

Supporting Statement A for

Cancer Genomics Cloud Pilots (CGC)(NCI)

Date: March 15,2016

Name: Anthony Kerlavage, PhD

Address: 9609 Medical Center Drive, Rockville, MD 20850

Telephone: 240-276-5190

Fax:

Email: Anthony.kerlavage@nih.gov

Check off which applies:

- New
- Revision
- Reinstatement with Change
- Reinstatement without Change
- Extension
- Emergency

- Existing

### **Table of Contents**

A.1	Circumstances Making the Collection of Information Necessary	3
A.2	Purpose and Use of the Information Collection	4
A.3	Use of Information Technology and Burden Reduction	4
A.4	Efforts to Identify Duplication and Use of Similar Information	5
A.5	Impact on Small Businesses or Other Small Entities	5
A.6	Consequences of Collecting the Information Less Frequently	5
A.7	Special Circumstances Relating to the Guidelines of 5 CFR 1320.5	5
A.8	Comments in Response to the Federal Register Notice	5
A.9	Explanation of Any Payment of Gift to Respondents	6
A.10	Assurance of Confidentiality Provided to Respondents	6
A.11	Justification for Sensitive Questions	6
A.12	Estimated Annualized Burden Including Annualized Hourly Costs	6
A.13	Estimate of Other Total Annual Cost Burden to Respondents or Record Keepers	7
A.14	Annualized Cost to the Federal Government	7
A.15	Explanation for Program Changes or Adjustments	8
A.16	Plans for Tabulation and Publication and Project Time Schedule	8
A.17	Reason(s) Display of OMB Expiration Date is Inappropriate	8
A.18	Exceptions to Certification for Paperwork Reduction Act Submissions: None	8

## **List of Attachments**

Attachment A: Survey Screenshot

Attachment B: Consultants

Attachment C: Invitations

Attachment D: Privacy Act Memo

## A. Justification

CBIIT, in collaboration with the Center for Cancer Genomics, is coordinating a program to develop three [Cancer Genomics Cloud Pilots \(CGC Pilots\)](#) to help meet the research community's needs to access and analyze high quality, large-scale cancer genomic data and associated clinical information. The goal of this effort is to develop an innovative, cost-effective model for computational analysis of biological data and provide broader yet secure access to genomic data that NCI generates. Cloud computing will be a valuable tool to support studies related to the mechanisms of cancer. This capability will be equally valuable to other NCI scientific areas, including clinical trials and other types of patient-focused research. Three contracts were awarded to develop the CGC Pilots, to the Broad Institute, the Institute for Systems Biology (ISB), and Seven Bridges Genomics. Each of these groups is developing infrastructure and a set of tools to access, explore, and analyze molecular data. Key design principles for the clouds include: APIs for secure tool and data access, usability for biologists and clinicians as well as bioinformaticists and application developers, scalability, sustainability, extensibility to new data types without major refactoring, and open source, non-viral software licenses. All three CGC Pilots have chosen to implement their systems through commercial cloud providers and are collaborating on adopting common standards. Beyond these commonalities, the three project teams have distinct system designs, data presentation, and analysis resources to serve the cancer research community.

### A.1 Circumstances Making the Collection of Information Necessary

The Public Health Service Act (PHSA), Section 413 (42 USC § 285a-1) authorizes the Director of the National Cancer Institute (NCI) to support information programs that collect, identify, analyze, and disseminate information on cancer research and treatment to the public. The Director of the Institute may take such action as may be necessary to insure that all channels for the dissemination and exchange of scientific knowledge and information are maintained between the Institute and the public and between the Institute and other scientific, medical, and biomedical disciplines and organizations nationally and internationally. Further, the PHSA, Section 413 (42 USC § 285a-2) authorizes the Director to utilize information systems to disseminate such information. Specifically, the mission of the Center for Bioinformatics and Information Technology of NCI is to plan, provide, and coordinate technology, standards, and scientific computing in support of the NCI mission to speed discovery, facilitate open science, and progress towards precision treatment in cancer care and a learning healthcare system. In order to understand the utility and value of the tools being developed, the NCI has developed a survey instrument to capture feedback from the cancer research community.

## A.2 Purpose and Use of the Information Collection

The information collected as part of this survey process will be used exclusively by the NCI to determine future funding of cloud technology projects. The purpose of the Cloud Pilots is to provide a means to understand and test the utility of genomic analysis tools and infrastructure to the cancer research community. The feedback collected will allow NCI to understand that utility and plan specific follow-on projects that will provide the most value to cancer researchers. Researchers will be asked what functionality they find useful, what they would like to see added, what functions they performed, and whether they would continue to use the cloud platforms if they remain available.

