

Formative Research to Develop an Open Data Standard for Reporting Environmental Health Data

OMB Control No. 0920-1154

Supporting Statement Part A –
Justification

Brian Hubbard, MPH
Health Scientist
National Center for Environmental Health
Division of Environmental Health Science and Practice
Water, Food, and Environmental Health Services Branch
Phone: (770) 488-7098
Email: bnh5@cdc.gov
Fax: (770) 488-7310

Date: May 23, 2019

Table of Contents

A.1. Circumstances Making the Collection of Information Necessary	5
A.2. Purpose and Use of the Information Collection	6
A.3. Use of Improved Information Technology and Burden Reduction	7
A.4. Efforts to Identify Duplication and Use of Similar Information	8
A.5. Impact on Small Businesses or Other Small Entities	8
A.6. Consequences of Collecting the Information Less Frequently	9
A.7. Special Circumstances Relating to the Guidelines of 5 CFR 1320.5	9
A.8. Comments in Response to the Federal Register Notice and Efforts to Consult Outside the Agency	9
A.9. Explanation of Any Payment or Gift to Respondents	9
A.10. Protection of the Privacy and Confidentiality of Information Provided by Respondents	10
A.11. Institutional Review Board (IRB) and Justification for Sensitive Questions	10
A.12. Estimates of Annualized Burden Hours and Costs	10
A.13. Estimates of Other Total Annual Cost Burden to Respondents and Record Keepers	11
A.14. Annualized Cost to the Federal Government	11
A.15. Explanation for Program Changes or Adjustments	11
A.16. Plans for Tabulation and Publication and Project Time Schedule	11
A.17. Reason(s) Display of OMB Expiration Date is Inappropriate	12
A.18. Exceptions to Certification for Paperwork Reduction Act Submissions	12
References	12
List of Attachments	13

Part A. Justification

Goal of the study: The goal of this information collection request is to conduct formative research to identify barriers and strengths associated with making aquatic facilities' environmental health data usable and more easily accessible to the public.

Intended use of the resulting data: The information collected through this GenIC will inform the development of an open data standard for existing, taxpayer funded environmental health data, including the standard's content, implementation, and adoption.

Methods to be used to collect: Individual in-depth interviews.

Subpopulation to be studied: Respondents include open data standards experts, municipal government information technology administrators, environmental health personnel, and technology providers.

How data will be analyzed: Descriptive analyses and thematic or grounded theory analysis of qualitative data.

A.1. Circumstances Making the Collection of Information Necessary

The purpose of this new GenIC is to conduct up to 43 key informant interviews to collect information that will support decision-making for sharing environmental health inspection data through an open data standard. The European Commission website describes open data as accessible data that anyone can use and share in any way they want (European Commission 2019). Open data has been used to report restaurant inspection scores in conjunction with for-profit entities (e.g., Yelp) and in other governmental settings (e.g., Louisville, Kentucky and San Francisco, California) to influence the health decision making of the public^{1,2,3}. Open data portals and open data standards have not been widely used as a method for publicly sharing environmental health inspection data. Currently, most environmental health inspection data

¹ O'Reilly T. Open Data and Algorithmic Regulation. In: Goldstein B, Dyson L, ed. *Beyond Transparency: Open Data and the Future of Civic Innovation*. San Francisco: Code for America Press; 2013:

https://s3.amazonaws.com/academia.edu.documents/34686172/BeyondTransparency.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1553633287&Signature=%2FY97V%2BCmJFV%2F8SxIXJCGC5xGkEw%3D&response-content-disposition=inline%3B%20filename%3DTable_of_Contents_Preface.pdf. Accessed: 03/26/19 p.294

² City of Louisville, KY. Louisville Metro Open Data Portal. <https://data.louisvilleky.gov/>. Accessed: 03/26/19

³ City of San Francisco, CA. DataSF. <https://datasf.org/opendata/>. Accessed: 03/26/19

remains inaccessible after collection due, in part, to significant variability in which the information is formatted. Developing best practices and examples of how environmental public health programs use open data standards has the potential to make inspection data more accessible and usable by the public and researchers, alike.

