

# *The Researcher Handbook*

U.S. Bureau of the Census  
Center for Economic Studies  
Research Data Center's  
Handbook for Researchers

November 2009

Intentionally blank page.

## Preface

Welcome to the Census Bureau's Research Data Centers, managed and operated by the Census Bureau's Center for Economic Studies (CES) and its RDC partner institutions. We hope that your research project goes smoothly and is successful. To help accomplish this, we provide this manual so that you will know what to expect at each stage your RDC project.<sup>1</sup> You are responsible for knowing the rules, regulations, and policies explained in this manual.

*Your research project at a Census Research Data Center (RDC) affords you opportunities to carry out unique research arising from the ability to access and explore confidential microdata, but the opportunities come at a price. Your research will take place under a set of rules and limitations that will be considerably more constraining than those prevailing in typical research environments. An important purpose of this manual is to explain these constraints to you.*

*These constraints stem from the legal requirements to meet the conditions of your appointment as a researcher at the RDC—particularly the requirement to maintain confidentiality of the underlying microdata files to, which you have been granted access, and the requirement that your project benefit Census Bureau data programs.*

This manual is organized as follows.

Chapter 1 explains the legal requirements to maintain confidentiality of the microdata at the RDCs and the legal basis for your obtaining access to the data. In some cases, these requirements come not only from the Census Bureau, but also from other agencies that provide data to the Census Bureau or sponsor surveys. These requirements underlie all the policies described in the rest of the manual.

Chapter 2 explains the life cycle of an approved research project at CES and summarizes the policies and procedures you must adhere to at each stage—from your first contact with us to your project completion and beyond.

Chapter 3 explains, in detail, the policies and procedures you must follow in obtaining release of the research output you wish to remove from the secure RDC laboratory. Most importantly, it lays out policies related to protecting the confidentiality of the data and outlines the procedures you must follow to obtain release of your research results.

Several appendices provide supplemental material. Most of these are simply copies of documents that are available for download from the CES website or that will be provided to you during the course of establishing your research project account. The exception to this is Appendix G, which provides summaries of and excerpts from the laws underlying the policies in this manual.

---

<sup>1</sup> For purposes of this handbook, the term "Research Data Center (RDC)" includes all of the restricted-access facilities operated by the Census Bureau's Center for Economic Studies and its RDC partners, including the facility operated by CES at Census Bureau headquarters.

