

## **Supporting Statement: Request to Solicit Public Input to Nominate Non-federal Health and Health Care Data Sets and Applications for Listing on Healthdata.gov**

This document provides a basis for the application for the Information Collection Request being established to enable a web portal, healthdata.gov, to include information that will enable the public to nominate datasets and applications.

### **Part A. Justification**

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

As a means to improve health and health care performance, the U.S. Department of Health and Human Services (HHS), as part of the Administration's Open Government Initiative, has been making federal data sets on health indicators available to data.gov. Through access to these datasets, users are able to develop applications of the data and apply them to a broad variety of challenges and opportunities to improve health and health care at the community level. This effort to obtain input from the public is in compliance with the OMB Guidance to Agencies on the Administration's Open Government Directive.<sup>1</sup>

A focused activity known as the Community Health Data Initiative (CHDI), a public-private partnership to foster better uses of data, is engaging a broad range of stakeholders to focus on health and health care problems by using data to create applications in new and improved ways. Earlier this year, HHS integrated a variety of HHS datasets on public health, services, and administrative transactions through an interim web-based workpage that has become a one-stop location for health and health care data users ([http://www.cdc.gov/nchs/data\\_access/chdi.htm](http://www.cdc.gov/nchs/data_access/chdi.htm)).

In December 2010, HHS will launch a data warehouse that includes these datasets, new data sets, analytical tools, and application programming interfaces (APIs) that will enable automated computer program downloading and use of the data. The *Health Data Indicators Warehouse* will become a common site for users to access federal data resources.

The health data indicator sources in the warehouse will also be exposed to data.gov, where there may be enhanced applicability for users of and other health data sets that are supported in the data.gov environment. For example, data on environmental factors affecting human health and nutritional resources are found in data.gov through data resources provided by the Environmental Protection Agency and U.S. Department of Agriculture, respectively. At the same time that the HHS data warehouse is being released, the General Services Administration (GSA) will be launching a new "health community" in data.gov that will bring together all of the federal health datasets, tools, and applications that address health. This resource, "healthdata.gov" will also be enhanced to create a more flexible and easily accessible data environment for social interaction (commenting, rating, and technical discussions). Healthdata.gov will be

---

<sup>1</sup> <http://www.whitehouse.gov/open/documents/open-government-directive>

hosted on the GSA services, but will be maintained, monitored, and administered by HHS.

With the launch of healthdata.gov, there is interest in establishing linkages to non-federal health and health care datasets and applications that may have high value to the resource user community. This Information Collection Request (ICR) is to enable input from the public on non-federal data sources that may be useful and appropriate for use in applications to improve health and health care. Examples of these non-federal health indicator sets may include resources developed by state or local government surveys, private industry data resources on surveys, or health indicator sets developed by professional organizations, universities, or others. These resources could be very helpful, alone or in combination with federal datasets, to achieve progress in health and health care. Federal resources are not sufficient nor would it be appropriate for the federal government to directly collect information to address all possible areas of possible data application.

Building on the successful use of federal health and health care data through the CHDI, HHS is now looking for ways to engage the private sector, researchers, and members of the public. This request is to enable public members to identify high-value datasets and applications that are appropriate for broad use.

Details of the proposed types of information to be collected are provided in Appendix A. As these are proposed to be used through the healthdata.gov website, images of a draft series of screens is presented in Appendix B.

## 2. Purpose and Use of Information Collected

The ICR will specifically be soliciting public input through the healthdata.gov website which is hosted on GSA servers, but will be monitored and maintained by HHS staff. To enable the maximal capability of the federal data sources and applications, the ICR will be made to solicit input from the public on two domains of content:

- a request to the public on the data.gov/health site requesting submissions for linkages to non-federal health data sets and,
- a request for to the public on the data.gov/health site requesting submissions of applications using health datasets.

The information collected from the public will be used in the following manner. The primary purpose of the solicitation is to identify datasets and applications that would be suitable and appropriate for establishing a weblink from the appropriate section (designated as non-federal sources of datasets and applications) to the site where the data or applications are hosted. The user of healthdata.gov will be notified that they are leaving a federal website when the access that link.

The submitted material will be collected by a system administrator of healthdata.gov (an HHS employee) and collected for completeness. Once completed, the evaluation of the suitability and appropriateness for listing on the non-federal dataset and applications section of the healthdata.gov site will be performed by the system administrator. The system administrator will review the submission of nominated data from the public to ensure that all required documentation has been included. Nominated data set

submissions will complete the information on the metadata template for non-federal data (see attached “metadata template non federal.xls” file).

An example of how this process may work is as follows. A researcher interested in developing measures to assess community-level access to healthy and nutritious foods in schools may know of survey data collected by food service suppliers organizations. The researcher submits information about a private, not-for-profit organization that has developed a web-based data system that enables local and regional local food retailers to enter data in about school food services. The researcher can submit the name of the contact, and pertinent information about the data resource through the submission portal on healthdata.gov. Next, the information is collected by the HHS System Administrator, and reviewed for completeness and correctness. The System Administrator will verify completeness of the information, compliance with federal regulations, and overall value to the healthdata.gov community users.