The CDC has awarded funding to the National Environmental Health Agency (NEHA) (CDC-RFA-OT18-1802) to support the development of an open data standard for aquatic facilities. This project includes conducting individual interviews with open data standards experts, municipal government information technology administrators, environmental health personnel, and technology providers to identify barriers and strengths associated with making facilities' environmental health data usable and more easily accessible to the public. For this reason, the Centers for Disease Control and Prevention (CDC) requests approval for a new GenIC, "Formative Research to Develop an Open Data Standard for Reporting Environmental Health Data" under OMB Control No. 0920-1154.

A.2. Purpose and Use of the Information Collection

The purpose of this information collection is to conduct formative research to inform the development of an environmental health open data standard (Attachment 1). This open data standard will serve as an organizational model for aggregating and publishing environmental health inspection data in a manner that allows free use, distribution, and analysis. The data standard that will be developed is a defined set of variables (i.e., environmental health information) that must be provided by a facility that wishes to publish its data through the open data portal, and for that data to have meaningful utility for external stakeholders.

Information resulting from key informant interviews will potentially reveal how open data experts use other data standards related to aquatic facility inspection data. NEHA's subject matter expert (SME) will conduct up to six key informant interviews with open data experts from both Europe, Canada, and the United States. Key informant information may also provide valuable insight into how similar efforts have succeeded and identify actions that yielded maximum results for other open data efforts. The possibility exists to identify standard terminology for data standards used by offices or agencies of the U.S. government that will make the long-term adoption and success of the standard much more likely.

As a part of these interviews, governments will be asked how much they have been told exporting their data might cost. Typically, they have engaged with their database providers to explore costs. Cost estimates are important considerations when working with extraction, transformation, and load tool providers that enable jurisdictions to share inspection data in a standardized format. NEHA's SME will conduct up to seven key informant interviews with technology providers. Likewise, NEHA will need cost estimates to contract open data portal providers in the event a municipal pilot site would need an open data portal tool. All cost estimates and information about pricing will assist NEHA and CDC in an evaluation process for selecting pilot sites to implement open data platforms and use open data standards.

The NEHA SME will also conduct key informant interviews with up to 20 environmental health personnel (managers and practitioners) that manage jurisdictional inspection database tools. These interviews will identify barriers for sharing data in open data portal tools. Barriers might include

- Restrictions imposed by data vendors that prohibit sharing data in machine-readable formats.
- Frequency and staff hours required to organize inspection data for use in an open data portal.
- Efficiency of tools used to manage and share aquatic facility inspection data.

Interviews will be initiated through direct contact to an individual key informant via email, phone, or Twitter direct message (similar to email), depending on how the key informant is publicly (or through existing contacts) known to be reachable (Attachment 2). One to two days before the scheduled interview, a reminder email will be sent to the respondent (Attachment 3).

CDC works with 25 local jurisdictions conducting surveillance to systematically collect, analyze, and interpret data on the results of routine inspections of public aquatic venues. In the past, environmental health practitioners from these jurisdictions sent their inspection data to CDC annually. The periodicity of data collection is less frequent now because the burden of collecting and organizing the data has grown. This current data collection will inform the development of a new way of accessing, sharing, and using aquatic facility inspection data. Environmental health inspection data made available through open data portals and data standards will not only be available for government researchers, but more importantly, these data have the potential to increase citizen participation in government, create opportunities for economic development, and inform decision making in both the private and public sectors.⁴

CDC's Water, Food, and Environmental Health Services Branch will use the information collected during key informant interviews to describe how environmental health agencies that are considering the use of open data standards can make their inspection data sharable and usable by anyone. CDC and NEHA will share information and lessons gathered from key informant interviews in webinars and practice-based journals to support the practice of environmental public health.

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

⁴ U.S. General Services Administration (GSA). Open Government. <https://www.data.gov/open-gov/>. Accessed 03/19/19

The National Environmental Health Association contracted an open data SME to identify key informants with high value information. The SME will conduct one-on-one interviews by phone, or in person when feasible.

The NEHA SME designed the unstructured interview questions to address information needed from each respondent category and the number of questions are seven or fewer for each respondent category.