## Table of Contents

**PREFACE..... I**

**CHAPTER 1 : INTRODUCTION TO GENERAL REGULATIONS AND POLICIES..... 1**

THE LEGAL REQUIREMENT TO MAINTAIN CONFIDENTIALITY OF DATA AT AN RDC..... 1

CONDITIONS FOR RESTRICTED ACCESS TO MICRODATA—YOUR SPECIAL SWORN STATUS ..... 2

**CHAPTER 2 : THE LIFE CYCLE OF RESEARCH AT THE RESEARCH DATA CENTERS .....3**

USING THE RDC..... 3

*Obtaining Access to the Secure RDC Office..... 3*

*Obtaining Access to RDC Computer Facilities ..... 3*

*Useful Documents..... 4*

*Scheduling Lab Time ..... 4*

*Annual Renewals: Special Sworn Status (SSS) and Data Security Training ..... 4*

*Annual Project Progress Reports ..... 5*

*Maintaining Physical, Computer, and Data Security..... 5*

*Reporting Data Breaches ..... 5*

*Maintaining Confidentiality and Obtaining Release of Research Output ..... 6*

*Getting Support ..... 6*

*Adding New Researchers to Project ..... 6*

*Adding an Additional RDC site to the Project..... 6*

*Obtaining New Data or Carrying Out Different Research..... 7*

*Seminars and Conference Presentations ..... 7*

*Producing CES Discussion Papers ..... 7*

*Providing Papers and Research Reports..... 7*

*Requesting Extensions on Projects..... 7*

EXIT PROCESS ..... 8

*Documenting Benefit to the Census Bureau ..... 8*

*Writing the PPC..... 8*

*Documenting Your Research..... 8*

*Backing Up and Archiving Data and Programs ..... 8*

*Returning Badge or Keys..... 9*

AFTER EXIT..... 9

*Returning for Follow-up Work ..... 9*

*Revise and Resubmit..... 9*

*New Project..... 9*

*Exit Interview..... 9*

*Providing Papers..... 10*

**CHAPTER 3 : DATA ACCESS, CONFIDENTIALITY, AND DISCLOSURE ANALYSIS**

**POLICIES AND PROCEDURES ..... 11**

GENERAL POLICIES ON MAINTAINING CONFIDENTIALITY OF DATA AND CLEARING RELEASE OF RESEARCH OUTPUT ..... 11

DISCLOSURE AVOIDANCE REVIEW: RULES AND CONSIDERATIONS..... 13

CES POLICY ON QUALITATIVE DISCUSSION OF REGRESSION RESULTS OUTSIDE RDCS AND BY TELEPHONE..... 15

DISCLOSURE ANALYSIS ..... 15

*Basics of Disclosure Analysis..... 16*

*Concentration Rules ..... 16*

*Cell Suppression in Primary and Secondary Disclosure Analysis ..... 17*

DISCLOSURE ANALYSIS RULES FOR CES AND THE RDCS ..... 19

*Disclosure Analysis Rules for Tabular Output ..... 19*

Tables From Economic (Business and Establishment) Data .....	20
Tables From Households and Individuals (Demographic Data) .....	22
Decennial Census and American Community Survey.....	22
<i>Disclosure Analysis for Other Distributional Characteristics</i> .....	23
Quantiles .....	23
Histograms, Kernel Densities, and Scatterplots .....	24
Maps .....	24
<i>Disclosure Analysis for Model Output</i> .....	24
<i>Disclosure Analysis for LEHD (Mixed Economic and Demographic) Data</i> .....	25
PREPARING FILES FOR DISCLOSURE AVOIDANCE REVIEW .....	26
CLEARING RESEARCH RESULTS.....	26
CLEARING OTHER FILES—PROGRAMS AND RESEARCH NOTES .....	28
<i>Roles and Responsibilities</i> .....	30
<i>File Preparation for Releasing Files</i> .....	30
<i>Guidelines for preparation of files</i> .....	31
PREPARING THE CLEARANCE REQUEST MEMO .....	32
REQUESTING FILE CLEARANCE USING THE DISCLOSURE AVOIDANCE REVIEW TRACKING SYSTEM (DARTS) IN THE CES PROJECT MANAGEMENT SYSTEM .....	34
DECISIONS AND RELEASE OF OUTPUT .....	36
RELEASE OF OUTPUT FROM OTHER AGENCIES’ DATA .....	36
<i>Release of Output Using AHRQ or NCHS Data</i> .....	36
<i>Projects using AHRQ or NCHS data exclusively</i> .....	36
<i>Projects using Combinations of Census, AHRQ and/or NCHS data</i> .....	36
BIBLIOGRAPHY .....	38
<b>APPENDIX A : GLOSSARY</b> .....	<b>40</b>
<b>APPENDIX B : ADRF</b> .....	<b>42</b>
<b>APPENDIX C : DATA BREACH MEMORANDUM</b> .....	<b>50</b>
<b>APPENDIX D : CLEARANCE REQUEST MEMO</b> .....	<b>53</b>
<b>APPENDIX E : USER PROVIDED DATA MEMO</b> .....	<b>56</b>
<b>APPENDIX F : APPLICATION TO ADD PERSONNEL TO AN APPROVED PROJECT</b> .....	<b>57</b>
<b>APPENDIX G : LEGAL INFORMATION</b> .....	<b>58</b>
CENSUS BUREAU LEGISLATION—TITLE 13 U.S. CODE.....	58
<i>General Information</i> .....	58
<i>Protection of Confidential Information—Sections 9 and 214</i> .....	58
<i>Authority for Temporary Staff (Special Sworn Status)—Section 23(c)</i> .....	59
<i>Data Programs Authorized by Title 13 Chapter 5</i> .....	59
<i>Other Title 13 Data Programs</i> .....	60
PROTECTING FEDERAL TAX INFORMATION (FTI)—TITLES 26 AND 18, U.S. CODE.....	62

## Chapter 1: Introduction to General Regulations and Policies

This chapter explains the legal requirements you must meet to maintain the confidentiality of microdata at the RDC, and the legal basis for your obtaining access to the data. You will see that in some cases the requirements are not only those of the Census Bureau, but also those of other agencies that provide data to the Census Bureau or sponsor surveys. These requirements underlie all the policies described in the rest of the manual.

### ***The Legal Requirement to Maintain Confidentiality of Data at an RDC***

The microdata underlying the Census Bureau's data programs are confidential under Title 13 of the U.S. Code (U.S.C.). These microdata were collected directly by the Census Bureau in various surveys and censuses, and they include statistics that quantify the activities of individual business establishments, firms, households, or individuals. Disclosure of confidential information obtained from Census Bureau surveys or censuses is illegal under Title 13. Officially or otherwise, the Census Bureau may not release publicly any data that reveal the identity of individual respondents or information about them. The Bureau gives a pledge of confidentiality to the respondents, the purpose of, which is concern for respondent privacy. In addition, there are practical reasons to ensure confidentiality: publication (even inadvertent) of confidential microdata records would most likely reduce cooperation with data collection programs.

To become a researcher at an RDC, the Census Bureau must grant you Special Sworn Status (SSS), as described below. Under SSS, you are subject to the same legal penalties as regular Census Bureau employees for disclosure of confidential information. The penalties are a fine of up to \$250,000, imprisonment for up to five years, or both. Moreover, your obligation to maintain confidentiality of any microdata you have accessed continues for the rest of your life.

Data collected under Title 13 include virtually all Census Bureau microdata sets available at the RDCs; however, the laws of other agencies may also be applicable, and in such cases you will be required to abide by the legal requirements and confidentiality procedures of *both* agencies. Two examples follow.

Your project may be approved to use administrative data that another agency supplies to the Census Bureau under legal agreements. A common case of this is Federal Tax Information (FTI). FTI is provided by the Internal Revenue Service to the Census Bureau for use in carrying out programs authorized under Chapter 5, Title 13 U.S.C. These programs are the Economic and Decennial Censuses and their related survey programs. The relevant laws are Titles 18 and 26 U.S.C. Appendix G provides a brochure that summarizes requirements that must be met in carrying out projects that use FTI; it also provides references to more detailed information. *All* researchers at an RDC must undergo training in the proper use and protection of FTI, whether or not their projects use FTI, because in an RDC researchers might inadvertently see FTI.

