

**SUPPORTING STATEMENT FOR
SECURE TRANSFER, RESEARCH-USE DATA LAKE**

OMB CONTROL NO. 1290-0NEW

This is a new information collection request.

A. JUSTIFICATION

1. Explain the circumstances that make the collection of information necessary. Identify any legal or administrative requirements that necessitate the collection. Attach a copy of the appropriate section of each statute and regulation mandating or authorizing the collection of information.

Summary

The Chief Evaluation Office (CEO) of the U.S. Department of Labor (DOL) intends to design and implement a data lake that will safely promote and expand restricted-use DOL data access to facilitate timely, accurate, and informative analysis, research, and program evaluation. In brief, the project involves: (1) developing a data-sharing infrastructure named Secure Transfer, Research-Use Data Lake (STRUDL); (2) supporting, onboarding, and training of approved STRUDL users for their research; and (3) providing privacy and statistical expertise to evaluate and ensure that research products from STRUDL are protected against disclosure risks and are publicly released in a timely manner.

Several laws and policies (Pub.L. 115-435)¹ compel agencies to make the data they collect into public data assets, and to do so in a manner that minimizes the challenges for the public in finding, accessing, and reusing that data. These mandates are echoed in the goals of the Federal Data Strategy and DOL's own Enterprise Data Strategy. For the strategic value in Federal data to be realized, it must be made available to people who can explore and analyze it, identify promising trends and patterns, and suggest innovative use cases. DOL is moving to come into compliance with these mandates and meet the goals articulated in the data strategy documents that guide us.

2. Indicate how, by whom, and for what purpose the information is to be used. Except for a new collection, indicate the actual use the agency has made of the information received from the current collection.

STRUDL's Predominant Purpose Statement requests information about the proposed project and why the applicants need access to protected DOL data (that is, why publicly available data are not sufficient for their purpose). The Biographical Sketch will be used to determine the

¹ Title II of the Foundations for Evidence-Based Policymaking Act of 2018 <https://www.congress.gov/bill/115th-congress/house-bill/4174>

qualifications of the STRUDL applicants to ensure that they have the technical expertise needed to operate in the secure environment and complete the proposed research in the allotted time. The Biographical Sketch will not request any personal information. Disclosure Review Forms will be used to evaluate the disclosure risks of proposed projects by asking researchers to identify the most sensitive variables used in the analysis. These forms will be available to applicants to ensure transparency of STRUDL's application process. This collection of information is for internal recordkeeping purposes. In accordance with DOL approved records schedule group 174, job number N1-174-06-1, filled out forms will be closed after 60 days and retained for 3 years, unless matter is ongoing. DOL will not disclose any individual information being collected to the public. DOL will not use this information for outreach.

3. Describe whether, and to what extent, the collection of information involves the use of automated, electronic, mechanical, or other technological collection techniques or other forms of information technology, e.g., permitting electronic submission of responses, and the basis for the decision for adopting this means of collection. Also, describe any consideration of using information technology to reduce burden.

The DOL staff working with STRUDL will receive applicant materials through email and review them.

4. Describe efforts to identify duplication. Show specifically why any similar information already available cannot be used or modified for use for the purposes described in Item A.2 above.

DOL has done due diligence in seeking to understand if there are proxy values that could be used and has not found anything available. There is no way to access, obtain, or gather this information beyond asking the users to provide it.

5. If the collection of information impacts small businesses or other small entities, describe any methods used to minimize burden.

The information collection does not target small businesses or entities.

6. Describe the consequence to federal program or policy activities if the collection is not conducted or is conducted less frequently, as well as any technical or legal obstacles to reducing burden.

Failure to collect applicant information would render hinder DOL's effort to implement a data lake (STRUDL) that will safely promote and expand restricted-use DOL data access to facilitate timely, accurate, and informative analysis, research, and program evaluation.

7. Explain any special circumstances that would cause an information collection to be conducted in a manner:

- **requiring respondents to report information to the agency more often than quarterly;**
- **requiring respondents to prepare a written response to a collection of information in fewer than 30 days after receipt of it;**
- **requiring respondents to submit more than an original and two copies of any document;**
- **requiring respondents to retain records, other than health, medical, government contract, grant-in-aid, or tax records for more than three years;**
- **in connection with a statistical survey, that is not designed to produce valid and reliable results that can be generalized to the universe of study;**
- **requiring the use of statistical data classification that has not been reviewed and approved by OMB;**
- **that includes a pledge of confidentiality that is not supported by authority established in statute or regulation, that is not supported by disclosure and data security policies that are consistent with the pledge, or which unnecessarily impedes sharing of data with other agencies for compatible confidential use; or**
- **requiring respondents to submit proprietary trade secret, or other confidential information unless the agency can demonstrate that it has instituted procedures to protect the information's confidentiality to the extent permitted by law.**

There are no special circumstances for the proposed data collection.

