

A. Justification

1. Necessity of Information Collection

The Data.gov website is inspired by the President's program for "Open Government" and "Transparency". In response to the President's direction to improve the transparency of government, the Federal Chief Information Officer (CIO) Council is working to create a website/portal that will improve public access to a wide variety of U.S. Government data. Data.gov is a public friendly website that provides descriptions of the federal datasets, information on how to access the datasets, contact mechanisms, metadata information, and links to publicly accessible applications that leverage the datasets.

This emergency information collection request is being submitted in order to fulfill the public engagement and feedback aspects of this important initiative. The end users will be provided opportunities to provide information feedback and ratings in the spirit of the President's open government and transparency agenda. This emergency information collection request will enable these capabilities to be included on Data.gov for the planned initial launch around the third week of May.

**What will data.gov do?** Data.gov will create a portal for access, download, and use of free data including databases, standardized data feeds, maps, graphics, and other data visualizations generated and held throughout the Government that the public, including researchers and industry, can use. The website will permit users to download data to produce analyses and reports in ways that satisfy their needs. The public from across this country can access, view, and use these federal datasets for business, research, analysis, and their own pursuits of knowledge. It will also permit informed users to combine datasets in new ways that answer important questions without doing new primary data gathering. The target audiences include the general public, data analysts, data aggregators, and web developers.

**Phased Launch and Need for Emergency ICR and Subsequent Generic ICR.**

Data.gov is being launched in phases, and will rely on feedback from government, industry, academia and the general public to make the site better as time goes on, with the goal of becoming a central place for easily finding, accessing, and downloading Government data that are not sensitive. The first phase of the release will feature data from Executive Branch agencies, including administrative, research, and statistical data.

The CIO Council began this endeavor in late February and has established a planned timeline for the delivery of Data.gov by May 21, 2009. The emergency information collection request is required to meet the planned timeline of the initial launch of the Data.gov website on May 21, 2009. As the launch of the website is phased, additional feedback mechanisms will be deployed. Therefore, a generic ICR will be submitted to the Office of Management and Budget to replace this emergency ICR before it expires.

The three feedback mechanisms addressed in this emergency information collection request to accommodate the features of the initial launch of Data.gov are: (1) a five-star rating system to rate the datasets on the Data.gov site, including a text box, “Comment on this Data,” (2) a “Suggest Other Data Sets” entry page for the public to submit ideas for additional datasets with an optional contact email address, and (3) a “Contact Us” data entry page for comments or questions with an optional contact email address.

## 2. Needs and Uses

The ratings information collected from the public will help the General Services Administration, the Office of Management and Budget, and the Data.gov Project Team to understand how the datasets presented on Data.gov are valued by respondents. It is the intention of the Data.gov team to leverage the rating responses to provide an average response back to the public through the use of stars on the catalog page. The information collected will not be characterized as ‘representative’ of either the initial visitors to the site or the general public.

**Five-Star Rating System.** The General Services Administration plans to include a rating capability within the Data.gov website to allow the public to rate datasets that are provided by the federal government on Data.gov. A short text box will also be included on the 5-star rating page to allow a visitor to “Comment on this Data.” The rating capability will allow for a more open government to public interaction in alignment with the President’s memorandum titled “Transparency and Open Government”, dated January 21, 2009.

The rating system seeks to give visitors information about which datasets other visitors found most useful and interesting. The ratings provided by users would solicit the users’ reactions to the datasets in four specific categories described below on a scale of 1 to 5. A rating of 1 star will indicate a poor experience with that dataset while a ranking of 5 stars will indicate an excellent experience. Site visitors will be able to choose to view the top datasets ranked by other users along each set of criteria. For example, a visitor will be able to view the top five highest rated catalog entries for Usefulness of Metadata and Instructions or the top five highest rated catalog entries overall. The rankings will not be used to determine whether a ‘poorly rated’ dataset should be taken off Data.gov; however, it might be used to determine whether additional documentation or quality assurance is required. Similarly, a high rating for a particular dataset would not be the sole basis for adding similar datasets, though it might be a trigger to investigate the feasibility of adding datasets with the desired characteristics (e.g., the metadata or data quality).

1. Overall Ranking (1-5 stars): How was the overall experience of selecting a dataset, learning about it via metadata and instructions, downloading the file, and using the data?
2. Data Utility (1-5 stars): Does the dataset provide useful information? Does the user

find the data to be in a practical and convenient structure/format?

3. Usefulness of Metadata and Instructions (1-5 stars): Did the metadata answer the questions the user had about the dataset? Were the instructions for using the dataset clear?

