

**Secure Transfer, Restricted-Use Data Lake  
Handbook  
United States Department of Labor**

Version 1.0

February 2024

1

# Table of Contents

Table of Contents.....	2
Acronyms and Abbreviations.....	5
Preface and Statement of Purpose .....	6
Intended audience and recommended sectional reading .....	6
Academic researchers.....	6
Contractors .....	7
Departments .....	7
Chapter 1: Introduction to the Secure Transfer, Restricted-Use Data Lake .....	8
About STRUDL .....	8
Purpose .....	8
The legal requirements and moral obligation to protect DOL data.....	8
The conditions of access .....	9
STRUDL applicant requirements .....	9
Responsibilities of the STRUDL applicants and recipients.....	10
Expectations for publishing research findings .....	11
Chapter 2: Project Lifecycle .....	12
Joining STRUDL.....	12
Who can apply to STRUDL .....	12
Criteria for an approved project.....	12
STRUDL application process .....	13
Onboarding to STRUDL.....	15
Clearance process .....	15
PIV Card .....	15
Acquiring a DOL Computer .....	16
STRUDL Cloud Computing Environment .....	16
Using STRUDL.....	16
Maintaining computer and data security.....	16
Reporting data breaches .....	16
Getting support .....	16
Requesting modifications on projects.....	17
Exiting STRUDL .....	17
Documenting the benefits of DOL.....	17
Documenting your research .....	18

Backing up files .....	18
Returning materials .....	18
After STRUDL.....	18
Returning for follow-up work.....	18
Data access for R&Rs for academic journals.....	18
Seminars and conference presentations .....	19
Providing links to final dissemination products.....	19
Chapter 3: Access, Confidentiality, and Disclosure Analysis .....	20
General policies on protecting confidentiality and clearing results for release.....	20
Be a trusted researcher.....	20
Be mindful of DOL information .....	20
Be a responsible data steward.....	20
Be familiar with the DOL disclosure review process.....	20
The disclosure review process .....	21
Outline of the disclosure review process.....	21
Disclosure review documents .....	23
Best practices for disclosure avoidance .....	23
General practices.....	23
Practices for frequency tables.....	24
Practices for moments and distributional tables .....	24
Practices for other distributional output.....	25
Practices for model output .....	26
Practices for other output.....	26
Alternative statistical disclosure control approaches .....	26
Other general best practices to consider .....	26
Disclosure avoidance review and disclosure review package submission.....	27
Practices for publications .....	27
Clearing results .....	27
Clearing other files .....	27
Disclosure review approval process.....	28
Dissemination review approval process .....	28
Appendices .....	29
Glossary .....	29
STRUDL terminology.....	29

Technical terminology..... 29

## Acronyms and Abbreviations

DOL	Department of Labor
NDA	nondisclosure agreement
OASAM	Office of the Assistant Secretary for Administration and Management
PI	Principal Investigator
PII	personally identifiable information
PPS	predominant purpose statement
R&R	revise and resubmit
SME	subject matter expert
STRUDL	Secure Transfer, Restricted-Use Data Lake

## Preface and Statement of Purpose

The U.S. Department of Labor (DOL)'s research and program evaluation activities are key to performing its mission to foster, promote, and develop the welfare of the wage earners, job seekers, and retirees of the United States; improve working conditions; advance opportunities for profitable employment; and assure work-related benefits and rights. To perform research" and program activities responsibly and effectively, DOL must safely and quickly grant approved researchers access to these data while minimizing the disclosure risks that these activities bring with them. That is, the data must be protected while they are accessed, the access must be for approved purposes only, and any confidential information must be protected, including when results are published.

The need to balance data access and security motivated DOL to develop the Secure Transfer, Restricted-Use Data Lake (STRUDL) as a secure process for approved researchers to access confidential data for approved purposes and to publish their results after they have been carefully screened for disclosure risk. The purpose of STRUDL is to safely promote and expand restricted-use DOL data access to facilitate timely, accurate, and informative analysis, research, and program evaluation. This handbook will serve as an introduction to STRUDL, its role in a project lifecycle, and summary of the policies and practices to applying disclosure avoidance techniques. The goal of the handbook is to increase transparency and understanding of data governance and disclosure avoidance practices at DOL. This handbook is also a living document that will be continuously updated to ensure that STRUDL remains up to date with the latest changes in safe data access.

**IMPORTANT NOTE:** Potential applicants, or individuals interested in participating in STRUDL, who plan to apply for STRUDL should know that the disclosure review process is throughout a project life cycle, from study design to data termination. In particular, the DOL staff and contractors who are responsible for disclosure review expect approved researchers (those who successfully applied for the STRUDL) to submit all final products for public release for review at least 90 days before end of contract or the desired release date, whichever is earlier.

### Intended audience and recommended sectional reading

This handbook is written for three DOL groups: academic researchers, contractors, and departments. We define each group below and outline which sections of the handbook each group should review. Although anyone interested in STRUDL is encouraged to read the entire handbook, there are some sections that are not relevant to certain groups.

#### *Academic researchers*

Academic or external researchers are subject matter experts, analysts, or practitioners that use and apply their expertise to DOL data that are from an academic institution or nonprofit organization. These individuals are the main audience for this handbook and are recommended to read its entirety to ensure they understand the requirements and expectations of STRUDL.

### *Contractors*

Contractors are subject matter experts, analysts, or practitioners that DOL hired to complete tasks outlined in their respective contracts. These individuals are recommended to review the following sections of the handbook:

- Chapter 1 – all sections
- Chapter 2
  - Onboarding to STRUDL
  - Using STRUDL
  - Exiting STRUDL
    - Backing up files
    - Returning materials
- Chapter 3 – all sections

### *Departments*

Departments are entities that seek to access and/or share DOL data with other DOL departments through STRUDL. Staff from this group are recommended to review the following sections of the handbook:

- Chapter 1
  - About STRUDL
- Chapter 2
  - Onboarding to STRUDL
  - Using STRUDL
- Chapter 3 – all sections

# Chapter 1: Introduction to the Secure Transfer, Restricted-Use Data Lake

## About STRUDL

DOL's Secure Transfer, Restricted-Use Data Lake (STRUDL) provides authorized external and internal researchers secure access to confidential data. The program aims to support approved social science research projects that align with DOL's mission. In summary, STRUDL grants researchers the ability to address important research questions of interest to the academic community, the general public, and DOL. These data have already been paid for by the public, allowing approved researchers to reuse the data rather than discard them. Storing and carefully reusing the data ensures good stewardship.

