the CPTC data release policy. If data which has passed quality checks and has been posted is subsequently determined to not meet quality standards, then the Contractor shall immediately remove it from the FTP site and post an explanatory statement to the site.

TASK 11 - caBIG® COMPLIANCE

NCI's Cancer Biomedical Informatics Grid (caBIG®, http://cabig.nci.nih.gov) requires that data be represented as aggregations of Common Data Elements (CDEs) that are registered in the NCI's Cancer Standards Data Repository (caDSR, http://cadsr.nci.nih.gov). These data elements are, in turn, defined using concepts codes that are maintained in the NCI's Enterprise Vocabulary System (EVS). Data submitted to the DC must be unambiguously analyzable into CDEs. Existing formats shall continue to be used, although it may be necessary to define additional CDEs and develop new formats to accommodate new kinds of outputs from the PCCs.

A complete set of Compatibility Guidelines can be found on the caBIG® Community Website at
https://cabig.nci.nih.gov/guidelines_documentation/?pid=primary.2006-07-07.4911641845&sid=compatability_certification&status=True

TASK 12 - DATA USE, DISCLOSURE, AND HANDLING OF SENSITIVE INFORMATION

TASK 12.1 - Security Controls

The contractor shall comply with the security controls stated in the contract.

TASK 12.2 - Privacy Standards

The Contractor shall address the potential sensitivity of the information collected, information security issues, local Institutional Review Board (IRB) requirements and the Health Insurance Portability and Accountability Act (HIPAA) of 1996 in its design of the system in question. The system shall accommodate the needs and actual uses, related to these laws and regulations, of caBIG® participants.

Final regulations issued by the Department of Health and Human Services (DHHS) provide privacy and security standards that shall be observed in the handling of patient data resulting from biomedical research. HIPAA privacy standards shall be used to establish safeguards and restrictions for the use and disclosure of research records. HIPAA security standards shall be used to help cancer centers implement administrative, physical, and technical safeguards to protect electronic health information. Improper use or disclosure of sensitive information under the rules may be subject to criminal or civil sanctions prescribed in HIPAA.

TASK 13 - DOCUMENTATION OF SOFTWARE

If required by the NCI COTR , the Contractor shall prepare the following documentation;

- Application Developers Guide
- System Administration Guide

The Contractor shall follow the guidelines specified in caBIG® Documentation and Training Review Process;
(https://cabig.nci.nih.gov/working_groups/Training_SLWG/Documents/).

TASK 14 - TRANSITION PLAN

The Contractor shall develop software code to implement DC and Portal functions and shall conduct unit and integration testing. In order to manage the development activities, the Contractor shall maintain a Product Backlog detailing the functions to be developed, function priority, and estimated development time per function. The Contractor shall coordinate with the Quality Management Systems Contractor and NCI COTR, who will review and assess the DC Contractor's testing and quality assurance procedures. Where existing open source software is available to perform a function (particularly existing caBIG®-compatible software), the Contractor shall adapt and use existing software rather than developing new software, unless there are compelling reasons to the contrary which must be approved by the NCI COTR and Contracting Officer.

6 Travel

It is expected that the Contractor will need to travel to the site of the RC (location(s) to be determined) on occasion during the requirements phase, and possibly later. The contractor shall also travel to each of the PCCs (site announcement forthcoming). CPTC is expected to consist of 6 to 8 PCCs.

7 Deliverables

All Deliverables shall be submitted in accordance with Sections C., D., E., and F. of the contract and as outlined in the SOW.

7.1 Deliverable List and Due Dates

The following table contains a list of deliverables with notation of due dates.

Acceptance criteria for each of the primary deliverables are given below. Inspection and acceptance also will be performed in accordance with the terms of the contract under Section E. (The contract will state that acceptance is presumed within 30 calendar days unless notified in writing by the CO – standard language.)

