U.S. General Services Administration

PRIVACY IMPACT ASSESSMENT

Data.gov

May 2011

Prepared by:
Office of Citizen Services and Innovative Technologies
General Services Administration

PART II. SYSTEM ASSESSMENT

A. Data in the System

Question	Explanation/Instructions
1. Describe all information to be included in the system, including personal data.	Data.gov is a website providing a catalog of datasets from agencies across the federal government. Users can find detailed metadata about each dataset and the URL taking them to the originating agency's website. Data.gov also has a platform using a software-as-a-service (SaaS) solution for data hosting, so that agencies' data can be viewed interactively on Data.gov. There are also "Communities" on Data.gov with content, forums and other features centered on a specific topic, such as Health, Law and Semantic Web. Use of the main features of Data.gov, to discover, view, and visualize data do not require a sign in or any other information from the user. For certain features, such as joining a community, submitting comments and questions, saving a user-created view of a dataset, etc., a sign in with an email address and user name is required. Email addresses are not made public, and users can pick any user name. Users have the option of uploading an image associated with their profile, which may contain a photograph of themselves, or any other content. Users have the option of associating additional profile information, such as description, location, tags, links to home pages and other social profiles, etc.
1.a. What stage of the life cycle is the system currently in?	Operation/Maintenance
2.a. What are the sources of the information in the system?	Information is provided by the user seeking access to the additional features of the site.
2.b. What GSA files and databases are used?	None.
2.c. What Federal agencies are providing data for use in the system?	None.
2.d. What State and local agencies are providing data for use in the system?	None.

2.e. What other third party sources will the data be collected from?	None.
2.f. What information will be collected from the individual whose record is in the system?	Email address, username.
3.a. How will the data collected from sources other than Federal agency records or the individual be verified for accuracy?	Users of the site will be responsible for accurately submitting their information. Email addresses will be verified for complete format only. Before access to the additional features is granted, a return email will be sent to verify that the email address is a functioning address.
3.b. How will data be checked for completeness?	Emails addresses will be verified for complete format only. Before access to the additional features is granted, a return email will be sent to verify that the email address is a functioning email address. This is a spam reduction measure.
3.c. Is the data current? How do you know?	The features on Data.gov requiring a log in are made available for voluntary use. Users will be responsible for accurately submitting an email address. Before access to the additional features is granted, a return email address will be sent to verify that the email address is a functioning email address. This is a spam reduction measure.
4. Are the data elements described in detail and documented? If yes, what is the name of the document?	The privacy statement will display messaging confirming to the user the limited use of the email for authentication purposes. The privacy policy will also indicate that a log in is not required for most uses of the site. The policy will state that, when making comments, asking questions, saving views, etc., the registrant's username will be displayed publicly by default. The policy will indicate that providing a profile image is optional. In addition, the policy will clearly state that this is not a Privacy Act System of Record (see Question 1.)

B. Access to the Data

Question	Explanation/Instructions
1. a. Who will have access to the data in the system?	All users and the general public will have access to comments and saved views, associated with the user name, as this is a public platform intended for transparent uses. Data.gov administrators will moderate comments and other content created by users before making them publicly available. Data.gov administrators can delete inappropriate comments or content.
1.b. Is any of the data subject to exclusion from disclosure under the Freedom of Information Act (FOIA)? If yes, explain the policy and rationale supporting this decision.	Email addresses, user names and profile images are excluded from disclosure under FOIA.
2. How is access to the data by a user determined? Are criteria, procedures, controls, and responsibilities regarding access documented?	Access to data on user names and email addresses is limited to Content Administrators (GSA) and System Administrators/Developers of Code (Vendor).
3. Will users have access to all data in the system or will the user's access be restricted? Explain.	**Individual Access**: The only data collected is voluntary. The username (which does not have to be a real name) is visible to the public, and this is highlighted in the privacy policy provided to the public on the website. Where a username is displayed, clicking it may display that user's profile including the optional profile image. **Content Administrators (GSA) and System Administrators/Developers of Code (Vendors)** have access to data as appropriate to fulfill their roles, within the conditions spelled out in this PIA and the site's privacy policy. As content administrators, the government does not own the database in which registrants' information will reside. **Vendors**' access to the system is controlled by the vendors and is dictated by duties and requirements of their positions and by the terms of the service agreement. **Citizens** using the platform will have access only to information that is made public.
4. What controls are in place to	GSA and all Data.gov vendors are operating under the

prevent the misuse (e.g. browsing) of data by those having access?	same rules of behavior described in the Data.gov privacy policy in terms of protecting the privacy of others and not using information in the system for personal gain or to the benefit of others. Passwords and segmentation of functions provide adequate protections.
5.a. Do other systems share data or have access to data in this system? If yes, explain.	No other system has access to this data, other than Google Analytics, which has access only to the anonymized, aggregate analytics data which it retrieves from cookies. Other systems will also be able to pull data from this system via an API, but that API will only feed data that is already publicly available (i.e., no PII).
5.b. Who will be responsible for protecting the privacy rights of the clients and employees affected by the interface?	Not applicable.
6.a. Will other agencies share data or have access to data in this system (International, Federal, State, Local, Other)?	No.
6.b. How will the data be used by the agency?	The information will not be used by the Agency. The vendor providing the Data.gov platform uses email addresses only as a means of authenticating a user of the platform.
6.c. Who is responsible for assuring proper use of the data?	The Program Manager.
6.d. How will the system ensure that agencies only get the information they are entitled to?	Not applicable. Email addresses are used only for authentication purposes, except where access is explicitly authorized, required for law enforcement reasons, or mandated by statute.
7. What is the life expectancy of the data?	The Data.gov platform will retain users' registration details permanently, unless the user deletes their account, which they may do at any time via the platform's online interface or by email to a support address. This is a standard feature of websites in which the creation of a unique, persistent user account is a requirement for participation.