Approved researchers, those who submit a successful STRUDL application, can access, link, and utilize restricted-use microdata from a variety of DOL sources. Prior to public release, staff from DOL's disclosure review process will review the analyses and statistics on DOL data. Once cleared, researchers are permitted to publish their results in reports and other professional outlets. By informing the public and policymakers on practical worker issues they contribute to fulfilling DOL's mission.

### *Purpose*

DOL collects survey and administrative data for many reasons, such as program administration, conducting evaluations of DOL programs, and other related activities. The collected data may contain personally identifiable information (PII) or other potentially sensitive data that cannot be publicly released for ethical and legal reasons. DOL also grants approved researchers access to the restricted data who provide a compelling research justification and agree to DOL guidelines, such as signing a nondisclosure agreement, to ensure safeguarding the data.

DOL's STRUDL program is a rapid, cost-effective way to generate applied knowledge, foster collaboration, responsibly manage taxpayer resources, and advance DOL-wide priorities. STRUDL allows approved researchers access to DOL datasets for approved research projects, providing more researchers safe access to DOL's data. STRUDL also advances DOL's mission to

“...foster, promote, and develop the welfare of wage earners, job seekers, and retirees of the United States; improve working conditions; advance opportunities for profitable employment; and assure work-related benefits and rights.”

Further, STRUDL's purpose aligns with the initiatives from the 2019 Foundations for Evidence-Based Policymaking Act for federal agencies to improve and expand data access for better evidence-based public policy decision-making.

## **The legal requirements and moral obligation to protect DOL data**

STRUDL complies with the physical and information security requirements and protocols for accessing restricted-use microdata from relevant agencies. To participate in STRUDL, academic researchers' respective organizations (hereafter referred to as "the recipient") will be required



to sign a nondisclosure agreement (NDA) between DOL and the organization. Potential applicants, or individuals interested in participating in STRUDL, should consult their institution's contracts offices and identify a restricted-use NDA administrator to review the NDA prior to applying for STRUDL. This will ensure that applicants can adhere to NDA requirements to safeguard data before applying for STRUDL.

In addition to the legal requirements, applicants have a moral and ethical obligation to protect the data. DOL collects only necessary information on U.S. residents to carry out its mission. Therefore, applicants should only use the data for only legal, ethical, and professionally accepted research that advances DOL's mission and benefits the broader public.

DOL will conduct a thorough review of all research outputs. Some examples of research outputs include regression coefficients, moments of distributions, and other statistical inferences. The objective of the review is to assess potential disclosure risks, such as unintentional release of PII, and to ensure the outputs are within the approved project's scope before authorizing for publication. Moreover, this review process aims to protect sensitive data and ensure ethical and legal standards.

## **The conditions of access**

This section outlines the conditions of access that need to be considered for STRUDL: applicant requirements, responsibilities of the STRUDL applicant, and expectations for publishing research findings. Potential applicants must carefully review this section to ensure they are eligible for STRUDL.

### ***STRUDL applicant requirements***

Applicants must satisfy the following requirements:

- Applicants must be individuals or groups of individuals, not organizations.
- Applicants may use DOL restricted data for general research purposes that can benefit DOL's mission and the broader public.
- Applicants cannot use DOL restricted-use data for purposes that do not meet DOL's guiding principles, such as conducting market research.
- Applicants must publicly publish results based on DOL restricted data.
- Applicants and the applicants' organizations are required to sign an NDA with DOL to access and use restricted-use data. Applicants must scan a signed version of the NDA or use verified e-signature software to sign the NDA. Applicants and recipients must agree to abide by the rules outlined in the NDA for using the data. Unauthorized use or disclosure of DOL information or information systems may result in cancellation of the approved researchers' access to controlled unclassified information, and any other appropriate action, including, but not limited to, criminal sanctions under the Privacy Act of 1974, the Trade Secrets Act, or 18 U.S.C. 641. In addition, the approved researchers' organization may be suspended from work involving remote access of DOL data due to a failed surrender of such materials or documents listed before.

Applicants are encouraged to read the "STRUDL Application FAQ," which addresses frequently asked questions from applicants and provides an overview of STRUDL.

### *Responsibilities of the STRUDL applicants and recipients*

The NDA outlines several responsibilities of the STRUDL applicant and the recipient. This section highlights a few, but not all, of the responsibilities that applicants must comply with. As previously stated, potential applicants are encouraged to consult their respective institution's contracts offices and identify a restricted-use NDA administrator to review the NDA prior to applying for STRUDL.

The DOL may revoke an approved researcher's access to the data at any time and without advance notice. The recipient may request that a current approved researcher's access be revoked. The recipient and its researchers shall notify the DOL Restricted Use Officer whenever a researcher is no longer associated with the recipient or where the continuation of a researcher's data access may endanger the confidentiality of data. All researchers will perform activities subject to this agreement under the supervision and control of the DOL Restricted Use Officer or any other DOL official that the DOL designates. The DOL will notify the recipient of any such designations.

Approved researchers must complete confidentiality training provided by the DOL both at the start of their project and on an annual basis thereafter for the duration of this agreement.

Approved researchers must agree, in writing, to comply with all provisions of law that affect information acquired by that agency. Approved researchers who improperly disclose confidential information may be subject to appropriate sanctions.

The recipient will review all laws applicable to the confidentiality of data provided under this agreement and ensure that all researchers designated under this agreement review such materials and are fully familiar with their obligations to safeguard confidential data.

Only approved researchers may access STRUDL. Approved researchers may not use confidential information for any purpose other than a statistical purpose, such as using the confidential information for the purpose of identifying in any way persons or business establishments. If the identity of a person or business establishment is inadvertently discovered, approved researchers will make no use of this knowledge and will hold the identity of the person or business establishment in confidence.

Unless specifically authorized in the NDA, approved researchers will not attempt to link DOL confidential information with individually identifiable records from any DOL or non-DOL data set. Approved researchers agree that it is able legally to extend to the DOL permission to access all non-DOL data for the purposes of loading and maintaining the data on DOL information technology systems, conducting confidentiality reviews of outputs derived from the linked data, and other purposes in support of this agreement. All new data sets created from linking DOL confidential information with other data are protected by the Confidential Information Protection and Statistical Efficiency Act and must not be removed from the DOL.

All outputs prepared using confidential information must be submitted to the DOL Restricted Use Officer for confidentiality and subject matter expert (SME) review 90 days prior to the desired release date and will not be removed from the place of performance or published or released to any individual other than an authorized DOL employee or DOL [disclosure avoidance](#)

contractor until cleared by the DOL Restricted Use Officer. SMEs review the paper for factual accuracy only. SMEs do not review the authors' tone, framing, or conclusions. Approved researchers will be bound by the determinations of the DOL Restricted Use Officer or other DOL officials designated by the CEO to clear such outputs. The approved researchers will provide DOL with a final version of the paper, as it will be published, prior to publication.