DELIVERABLE	DUE DATE
Cross Project Deliverables	
Project Management Plan	Due twenty (20) calendar days after the contract effective date
Revised Project Management Plan	As required, seven (7) days after approval by NCI COTR
Software Requirements Specification Document	Due no later than 160 days after the contract effective date
System and Software Design (SSD) Document	Due no later than 180 days after the contract effective date
SSD Design Review Document	Due fifteen (15) days after submission of the SSD Document
Monthly Status Report/Financial Report	Due the 10th calendar day of each month
Project Summary Report	Due ten (10) calendar Days prior to the expiration date of the contract
Software Code and Documentation	Due no later than ten (10) calendar days after completion of each development iteration
System Deployment Plan	As required
caBIG Compliance Review Submission Package	As required
Application Development Guide	Due fifteen (15) business days after each software update, and ten (10) business days prior to the expiration date of the contract
System Administration Guide	Due fifteen (15) business days after each software update, and ten (10) business days prior to the expiration date of the contract
Final PMP	Due fifteen (15) business days prior to the expiration date of the contract
Final Software Code, Documentation, and Data Files	Due ten (10) business days prior to the expiration date of the contract
CPTC-Portal Deliverables	
Product Backlog	Monthly and as required

DELIVERABLE	DUE DATE
Use case document	As required
Functional specification document.	As required
User interface design, consisting of a site map and wire-frames	As required
Software architecture specification.	As required
Physical database model.	As required
Code to implement Portal functions	As required
Code to implement the data extraction, transformation and load procedures.	As required
CPTC-DC Deliverables	
Product Backlog	Monthly and as required
Report detailing information model for each category of data, accompanied with the models themselves in suitable file formats.	As required
Report detailing specific caDSR-registered data elements, to be used as the basis for specification of data processed and distributed by the DC.	As required
Report detailing the terminology sources to be used for controlled vocabulary.	As required
SOPs for the collection and processing of data from the RC and PCCs, including transmission of data to the DC, quality assurance procedures, and upload for FTP access	As required
Test Plans for data collection and processing SOPs, including transmission of data to the DC, quality assurance procedures, and upload for FTP access.	As required
Validated script and software programs needed to implement data processing SOPs, including transmission of data to the DC, quality assurance procedures, and upload for FTP access.	As required
Test Results for SOPs for data transmission, quality assurance and upload for FTP access	As required
Monthly progress reports on the data collection and processing activities for each PCCs and the RC	Due by the 15th calendar day of each month
QA Plan	As required
QA and Productivity Reporting Metrics	Due by the last day of each calendar month
Documentation of bulk file structures, including description of correspondence between data fields and registered data elements in the caDSR	As required

DELIVERABLE	DUE DATE
Instantiated system for hosting data for FTP access	As required
Regular data loads, depositions, and bulk uploads on an ongoing basis	Recurring, beginning 4 months after the effective date of the contract and continuing on a monthly basis at the end of each calendar month at a minimum

7.2 Format and Delivery

Monthly Reports and any documents containing sensitive information such as financials shall be sent via email directly to the NCI COTR. Documentation deliverables other than those described above shall be in MS Office Format and uploaded to the controlled wiki pages associated with the project. A notification of upload or subsequent updates to any deliverables on the wiki site shall be forwarded via email to the NCI COTR.

Delivery of the final software code, documentation, and data files shall be delivered to a designated site to be named by the NCI COTR.

7.3 Deliverable Acceptance Criteria

Specific Acceptance Criteria

The following specific criteria will be used when evaluating deliverables listed above.

Deliverable: Project Management Plan

Acceptance Criteria: The project management plan shall include the following;

- Project Schedule; The Project schedule shall be delivered in MS Project format. The document shall include a hierarchical set of tasks representing the Work Breakdown Structure (WBS) required for the Contractor to successfully complete the tasks given in this statement of work. Task dependencies, task dates, resource loading, and updated milestones and deliverable dates shall all be included.
- Management Approach; The Management Approach shall be an MS Word® formatted document describing the organizational resources and management controls employed to meet the cost, performance, and schedule requirements as outlined in the project schedule. The required sections of the Management Plan are outlined in TASK 1.1 - This document shall detail the products, methods for producing deliverables, allocation of staff and other resources necessary to produce deliverables, and timelines, and any Government Furnished Information/Equipment expected from the Government.

