

# Data Protection Toolkit (DPT) Resource Collection

## **Volume I** **Supporting Statement**

**OMB #1850-0803 v.261**

Submitted by  
National Center for Education Statistics (NCES)  
U.S. Department of Education

February 2020

## Contents

Submittal-Related Information.....	1
1. Background and Study Rationale.....	1
2. Study Design and Sample Characteristics.....	2
3. Data Collection.....	2
3.1 Recruitment.....	2
3.2 Information Collection Methods.....	3
4. Consultation Outside the Agency.....	3
5. Assurance of Confidentiality.....	3
6. Estimate of Costs for Recruiting and Paying Respondents.....	3
7. Estimate of Hourly Burden.....	3
8. Schedule.....	4
9. Cost to the Federal Government.....	4

## Attachments

Attachment 1: Communication Materials

Attachment 2: Instrument

## Submittal-Related Information

The following material is being submitted under the National Center for Education Statistics (NCES) generic clearance agreement (OMB# 1850-0803), which provides NCES the capability to improve data collection instruments by conducting testing—such as usability tests, focus groups, and cognitive interviews—to improve methodologies, survey questions, and/or delivery methods.

### 1. Background and Study Rationale

The Foundations for Evidence-Based Policymaking Act of 2018 (Public Law No: 115-435) requires each government agency to develop an inventory that includes basic technical information to facilitate an improved understanding of available information resources and how to access them. Furthermore, the President's Management Agenda (PMA) 2018-2021 includes data as one of the three key drivers for modernizing the Federal Government. Recognizing the critical importance of data to the U.S. economy, one of the PMA Cross-Agency Priority (CAP) Goals involves leveraging data as a strategic asset. Development of the Data Protection Toolkit is part of a U.S. Department of Education ESSIN Task Order 36 to develop an automated tool that leverages agency Information Collection Review (ICR) processes and documentation under the Paperwork Reduction Act (PRA) to populate metadata, or information about each dataset, on agency data inventories based on the tenets of a 2013 Executive Order and related Office of Management and Budget (OMB) guidance on managing Information as an asset. The work conducted as a result of the award will be performed for NCES within the Institute of Education Sciences.

Today, each agency that releases de-identified confidential data must decide for itself how it will assess the risk of re-identification and the resulting potential for harm to individuals from misuse of individual information. Privacy experts are increasingly concerned about the "mosaic effect," that is, the increased risk of re-identification due to the proliferation of electronic databases with records of individuals and enterprises and increases in computational power. The Commission on Evidence-Based Policymaking found that almost half of agencies releasing confidential data do not have a formal or documented method for measuring or mitigating the risk of re-identification. Even best-practice statistical agencies are struggling to keep pace given the growing risk due to the growing amount of information about individual people and businesses available publicly.

This task will build upon the work conducted to date by NCES by supporting the next phases of work associated with the development of a re-identification risk assessment toolkit for federal agencies that will include templates based on best practices for assessing, managing, and mitigating the risk that the identify of individuals or enterprises are re-identified from the release of confidential federal data.

The development of the toolkit will include additional discovery and synthesis of current federal best practices, developing the risk assessment framework and template(s), and developing and beginning to implement a dissemination and training strategy to help federal agencies adopt best practices—including developing a useful and user-friendly website with resources supporting risk assessment.

Anticipated activities that will occur within each of these tasks include but are not limited to:

- Conducting research on risk assessment best practices including current methods of risk assessment, management, and mitigation
- Identifying alternative methods to support increased data sharing and user access
- Facilitating the discovery of current federal best practices by working with a variety of stakeholders
- Synthesizing findings and summarizing them for documentation and implementation
- Developing risk assessment framework and template(s)
- Developing templates and tools that can be utilized by a variety of stakeholders
- Compiling a resource library
- Developing a useful and user-friendly website
- Developing a dissemination and training strategy to help federal agencies adopt best practices

NCES is the primary federal entity for collecting and analyzing data related to education in the U.S. and other nations. NCES is located within the U.S. Department of Education and the Institute of Education Sciences. NCES fulfills a Congressional mandate to collect, collate, analyze, and report complete statistics on the condition of American education; conduct and publish reports; and review and report on education activities internationally.

The NCES Statistical Standards and Data Confidentiality Staff (SSDCS) provide methodological and statistical support to NCES, as well as to federal and nonfederal organizations that engage in statistical work in support of NCES' mission. SSDCS takes the lead in developing and maintaining the statistical and survey methodology standards that ensure the quality of statistical surveys, analyses and products (2012 NCES Statistical Standards - <http://nces.ed.gov/statprog/2002/stdtoc.asp>); consults and advises on the implementation of standards for all NCES projects; conducts internal technical reviews for all NCES publications and products (see the NCES electronic publication catalog at <http://nces.ed.gov/pubsearch/> for examples of existing NCES publications and products); and coordinates the revision of the NCES Statistical Standards. SSDCS also monitors and administers confidentiality procedures (see <http://nces.ed.gov/statprog/confproc.asp> for information on the NCES confidentiality program) and related restricted-use data licenses (see <http://nces.ed.gov/statprog/instruct.asp> for information on the licensing process) for NCES data products. In addition, SSDCS consults and advises on emerging statistical issues and initiates, and monitors or participates in long-term statistical and methodological research projects, making that group well-suited to carry out and coordinate the work for this project.