Your project may be approved to use confidential microdata collected by the Census Bureau under Title 13 for another agency. A common case of this is the Current Population Survey, which is collected by the Census Bureau under Title 13 for the Bureau of Labor Statistics. Researchers may legally use these data for statistical and research purposes only as stated in Title 42, U.S.C., Section 3789g.

We will inform you of any additional requirements.

## **Conditions for Restricted Access to Microdata—Your Special Sworn Status**

The Census Bureau has granted you Special Sworn Status (SSS) to carry out a specific research project that has been approved by the Census Bureau. As a researcher with SSS, you are given restricted access to the specific data required for your approved project. Under Title 13, these data may be used only for *statistical purposes*; that is, to generate an aggregate description of a group of persons or businesses. Data collected for statistical purposes may not be used to make decisions regarding a specific individual or business. For more information, see Appendix A on terminology. SSS is authorized by Title 13 U.S.C. 23 (c) “to assist the Bureau of the Census in performing the work authorized by this title.” The Census Bureau may provide SSS to an individual when...

- an individual has expertise or specialized knowledge that can contribute to the accomplishment of Census Bureau projects or activities or engages in a joint project with the Census Bureau, or
- an individual is employed by an agency/organization performing a service for the Census Bureau under contract or providing information to the Census Bureau for statistical purposes, or
- Federal law requires an individual to audit, inspect, or investigate Census Bureau activities.

*It is important to understand that the Census Bureau's use of researchers with SSS is to further its agenda. The discretionary authority to swear in researchers with SSS rests with the Census Bureau. No individual has a “right” to appointment nor is the Census Bureau obligated to appoint a person who may believe the Census Bureau needs his or her expertise. In approving your research project, the Census Bureau has made the judgment that your activities can contribute to the success of Census Bureau data programs or lead to improvements in them. We have found that granting access to microdata by qualified researchers for specific projects results in mutual benefits to the Census Bureau, the researchers, their scientific fields, and provides information useful to decision makers.*

## Chapter 2: The Life Cycle of Research at the Research Data Centers

This chapter provides a summary and overview of what you can expect in doing research at the Census RDCs. One of the purposes of the chapter is to give you an understanding of the perhaps unaccustomed constraints on your activities, particularly if you are used to carrying out research in an academic environment. As explained in the previous chapter, these constraints arise mostly from two legal requirements: to protect confidentiality of the data and to produce benefits to Census Bureau data programs.

### ***Using the RDC***

The following summarizes policies governing your use of the RDC: trainings, renewing SSS, maintaining security, maintaining confidentiality, obtaining release of research output, obtaining new data or carrying out different research, progress reports, and giving seminars or conference presentations.

### **Obtaining Access to the Secure RDC Office**

The RDC office is tied to a security system that meets Census Bureau specifications. You will be given the proper items and procedures for entry (e.g., an alarm code, badge, and the like). Your RDC Administrator (RDCA) will guide you through the security and access procedures particular to your RDC.

You will be provided a badge issued by the Census Bureau that will give you access to the main RDC door. This badge is the property of the U.S. Government. As stated on the back of the badge, “This credential is the property of the United States Government. Counterfeiting, altering, or misusing violates section 499, Title 18 of the U.S. Code.” If you lose your badge, immediately notify your RDCA. Moreover, at the end of the project (or once you stop working on a project), you will be responsible for returning your badge.

### **Obtaining Access to RDC Computer Facilities**

The standard computing equipment at RDCs are X-Terminals (thin-client computing devices) that are used to communicate with secure servers housed in the Census computer center. Through the thin client, you will have remote access — only from within the RDC secure facility — to the CES servers containing your project data and files. All data storage and data processing are performed on these servers. No Title 13 or Title 26 data files are physically on site at any RDC.

The RDC computing environment has the following characteristics.

- Data are stored in a format determined by CES to be best suited to support research activities (e.g., as SAS data sets)
- A limited range of statistical software is available for researcher use
- Access to the central servers is accomplished via a thin client located in the RDC
- All data process work is conducted on the central servers, and no actual data processing is performed on the RDC thin client
- All thin clients are currently running Red Hat Enterprise Linux
- No portable media devices (e.g., floppy, CD/DVD, or USB drives) are permitted in the lab

- The RDC computer network is isolated from other networks, which means that within the laboratory you will have no access to e-mail or the internet
- You will be accountable for your computer use through the use of passwords and system logs

Each researcher will get an individual project account on the server. You should have already received an Account and Data Request Form, signed it, and sent it to your RDCA. Your administrator will coordinate your account activation with you and the CES data staff. ***Please note that your project's duration will officially begin when the first person on your project logs into his/her project account for the first time.***

## Useful Documents

You will receive a packet of materials, which includes a copy of this manual and a copy of the *Thin Client User Guide* to help you become oriented to the system.

Many useful working papers and data documentation can also be found online at the CES website at [www.ces.census.gov](http://www.ces.census.gov). Here you will also find templates of forms available for download. These include the *Clearance Request Memo*, which is used to release information from the RDC (see Appendix D) and the *User Provided Data Memo*, which must accompany any outside data you would like to bring into the lab (see Appendix E). (Any uploaded data must be part of your approved project, or required to help you obtain the benefits proposed in your approved project.)