*Expectations for publishing research findings*

Per NDA requirements, approved researchers will use reasonable efforts to ensure that the end product of the study produced pursuant to this agreement is published or otherwise made available to the public. The approved researchers shall transmit to DOL copies of any final reports, charts, research articles, or other media upon its publication or release. The DOL reserves the right to make any final reports, charts, research articles, or other media publicly available on its website or through other means of dissemination.

Approved researchers will also ensure that the DOL Restricted Use Officer receives a 200- to 300-word abstract summarizing any published works resulting from research conducted under this agreement. The DOL may make the abstract publicly available on its website or through other means of dissemination.

Approved researchers will include in all final publicly released reports, work products, or research outputs, the following disclaimer: "This research was conducted with restricted access to U.S. Department of Labor (DOL) data. The views expressed here are those of the author and do not reflect the views of the DOL. This research was performed under DOL Secure Transfer, Restricted-Use Data Lake Project Number XXXX-XXX."

## Chapter 2: Project Lifecycle

This chapter offers an overview of the complete project lifecycle within STRUDL, focusing on the steps involved for potential applicants who are researchers seeking access to DOL data. Potential applicants will learn about STRUDL in detail, which includes joining STRUDL, conducting research within the STRUDL environment, exiting the program, and following up on any dissemination products resulting from STRUDL.

### Joining STRUDL

This section outlines who can apply, the criteria for an approved project, and the general policies governing the process of joining STRUDL. Applicants must carefully review this section before submitting their applications to ensure that the potential project meets the STRUDL standards, and the application materials are completed accurately and in accordance with the requirements. Please note that incomplete or improperly filled out applications will be automatically rejected.

#### *Who can apply to STRUDL*

Applicants must demonstrate to DOL that they have the appropriate knowledge, skills, and resources to meaningfully work with the data to accomplish their research objectives while adhering to the NDA. As part of the application process, all senior personnel, key personnel, and other significant contributors must each complete a DOL biographical sketch (i.e., the “DOL Biographical Sketch” document) to demonstrate their qualifications to complete the proposed work. Applicants must also be based in and access STRUDL data from the United States.

DOL only grants access to restricted-use data to individuals or groups of individuals and not organizations. Any applicants who intend to access or interact with the data or statistics that have not undergone disclosure review in any way before publication clearance from DOL must be listed in the research application proposal. These individuals and groups represent their respective organizations (i.e., recipients). Violation of DOL agreements can result in applicable legal penalties as well as suspension from accessing DOL data and systems for individuals and their organizations (see DOL STRUDL NDA for further details).

If additional project personnel are required after the project commences, then the NDA may be eligible to be amended. Members of the project should contact their restricted-use NDA administrator to add or remove any individuals from the NDA.

Researchers or potential applicants will begin the process by contacting the Restricted Use manager at [STRUDL@dol.gov](mailto:STRUDL@dol.gov).

#### *Criteria for an approved project*

DOL provides access to confidential data to approved researchers who propose a tractable project that benefits DOL and its customers. Applicants must explain why access to restricted-use data is necessary and that public-access data or alternative sources of data are insufficient to answer their research questions. The proposal must also describe how the intended project can benefit DOL’s mission and the broader public. Applicants must clearly describe how the

question is of interest to DOL and its customers as well as commit to publicly publishing their research results. Researchers cannot use DOL restricted data that do not meet DOL's guiding principles, or for fiduciary purposes, such as conducting market research.

### *STRUDL application process*

This section provides an overview of the STRUDL application process, as well as a swimming lane process map of the process for all applicants (see figure 1).

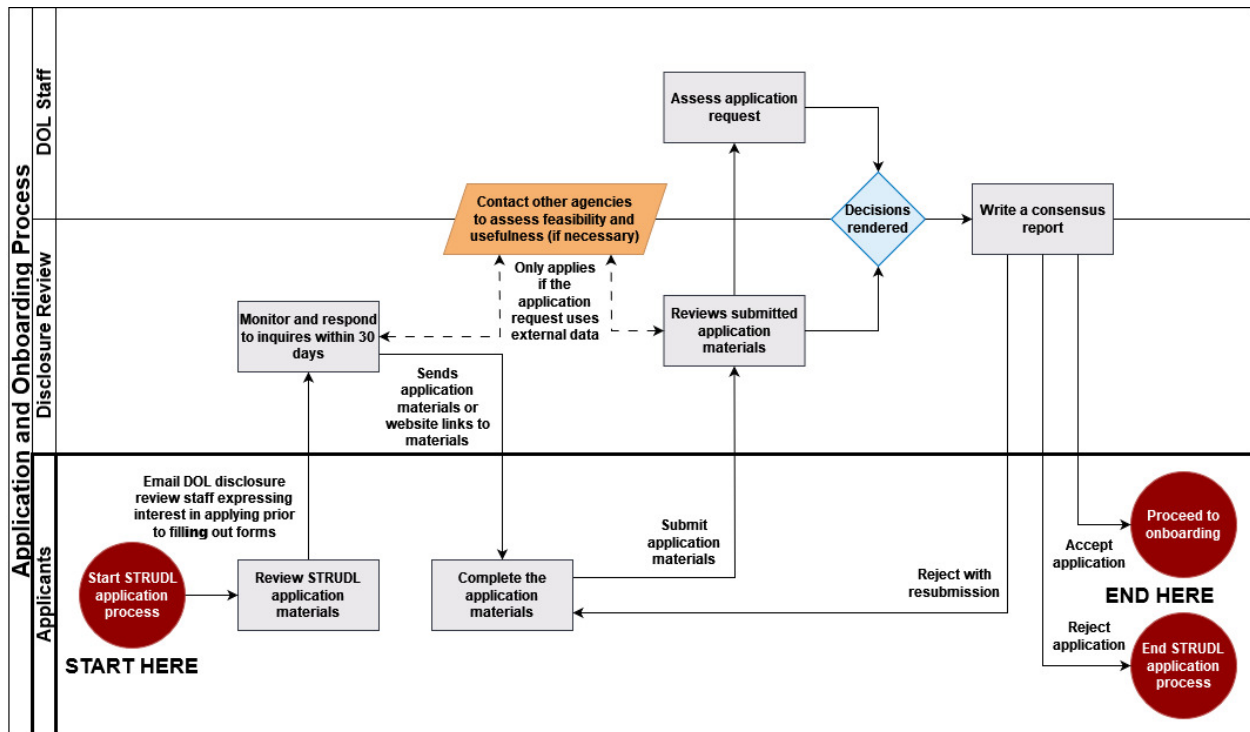


Figure 1: A swimming lane process map for the STRUDL Application Process throughout the lifecycle of an applicant applying to STRUDL. The three swimming lanes are DOL staff, DOL disclosure review staff, and researchers (applicants). The colors/shapes indicate a type of step in the process map. A red circle is a process start/end (event or trigger), a grey rectangle is a process step, a light blue diamond is a decision point, and a yellow rhomboid is an optional process step.

