

**Supporting Statement for Paperwork Reduction Act Submission
3090-0284 – Data.gov Feedback Mechanisms**

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 created a website/portal to 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 information collection request for a three-year generic clearance with an annual budget of 400 burden hours and 16,200 responses 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.

What will data.gov do? Data.gov is 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 permits 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 also permits 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 Generic ICR. Data.gov is being launched in phases and relies 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 initial phase of the release features data from Executive Branch agencies, including administrative, research, and statistical data.

The CIO Council began this endeavor in late February and established a planned timeline for the delivery of Data.gov by May 21, 2009. GSA sought and obtained approval of an emergency information collection request to meet the planned timeline of the initial launch of the Data.gov.

A few feedback mechanisms were addressed in the 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, (2) including a text box, “Comment on this Data,” (3) a “Suggest Other Data Sets” entry page for the public to submit ideas for additional datasets with an optional contact email address, and (4) a “Contact Us” data entry page for comments or questions with an optional contact email address.

Additional feedback mechanisms will be deployed as the website matures and as visitor feedback and stakeholder needs dictate. Therefore, this generic ICR is submitted to the Office of Management and Budget to replace the emergency ICR and to seek a burden hour budget for additional feedback mechanisms to be submitted as they are planned and defined.

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 team to understand how the data presented on Data.gov is 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. A rating capability is included in 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 is also included on the 5-star rating page to allow a visitor to “Comment on this Data.” The rating capability allows 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 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 indicates a poor experience with that dataset while a ranking of 5 stars indicates an excellent experience. Site visitors are able to choose to view the top datasets ranked by other users along each set of criteria. For example, a visitor is 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 are not 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 is not 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).

a. 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?

b. 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?

c. 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?

d. 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 is not and will not be used for statistical analysis nor does it or will it require formal sampling but the information itself is routed to agencies for their consideration. The number of times a given dataset is suggested, however, is not to be the sole basis for either a) determining what additional data is 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 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 is used to gather the public comments and questions that are provided to the appropriate agency or agencies. The email address field is optional so that the end user can submit comments without an email address. Data.gov does not disclose publicly email addresses that are provided. The Data.gov team may, for example, provide compilations of comments without disclosing email addresses.

A Social Media Venue. A url will be posted at the Data.gov site to guide visitors to a social media site to enable conversation with the public throughout the evolution of Data.gov; e.g., posting draft documents for the public's review, obtaining feedback to posted questions, obtaining suggestions in general. Initially, documents will be posted for review only; then for comment, and ranking of comments. Eventually, this site may be used to receive the public's suggested datasets and other recommendations for Data.gov instead of the current menus at Data.gov or in addition to those menus.

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.

Minimization of public burden will be taken into consideration when selecting the social media venue.

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 offerings of 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 and social media venue.

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 collects the information in a manner that complies

with OMB guidelines.

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, social media url, and the fields on the Suggest Other Datasets page are all desired, within scope of Data.gov, and 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 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 is 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 is routed to datagov@gsa.gov at a federally managed email system when the user clicks "submit." Suggestions, Comments, and Questions are sorted once received at 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 are posted to answer general questions. Misdirected e-mail is 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

This ICR for a generic clearance includes clearance of the existing feedback mechanisms approved via the Emergency ICR and additional yet-to-be implemented feedback

mechanisms.

IC's are included in this package for the following feedback mechanisms.