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

NEHA's open data SME conducted an ecosystem scan to assess the current open data environment and identify jurisdictions and vendors that are involved in open data reporting. Part of the data collection strategy will be to identify informants that have direct experience in establishing open data standards in other sectors and countries. By defining the open data environment, CDC and NEHA are hoping to build on the strength of past open data efforts at state and county environmental and public health agencies and avoid any barriers that prevent data from being usable and shareable without restrictions.

The ecosystem scan included literature searches to pull together an open standards reading list with guidance for collaborating across government agencies. NEHA conducted an online scan for aquatic facility inspection data. The data searches helped NEHA to develop a list of the most commonly used data and inspection vendors, and a list of 14 jurisdictions that use open data. More importantly, the scan identified six states that post aquatic inspection data through open data portals.

Although the ecosystem scan produced meaningful and immediately useful links to existing open data sources, the scan did not identify some of the most commonly encountered barriers faced by jurisdictions that are trying to make their data open and accessible. In certain instances, data vendors or jurisdictions describe data and an open data source, but data use restrictions prevent the data from meeting the definition of truly open data. For example, data may not be machine-readable, which limits how the public and private sectors can use the data. Duplication of this effort or the existence of similar information was not identified during the ecosystem scan process.

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

The NEHA SME will contact technology vendors that may meet the definition of a small business. However, the key informant interview questions will be limited to fewer than ten

questions and focus on obtaining pricing information and service descriptions because many of these vendors do not publicly list their prices.

The collection of information that could affect small businesses will be limited to at most seven hours of time burden. However, the limited number of questions and focus on cost estimates will likely require less than the total seven hours of project time burden.

A.6. Consequences of Collecting the Information Less Frequently

The NEHA SME will contact key informants for one interview each. This request is for a one-time data collection. If NEHA does not conduct the key informant interviews, CDC and NEHA will not have the information available to make decisions for how to work with public health jurisdictions to develop the open data platform and standard. If the key informant interviews are not conducted, the open data platform would have little chance at meeting the needs of the stakeholders it aims to serve.

There are no technical or legal obstacles to reducing burden.

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

This request complies with the guidelines of 5 CFR 1320.5.

A.8. Comments in Response to the Federal Register Notice and Efforts to Consult Outside the Agency

The *Federal Register* notice was published for this collection on July 18, 2016, Vol. 81, No. 137, pp. 44680. No public comments were received.

CDC project staff are working with NEHA on the study design, screening instruments, and data collection instruments. NEHA has contracted with a SME in open data, Sarah Schacht, to conduct the key information interview.

A.9. Explanation of Any Payment or Gift to Respondents

NEHA will not compensate the respondents.

A.10. Protection of the Privacy and Confidentiality of Information Provided by Respondents

The semi-structured key informant interview instruments do not require respondents to provide identifying information. NEHA knows most of the key informants through their professional network. Information such as name, occupation, and contact information will be used and is publicly listed as they are either a) public employees, or b) employees of well-known and documented civil society organizations who specialize in open data. Finally, NEHA will contact informants directly via email and through Twitter's Direct Message feature (similar to email).

NEHA will collect the key informant information by taking hard copy notes on their discussions. Afterward, the information will be logged into a spreadsheet, de-identified, organized, and will be provided to the CDC in a summary of results. At no time will CDC have access to any key informant interview transcripts.

NEHA will use physical and procedural safeguards to protect the integrity of key informants' information. These safeguards include storing notes and notebooks in locked cabinets in locked rooms, and safeguarding electronic data using password protection and anti-virus software. Only NEHA will have access to the raw data.

A.11. Institutional Review Board (IRB) and Justification for Sensitive Questions

The NCEH/ATSDR Human Subjects Advisor has reviewed this project and determined that it does not meet the definition of research under 45 CFR §46.102(l) (Attachment 4).

No information will be collected that are of personal or sensitive nature.

A.12. Estimates of Annualized Burden Hours and Costs

Burden estimates were calculated using the mean hourly rate of the respondent type from the Bureau of Labor Statistics' May 2017 Occupational Employment Statistics⁵, the number of interviews planned, and a one-hour length of interview. Each interview will occur one time, for a total burden of 43 hours.