The general process is as follows:

1. Potential applicants are advised to thoroughly review the STRUDL application materials before submitting their STRUDL applications. The latest information and documentation can be found on the DOL STRUDL website. For any questions about STRUDL, potential applicants are encouraged to contact [STRUDL@dol.gov](mailto:STRUDL@dol.gov). Below is the list of the required application documents, along with details about each document.
  - a. **DOL Nondisclosure Agreement (NDA)** – The DOL NDA is a mandatory requirement for the applicants and applicants' organizations to sign and the contents are not negotiable. Applicants must scan a signed version of the NDA or

use verified e-signature software to sign the NDA. It is important to note that any violation of the NDA agreement will lead to the applicants and potentially the applicants' institution being permanently barred from participating in STRUDL and may result in additional penalties.

- b. **Predominant Purpose Statement (PPS)** – The purpose of this document is to provide DOL with information regarding the proposed project and to explain why the applicants require access to DOL data. Applicants should also review the "Predominant Purpose Statement Example" document, which demonstrates a completed PPS.
  - c. **DOL Biographical Sketch** – The purpose of this document is for applicants to provide detailed information about their qualifications to be considered as approved researchers for STRUDL. This form must be completed by applicants who are senior personnel, key personnel, or other significant contributors to the proposed project. It is advised that applicants review the accompanying "Biographical Sketch Details" document, which provides additional guidance on how to complete the "Biographical Sketch" section. Applicants must note that the inclusion of additional information beyond that specified in the "Biographical Sketch" may result in the application being returned without review. Applicants must not submit any personal information in the "Biographical Sketch." This includes items such as home address; home telephone, fax, or cell phone numbers; home email address; driver's license number; marital status; personal hobbies; and the like. Such personal information is not appropriate for the biographical sketch and is not relevant to the merits of the application. DOL is not responsible or in any way liable for the release of such information.
  - d. **Disclosure Review Checklist** – The purpose of this document is to evaluate the anticipated level of effort and identify high-risk variables or features in the data for the disclosure review of datasets. Researchers are required to complete the checklist both during their initial application to access DOL data and after they have been granted access. Questions that are applicable only after researchers gain access will be clearly indicated as such on the form.
2. Upon reviewing the application materials, potential applicants should express their interest in applying by sending an email to [STRUDL@dol.gov](mailto:STRUDL@dol.gov). In the email, applicants should briefly describe their proposed research idea, state the DOL data they hope to access, and indicate if they intend to use external data along with DOL data. Within 30 days, a DOL disclosure review staff member will respond to the inquiry and provide the latest version of the application materials. If the potential applicant plans to incorporate external data, DOL staff may contact other agencies to evaluate the feasibility and potential value of combining these datasets.
  3. Once the application materials have been completed, applicants must submit the forms to [STRUDL@dol.gov](mailto:STRUDL@dol.gov). Within 30 days, the DOL disclosure review staff will acknowledge receipt of the application and collaborate with other DOL personnel to assess the feasibility and disclosure risk associated with the proposed project. The DOL staff will complete the following two forms, generating a consensus report indicating whether the proposed project is "approved," "rejected with resubmission," or is "rejected."

- a. **STRUDL Application Determination** – The purpose of this document is for the DOL staff to make a final determination regarding granting data access after assessing the feasibility and usability of the proposed project.
- b. **Disclosure Review Determination** – The purpose of this document is for the DOL staff to make a final determination regarding granting data access, releasing project outputs, or approving dissemination products after conducting a disclosure review at various stages of the project.

## **Onboarding to STRUDL**

This section outlines how approved researchers will be onboarded to STRUDL and access the restricted data. Specifically, the process involves how approved researchers undergo DOL clearance, obtain security credentials, acquire a DOL laptop, and access DOL data through the STRUDL Cloud Computing Environment.

### *Clearance process*

This section covers the general DOL clearance procedure for approved researchers in STRUDL.

An approval email detailing the conversion of applicants into approved researchers will contain four forms. The approved researchers must fill out these forms to begin the clearance process:

- Declaration for Federal Employment (Optional Form 306)
- Information Systems Security and Privacy Awareness Training for New DOL Users
- U.S. Department of Labor Office of the Assistant Secretary for Administration and Management Office of the Chief Information Officer User Agreement (cyber security safety agreement)
- New UpSTART Request Form

These forms contain PII, so applicants should encrypt the files before emailing them to [STRUDL@dol.gov](mailto:STRUDL@dol.gov), and then send the password by other means, such as a separate email.

After submitting these forms, approved researchers will receive an email from the Office of Personnel Management with a unique e-QIP registration code. Around the same time, staff from the DOL Office of the Assistant Secretary for Administration and Management (OASAM) will contact approved researchers with information to fill out and submit the e-QIP.

The e-QIP includes numerous questions about topics including educational history, residential history, employment history, and international travel history. Completing the form may require tracking down personal information and contact information. The e-QIP may also require Google Authenticator for multifactor authentication. Approved researchers should be prepared to answer follow-up questions from OASAM in a timely manner after submitting the e-QIP.

### *PIV Card*

Once the approved researchers receive DOL clearance, they will receive an email from the General Services Administration about credential pickup. The email should contain detailed instructions about the pickup location, scheduling a pickup appointment, and the materials to bring to the appointment. The pickup location is determined by residential address. Acquiring a PIV card may require fingerprinting.

### *Acquiring a DOL Computer*

DOL staff will coordinate mailing a DOL laptop to the residential addresses of the approved researchers. Approved researchers should work with the DOL Enterprise Service Desk to onboard a DOL computer.

### *STRUDL Cloud Computing Environment*

Approved researchers will receive a welcome email at their DOL email address with access instructions for the STRUDL Cloud Computing Environment. This environment is how approved researchers will access DOL data. The environment supports analyses with R accessed through Posit Workbench, Python accessed through Posit Workbench, Stata 16, and SAS Viya. Approved researchers also have access to Posit Connect.

All systems are accessed through a web browser and can only be accessed through a DOL-provided laptop. Additional documentation is available on DOL systems once access is granted.