Other useful documents available in the RDC may include data documentation, survey forms, software manuals, and/or SAS or Stata programming books. These materials will vary across RDCs. Ask your RDCA what is available at your RDC. Some data documentation is available in the directories for the corresponding data sets on the RDC server.

## Scheduling Lab Time

At some RDCs, you may need to schedule your laboratory time. In your proposal, you will have specified the term of your project and some pattern of use.

- If your lab use will be intermittent, you should schedule your visits well in advance. If you do not, it may cause problems in scheduling other researchers and could result in your not having access to a thin client at your desired times.
- It is especially important for you to be aware of when the term of your agreement will expire. You can check this date by looking at your project record in the CES online management system. Your RDCA can also be able to provide you with this information.

***We will give priority to researchers who adhere to their schedules and schedule visits in advance.***

## Annual Renewals: Special Sworn Status (SSS) and Data Security Training

It is required that your SSS be renewed every year. You will not be able to enter the RDC without appropriate training.

The following Awareness Training modules need to be completed annually when notified by your RDCA that the new fiscal year training is available; this is typically in late autumn. The CES online management system will also send reminders to renew each of the following training

modules thirty days before they are due. Please remember to keep your e-mail address up to date in the management system so that you can receive these reminders and other important notices.

- Title 13 Awareness Training
- Title 26 Awareness Training
- Information Technology Security Awareness Training
- Rules of Behavior for Information Technology Systems Awareness Training

## Annual Project Progress Reports

You must submit annual progress reports on your project to your RDCA. The report should describe the progress that is made towards achieving the benefits to the Census Bureau proposed in your research project. This report should also include any new benefits that may have been realized as a result of your work on your RDC project. Your RDCA will provide you with the template.

## Maintaining Physical, Computer, and Data Security

The RDC lab has physical security procedures, computer security procedures, and a related set of policies that you must follow. The security and related policies stem from the need to meet the conditions under, which you have been granted access to the data for your project and the need to protect the confidentiality of the data. The official CES policies on Office Security, Computer Security, Data Access, and Confidentiality are listed in the Account and Data Request Form that you signed. A generic version of this form is included here as Appendix B.

The RDC office has a security system that meets Census Bureau specifications. Your RDCA will give you the proper items and procedures for entry. Your RDCA will also review with you the security policies specific to that RDC, including what to do and who to call in an emergency, what to do if you accidentally set off an alarm, building evacuation routes, etc.

## Reporting Data Breaches

In the event of a data breach or if you *suspect* a data breach, you must notify the Census Bureau immediately. A complete definition of a breach and the appropriate actions are shown in Appendix C on p. 50. In summary, there are two types of data breach: a breach involving personally identifiable information (PII) and breaches of all other types.

In the event of a PII breach (or a *suspected* PII breach), you must notify the Census Bureau's Computer Incident Response Team at either [BOC.CIRT@census.gov](mailto:BOC.CIRT@census.gov), 1-877-343-2010, or 1-301-763-5141.

In the event of any other type of breach (or *suspected* breach), you must notify a CES management official. Contact should be made in the order listed, contacting the next only if the previous is not available.

- CES Division Chief: Ron Jarmin  
ron.s.jarmin@census.gov or 301-763-1858
- Assistant Division Chief for Research: Lucia Foster  
lucia.s.foster@census.gov or 301-763-2861
- Assistant Division Chief for Research Support: Mark Mildorf  
mark.f.mildorf@census.gov or 301-763-6449

- Lead RDC Administrator: Lynn Riggs  
tammy.l.riggs@census.gov or 301-763-8840
- Your local RDC Administrator: contact information is available at  
[http://www.ces.census.gov/index.php/ces/contact\\_us\\_ces](http://www.ces.census.gov/index.php/ces/contact_us_ces)

## Maintaining Confidentiality and Obtaining Release of Research Output

Chapter 3 details CES policies on maintaining confidentiality of the data. It also provides instructions for obtaining release of research output. CES policies regarding disclosure analysis are also listed in the Account and Data Request Form (Appendix B).

## Getting Support

You may have data or subject matter questions during the course of your project. **Please begin by addressing questions to your RDC Administrator.** The administrator may be able to answer your question or may put you in touch with a CES or other Census Bureau employee who can answer the question(s). We believe that having contact with CES and Census Bureau experts can provide benefits both to your project and to the Census Bureau's data programs; however, we need to manage this process so that you obtain the support you need in a way that does not overburden the staff, and so that we can maintain a record of support given and of your suggestions for improvements to Census Bureau data programs.

You will have access to the lab during non-business hours (including holidays), but support during these times will be limited or non-existent. Sometimes support will be limited during business hours, as well, for example, if the government is shut down due to severe weather conditions.

## Adding New Researchers to Project

The Principal Investigator on a project can request addition of new researchers to the project. Justification for such additions will be formally itemized by the PI in the *Application to Add Personnel to an Approved Project* form (see Appendix F); this form will be e-mailed to you by your RDCA. New researchers must have accounts on the CES website before this form is formally submitted. Approval of additions is the discretion of CES. If the addition is approved, then the new researcher will apply for Special Sworn Status with the assistance of your RDCA.

