FIRST THINGS FIRST

Welcome to the DAP-TS. This system is intended to walk users through the process of preparing a DAP that can fulfill the requirements for data sharing and data dissemination of IES study data.

Purpose of the Disclosure Analysis Plan (DAP)

Any data collected for the Department of Education must, by law, ensure the privacy of respondents. Therefore, any data collection intended for data dissemination (microdata, data tool, report, tables, etc) must first be reviewed and analyzed to determine whether the data require any data perturbation (masking, suppression, collapsing, swapping, etc). Therefore, the IES Disclosure Review Board requires a Disclosure Analysis Plan (DAP) prepared that explores the sources and potential for disclosure risk as well as a solution to reduce or mitigate for disclosure risk.

A DAP reports on sources of disclosure risk and proposes a solution to eliminate, or reduce risk to an acceptable level. The DAP must include a section that provides the analytical motivation for the solutions being proposed. For example, cross-tabulations can be used to illustrate the need for recoding, and lists of the largest values of a highly skewed distribution may show the need for top-coding. The confidentiality analysis proposal should include a discussion of external files, whether or not they exist. The intention to provide post-SDC edit checks that include frequencies, etc., to evaluate the safe file should be stated. Finally, the DRB requires that all of the steps to mitigate risk also limit data distortion. Thus an analysis of pre and post data perturbation is normally conducted to satisfy the need to maintain the integrity of the data while safeguarding privacy.

Typically, all studies require disclosure analysis with perturbation (PII suppression and data swapping at a minimum) though there are possible exceptions:

- Field test, pilot, and feasibility studies are often exempt from confidentiality concerns. The purpose of these studies is to inform for the main studies. The data and results of the study are not shared in a public arena. Nevertheless, safeguards, such as PII suppression are still enforced;
- A study that has previously undergone IES approved SDC treatment that collects supplemental data that provide no additional identifying data may be exempt from new SDC treatment; and
- A special study intended solely for federal internal use may be exempt from disclosure analysis.

Even if you are conducting one of the above studies, you should still inform the chair of the DRB and prepare a summary of why you believe a DAP is not necessary. There may be feasibility studies, or any study covered above that may actually require SDC treatment due to the particular circumstances with these studies.

The next section deals with the preparations you should undergo before using the DAP-TS. Gather materials: your winning proposal, the sampling and weighting plans, the analysis plans, and the dissemination plans. The DAP-TS encourages cutting and pasting documents into the DAP study template in an orderly, consistent manner for standardized DAPs for DRB review and approval.

Potential techniques for Statistical Disclosure Control

Typically, the DRB would prefer the standardized statistical disclosure control (SDC) approaches to analyzing the data for risk, and for implementing the SDC tools for handling the risk. Currently, there are a number of acceptable SDC techniques and a few approved software programs that should be considered for incorporating into the DAP and implementing for the analyses. ***FRIL*** is a probabilistic matching software previously supported by the CDC. The IES DRB has a FRIL manual available to assist users. ***InitialRisk (Initial Risk Analysis)*** software provides a user-friendly software that facilitates the review of the data for risky variables and combinations of variables. ***DataSwap*** software is used for random swapping of the data and producing a report that measures the impact of the swapping on the data in order to minimize data distortion.

The following list are the common approaches used to safeguard privacy:

- Matching against external files to identify risk (*FRIL*) and then masking data that do match;
- Initial Risk Assessment software (*InitialRisk*) to identify risky variables;
- Random data swapping routine (*DataSwap*) to add a level of uncertainty to the data;
- Variable suppression to ensure direct identifiers and sometime indirect identifiers cannot be used to identify respondents;
- Collapsing categories to ensure small cells cannot be created to identify respondents;
- Top and bottom-coding to eliminate outlier data; and
- Record suppression in cases where their inclusion will be too risky.

Materials for the DAP-TS

1. The **proposal** that describes the **SDC measures** intended to use for the confidentiality of the study data: This is a good starting point to inform how the SDC measures were originally conceived. This approach, though, may have been preliminary and more specific nature of the study would require changes to what had first been considered;

2. A **description** of the study: Don't assume that the members of the DRB are familiar with all of the IES studies. The more the DRB understands what the study is about, the better they can determine whether the SDC measures proposed are adequate. Specific aspects of the study follow in items 3 – 6;
3. **Sampling plan** and **weighting** procedures: This information may also be found in the proposal or SOW. Depending on nature of sampling and weighting, this can determine whether any aspect of these procedures may add disclosure risk;
4. **Data collection** procedures: These materials should be available in your proposal and/or statement of work. Online, telephone, pencil and paper, CAPI, etc. collections all have certain confidentiality risks that need to be addressed;
5. **Data sharing**: Use information from agreements made with any entity that will have access to the data during the life cycle of the study. Data flow, including through the internet need to provide security protocols for the systems involved; and
6. **Data reporting** and **dissemination plans:** May also be found in proposal and/or SOW. Data reports, data tools, and microdata should all described in terms of disclosure risk.

Any information found in proposals, SOWs and additional study plans should be updated reflecting any changes from the original documents.