## **Using STRUDL**

This section summarizes how approved researchers should maintain computer and data security, report data breaches, receive support, and request modifications for STRUDL projects. Approved researchers should review this section to ensure they adhere to the safety and security of DOL data.

### *Maintaining computer and data security*

All approved researchers for the project must be granted access to the data using their own credentials. Discussions about data or results outside of the DOL system are not allowed. Data cannot be shared with any unapproved researchers. Results cannot be shared with unapproved researchers unless it has gone through the disclosure review process to be released publicly.

### *Reporting data breaches*

All data breaches or suspected data breaches must be reported immediately to DOL. Complete steps for reporting data breaches or suspected data breaches are available under the PII Breach Notification Plan on LaborNet (intranet) in DOL systems.

### *Getting support*

Approved researchers should send all questions to [STRUDL@dol.gov](mailto:STRUDL@dol.gov). Approved researchers may be directed from there to specific DOL staff, who can provide the answer to the questions. This process ensures that DOL staff are not overburdened directly with questions related to STRUDL projects.

For IT and technical problems with DOL laptops, please contact the DOL Enterprise Service Desk. This contact information can be found on your laptop desktop and the DOL LaborNet (intranet) platform.



### *Requesting modifications on projects*

Approved researchers that need to request additional time, change project staff, change data, or update the research plan must submit a "STRUDL Project Modification Request" form, stating the reason for the modification. Requests are reviewed by DOL staff, who will either approve or deny modifications. If denied, approved researchers must continue their initially approved research plan and close-out project before the end of the research period. Requests for modifications should be submitted at least 90 days in advance of the project closing.

Modification requests must include rationale and reason for the modification. Modifications should include a statement of the proposed benefits of the research to DOL. Below are the accepted reasons for modification:

- **Need more time to complete deliverables** – The Principal Investigator (PI) on the approved project must provide an explanation for the time extension. The explanation must include reasons for the potential delay and how completing the deliverables will still benefit DOL.
- **Changing project personnel** – The PI on the approved project must make a request to add new researchers to the project. The new researcher will submit required documentation for access to STRUDL. See the "Onboarding to STRUDL" section. DOL has discretion on approval of access to STRUDL. Once approved, the new researcher and PI are notified and the new research is provided with the credentials for access.
- **Changing data** – The PI on the approved project must provide an explanation for the data change. The explanation must include why the current data access does not meet the approved project goals and how the change in data access will benefit DOL.
- **Update Analysis Plan** – The PI on the approved project must provide an explanation for the update in analysis plan. In addition to including a reason for the change in the "STRUDL Project Modification Request" form, the PI must complete an updated PPS, highlighting the areas that have changed.

### **Exiting STRUDL**

This section provides details on how to exit STRUDL, which involves documenting the benefits of the project to DOL and the project outcomes, backing up files, and returning materials. Approved researchers who fail to complete the exiting process may be suspended from future work through STRUDL involving remote access of DOL data.

#### *Documenting the benefits of DOL*

At the conclusion of the project, approved researchers must complete the close-out checklist and submit it to DOL. The close-out checklist should:

- attach the original PPS,
- summarize the key research findings (no more than 400 words),
- revisit whether and how the project accomplished the research benefits proposed in the PPS,
- document any changes from the PPS,
- list the key committed deliverables for the project (e.g., white paper, seminar, or document on the agency benefits),

- verify that all the deliverables have been submitted and accepted by DOL staff and that all materials, such as reports and statistics, have been reviewed for disclosure risks,
- ensure all workspaces and materials (e.g., laptops) closed or returned in accordance with the nondisclosure agreement, and
- confirm that all nonpublic data, programs, and statistics have been destroyed, returned, or archived in accordance with the nondisclosure agreement.

### *Documenting your research*

Approved researchers must document all datasets and programs used to produce any promised benefits to DOL and research output. Approved researchers must provide all programs, program logs, and outputs from the research, along with the initial datasets and any versions of the dataset that were used to produce output to DOL. File names and datasets should be sufficiently detailed so that other researchers or DOL staff members can reproduce the work.

### *Backing up files*

Additional information about backing up files will be communicated to approved researchers.

### *Returning materials*

Approved researchers must go through a complete offboarding process upon the completion or termination of their project. Researchers should contact [STRUDL@dol.gov](mailto:STRUDL@dol.gov) for an up-to-date offboard checklist. Please note that researchers may be responsible for packing and shipping computing equipment and PIV cards to return them to DOL.

## **After STRUDL**

This section summarizes how approved researchers may return to do follow-up work on past STRUDL approved projects, access DOL for revise and resubmit (R&R) of academic papers, and provide information on final project outputs.

### *Returning for follow-up work*

Approved researchers interested in carrying out follow-up work after the project term expires (other than in the instance of a R&R), must submit a new proposal under the standard proposal review process. Approved researchers may contact their DOL point of contact to develop the new proposal.

### *Data access for R&Rs for academic journals*

To the extent possible, R&Rs should be accommodated within the approved duration of the project. However, understanding that this is not always possible, approved researchers may submit **up to two** requests for readmission after the STRUDL project term expires. Requests can only be made for reasonable amounts of follow-up work. No new data will be approved, new output should be similar to previously approved output, and all new output will be subject to disclosure review process.

The request for follow-up work must include the following:

- A letter requesting re-admission to the system for follow-up work, including a statement of the new benefits to DOL of the work
  - The letter must include project number, title, and PI name, name and affiliations of all people who will need access to the data, list of all the DOL datasets required for reactivation, the specific tasks that will be undertaken, and the new specific benefit to DOL resulting from the additional analysis
- The manuscript that was submitted to the journal
- The letter from the journal editor and the comments

### ***Seminars and conference presentations***

Only final products that have gone through the disclosure review process (see Chapter 3) may be shared at seminars and conferences. Presentations should include the following disclaimer:

***Disclaimer:*** This research was conducted with restricted access to U.S. Department of Labor (DOL) data. The views expressed here are those of the author and do not reflect the views of the DOL. This research was performed under DOL Secure Transfer, Restricted-Use Data Lake Project Number XXXX-XXX.

### ***Providing links to final dissemination products***

Approved STRUDL researchers must email [STRUDL@dol.gov](mailto:STRUDL@dol.gov) with the URLs and DOIs (if applicable) for all products based on restricted-use data accessed through STRUDL.

## Chapter 3: Access, Confidentiality, and Disclosure Analysis

This chapter outlines the policies and procedures that approved researchers must adhere to when accessing DOL data through STRUDL to ensure the confidentiality of DOL data. Specifically, approved researchers will learn about general policies on protecting confidentiality and clearing results for release, the disclosure review process, best practices for disclosure avoidance, and disclosure avoidance review and disclosure review package submission.