In making recommendations for non-federal data, sponsors will provide all information requested on the Metadata Template for Datasets and Tools. Submissions will be reviewed using the criteria/categories contained on the template. Information on the Metadata Template provides a transparent quality profile of the nominated non-federal data for data users and provides information on data collection and design features regarding data utility, accuracy, and quality.

No data sets will be posted that contain identifiable information on individuals, whether the information directly or indirectly identifies an individual. An attestation will be required of all submissions stating that the nominated data set does not have personally identifiable information.

Nominated data sets must be available at no cost to the government.

Non-federal data posted on the site will not be moderated or monitored by HHS, and posting does not indicate U.S. Department of Health and Human Services endorsement of the dataset or its collection source.

### 3. Use of Improved Information Technology and Burden Reduction

This information collected under the ICR will be using web-based technology to enable the public easily accessible and usable tools for the data collection. This is similar to other web 2.0 practices that enables the user controlled domains and prompts to address the information submission.

The users will access the information request submission site on the healthdata.gov site itself, thus identifying the requesting site with the ultimate home for the links that may be established to the non-federal datasets and applications.

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

In some cases, a limited (less than a 10) public dataset sources will be known to the government that have been identified through comments provided at meetings,

conferences, and personal contacts. There are no other suggested mechanisms to identify these information sources. This request will not be duplicative of other federal information requests.

5. Impact on Small Businesses or Other Small Entities

This information request places no specific burden on small businesses. The information collected through this site may be voluntarily offered by small business (i.e., their applications developed using the data). Small businesses or other small entities are commonly users of the datasets and may find the resources developed from the non-federal dataset submissions to be highly valued in their business opportunities. Overall, we would anticipate a positive value for the small business community by helping identify data resources and making it more easily accessible to them.

6. Consequences of Collecting the Information Less Frequently

Without collecting this information, there will be a lack of more complete data to provide to enhance the applications development to improve health and health care. The information collection system through uses a webportal that requests information on a voluntary basis and is directed only to those individuals interested in the uses of health and health care data. The impact of the request will fall only on those accessing the portal. The request will be maintained continuously on the website.

We anticipate that the majority of those accessing the site will be private sector entities (such as information technology developers, health plans, and applications developers) and members of the public (including academia, researchers, and public advocacy groups).

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

There are no special circumstances regarding 5 CFR 1320.5

8. Comments in Response to the Federal Register Notice.

The 60-day notice was filed in the Federal Register on Wednesday, November 17, 2010 volume 75, page 70264-5, FR Doc. 2010-28925. No comments were received in response to this submission. Prior to and since this posting, Health and Human Services has discussed the concept of this request in many public meetings including state and local regional governments, industry conferences, and health care meetings.

9. Explanation of any Payment/Gift to Respondents

There is no payment of any kind to parties offering information.

10. Assurance of Confidentiality Provided to Respondents

The request for information through healthdata.gov is intended to acquire information to be made public. The web portal will include language that makes it clear to the submitter that the information that they are providing is intended for the public to see. The purpose

of commenting and providing information is to make it publicly available. Therefore, no assurance of confidentiality is to be provided.

11. Justification for Sensitive Questions

There are no sensitive questions posed for this information request.

12. Estimates of Annualized Hour and Cost Burden

A. Estimated Annualized Burden Table

<b>Forms</b>	<b>Type of Respondent</b>	<b>Number of Respondents</b>	<b>Number of Responses per Respondent</b>	<b>Average Burden (in hours) per Response</b>	<b>Total Burden Hours</b>
One (computer webpage with 6 subpages)	data specialist /technologist from public (government, individuals, private organizations), or private entities, (research organizations)	40	1	15/60	10

A. Estimates of annualized cost to respondents for the hour burdens for collections of information, identifying, and using appropriate wage categories.

<b>Type of Respondent</b>	<b>Total Burden Hours</b>	<b>Hourly Wage Rate</b>	<b>Total Respondent Costs</b>
data/technology specialist	40	\$75.00	\$3000

13. Estimates of other Total Annual Costs Burden to Respondents or Recordkeepers/Capital Costs

There are no annual costs or additional burdens to respondents.

14. Annualized Cost to Federal Government

There is no anticipated cost to the government for the collection of the information. The public provides the information through a nomination platform provided by the healthdata.gov web portal. The maintenance of the portal is estimated at \$2,500 per year.

15. Explanation for Program Changes or Adjustments

This is a new data collection request. There are no program changes or burdens created as a result of requesting this information.

16. Plans for Tabulation and Publication and Project Time Schedule

The information submitted through the webportal by the nominator will be assimilated by the system administrator for review. Once the information submitted has been approved and reviewed, and descriptive information about the resource will be published and a link established to the resource from the non-federal dataset and applications resource site on the healthdata.gov webportal. The timeline for this project is aimed at aligning this functionality to healthdata.gov that was released in February 2011. Rapid processing of this request will facilitate the accommodation of user input that is a highly desired attribute of the porta.

17. Reason Display of OMB Expiration Date is Inappropriate

The expiration date will be posted.

18. Exceptions to Certification for Paperwork Reduction Act Submissions

There are no exceptions to the certification.

**Part B. Statistical Methodology**

This ICR pertains to information about datasets and applications and not the collection performance of studies or surveys. Therefore, there are no statistical or methodological issues that are directly applicable.