## Adding an Additional RDC site to the Project

If one of the researchers on your project lives or works closer to a different RDC than the rest of the project team, then you can request an additional RDC be a secondary work site for your project. Please note that the project must show that there is a "substantial presence" at the secondary site, meaning that at least one of the project researchers will be accessing the project at the secondary site for a significant period of time. Secondary sites will **NOT** be approved for brief periods of access, such as two weeks of access while on a vacation or research trip. The RDCA at the project's *primary* site is responsible for administration of the project. **Moreover, fees may apply at each RDC.**

The PI on the project needs to send the request to the RDCA of the primary site in order to add a secondary site. The primary site RDCA will notify the RDCA at the new site. The Executive

Directors at both sites need to approve of the addition. If the request is approved, the RDCAs will make the necessary arrangements for researcher access.

### **Obtaining New Data or Carrying Out Different Research**

Under your Special Sworn Status you may carry out only your approved research project using the data specified in your approved research proposal and your project agreement. If you wish to conduct research using a different/additional Census data set, you must submit a new proposal to be sent through the regular review process. The same applies to a new line of research that is not covered in your approved research proposal. If you have a new idea, please discuss it with your RDC Administrator. The RDCA will work with CES to determine if the idea falls within the scope of your existing project. *CES will not release output that does not fit in the current scope of the project.*

### **Seminars and Conference Presentations**

We encourage you to present your work in seminars and at conferences, including seminars in your local area; however, all output presented to persons without SSS status must undergo disclosure review and clearance. Any presentations or reports must contain the standard disclaimer shown on page 13.

### **Producing CES Discussion Papers**

You are required to submit at least one research paper produced as part of the project for inclusion in the CES Discussion Paper series (or another appropriate Census Bureau series as approved by CES). *Inclusion in such series does not preclude inclusion in any other paper series (e.g., university or NBER working paper series).* For submission procedures, contact your RDC Administrator.

### **Providing Papers and Research Reports**

You must provide copies to CES (via your RDCA) of all published research papers and reports resulting from your RDC research project. These are crucial parts of your project records.

*You will be allowed to start a new project only when you have provided us with all available project reports and papers from any previous projects.*

### **Requesting Extensions on Projects**

Extensions are strongly discouraged and are granted only in rare circumstances. CES considers these requests on a case-by-case basis. You must request a project extension at least 90 days in advance of your project's end date. Your extension request must contain a rationale that focuses on why the additional time is needed to provide the proposed benefits to the Census Bureau. In addition, you must submit a document that summarizes the benefits that have been realized so far in your project.

## **Exit Process**

### **Documenting Benefit to the Census Bureau**

As explained previously, as a researcher with Special Sworn Status, you are carrying out a project that aids the Census Bureau in its mission. You described these benefits in your proposal and benefits statement (also called a Predominant Purpose Statement). Therefore, you will be required to submit a Post-Project Certification (PPC) summarizing the key findings pertinent to Census Bureau Chapter 5, Title 13 programs, including whether/how the accomplished benefits differed from those proposed and describing their potential application. Often the PPC is accompanied by a technical memorandum, which explains the benefits in detail. This technical memorandum may contain confidential information that will not be released to the public.

### Writing the PPC

For the vast majority of projects, a large portion of the PPC will simply be the Predominant Purpose Statement (PPS) written in the past tense since most projects achieve their proposed benefits with the methodology itemized in the original proposal documents. Common sense will guide you in deviating from this basic formulation: if you had to change your methodology, explain why and describe the replacement methodology; if you were unable to produce a benefit, so state and explain why; if you produced a benefit during the course of your project that wasn't itemized in the original PPS, then document it in the PPC; and so forth. The PPC should also include your assessment regarding the potential use by the Census Bureau of your findings, methodology, and benefits.

*We will not release your final research output before you have provided the PPC for your project. Moreover, you will not be allowed to return as a researcher until the PPC is submitted.*

### **Documenting Your Research**

You must provide documentation of data sets and programs for any output you use in a report, memo, or paper—whether you remove them from the Bureau or not. For a variety of reasons, it is important that CES and perhaps other researchers (including yourself) be able to reproduce any results that appear in print. Therefore, for any tabulation or model output you have released, you must provide the program, the log (if the statistical software produces a log), and the output. We will ask you for this information as part of the process of clearing your research output for release.

Moreover, and very importantly, you should trace the evolution of your data sets from the initial data sets we provide you (and that you may bring to the project) to the data sets that give your research results and the statistical output generated. This trail should include program and data set names, and should be sufficiently detailed that CES staff members could reproduce your work. See Chapter 3 for more details.

### **Backing Up and Archiving Data and Programs**

The thin client server is backed up regularly. If you lose a file and wish to have it restored, it may be able to be restored; contact your RDCA.

If your project team wishes to archive its work and/or data, the team must create a directory named /archive and move all relevant files to that directory before the end date of the project. This directory must be located at the top of the project directory tree (along side /data, /programs and /disclosure). Only the files that are placed in /archive will be saved at the end of the project. The current CES data archive policy will save the files in /archive for **seven** years. Note that this clock is reset in the event of a successful reactivation of the project in response to a revise and resubmit request from a journal referee.

## Returning Badge or Keys

We will require you to return your Census badge and any other security items when you exit or whenever you will not be on site for a significant amount of time. Any codes you were given for a door or for an alarm will be disabled at that time.