### **General policies on protecting confidentiality and clearing results for release**

This section outlines the general policies for protecting confidential data from the DOL. Failure to comply with these guidelines may result in the removal of approved researchers from STRUDL and other individuals affiliated with the researcher's institution (see DOL STRUDL NDA).

#### *Be a trusted researcher*

Approved researchers have a legal, ethical, and moral obligation to protect the confidentiality of DOL's data. Any data or results derived from DOL data must undergo the formal STRUDL disclosure review process before leaving DOL systems. Additionally, information regarding the presence or absence of a record in the data, as well as the parameters of the disclosure review process when applied to the approved researcher's project in STRUDL, are considered confidential.

Discussions or notes about data or results outside of the DOL system are prohibited. Examples include handwritten notes, email, research papers, phone calls, fax, or casual conversations outside of DOL facilities. If approved researchers wish to release or share notes with external partners or colleagues, they must first type them up the notes and submit them through the formal disclosure review process.

#### *Be mindful of DOL information*

Approved researchers must understand that multiple pieces of nondisclosive information can create a disclosive piece of information. In other words, approved researchers may have nondisclosive information from a different project that they could combine with nondisclosive, DOL information, such as industry list and audit measures. The combined information could be disclosive. Therefore, approved researchers should be mindful of what information they share with people outside of the DOL system.

#### *Be a responsible data steward*

Approved researchers must be responsible stewards and protectors of DOL data. Approved researchers are required to adhere to disclosure review guidelines and processes.

#### *Be familiar with the DOL disclosure review process*

Approved researchers must be familiar with DOL's expectations for the DOL disclosure review process. The process is involved in three parts of STRUDL: application to STRUDL (i.e.,

"Predominant Purpose Statement" document), project results review (i.e., "Disclosure and Dissemination Review Determination" form), and project results dissemination (i.e., "Disclosure and Dissemination Review Determination" form). Approved researchers should be familiar with this process.

## The disclosure review process

This section provides an overview of the Disclosure Review Process, as well as a swimming lane process map of the process for all STRUDL approved research projects (see figure 2).

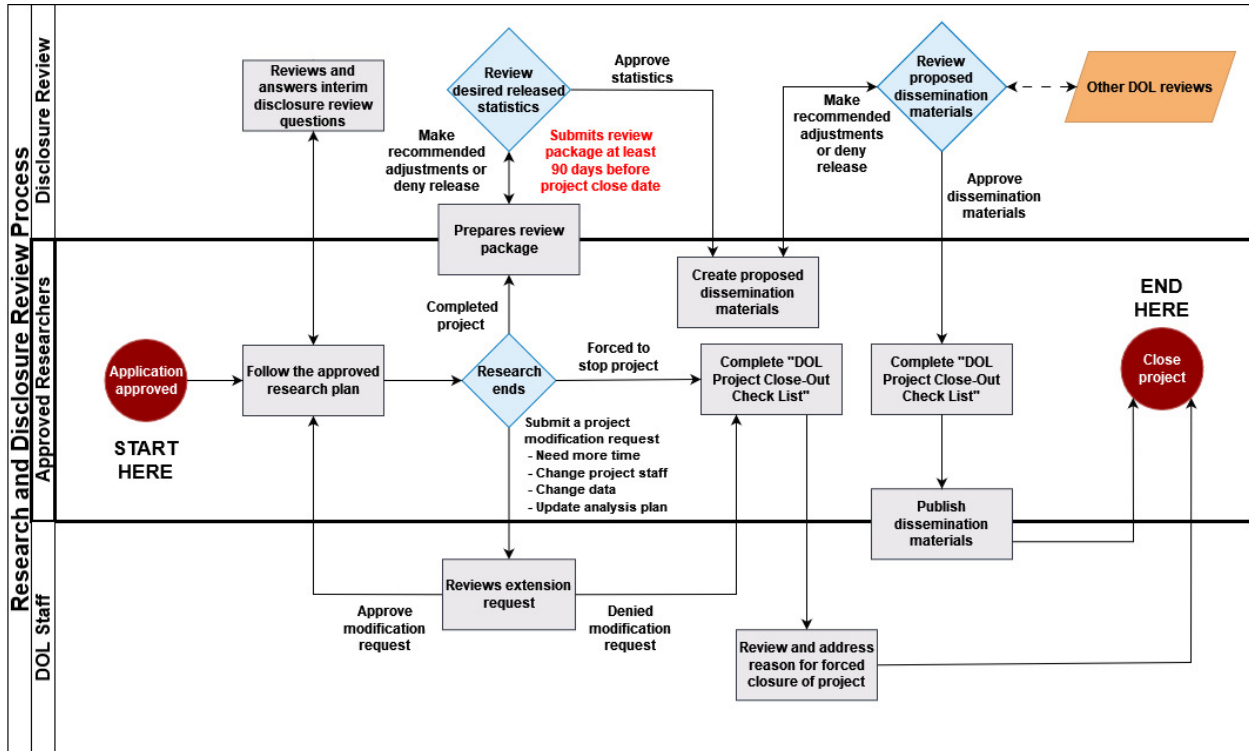


Figure 2: A swimming lane process map for the Disclosure Review Process throughout the lifecycle of a STRUDL approved research project. The three swimming lanes are DOL disclosure review staff, approved researchers, and DOL staff. The colors/shapes indicate a type of step in the process map. A red circle is a process start/end (event or trigger), a grey rectangle is a process step, a blue diamond is a decision point, and a yellow rhomboid is an optional process step.

### Outline of the disclosure review process

The general process is as follows:

1. Once researchers are approved and gain access to DOL data, they should adhere to the research plan submitted during the application process and update the "Disclosure Review Checklist." Any changes made to the original "Disclosure Review Checklist" should be highlighted in the updated checklist and submitted to DOL for review.

Additionally, researchers are encouraged to seek guidance and ask disclosure review questions throughout the course of their research project.

2. A STRUDL research project ends in three ways:
  - a. Completed the project deliverables (see steps 3 to 7).
  - b. Submitted a modification form via the "STRUDL Project Modification Request" Form (see step 8).
  - c. Forced to stop project by DOL (see steps 9 to 10).

