

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 at two points in the Secure Transfer, Restricted-Use Data Lake process: during their initial application to access U.S. Department of Labor data and after they have been granted access. Questions on the form that are only applicable after applicants gain access as approved researchers will be clearly indicated as such on the form.

Please X the appropriate box and fill in today's date:

Application | Date of Submission:

Approved Project for Disclosure Review | Date of Submission:

Briefly describe the project and data, such as the motivation and goal analyses.

If possible, list all variables intended for public release and rank those variables as high, medium, and low priority for the goal analysis. If there are too many variables to list, provide a summary of the variable groups for each priority rank, and why they are essential (or not) for the goal analysis.

Note: Please provide full descriptions for each variable or provide a data dictionary that properly describes the variables. If there are dummy variables, combine the variables and describe the new variable.

List any variables in the previous question that would be considered high disclosure risk if released on their own and if they are ranked as high or medium priority for the goal analysis (e.g., labor violations).

List any known public datasets that have overlapping or similar information to the list of variables intended for release.

Variable Information	<i>Note: Fill out this section on variables for public release.</i>
Type of data (e.g., microlevel data or summary statistics)	
Number of variables	
Number of categorical variables (and values and assumed bounds)	
Number of continuous variables (and range of values and assumed bounds)	
Survey weights	
Longitudinal information (i.e., time information)	

Spatial (i.e., location information)	
Conditional variables (i.e., any variables that are dependent on other variables)	

Privacy Threat	<i>Note: Only fill this section out after data access is granted.</i>
Number of uniques from each categorical variable	
Variables that would make the high-risk variable listed (if any) more disclosive? (e.g., labor violations and demographic details)	
Number of uniques from the combination categorical variables in the previous question	
Number of uniques from a full-cross tabulation of categorical variables (Note: Only if the dataset is of reasonable size)	
Distribution of continuous values (i.e., Are any of the variables highly skewed?)	

Data Collection Instrument Disclosure with Record Retention Periods

OMB No.: 1290-0NEW

Expiration: TBD

CEO Form No.: ST-133

Public reporting burden for this form is estimated to average 2 hours per response. The burden estimate includes the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and submitting the form. This collection of information is required to obtain or retain benefits and will be retained for 60 days. You are not required to respond to this collection of information unless it displays a valid OMB control number. Please send comments regarding the burden estimate or any other aspect of this collection of information to the U.S. Department of Labor, OASP S-2218, 200 Constitution Ave NW, Washington, DC 20210 or STRUDL@dol.gov and reference OMB control number 1290-0NEW. **NOTE: Please do not send your completed form to this address.**