The survey will be used to understand:

- Usefulness and value of the CGC Pilots and associated functionality
- Whether platforms such as these should be maintained or further developed
- If specific functionality should be incorporated into future version of a cloud platform for cancer research

Information will be collected through an online survey (Attachment A) that provides simple collating and reporting functionality. The respondents will include NCI researchers, NIH researchers, Principal Investigators at research institutions, informaticists and software developers at research institutions, and scientists who are performing genomics researchers. The survey will be available online, and respondents will be invited (Attachment C) to complete the survey through a variety of ways:

- Email lists to which researchers are currently subscribed:
  - An email will be sent to people who attend a workshop NCI is holding in May, 2016. The workshop will be conducted by the Cloud Pilot teams, and will including demonstrations and hands-on work with the Cloud Pilot systems. After the workshop is completed, we will email each participant asking for them to participate in the survey.
  - We will be emailing people who are registered with one or more of the Cloud Pilot systems, asking them to describe their experience and provide feedback by completing the survey.
  - We have several general email lists, such as the one associated with TCGA research, through which we will request feedback by completion of the survey.
- Links in the NCIP Twitter feed: Using our NCIP Twitter account, we will tweet requests to the NCIP community asking them to provide feedback on the Cloud Pilot systems by completing the survey. The Twitter account is currently used to communicate with the community about the Cloud project. Because of the nature of Twitter, we have composed several different tweets we will use at different times to request survey completion.
- On the NCI CBIIT and Cloud Pilots web sites
  - As people are using the Cloud Pilot systems, we will present them with a pop-up window or a link on those sites asking them to describe their experience and provide their feedback by completing the survey.

- o The CBIIT/NCIP page on the cancer.gov site has a section dedicated to the Cloud Pilots. When the survey is available, we will place a link on that section of the site asking for feedback by completing the survey.
- At conference presentations and workshops: The Cloud Pilot teams will be attending conferences throughout the year. We plan to have a flyer and other materials to hand out about the Cloud Pilots. We will include the text below in the flyer / materials, requesting feedback by completion of the survey.

The survey results will be compiled in a document and reviewed by federal staff to determine future funding and direction for the cloud platforms and functionality.

### **A.3 Use of Information Technology and Burden Reduction**

The survey is web-based and administered using Survey Monkey. This will make the process extremely simple for the respondents. The link to the Survey Monkey page will be provided in several ways:

- As links from the CGC Pilot web sites
- As links from the cancer.gov CBIIT web site
- In email distribution, to those who have subscribed to listservs related to the project
- During a workshop planned for May, 2016
- At conference presentations about the CGC Pilot

Researchers will have the choice to participate or not. The survey will be primarily multiple choice, so it will be quick and easy to complete.

### **A.4 Efforts to Identify Duplication and Use of Similar Information**

There is no possibility of duplication. The NCI is funding this project and is the only entity requesting formal feedback to the government on the CGC platforms. No other IC or agency is involved with this project. It was launched in October 2014, and no information request was issued until this time.

### **A.5 Impact on Small Businesses or Other Small Entities**

This information collection activity will have no impact on small businesses or entities.

### **A.6 Consequences of Collecting the Information Less Frequently**

If this information collection activity is not conducted, it will greatly undermine the intent of the program. Without input from the cancer research community, the government will have no way to fully understand the benefits of the technology and which functions and features are most valuable. The survey is designed to be completed by each researcher once for each Cloud platform he/she uses. This will not be an on-going information collection activity. Researchers will have the opportunity to use the platforms, and after they

have had experience with them, they will be requested to complete the survey to provide NCI feedback on which tools are most useful.

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

There are none.

#### **A.8 Comments in Response to the Federal Register Notice**

A 60 day Federal Register Notice was published in the Federal Register on January 13, 2016, Vol.81; pp. 1633. No comments were received.

##### **A.8.2 Efforts to Consult Outside Agency**

The three contractors who are developing the CGC platforms, Broad, ISB, and SBG,(Attachment 2) were consulted for their feedback on the appropriate survey questions. The general research public has not yet had access to any of the CGC platforms, and therefore is not in a position to provide input as to the survey questions. The survey itself will have two open-ended questions that will allow researchers to provide a free text response that could include information not specified in the multiple choice questions.