## After Exit

### Returning for Follow-up Work

#### Revise and Resubmit

Our policy is to allow researchers to return to the primary RDC without additional charge for reasonable amounts of follow-up work in response to a journal referees' comments for a "revise & resubmit." The typical situation involves carrying out additional model estimations. A "reasonable" amount of work uses no new data or research samples, asks for release of a small amount of output that is very similar to output previously released, and otherwise does not cause disclosure problems, *and* is carried out in less than 3 weeks.

To access the RDC lab for work after your project term expires (follow-up work), you must submit...

1. a letter requesting re-admission for follow-up work,
2. the letter from the journal editor and the referees' comments,
3. the Post-Project Certification (PPC) document, *and*
4. a statement of new benefits to the Bureau from the follow-up work.

The PPC for the original project must be approved before the request for reactivation of the project will be reviewed by CES. In addition, projects accessing Federal Tax Information will require IRS approval of both the PPC and the reactivation request.

#### New Project

To carry out follow-up work after your project term expires (other than that required for a revise and resubmit), you must submit a new proposal under the standard proposal review process—regular access fees will apply. Contact your RDCA to develop your new proposal.

## Exit Interview

At the end of your project, you must go through an exit interview. At this time you will be reminded of your obligations, including the fact that the Oath of Confidentiality is a lifetime obligation.

## **Providing Papers**

You are required to submit a working paper to a relevant Census Bureau series (as approved by CES) and we strongly encourage you to submit at least one working paper for inclusion in the CES Discussion Paper series; this does not preclude inclusion in any other working paper series.

Also, you must provide us copies of **all** published research papers and reports resulting from your RDC research project, even if it is long after your project ends. These are crucial parts your project records.

## Chapter 3: Data Access, Confidentiality, and Disclosure Analysis Policies and Procedures

This chapter presents policies and procedures you must follow in protecting the confidentiality of Census Bureau data supplied for your project, and in clearing your research results for release.

*This chapter gives policies and procedures concerned with protecting confidentiality of data supplied to you by the Census Bureau under Title 13, U.S.C., and possibly also protected by Title 26 U.S.C. (Federal Tax Information provided by IRS). The chapter does not discuss the rules or the information required for releasing output based on data supplied by the National Center for Health Statistics (NCHS) or the Agency for Health Care Research and Quality (AHRQ).<sup>2</sup> It does, however, cover the process for the submission of requests to these agencies (see the section “Release of Output From Other Agencies’ Data”). If you have questions about what is required for review, please contact these agencies for more information.*

### General Policies on Maintaining Confidentiality of Data and Clearing Release of Research Output

- *Absolutely no data and/or research results may leave the facilities without being examined and approved by the CES Disclosure Officer or a designated RDC Administrator (RDCA). This applies especially to confidential information on individuals, households, establishments, and firms—including microdata; sensitive identifiers such as names and addresses; and the presence or absence of an establishment, firm, household, or individual person in a survey.*
  - *CES releases approved research output only electronically; we do not release output in paper form.*
  - *Handwritten notes may not be removed from the RDC. Handwritten notes should be treated as confidential materials. If you want notes released, type them up and request disclosure review. Do not put undisclosed results (from memory) into any e-mail, even to Census Bureau employees and even if you think they would pass disclosure review. You must not discuss individual respondents or specific results that have not been cleared—in your research papers; via e-mail, phone, or fax; or in casual conversation—with persons outside Census Bureau facilities. This includes discussion with other project researchers with Special Sworn Status when you are not in a secure Census Bureau facility (e.g., inside the RDC). Qualitative discussions are allowed **under certain conditions**. For more about these conditions, see “*
    - *CES Policy on Qualitative Discussion of Regression Results outside RDCs and by Telephone.”*
    - *Even revealing that an entity is in the data is a disclosure. In IRS terms, this is revealing the fact-of-filing.*
    - *The Census Bureau considers it important not only to avoid disclosure of confidential information, but also to avoid the **perception** of disclosure. If respondents perceive that*

<sup>2</sup> An exception: The Census Bureau carries out the Medical Expenditure Panel Survey–Insurance Component (MEPS–IC) for AHRQ under Title 13 U.S.C. This survey is supplied by the Census Bureau and is covered here.

- their confidential data is being or will be disclosed, they will be less likely to participate in Census Bureau data collections, whether or not there is any actual disclosure.
- *The parameters in certain disclosure rules themselves are confidential.* Revealing these rules is considered a disclosure and violators will be dealt with accordingly. This policy is explained in the section “Disclosure Analysis.”
  - *It is your responsibility to protect the data:*
    - *You must meet with the RDC Administrator at the **beginning** of your project* to discuss confidentiality and security policies and procedures.
    - *Use data only as authorized, only for project-related purposes.* Confidential Census data may be used only for purposes authorized by the U.S. Census Bureau, and IRS data may be used only for purposes authorized by the Internal Revenue Code. **Casual browsing of confidential data is illegal.** For more detail about protecting IRS data, see Title 26, Appendix B, U.S. Code.
    - *Keep the RDC secure at all times.*
      - Make sure the door is secured whenever you leave.
      - Store confidential data in a secure, enclosed cabinet when not in use.
      - Ensure that listings and computer screens with confidential data are never visible to visitors or other non-authorized users.
      - Log off the thin client when you leave.
    - *Dispose of all confidential data properly.* Shred or lock up project-related materials when you are not using them.
    - *All RDC printers are configured to place a watermark on all printouts.* It reads: “Disclosure Prohibited—Title 13 U.S.C. and Title 26 U.S.C.” If this watermark does not appear, please report this to your RDCA immediately.
    - *The creation of any printout with FTI must be recorded in one of the CES FTI Printout and Disposal Logs.* Ask your RDCA where the log is located. The destruction (shredding) of any printout with FTI must also be logged. Only printouts with actual FTI need to be recorded in the logs. **FTI printouts that need to be logged are those printouts for economic and SSA projects that would not pass disclosure** (even if you do not submit it for disclosure). If you are in doubt, log it.
  - *Results must fall within the scope of the approved project.* CES will not release results that are determined to be outside the scope of the approved project topic. If your research idea has changed or you have a new research idea, contact your RDCA to discuss whether or not the new research falls within the scope of the approved project.
  - *The release of "intermediate output" (i.e., output that you know will not appear in a publication) is STRONGLY DISCOURAGED.* Producing intermediate output for examination inside the RDC is not problematic—removing it, however, can be. Since intermediate output often consists of detailed tables of preliminary descriptive statistics, or large numbers of similarly specified regression models, and are often based on “thin” samples, releasing this type of output can increase disclosure risk. For example, releasing similar tables based on changing samples (adding or dropping small numbers of observations) causes complementary disclosure problems, since comparison of the released data tables may reveal information on who is in the sample as well as their characteristics.