## 2. Study Design and Sample Characteristics

This study consists of the gathering of best practices from leaders in the field of data confidentiality and disclosure avoidance, whether those leaders are inside or outside of the federal government. Initially, federal government employees knowledgeable in the data disclosure protection practices of their agency will be consulted. As the information is being collected from federal government employees, it is not subject to OMB-PRA. That collection, which was sent to Interagency Council on Statistical Policy (ICSP) members via email on January 17, 2020, of best practices will serve as the primary source of information used to build the materials in the toolkit. Agency heads were asked to forward the data call to someone on their staff capable of coordinating their agency's response by February 5, 2020. Following that initial effort with federal employees, this study attempts to obtain additional information from universities and organizations/businesses that deal with the collection and dissemination of data, sometimes even assisting the federal government through contractual agreements. The information from both of these collection efforts will be combined in an effort to document the policies, tools, templates, and other resources used to protect privacy in public data releases and to promote responsible access to data for research and evidence-building by the government and leaders in the field.

The Confidentiality and Data Access Committee (CDAC) of the Federal Committee on Statistical Methodology (FCSM) has identified individuals from universities and businesses who are the forefront of data confidentiality and disclosure avoidance. These organizations and individuals were selected in a non-random and non-statistical manner and have been recommended for participation because of their unique insight and depth of experience in the field. In total, there are 21 individuals (12 in the U.S. and nine internationally) that will be asked for input relating to existing resources for statistical confidentiality. Their role is to provide resources (links to materials, tools and manuals) via the data collection instrument discussed below. Among the 12 individuals within the U.S., there are five individuals from federal contracting organizations/businesses that are practitioners in statistical confidentiality. They partner with Federal agencies in the dissemination of Federal data. The contractors may be able to provide insights into resources used during the implementation of Federal guidelines. There are seven university professors who have published journal articles on statistical confidentiality and are recognized as leaders in the field. Among the participants from international community, six individuals are identified from national statistical institutes or other organizations who have established guidelines, or developed automated tools, and three university professors who have publications and experience in dealing with data disclosure issues have also been identified. Among the 21 individuals, seven are members of the American Statistical Association Privacy and Confidentiality committee.

## **3. Data Collection**

### **3.1 Recruitment**

Initial recruitment will consist of an email. Two weeks after the initial email, if no response is received, a follow-up email will be sent to the participant. A week after the follow-up email, if no response is received, a final follow-up email will be sent to the participant. No additional follow-up is planned for non-participants.

All communication materials are included in Attachment 1.

### **3.2 Information Collection Methods**

The information collection materials consist of a set of instructions, a glossary, a reference guide, and the Request Form for Existing Resources. The reference guide serves to remind the participant of the types of information that are being requested, while the glossary of terms is meant to ensure that the participants have the same understanding of terms as is intended. Finally, the Request Form asks respondents to provide URLs for any materials for which they are responsible. If they are not able to provide URLs, participants are instructed to attach the actual materials. If neither URLs nor documents are able to be provided, the participant is instructed to provide a summary of the information.

The full instrument is included in Attachment 2.

## **4. Consultation Outside the Agency**

Given the nature of this project, NCES has been in consultation with outside federal and non-federal organizations. Of particular note is the interaction with and direction of the FCSM. The FCSM is the driving force for this project and has been consulted at every stage of the process in defining the scope of the project as well as collaborating on all of the materials for this study.

The project team assembled for this effort reflects the interagency collaborative of the FCSM. Government representatives from both the Census Bureau and NCES bring expertise in data confidentiality and disclosure avoidance. The contractors selected to support this effort also are recognized in the web solutions and disclosure avoidance space. Sanametrix has a proven track record in providing customized web data collection and dissemination for federal statistical agencies, particularly NCES. Sanametrix subcontracting partner Westat has developed expertise in statistical confidentiality, including disclosure risk analyses and applying confidentiality treatments prior to delivering data to government agencies. The Sanametrix team will support this project from data collection through analysis and reporting via the Data Protection Toolkit.

## **5. Assurance of Confidentiality**

All participants will be assured that their participation is voluntary. However, the purpose of this information collection is to create a repository of best practices for data disclosure avoidance. As such, the information provided by participants will not be confidential. This lack of confidentiality will be communicated to the potential participants in all communications materials (see Attachment 1).

## **6. Estimate of Costs for Recruiting and Paying Respondents**

No monetary or in-kind payments will be made for the purposes of recruitment or completed interviews.

## **7. Estimate of Hourly Burden**

The list provided by the FCSM is a collection of single individuals from a total of 21 entities. It is estimated that it will take approximately 5 minutes to read the invitation and associated materials and decide whether or not to participate. It is further estimated that once having decided to participate, the information collection process will take approximately 20 minutes. Through email follow-up efforts, the estimated response rate is 48%, resulting in approximately 10 completed information collection requests. Table 1 below shows burden estimates for (a) recruiting participants; and (b) providing the requested information.

**Table 1. Estimate of hourly burden for Data Toolkit Information Collection**

<b>Activity</b>	<b>Number of Respondents</b>	<b>Number of Responses</b>	<b>Burden Minutes per Respondent</b>	<b>Total Burden Hours</b>
Recruitment	21	21	5	1.75
Information Collection	10	10	20	3.33
<b>Total</b>	<b>21</b>	<b>31</b>	-	<b>5.08</b>

## **8. Schedule**

Recruitment/collection for the data toolkit information will begin in March 2020. Collection will continue until all responses are received or until the beginning of April when development of the toolkit website needs to commence. Table 2 below provides the overall schedule for the project.

**Table 2. Schedule of high-level activities for Data Toolkit Information Collection**

<b>Activity</b>	<b>Start Date</b>	<b>End Date</b>
Recruitment	March 2	March 23
Information Collection	March 2	April 3

## **9. Cost to the Federal Government**

The estimated cost to prepare for, collect, and collate the results of the information collection is approximately \$7,500. The cost includes salaried labor for contractor staff and other direct costs associated with the organization of the information collection.