4. Ease of Access (1-5 stars): This ranking assesses the data user's experience from choosing the dataset from the catalog to actually having a data file in hand. Was the data user satisfied with the access speed of the dataset? If the user was redirected to another site to download a dataset, was it easy to find the dataset and download it from that site?

**Suggest Other Datasets.** The information collected regarding dataset suggestions will not be used for statistical analysis or require formal sampling but the information itself will be routed to agencies for their consideration. The number of times a given dataset is suggested, however, will not be the sole basis for either a) determining what additional datasets are put on the site or b) whether a suggested dataset should be considered. Rather, Data.gov hopes to maximize the number of datasets exposed through Data.gov, considering such factors as the sensitivity of the data requested. To that end, agencies will review the comments and potentially act on comments to better expose high quality datasets that are not sensitive through Data.gov.

**Contact Us.** The Contact Us data entry page will be used to gather the public comments and questions that will be provided to the appropriate agency(ies). The email address field will be optional so that the end user can submit comments without an email address. Data.gov will not disclose publicly email addresses that are provided. The Data.gov team may, for example, provide compilations of comments without disclosing email addresses.

### 3. Use of Improved Technology to Minimize Burden

To minimize the respondent burden, the purely voluntary five-star rating is designed to enable an intuitive, simple, and speedy evaluation of the datasets through selection of one through five stars in the four clear-cut categories described above. Explanations of the four evaluation categories are also provided to afford a degree of comparability of responses, and thus support aggregation. The text box to Comment on this Data enables a short, optional text input.

To limit the public burden when using the optional Suggest Other Datasets page, the data entry fields are brief and make use of selection boxes to minimize the amount of text required for submission.

To minimize the public burden to use the Contact Us page, the page contains an open-ended text box with an optional email address field.

4. Efforts to Identify Duplication

Ratings questions relate specifically to the Data.gov datasets and do not duplicate what other agencies are doing.

The Suggest Other Datasets entry page is a new capability that does not exist within any other inter-agency dataset portals such as USA.gov.

The Contact Us data entry page is not requested anywhere else within the Data.gov domain and is optional for completion.

The rating categories, the Comment on this Data text box, the Suggest Other Datasets entry page, and the Contact Us entry page and information are not redundant and include only requests for information relevant to the specific datasets on Data.gov.

5. Minimizing Burden on Small Businesses

Small businesses are not specific targets of the Data.gov rating efforts or information collection efforts. However, if they leverage Data.gov they will be presented with the opportunity to provide dataset ratings, comment on specific datasets at the Comment on this Data, nominate datasets through the Suggest Other Datasets page, and provide comments through the Contact Us page.

6. Consequences of Less Frequent Collection

This clearance involves establishing a feedback mechanism to rate datasets on Data.gov, the ability for the public to suggest additional datasets, and the ability for the public to provide feedback. The voluntary responses to these information collection opportunities will determine the frequency of the information collection.

7. Consistency with OMB Guidelines

The General Services Administration will collect the information in a manner that complies with OMB guidelines. GSA requests a waiver to the Guidelines in 5 CFR 1320.5(d)(2) of the Emergency ICR FR Notice due to the short project timeline and imminent launch date of May 21, 2009.

8. Consultation Outside the Agency

The General Services Administration understands from the Office of Management and Budget that a ratings approach is desired, within scope of the Data.gov project, and is considered to be in alignment with the Administration's transparency and open

government memorandum. The General Services Administration has also conferred with the Office of Management and Budget to determine that the Comment on this Data text box, Contact Us page, and the fields on the Suggest Other Datasets page are all desired, within scope of Data.gov, and will provide value to the public. The Data.gov project team also includes representatives from the Department of the Interior, Department of Commerce, and the Environmental Protection Agency.

9. Explanation of Decision to Provide any Payment or Gift to Respondents

No payments or gifts will be made to respondents under this Emergency ICR.

10. Describe any assurance of confidentiality provided to respondents and the basis for the assurance in statute, regulation, or agency policy

While visitors to the Data.gov website are not assured of confidentiality, visitors are assured that identification is not necessary to use the feedback mechanisms. In addition to “optional” labeling on the “Contact Us” and “Suggest Other Datasets” feedback pages, a Privacy Policy will be prominently posted. The Privacy Policy (1) advises that no personally identifiable information is required, tracked, or stored through the Data.gov information collection capabilities, and (2) discloses how any personally identifiable information they may voluntarily provide will be handled. If a visitor provides an email address or other personal information through simple text boxes; i.e., the Comment on this Data, Suggest a Dataset, and/or Contact Us, the content will be routed to [datagov@gsa.gov](mailto:datagov@gsa.gov) at a federally managed email system when the user clicks “submit.” Suggestions, Comments, and Questions will be sorted once received at [datagov@gsa.gov](mailto:datagov@gsa.gov) through the GSA e-mail system and forwarded to the agency dataset owner relevant to the comment or to the Data.gov project management office for action, as appropriate. A set of FAQs will also be posted to answer general questions. Misdirected e-mail will be forwarded to USA.gov.