8. If applicable, provide a copy and identify the date and page number of publication in the *Federal Register* of the agency's notice, required by 5 CFR 1320.8(d), soliciting comments on the information collection prior to submission to OMB. Summarize public comments received in response to that notice and describe actions taken by the agency in response to these comments. Specifically address comments received on cost and hour burden.

Describe efforts to consult with persons outside the agency to obtain their views on the availability of data, frequency of collection, the clarity of instructions and recordkeeping, disclosure, or reporting format (if any), and on the data elements to be recorded, disclosed, or reported.

Consultation with representatives of those from whom information is to be obtained or those who must compile records should occur at least once every 3 years -- even if the collection-of-information activity is the same as in prior periods. There may be

circumstances that may preclude consultation in a specific situation. These circumstances should be explained.

The 60-day notice to solicit public comments was published in the Federal Register on August 24, 2023, (88 FR 57975). A correction was published on September 25, 2024, (89 FR 78337) No comments were received in response to the 60-day notice.

Multiple staff from the following federal agency's and contractors who have supported restricted-use data access programs were consulted to learn about their methods for recordkeeping, statistical disclosure, data privacy and security, and application processes of their respective programs: Census Federal Statistical Research Data Centers (CENSUS FSRDC), the Administration for Children and Families, the Urban Institute, and Bureau of Labor Statistics (BLS).

9. Explain any decision to provide any payments or gifts to respondents, other than remuneration of contractors or grantees.

DOL will not provide any payments or gifts to respondents to this proposed information collection.

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

Respondents are given no assurances of confidentiality. In each data collection activity, respondents will be informed that all data will be used for internal administration purposes, performance metrics, and to guide enhancements to services, data offerings, and documentation for STRUDL.

11. Provide additional justification for any questions of a sensitive nature, such as sexual behavior and attitudes, religious beliefs, and other matters that are commonly considered private. This justification should include the reasons why the agency considers the questions necessary, the specific uses to be made of the information, the explanation to be given to persons from whom the information is requested, and any steps to be taken to obtain their consent.

This collection does not include any questions of a sensitive nature. Respondents are explicitly advised to not provide such information when filling out the forms.

12. Provide estimates of the hour burden of the collection of information. The statement should:

- **Indicate the number of respondents, frequency of response, annual hour burden, and an explanation of how the burden was estimated. Unless directed to do so, agencies should not conduct special surveys to obtain information on which to base hour burden estimates. Consultation with a sample (fewer than 10) of**

potential respondents is desirable. If the hour burden on respondents is expected to vary widely because of differences in activity, size, or complexity, show the range of estimated hour burden, and explain the reasons for the variance. General, estimates should not include burden hours for customary and usual business practices.

- **If this request for approval covers more than one form, provide separate hour burden estimates for each form.**
- **Provide estimates of annualized cost to respondents for the hour burdens for collections of information, identifying and using appropriate wage rate categories. The cost of contracting out or paying outside parties for information collection activities should not be included here. Instead, this cost should be included in Item 13.**

The data collection for these activities will be ongoing, as new users learn of STRUDL and seek to gain access and use the service. DOL estimates that the service will add 15 new projects with approximately 5 program participants per project each year. The burden estimates below reflect the specific burden of individuals who are providing data as part of registering to use the service.

ESTIMATED ANNUAL BURDEN HOURS

Type of Instrument (Form/ Activity)	Number of Respondents	Number of Responses per Respondent	Number of Responses	Average Burden Time per Response (hours)	Estimated Burden Hours	Average Hourly Wage (\$) ³	Monetized Value of Time
Predominant Purpose Statement Form ST-132	15 ¹	1	15	3	45	\$49.76	\$2,239.2
Biographical Sketch and supporting documents Form ST-131	15 ²	5	75	1.5	112.5	\$49.76	\$5,598
Disclosure Review Forms Form ST-133	15 ¹	1	15	2	30	\$49.76	\$1,492.8
Total	15*		105		187.5 (188 rounded)		\$9,330

*= not cumulative

Assumes approximately 15 STRUDL applications over the calendar year.