### **Complete the project deliverables**

3. Approved researchers should prepare a disclosure review package for any outputs intended for dissemination (such as presentations or reports). See the "Disclosure and Dissemination Review Determination" form.
  - a. The package should be submitted at least 90 days prior to the project close date to allow for iteration and adjustments.
  - b. All necessary forms must be completed; failure to do so may result in automatic rejection.
  - c. Review times may vary depending on the number and types of outputs.
  - d. Only include outputs that will appear in disseminations; intermediate outputs are discouraged as they may lengthen the review process.
  - e. Clear documentation and formatted files will minimize review time.
4. If the disclosure review package is approved, the approved researchers may create the proposed dissemination materials for public release. Otherwise, the proposed outputs must be adjusted or denied for release.
5. Approved researchers should complete and submit the "Disclosure and Dissemination Review Determination" form along with the dissemination materials for review.
6. Upon approval, DOL staff will publish the research results on the DOL website and approved researchers can share the dissemination materials.
7. Approved researchers should complete the "DOL Project Close-Out Checklist."

### **Submit a modification request**

8. Approved researchers should complete and submit the "STRUDL Project Modification Request" form if the approved researchers need more time, change the data access, and/or update research plan.
  - a. If approved, the approved researchers may continue their research.
  - b. If denied, the approved researchers must follow the steps for a forced stopped project.

### **Force to stop project**

9. Approved researchers should complete the "DOL Project Close-Out Checklist."
10. DOL staff will review and address the reason for the forced closure of the STRUDL project.

## *Disclosure review documents*

This section provides an overview of forms required for the disclosure review process. Approved researchers should become familiar with these forms as they are required throughout STRUDL.

- **Disclosure Review Checklist** – The purpose of this document is to evaluate the anticipated level of effort and identify high-risk variables or features in the data for the disclosure review of datasets. Applicants and approved researchers are required to complete the checklist both during their initial application to access DOL data and after they have been granted access. Questions that are applicable only after applicants gain access as approved researchers will be clearly indicated as such on the form.
- **STRUDL Project Modification Request** – The purpose of this document is for approved researchers to formally request a modification from DOL.
- **Disclosure and Dissemination Review Determination** – The purpose of this document is for the DOL staff to make a final determination regarding granting data access, releasing project outputs, or approving dissemination products after conducting a disclosure review at various stages of the project.
- **DOL Project Close-Out Check List** – The purpose of this document is to outline the procedures for approved researchers to appropriately close their STRUDL projects and return all DOL materials and equipment, such as laptops. It is crucial to adhere to this procedure to avoid any potential violations of the NDA, which could lead to researchers being prohibited from STRUDL in the future. All outputs, including research papers and reports, must explicitly include the following disclaimer.
  - *Disclaimer:* This research was conducted with restricted access to U.S. Department of Labor (DOL) data. The views expressed here are those of the author and do not reflect the views of the DOL. This research was performed under DOL Secure Transfer, Restricted-Use Data Lake Project Number XXXX-XXX.

**IMPORTANT NOTE:** All approved researchers on the STRUDL project must complete the “DOL Project Close-Out Checklist.” Failure to do so may violate the NDA and result in being barred from future access to STRUDL.

## **Best practices for disclosure avoidance**

This section outlines the recommended best practices for protecting outputs derived from confidential data at DOL. Approved researchers are encouraged to maintain an open dialogue with DOL disclosure review staff throughout the course of their research project to ensure proper implementation of these best practices and statistical disclosure control methods (i.e., methods of data privacy and confidentiality).

Definitions and detailed information on the statistical disclosure control methods can be found in the training materials found on the STRUDL website.

## *General practices*

The following are general policies for any outputs based on confidential data:

- DOL will reject all results that are outside the scope of the original STRUDL application.

- When possible, approved researchers should prioritize model or distributional output rather than tabular output. DOL does not permit outputs derived from confidential data that contain individual record-level information.
- Approved researchers should ensure they do not reveal any information in their dissemination that could identify an individual, household, or establishment.

### *Practices for frequency tables*

The following are recommendations when releasing tables containing tabular or frequency outputs:

- When possible, prioritize using marginal tables over tabular outputs with more than two-dimensional tables of variables. In other words, a multi- or  $k$ -dimensional output, where  $k$  is any integer greater than two, poses a higher risk than marginal tables.
  - For example, the dataset includes the variables age, education level, and employment status. The approved researcher should consider reporting the marginal distributions of age and education level, age and employment status, and education level and employment status separately, instead of presenting a three-way table that enumerates the exact number of records for each combination of the three variables.
- High-risk continuous variables, such as age, should use appropriate break points in frequency output to ensure that counts and proportions are not based on fewer than  $k$  observations.
- Suppress counts and proportions based on fewer than  $k$  observations.
- Ensure that unweighted counts of suppressed subpopulations cannot be deduced from reported totals.
  - For example, consider a frequency table showing that a county has 7 health care and social assistance establishments, 10 accommodation and food services establishments, and 3 educational services establishments, with a total of 20 establishments. The approved researcher must suppress at least two of the four pieces of information, otherwise it is possible to deduce a suppressed count. For instance, an approved researcher could suppress the 3 educational services establishments and 20 total establishments or suppress the 3 educational services establishments and 7 health care and social assistance establishments.

### *Practices for moments and distributional tables*

The following are recommendations when releasing tables containing moments and distributional outputs:

- Do not include medians or percentiles in tabular output. Instead, consider using the average of the  $k$ -closest observations to median or percentile, or trimmed means.
- Suppress all statistics based on  $k$  or fewer observations.
- Top-code and bottom-code all high-risk continuous variables (e.g., age).



### *Practices for other distributional output*

The following are recommendations when releasing distributional outputs along with more specific distributional outputs, such as histograms.

- Provide sample sizes and disclosure statistics for each sample and subsample used to create estimates as well as for all implicit samples created from the project.
- Suppress any distributions in the categories below based on fewer than  $k$  observations.
- Investigate the min/max and medians of the distributions and ensure that (1) these exact values are not apparent, and (2) these cannot be linked to fewer than  $k$  observations. An exception to this min/max rule is structural bounds. For example, an age range starting at 0, or an income range capped at \$100,000 for the research question. It is possible that the data do not meet structural bounds because of data quality issues. Confirm that the data match expectations for structural bounds.
- **Histograms** – Ensure no bin contains fewer than  $k$  observations and does not reveal the exact median, minimum, and maximum. Specifically, consider aggregating the tails of the distributions.
- **Quantiles** – Suppress quantiles with fewer than  $k$  observations and do not reveal the exact median, minimum, and maximum. Approved researchers are recommended to round reported values to minimize specificity. The following are a couple suggested rounding rules, but approved researchers should consult with DOL disclosure review staff to ensure the appropriate implementation of disclosure avoidance measures.
  - When rounding to the nearest digit or significant figure, probabilistic rounding should be used. An example is a person's income of \$596 needs to be rounded to the nearest \$10. Because the significant figure is six, there is a 60 percent probability of rounding up to \$600 and a 40 percent probability of rounding down to \$590.
  - When deciding which significant figure to round, use the following recoding scheme:
    - \$0 rounded to \$0
    - \$1–\$999 rounded to nearest \$10
    - \$1,000–\$49,999 rounded to nearest \$100
    - \$50,000+ rounded to nearest \$1,000
- **Kernel Densities** – Do not reveal the exact median, minimum, and maximum values and consider aggregating the tails of the distributions. If the purpose of releasing a kernel density plot is to show the full distribution, the previously mentioned rules can be relaxed if BOTH of the following conditions are satisfied:
  - The plot does not include X-axis values. In other words, no specific value on the graph can be associated with a particular number.
  - No other summary output is released elsewhere that could be used to deduce the range of values (i.e., median and quantiles).
- **Scatterplots** – Scatterplots where each point corresponds to a single record, such as an individual or establishment, are not permitted because they present a disclosure risk. Consider grouping records based on other characteristics. If these groups include fewer than  $k$  observations, approved research should not include them.
- **Maps** – Point maps are not permitted if each point corresponds to a single record, such as an individual or establishment, because they present a disclosure risk. Consider grouping records based on other characteristics. If these groups include fewer than  $k$  observations, approved research should not include them.