- *CES requires projects to emphasize model output rather than tabular output.* Tabular output typically consists of small one- or two-dimensional tables of variables that describe the samples that appear in the models. CES will not approve removal of large tabulations that were not explicitly described in the approved proposal. Moreover, review of large amounts of tabular output requires significant resources on the part of Census Bureau staff and will be sent to the Disclosure Review Board for review.
- *You must provide the appropriate documentation and prepare files as described in the section “Preparing Files for Disclosure Avoidance Review.”* Inappropriately documented or formatted requests will simply be **NOT APPROVED** and another properly formatted and documented request will need to be submitted.
- *It takes time to conduct disclosure analysis, and you must account for this in your project planning.* CES will not make rushed clearance decisions for events such as conference deadlines or the end of a project. Inform your RDCA as soon as you know you will need release of research output and the approximate date you will need the output.
  - *Plan to meet as early as possible with your RDCA and/or the reviewer of your clearance request.* This will minimize learning time on both sides, avoid costly misunderstandings, and speed approval of your clearance requests.
  - *Allow for 3 weeks from the date of submission to receipt of the final cleared file(s) for a **typical** release of output*—this includes NCHS and AHRQ review of output. Properly documented and formatted files can help to minimize the agency review time.
  - *Allow for 3 business days for receipt of cleared files for a **typical** release of programs or research notes.* See the section “Clearing Other Files—Programs and Research Notes” for more information about having these files released.
  - *Review times can be longer in certain cases;* e.g., if the output has to go to the Census Bureau’s Disclosure Review Board (DRB) or if there is a large amount of output being requested.
- *You must use a standard disclaimer on all research papers and reports that indicates that your views/results do not represent the opinions or views of the Census Bureau.* Your research products are not considered official Census Bureau data products since they do not undergo an official Census Bureau review to ensure that they meet certain Bureau standards. The disclaimer may be modified to fit the number of authors and to include disclaimer information regarding other organizations. Please check with your RDCA to see if there are any RDC specific acknowledgements that should be included. The disclaimer:

**DISCLAIMER:** *Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.*

**For research output produced from 2004–2007,** please add the following acknowledgement for the NSF grant that provided funds for the entire RDC Network: ***Support for this research at the \_\_\_\_\_ RDC from NSF (ITR-0427889) is also gratefully acknowledged.***

## **Disclosure Avoidance Review: Rules and Considerations**

This section describes the rules for releasing output and programs from the Research Data Centers. The information included in this section provides basic information for typical requests for release in the RDC network. More complex or new types of output will require additional consideration. Moreover, the references at the end of this chapter can be used to obtain more detailed information about the disclosure rules and techniques discussed here.

In short, any information that is based on an individual respondent, on a small sample, or on a highly concentrated cell will **NOT** be released. The following table provides some general guidelines for different types of output that are not allowed as opposed to those that are allowed. These are illustrative examples only and do not cover all scenarios.

<b>Description of Output</b>	<b>Not Allowed ...</b>	<b>Allowed ...</b>
Minimum or maximum values	For individual-level observations.	For aggregated observations (e.g., county, industry) if the cell meets disclosure rules.
Summary Statistics	For highly concentrated industries or those based on an individual-level observation.	For cells that pass disclosure rules.
Regression Coefficients	For variables based on a small number of observations.	For continuous variables or binary variables based on a large number of observations where the cell sizes pass disclosure rules.
Scatter Plots	If points are based on individual observations that have not been smoothed.	If points are based on aggregated observations (e.g., county, industry) and if the cells meet disclosure rules.
Distributional Graphs	Histograms.	Kernel Density Plots (tails may need to be suppressed).
Graphs	If some or all points are based on small or highly concentrated cells.	If points based on small or highly concentrated cells and the graph has been smoothed or the axes have been suppressed.
Programs	If unique identifiers or other hard-coded numbers that have not been cleared are embedded within the program.	If the sources of the numbers are documented (to ensure that data are not embedded within the program).
Variable Listings	File layouts or proc contents. <sup>3</sup>	Short list of variables without variable labels.