	(In this case, it is a requirement only for certain optional uses of the site.)
8. How will the data be disposed of when it is no longer needed?	The Data.gov platform will retain users' registration details permanently, unless the user deletes their account, which they may do at any time via the platform's online interface or by email to a support address. This is a standard feature of websites in which the creation of a unique, persistent user account is a requirement for participation. (In this case, it is a requirement only for certain optional uses of the site.)

C. Attributes of the Data

Question	Explanation/Instructions
1. Is the use of the data both relevant and necessary to the purpose for which the system is being designed?	For the optional uses of the website, yes. Email addresses are necessary to authenticate a user of the platform, and to allow users to return and modify their comments, ideas, and saved views. Upon initial log-in, an email is sent back to the email address to confirm the address is working before allowing participation in the optional uses of the site. This is a spam prevention measure.
2.a. Will the system derive new data or create previously unavailable data about an individual through aggregation from the information collected?	No.
2.b. Will the new data be placed in the individual's record (client or employee)?	No.
2.c. Can the system make determinations about individuals that would not be possible without the new data?	No.
2.d. How will the new data be verified for relevance and accuracy?	Email addresses will be verified for complete format only.
3.a. If the data is being consolidated, what controls are in place to protect the data and prevent unauthorized access? Explain.	N/A
3.b. If processes are being consolidated, are the proper controls remaining in place to protect the data and prevent unauthorized access? Explain.	N/A

4. How will the data be retrieved? Can it be retrieved by personal identifier? If yes, explain.	Data containing email addresses submitted by users cannot be retrieved by the government. The system will retrieve registrants' records by username, which is a manufactured alias and does not constitute PII. Administrators can search by email address in order to grant user roles, but access to that search is limited only to Administrators and only on that page. It is not externally accessible.
5. What are the potential effects on the privacy rights of individuals of: a. Consolidation and linkage of files and systems; b. Derivation of data; c. Accelerated information processing and decision making; and d. Use of new technologies. How are the effects to be mitigated?	There are no known effects on the due process rights of individuals who avail themselves of the Data.gov platform. The system with the user email addresses and usernames is not linked to other files and systems. Participants will be presented with a clear disclaimer in the Privacy Policy that any submissions of email addresses are voluntary and that this is not a Privacy Act System of Record.

D. Maintenance of Administrative Controls

Question	Explanation/Instructions
1.a. Explain how the system and its use will ensure equitable treatment of individuals.	There are no known effects on the equitable treatment of individuals who avail themselves of the optional features on the Data.gov site. The system with usernames and email addresses is not linked to other files and systems.
1.b. If the system is operated in more than one site, how will consistent use of the system be maintained at all sites?	N/A
1.c. Explain any possibility of disparate treatment of individuals or groups.	There is no possibility of disparate treatment of individuals or groups.
2.a. What are the retention periods of data in this system?	The Data.gov platform will retain users' registration details permanently, unless the user deletes their account, which they may do at any time via the platform's online interface or email to a support address. This is a standard feature of websites in which the creation of a unique, persistent user account is a requirement for participation. (In this case, it is a requirement only for certain optional uses of the site.)
2.b. What are the procedures for eliminating the data at the end of the retention period? Where are the procedures documented?	Registration data will not be eliminated except voluntarily by the user who originally submitted it, which they may do at any time via the platform's online interface or email to a support address. This is a standard feature of websites in which the creation of a unique, persistent user account is a requirement for participation. Participation data will not be eliminated.
2.c. While the data is retained in the system, what are the requirements for determining if the data is still sufficiently accurate, relevant, timely, and complete to ensure fairness in making determinations?	The individual will be responsible for ensuring that the information is complete and accurate when they first use the platform via authentication of a valid email address. Thereafter, they may make changes to their user profile via the platform's online interface.

3.a. Is the system using technologies in ways that Federal agencies have not previously employed (e.g. Caller-ID)?	No.
3.b. How does the use of this technology affect individuals' privacy?	No effect on individual privacy. The only impact is the storage of an email address and username. The email address is submitted voluntarily for authentication purposes only, and after being provided with appropriate notices on the website. In addition, the persistent cookies used by Data.gov do not collect or store any PII, nor can they be used to track individual users' activities across other websites. Google Analytics provides reporting on "referring sites" but this data is only recorded anonymously and reported in the aggregate. Some cookies may be stored for up to two years.
4.a. Will this system provide the capability to identify, locate, and monitor individuals? If yes, explain.	No.
4.b. Will this system provide the capability to identify, locate, and monitor groups of people? If yes, explain.	No.
4.c. What controls will be used to prevent unauthorized monitoring?	The information that would be required for such monitoring is never solicited or entered into the system.
5.a. Under which Privacy Act System of Records notice (SOR) does the system operate? Provide number and name.	The Data.gov platform is not a Privacy Act System of Record.
5.b. If the system is being modified, will the SOR require amendment or revision? Explain.	N/A