- When creating choropleth maps for geographic areas smaller than the state level, follow these guidelines:
  - Suppress data aggregated to any category with fewer than  $k$  total observations, or statistics based on fewer than  $k$  total observations.
  - Report numeric data or statistics in sequential form (e.g., bins) instead of the continuous forms. If numeric data is not binned, consider rounding and adding noise to these values to minimize disclosure risk.

### *Practices for model output*

The following are recommendations when releasing model outputs:

- Ensure model outputs have at least  $k$  degrees of freedom.
- Categorical predictors must contain at least  $k$  observations in each class.
- Models should not use data exclusively from vulnerable subpopulations, such as individuals with veteran or disability status.

### *Practices for other output*

If the intended project output type is not covered in this section, approved researchers should ensure that their STRUDL project proposal outlines the planned statistical disclosure control methodology. Approved researchers are also encouraged to reach out to DOL disclosure review staff for guidance. For suggestions on other elements to consider, please refer to “Other general best practices to consider” subsection.

### *Alternative statistical disclosure control approaches*

Approved researchers may benefit from adopting state-of-the-art statistical disclosure control methods like synthetic data generation and formal privacy or differential privacy in certain situations. For example, an approved project that adopts differential privacy for extremely sensitive data may have a higher probability of approval. Another example is synthetic data generation may allow approved researchers to release pseudo microdata at the observation level, which is not allowed under the policies outlined above. Potential applicants and approved researchers are encouraged to consult the STRUDL website to learn more and contact DOL during the application process to learn if these methods are a good fit for the potential or approved STRUDL project.

### *Other general best practices to consider*

The following examples illustrate outputs that are comparatively safer and more likely to receive DOL disclosure review approval or expediting the approval process:

- Sample sizes that are at least an order of magnitude larger than the recommended thresholds.
- Subpopulations that are not considered sensitive (examples of sensitive populations are individuals with veteran or disability status).
- Results based on categorical data instead of continuous data.
- Data that do not include longitudinal data or data with a frequency greater than quarterly within a year.
- Nonspatial data, or spatial data presented at the state level or higher.

- Data that do not include survey weights information.

## **Disclosure avoidance review and disclosure review package submission**

This section covers how approved researchers should prepare their project outputs for disclosure avoidance review and complete their disclosure review package.

### *Practices for publications*

The following are recommendations when releasing research results in publications:

- All statistics referenced in the narrative must be included in tables and figures.
- Do not mention features of the data you observed while in the STRUDL environment unless this information was included in your proposal or has undergone disclosure review.
- Do not mention dates (including coarsened information such as seasons), geographical units, or unweighted sample sizes without written permission.
- Discuss inclusion and exclusion criteria for subpopulation analysis in a way that does not inadvertently identify small cells or extreme cases (fewer than  $k$  observations, extreme cases or outliers, and unweighted counts that could be subtracted from totals).

### *Clearing results*

The following are recommendations to increase the likelihood of a timely approval for disclosure review approval.

- Submit only the results that will be included in the final dissemination products. Intermediate outputs are discouraged as they may lengthen the review process.
- Be mindful of the number of outputs that require review. Do not submit separate disclosure review packages for the same project because the outputs are tracked cumulatively.
- Allow for a minimum of 90 days from project end date to receive the final cleared files for a typical release of the output.
- Account the disclosure review timeline when planning your project. DOL will not expedite the review process for most reasons, such as conference deadlines or the scheduled STRUDL project end.
- Maintain open communication with the DOL disclosure review staff to minimize the need for catching up on STRUDL project details, prevent misunderstandings, and expedite the approval process for the disclosure review package.

### *Clearing other files*

For projects that use data from external agencies, DOL must comply with the disclosure rules and policies of those respective agencies. Approved researchers with projects that involve data from external agencies are encouraged to contact DOL disclosure review staff early in the project to assess if there are additional review requirements.

### *Disclosure review approval process*

The disclosure review approval process requires researchers to submit their disclosure review packet form and associated documents. To ensure a timely approval process, approved researchers must submit a complete disclosure review package at least 90 days before the project's close date. A complete package should include the following:

- "Disclosure and Dissemination Review Determination" form
- "Disclosure Review Checklist" form
- Research outputs, accompanied by a clear "Read Me" document that outlines the research outputs requiring review

### *Dissemination review approval process*

Once the research outputs have been approved, approved researchers can proceed with creating their proposed dissemination materials. If approved researchers are uncertain about the final dissemination products, they should present examples of dissemination products they are considering, specifying the expected formats (e.g., presentations, journal publications, reports, figures, etc.). Along with submitting the proposed dissemination materials, approved researchers must also complete the "Disclosure and Dissemination Review Determination" form.

# Appendices

## Glossary

### *STRUDL terminology*

- **academic or external researcher** – subject matter experts, analysts, or practitioners that use and apply their expertise on DOL data that are from an academic institution or nonprofit organization
- **applicant** – researchers and/or data users who undergo or are interested in undergoing STRUDL application process
- **approved researcher** – applicants who successfully applied for STRUDL
- **recipient** – academic researcher’s respective organization (as defined in the STRUDL nondisclosure agreement)

### *Technical terminology*

- **bottomcoding or topcoding** – changes or encodes the information that is below a selected lower bound or above a selected upper bound to those bounds, respectively
- **noise infusion, adding noise, or sanitization** – adds or subtracts random values based on a probability distribution
- **rounding** – changes the values based on certain criteria that shortens the value
- **suppression** – removes information from the data based on certain criteria