Acceptable Project Management Plans shall include the above information, in the formats requested, in sufficient detail to cover all tasks stated in the SOW.

Deliverable: Monthly Status Report

Acceptance Criteria: The monthly status report shall be delivered in the template supplied by the NCI COTR with all fields completed. Changes need to be approved by the Contracting Officer via a contract modification. This document shall be sent via

email directly to the NCI COTR. Acceptable Monthly Status reports shall contain the complete set of requested information in the format requested.

Deliverable: System and Software Design and Implementation Plan
Acceptance Criteria: The System Design and Implementation Plan shall include at a minimum, the following;
Software Requirements Specification: The SRS will document the CPTC Project Requirements, Functional User Requirements, Non-Functional User Requirements, Performance and Design Requirements, and system Use Cases or Stories.
System and Software Design: Design documentation shall contain at a minimum design artifacts describing both the architecture and design of the products to be developed. These can be presented in the design framework of the Contractor's choice, although UML is preferred. Acceptable design documents shall include at least the following; assumptions and dependencies, general constraints, architectural strategies and development methods, system and subsystem architecture, detailed system and subsystem design – both static and dynamic. These shall be presented in sufficient breadth and detail to cover the complete set of requirements as outlined in this Statement of Work.
Preliminary Compatibility Review Package; For projects developing compliant products, a draft Compatibility package shall be included.
Implementation Plan; An acceptable Implementation plan shall describe, at a minimum, the technical environment within which the product (e.g. HW, OS, Middleware) is to be developed, the external dependencies, the teams and locations that will perform the development, the procedures to be used for the development (e.g. lifecycle model, CM) and a detailed schedule.

Deliverable: Software Code and Software Documentation
Acceptance Criteria: This deliverable is comprised of both the Software Code and Documentation
Software Code acceptance is contingent on the following;

- Test Results acceptance
- Evidence by inspection that all source code has been placed in CM
- Adequate commenting of source code (adequate will mean that at a minimum, each function point of the code has at least one comment)
- Build Documentation
- Demonstration that software retrieved from the delivered code base can be successfully built and accomplishes the agreed upon functions.
- Where appropriate, passes the caBIG® Compliance review.

Software documentation is contingent on the following;

- Software Code acceptance
- The documentation is complete from the perspective of the intended user, i.e. that it addresses all major end-user functions as defined by the system use cases.
    - reporting and responded to software defect reports from the Adopter.

7.4 Deliverable Review and Acceptance Process

For all deliverables, the COTR will perform an evaluation for each deliverable using acceptance criteria provided above. The Contracting Officer will provide written

notification of unacceptable deliverables to the Contractor within thirty (30) calendar days of receipt of the deliverable. The Contractor shall address each issue or state why the issue may not be addressed and resubmit the deliverable within fifteen (15) calendar days. The Government will not pay for revisions made to unacceptable deliverables. Resubmitted deliverables will be reviewed by the Government within fifteen (15) calendar days.

8 Place of Performance

The Contractor may propose that work be performed at the Contractor's site. If the Contractor's site is within the Washington DC vicinity, the project lead shall meet with the NCI COTR on a weekly basis. If the Contractor's site is not within the Washington DC vicinity, the project lead and NCI COTR shall have weekly teleconferences and a bimonthly in-person meeting. Additionally, the NCI COTR may ask that specific tasks, such as testing and demonstrations, be performed at other locations, to include locations within the Washington, DC area, or at any of the CPTC member institutions.

9 Period of Performance

The Contractor's period of performance shall extend one year from the date the contract goes into effect. The contractor may also be awarded 4 one-year options on an annual basis.