² Assumes approximately 5 program participants per application for approximately 15 STRUDL applications over the calendar year.

³ Hourly wage for program staff and partners reflects the May 2022 median hourly wage estimate for “data scientists” as reported by the U.S. Department of Labor, Bureau of Labor Statistics, Occupational Employment and Wage Estimates, 2022, “Occupational Outlook Handbook,” (accessed from the following web site as of December 19, 2023: <https://www.bls.gov/ooh/math/data-scientists.htm#:~:text=The%20median%20annual%20wage%20for%20data%20scientists%20was,projected%20each%20year%2C%20on%20average%2C%20over%20the%20decade.> While the likely users of this service will come from a much more diverse population such as journalists, academics, public health professionals, state regulators, and social science researchers, this estimate is conservative and likely to produce an upper bound on burden costs.

13. Provide an estimate of the total annual cost burden to respondents or recordkeepers resulting from the collection of information. (Do not include the cost of any hour burden shown in Items 12 and 14).

- **The cost estimate should be split into two components: (a) a total capital and start up cost component (annualized over its expected useful life); and (b) a total operation and maintenance and purchase of service component. The estimates should take into account costs associated with generating, maintaining, and disclosing or providing the information. Include descriptions of methods used to estimate major cost factors including system and technology acquisition, expected useful life of capital equipment, the discount rate(s), and the time period over which costs will be incurred. Capital and start-up costs include, among other items, preparations for collecting information such as purchasing computers and software; monitoring, sampling, drilling and testing equipment; and record storage facilities.**
- **If cost estimates are expected to vary widely, agencies should present ranges of cost burdens and explain the reasons for the variance. The cost of purchasing or contracting out information collection services should be a part of this cost burden estimate. In developing cost burden estimates, agencies may consult with a sample of respondents (fewer than 10), utilize the 60-day pre-OMB submission public comment process and use existing economic or regulatory impact analysis associated with the rulemaking containing the information collection, as appropriate.**
- **Generally, estimates should not include purchases of equipment or services, or portions thereof, made: (1) prior to October 1, 1995, (2) to achieve regulatory compliance with requirements not associated with the information collection, (3) for reasons other than to provide information or keep records for the government, or (4) as part of customary and usual business or private practices.**

There are no direct costs to respondents and no recordkeeping requirement imposed by STRUDL, or to be a respondent. The only burden on respondents is the time burden described in section 12.

14. Provide estimates of the annualized cost to the Federal Government. Also, provide a description of the method used to estimate cost, which should include quantification of hours, operational expenses (such as equipment, overhead, printing, and support staff), any other expense that would not have been incurred without this collection of information. Agencies also may aggregate cost estimates from Items 12, 13, and 14 into a single table.

The average, annualized cost to the Federal government over three years of managing STRUDL user information is estimated to be \$10,000 in staffing costs and \$2,500 in operations and maintenance. The capability and requirement to collect this data require no new development, testing or evaluation costs, and no costs associated with defining or creating a storage solution. The only costs DOL will incur are for preexisting staff to mail, store, manage, and access new information once this service becomes available to the public.

Description	Cost
Managing STRUDL user information staffing costs (ex: 8% of time for 1 GS-13 step 2) ¹	\$10,000
Operations and maintenance (AWS cloud computing platform)	<u>\$2,500</u>
Total	\$12,250

Using OPM's Salary Table 2024-DCB to calculate percentage of employee's time spent managing STRUDL user information. <https://www.opm.gov/policy-data-oversight/pay-leave/salaries-wages/salary-tables/24Tables/html/DCB.aspx>

15. Explain the reasons for any program changes or adjustments.

There are no program changes or adjustments. This data sharing program is a new collection that is not yet on the OMB Inventory. .

16. For collections of information whose results will be published, outline plans for tabulations, and publication. Address any complex analytical techniques that will be used. Provide the time schedule for the entire project, including beginning and ending dates of the collection of information, completion of report, publication dates, and other actions.

This collection will not be published online. Information from these forms are strictly administrative data and will only be used for internal recordkeeping purposes.

17. If seeking approval to not display the expiration date for OMB approval of the information collection, explain the reasons that display would be inappropriate.

No approval is sought not to display the expiration date for OMB approval of the information collection.

18. Explain each exception to the certification statement.

No exceptions are necessary for this information collection.