Table A.12.1. Estimate of Annualized Burden Hours

Type of Respondents	Form Name	Number of Respondents	Number of Responses	Average Burden	Total Burden
---------------------	-----------	-----------------------	---------------------	----------------	--------------

⁵ United States Bureau of Labor Statistics. May 2017 National Occupational Employment Statistics. https://www.bls.gov/oes/current/oes_nat.htm#45-0000/. Accessed: March 18, 2019.

			per Respondent	per Response (in Hours)	Hours
Open Data Standard Expert	Key Informant Interview Plan	6	1	1	6
Municipal Government IT Administrator	Key Informant Interview Plan	10	1	1	10
Environmental Health Personnel	Key Informant Interview Plan	20	1	1	20
Technology Providers	Key Informant Interview Plan	7	1	1	7
Totals					43

Table A.12.2. presents the calculations for respondents' time using average hourly wage information from the U.S. Department of Labor, Bureau of Labor Statistics website for May 2017. The total estimated annualized respondent cost is \$1829.61.

Table A.12.2. Estimate of Annualized Burden Costs

Type of Respondents	Form Name	Number of Respondents	Total Burden (in Hours)	Average Hourly Wage	Total Cost
Open Data Standard Expert	Key Informant Interview Plan	6	1	\$57.49	\$344.94
Municipal Government IT Administrator	Key Informant Interview Plan	10	1	\$71.99	\$719.90
Environmental Health Personnel	Key Informant Interview Plan	20	1	\$23.71	\$474.20
Technology Providers	Key Informant Interview Plan	7	1	\$41.51	\$290.57
Totals					\$1,829.61
Source for pay data: https://www.bls.gov/oes/current/oes_nat.htm#45-0000					

A.13. Estimates of Other Total Annual Cost Burden to Respondents and Record Keepers

There are no other costs to respondents other than their time.

A.14. Annualized Cost to the Federal Government

CDC works with NEHA through cooperative agreement #CDC-RFA-OT18-1802 to develop open data standard for aquatic facilities. The annualized cost to the government for this contract is \$3803.81.

Expense Type	Expense Explanation	Annual Costs (dollars)
Costs to the Federal Government	NEHA Project Director: 12.5 hours	\$728.81
	NEHA Project Coordinator: 25.0 hours	\$575.00
	NEHA Open Data Contractor SME: 25.0 hours	\$2,500.00
TOTAL COST TO THE GOVERNMENT		\$3,803.81

A.15. Explanation for Program Changes or Adjustments

This is a new generic Information collection.

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

Data collection will begin in April or May 2019, immediately after CDC receives PRA clearance to conduct the key informant interviews.

Project Time Schedule	
Activity	Time Schedule
Begin recruitment	Immediately after OMB approval
Interviews commence	1 month after OMB approval
Draft final report	12 months after OMB approval

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

The display of the OMB expiration date is appropriate.

A.18. Exceptions to Certification for Paperwork Reduction Act Submissions

There are no exceptions to the certification. These activities comply with the requirements in 5 CFR 1320.9.

References

City of Louisville, KY. Louisville Metro Open Data Portal. <https://data.louisvilleky.gov/>. Accessed: March 26, 2019.

City of San Francisco, CA. DataSF. <https://datasf.org/opendata/>. Accessed: March 26, 2019.

O'Reilly T. Open Data and Algorithmic Regulation. In: Goldstein B, Dyson L, ed. *Beyond Transparency: Open Data and the Future of Civic Innovation*. San Francisco: Code for America Press; 2013:
https://s3.amazonaws.com/academia.edu.documents/34686172/BeyondTransparency.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1553633287&Signature=%2FY97V%2BCmJFV%2F8SxIXJCGC5xGkEw%3D&response-content-disposition=inline%3B%20filename%3DTable_of_Contents_Preface.pdf. Accessed: March 26, 2019. p.294

United States Bureau of Labor Statistics. May 2017 National Occupational Employment Statistics. https://www.bls.gov/oes/current/oes_nat.htm#45-0000/. Accessed: March 18, 2019.

List of Attachments

Attachment 1. Key Informant Interview Plan

Attachment 2. Key Informant Interview Request Letter

Attachment 3. Key Informant Interview Reminder Letter

Attachment 4. NCEH/ATSDR Research Determination Form