#### **A.9 Explanation of Any Payment of Gift to Respondents**

No gifts or remuneration will be provided for completion of the survey.

#### **A.10 Assurance of Confidentiality Provided to Respondents**

All information will be kept private to the extent permissible of the law. Respondents will be asked to provide the name of the institution in which they work and their role at a very high level. The respondents will not be identifiable from this information.

The Privacy Coordinator is in the process of completing a PIA.

#### **A.11 Justification for Sensitive Questions**

No questions of a sensitive nature will be asked. PII is not being collected.

The Privacy Act is not applicable as determined by the Privacy Act Memo (Attachment D) provided by the NIH Privacy Officer.

This information collection is not research and is not subject to human subject regulations.

## A.12 Estimated Annualized Burden Including Annualized Hourly Costs

### A.12.1 Estimated Annualized Burden Hours

The estimated annualized total burden hours are 375. There are a total of approximately 1500 respondents. This is based on the number of PIs who are involved with the TCGA project, since the data contained in the Cloud Pilots is TCGA data. The survey questions, of which there are 19, are primarily multiple choice (16 out of 19), and each multiple choice question is expected to 30 second to complete. The 3 text questions are expected to take 2 minutes each. Therefore, the survey is expected to take approximately 15 minutes to complete.

#### A.12-1 Estimated Annualized Burden Hours

Form Name	Type of Respondent	Number of Respondents	Number of Responses per Respondent	Average Burden Per Response (in hours)	Total Annual Burden Hour
Cloud Pilot Survey	Principal Investigators	1500	1	15/60	375
Totals		1500	1500		375

### A.12-2 Annualized Cost to respondents

The annualized cost to the respondents is \$14,047.50 total. The wage rate is \$37.46 per hour. This was obtained for the most recent Bureau of Labor Statistics Occupational Employment Statistics, for 19-1029 Biological Scientists found on the Bureau of Labor Statistics Occupational Employment and Wages, May 2014, 19-1029 Biological Scientists website, <http://www.bls.gov/oes/current/oes191029.htm>.

#### A.12-2 Annualized Cost to the Respondents

Type of Respondent	Total Burden hour	Hourly Wage Rate*	Respondent Cost
Principal Scientists	375	\$37.46	\$14,047.50
Total			\$14,047.50

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

Cost to respondents of this survey is solely their time to respond to the questions. There are no capital or start-up components, no maintenance or services, and no record keeping required.

**A.14 Annualized Cost to the Federal Government**

The annualized cost to the federal government is \$21,557.65. The federal personnel will engage in review of the results. The contractor tasks are to track and compile results as they are collected, and to provide reports to the NIH Project Oversight Officer.

Annualized Cost to Federal Government Item	Salary	% Effort	Annualized Cost
NIH Project Oversight Officer - GS15-10	\$158,700.00	1.0%	\$1,587.00
One (1) in-house contractor staff	\$333,177.60	6%	\$19,990.65
Other costs, non-labor			0
<b>Total</b>			<b>\$21,577.65</b>

**A.15 Explanation for Program Changes or Adjustments**

This is a new collection of information.

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

Statistical analysis will not be performed in this information collection.

The Project Time schedule is provided in the table below. As soon as the survey is approved by OMB, it will be published online and publicized to the scientist community through emails and at public speaking engagements. The survey will remain open for several months and results will be compiled as they are collected. The results will be reviewed 10 - 12 months after the survey has been approved.

<i>A.16 - 1 Project Time Schedule</i>	
<i>Activity</i>	<i>Time Schedule</i>
<i>Survey published on Survey Monkey</i>	<i>Immediately on OMB approval</i>
<i>Survey announced at AACR Meeting</i>	<i>April 14, 2016 (if approved at this time)</i>
<i>Notice of survey tweeted</i>	<i>Following AACR Meeting (or approval, if after April 14, 2016)</i>
<i>Notice of survey added to cancer.gov site</i>	<i>Following AACR Meeting, (or approval, if after April 14, 2016)</i>

<i>Notice of survey emailed to listserv members</i>	<i>Following AACR Meeting, (or approval, if after April 14, 2016)</i>
<i>Analyses</i>	<i>10 - 12 months after OMB approval</i>
<i>Publication</i>	<i>12 - 18 months after OMB approval</i>

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

The expiration date will be displayed.

**A.18 Exceptions to Certification for Paperwork Reduction Act Submissions: None**

There are no exceptions.