For the five-star ratings, it is estimated that 5,000 ratings (responses) may be submitted each year. Each of the 5-star ratings is estimated to take 7.5 seconds per category, therefore, it is estimated that 625 minutes (10.5 burden hours) per year may be expended to submit the data set rating for each category provided. Initially, Data.gov offers four categories to rate at an estimated burden of 42 hours. It is estimated that 3 categories may be added each year, up to 10 categories; i.e., 42 hours in year 1; potentially 74 hours in year 2; and potentially 105 in year 3, which averages (266,400 seconds/4,440 minutes) 73 estimated burden hours per year (5,000 ratings X 7 categories). Therefore, 73 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 Comment on this Data text box on the 5-star rating page, from its May 21 launch through August 26, 2009, (a 3-month period), Data.gov received 110 comments on datasets. Therefore, it is estimated that 400 visitors (responses) may choose to use the text box instead of the 5-star rating system or use the text box in addition to completing the 5-star rating. Each Comment on this Data is estimated to take a minute, therefore, it is estimated that 400 minutes (approximately 7 burden hours) per year may be expended to Comment on this Data. Therefore, 7 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, from its May 21 launch through August 26, 2009 (a 3-month period), Data.gov received 931 dataset suggestions. Therefore, it is estimated that 4,000 nominations (responses) may be submitted each year. It is estimated that completing the Suggest Other Datasets entry page may take 2 minutes; therefore, it is estimated that 8,000 minutes (134 burden hours) per year may be expended to submit the nomination data entry page. Therefore, 134 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, from its May 21 launch through August 26 (a 3-month period), Data.gov received 446 contact us e-mails. Therefore, it is estimated that 1,800 contacts (responses) may be submitted each year. It is estimated that completing the Contact Us data entry page may take a minute; therefore, it is estimated that 1,800 minutes (30 burden hours) per year may be expended to Contact Us. Therefore, 30 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.

A Social Media Venue, a url will be posted at the Data.gov site to guide visitors to a social media site to enable conversation with the public throughout the evolution of Data.gov; e.g., posting draft documents for the public's review, obtaining feedback to posted questions, obtaining suggestions in general comments. It is estimated that 1,000 respondents (one

response each) may participate each year in the social media functionality to respond to posted documents. It is estimated that participating may take 2 minutes; therefore, it is estimated that 2,000 minutes (34 burden hours) per year may be expended on a Data.gov social media site. For visitors to view and rank the feedback of other visitors' comments, suggested datasets, recommendations, etc., through voting up or down using an arrow icon, it is estimated that approximately 1,000 votes (responses) may be submitted. Each vote may take approximately a minute; therefore, 1,000 minutes (16 burden hours) may be expended per year. Therefore, the combined IC for social media commenting and voting capability represents 1,000 responses at 3 minutes each for a total estimated burden of 50 hours.

Potential feedback mechanism ICs under consideration to be employed over the next few years include:

- Venues for visitors to advise how they leverage the datasets in new and different ways to build applications, conduct analysis, and perform research at an estimated annual burden of 83 hours;
- Venues for visitors to rate the benefit of the reported new solutions at an estimated annual burden of 17 hours; and

Therefore, the total burden budget requested for existing and future feedback mechanisms under this generic clearance is 400 hours and 16,200 responses.

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 part of the overall Data.gov project. The General Services Administration is incurring no additional costs for hosting the feedback mechanisms.

Suggestions, comments, or questions submitted through the text boxes on the "Contact Us" and/or "Suggest a Dataset" pages are routed to datagov@gsa.gov. Initially, they are manually sorted by the Data.gov project management office staff once received at datagov@gsa.gov through the GSA e-mail system. Those e-mails that are relevant to datasets are forwarded to agency dataset points of contact for consideration in providing additional datasets. The Data.gov project management office uses the general feedback not specific to a certain dataset to improve the website. The project management office

redirects misdirected e-mail to USA.gov. The types of feedback and their resolution has and will continue to be tabulated by the project management office and is exploring automation of the binning and responses. Because existing FTE and contractor resources are used, there is no additional cost to the feedback processing.

15. Reason for Change in Burden

This is a request for a generic clearance to replace the Emergency ICR provided for the initial feedback mechanisms. Additional hours are requested for a budget to cover the existing feedback mechanisms and potential feedback mechanisms to be implemented as Data.gov matures and as its visitors' and stakeholders' needs dictate.

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

GSA collects rating numbers to track the utility of the datasets and to improve the value of the site. Aggregation is simple coding to receive, calculate, and display the 5-star ratings chosen to other visitors as explained in our response to Question 2 above. The results are not to be characterized as representative of the user population, as many users may elect not to provide ratings.

17. Request to Not Display Expiration Date

GSA displays the OMB Control Number at Data.gov and continues to request 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.