11. Justification for Sensitive Questions

The feedback mechanisms provided at Data.gov do not ask questions or collect data of a sensitive nature.

12. Estimated Burden of Information Collection

For the five-star ratings page, it is estimated that 10,000 ratings (responses) will be submitted each year. Completing the ratings for four Data.gov Launch categories and optional use of the Comment on this Data text box is estimated to take 30 seconds; therefore, it is anticipated that 5,000 minutes per year will be required to submit the dataset ratings. 84 annual burden hours are requested based on these estimations. Based upon similar ratings methods used in government solutions, such as USA Services, we believe that this is an accurate estimate.

For the Suggest Other Datasets page, it is estimated that 1,000 suggestions (responses) will be submitted each year. It is estimated that completing the suggestion data entry page will take 5 minutes; therefore, it is anticipated that 5,000 minutes per year will be required to submit the Suggest Other Datasets data entry page. 84 annual burden hours are requested based on these estimations. Based upon similar data entry page methods, we believe that this is an accurate estimate.

For the Contact Us data entry page, it is estimated that 5,000 contacts (responses) will be submitted each year. It is estimated that completing the Contact Us data entry page will take 30 seconds; therefore, it is anticipated that 2500 minutes per year will be required to submit the contact data entry page. 42 annual burden hours are requested based on these estimations. Based upon similar Contact Us data entry page methods, we believe that this is an accurate estimate.

Therefore the total burden for this ICR is 210 hours (84 hours for ratings, 84 hours for Suggest Other Datasets data entry page, and 42 hours for Contact Us data entry page). The total number of respondents is 16,000, and the total number of responses is 16,000.

This Emergency ICR will be succeeded by a Request for Generic Clearance for these and additional feedback mechanisms to be employed during the phased Data.gov deployment, such as:

- a. pages for visitors to advise how they were able to leverage the datasets in new and different ways to build applications, conduct analysis, and perform research; and
- b. pages for visitors to rate the benefit of the reported new solutions.

13. Estimate of Cost Burden

We do not expect respondents to incur any costs other than that of their time expended. The information requested is of the type and scope normally provided online and without additional research time required. Therefore, respondents are not expected to incur any costs or burden for responding to this voluntary rating procedure.

14. Cost to Federal Government

Electronic feedback mechanisms are virtually cost-free as part of the overall website. The General Services Administration is incurring no additional costs for hosting the information collection pages and will incur no additional costs for tabulating and displaying the responses to the ratings.

Suggestions, comments, or questions submitted through the text boxes on the “Contact Us” and/or “Suggest a Dataset” pages will be routed to [datagov@gsa.gov](mailto:datagov@gsa.gov). Initially, they will be manually sorted by the Data.gov project management office staff once received at [datagov@gsa.gov](mailto:datagov@gsa.gov) through the GSA e-mail system. Those e-mails that are relevant to

datasets will be forwarded to agency dataset owners for consideration in providing additional datasets. The Data.gov project management office will use the general feedback not specific to a certain dataset to improve the website. The project management office will redirect misdirected e-mail to USA.gov. The types of feedback and their resolution will be tabulated by the project management office while the Emergency ICR is in place to explore contract vehicles to automate the binning and responses. There is no cost to the initial phase of feedback because existing FTE and contractor resources are being used to develop and launch the website.

15. Reason for Change in Burden

This is a new collection.

16. Plans for Tabulation, Statistical Analysis and Publication (Project Schedule)

GSA will collect rating numbers to track the utility of the datasets and to improve the value of the site. Aggregation will be simple coding to receive, tabulate, and display the 5-star ratings chosen to other visitors as explained in our response to Question 2 above. The results will not be characterized as representative of the user population, as many users may elect not to provide ratings. The General Services Administration plans to release the first version of Data.gov to the public in May, 2009.

17. Request to Not Display Expiration Date

GSA plans to display the OMB Control Number at Data.gov but requests permission to waive display of the expiration date to mitigate the risk of the public's misunderstanding of its meaning. A misunderstood expiration date may be detrimental to the President's vision of an open and transparent government. Acknowledging that most citizens may not be acquainted with the PRA and the reason for an expiration date, potential participants might be discouraged from participating if the experience is perceived to have an impending "expiration" date, especially a six-month expiration date provided for this Emergency ICR.

18. Exceptions to the Certification

No exception to the certification statement is being requested.