<sup>3</sup> Some zero observation files have been released for use in the Virtual RDC. These have undergone thorough review. Please check the Virtual RDC if you need access to file layouts or proc contents outside the RDC.

<http://www.vrdc.cornell.edu/news/>

## **CES Policy on Qualitative Discussion of Regression Results outside RDCs and by Telephone**

Under certain conditions, researchers with Special Sworn Status at Census Bureau RDCs may discuss the *signs and significance* of undisclosed, model-based coefficient estimates outside the RDC or over the phone without case-by-case Census Bureau review (or with expedited review of a notes document). Such discussions may include comparing the signs and significance of coefficients across different models—e.g., “in model A, the coefficient of income is larger than in model B, and its standard error is smaller;” or “both coefficients are significantly different from zero.” However, numeric results should not be discussed specifically. In general, these practices follow the standard rules for clearance of research notes. The discussions should be general and should not reveal any information about individual observations. The underlying sample should also pass disclosure. For example, researchers may discuss whether the model includes individual effects (e.g., person or firm effects) but may not identify whether any particular individual observations are included in the models. (For example, they may not discuss binary variables that represent particular firms or known groups of firms—e.g., the “big 3” in an industry.) The researcher is urged to be as general and qualitative as possible.

If the researcher requires notes for the discussion, then these need to be reviewed and released electronically; though, they will be given an expedited review (see “Clearing Other Files—Programs and Research Notes”). **Handwritten notes should NOT be removed from the RDC.** To be released, notes must be qualitative: discussion of signs and significance with no numbers. Notes must also be based on an underlying sample that would pass full disclosure review; however, full documentation (as is expected for release of numeric output) is not required. For discussions with non-SSS researchers, the researcher must document the following information either in an e-mail to the RDCA or in an electronic document in the disclosure folder (notify the RDCA of the location of the document):

- that the SSS researcher intends to discuss signs and significance of intermediate results from his or her project with a particular non-SSS researcher or group of researchers,
- that, based on the researcher’s knowledge of the data, the underlying sample includes enough units (firms, households, individuals) to pass a disclosure analysis, and
- that the estimates are only being discussed verbally (and the researcher will not leave any written materials with the audience if being discussed as part of a formal presentation).

If notes are being released for the discussion with non-SSS researchers, then these three items could also be included as an introductory paragraph in the release request document.

If you need something between this type of qualitative release and a full release of output, please discuss this with your RDCA. Just keep in mind that this type of request may take longer than the qualitative release. Moreover, nothing in the qualitative request policy should be interpreted as changing the requirements for formal release of model-based results for distribution outside the RDC.

### **Disclosure Analysis**

Disclosure analysis involves producing supporting statistics, in addition to the output you want released, that shows that your requested output reveals no confidential information about individual respondents to Census Bureau surveys or censuses.

All supporting statistics described here should be placed in the “support” subdirectory for your disclosure request, not in the “output” subdirectory. The organization of these disclosure files directories is discussed later in the section “Clearing Research Results.”

## Basics of Disclosure Analysis

At CES and the RDCs, disclosure analysis has two basic considerations:

- The output must not reveal any confidential information about any particular firm, establishment, household, or individual. Two general types of rules are used to determine whether a disclosure has occurred. We describe each in detail later.
  - *Threshold (count) Rule*: Any output cell containing a number of observations (contributors) below a certain threshold is considered a disclosure. These rules are usually used for both demographic (household and individual) data and economic (business firm and establishment) data.
  - *Concentration Rules*: The quantity displayed in the cell must not be too concentrated (i.e., dominated by a few firms). If it is too concentrated, it will be possible for an intruder to determine too closely the value of any contributor to the cell. We discuss two such rules below, the p% rule and the (n,k) rule. These rules are usually used for tables involving business (establishment and firm) magnitude data such as shipments or sales.
- Your requested output must not reveal that any particular firm, establishment, household, or individual has participated in a Census Bureau data program (survey or census). This consideration typically comes into play when combining Census Bureau data with outside data sets. For example, a 100% match rate to an external microdata set is a breach of this requirement—the matched external sample negates Census survey participation uncertainty.

## Concentration Rules

Concentration rules are usually applied to economic magnitude data. Economic magnitude data are typically presented as tables of aggregates of the values reported by individual census or survey respondents. Examples are sales (or value of shipments), employment, or inventory. The Census Bureau publishes a multitude of tables of this type, broken down by industry, geography, and other characteristics. These tables also include marginal and overall totals. The values in these tables are additive, in the sense that the value for an aggregate of cells (e.g., a marginal total) is the sum of the values in the cells. Though, much of the literature on disclosure focuses on release of aggregates, the same rules are applied to regression output. The basic idea to keep in mind when applying disclosure rules is that each statistic, coefficient, number, etc. to be released has underlying it a sample of original person or firm observations in the microdata. That underlying sample must be demonstrated to pass disclosure, as documented in supporting statistics that have applied these rules.

Disclosure analysis for this type of table is a complex topic, with a large and rapidly developing literature. Two summaries are in FCSM (2005, chapter IV) and Hundepool et al. (2007, chapter 4). These documents are in /public/disclosure as well as on the internet. The following is a brief summary designed to give researchers some basic background.

In economic data, the distributions of the variables are often highly skewed so that a few firms (even one) may account for most of the value in a cell; i.e., the cell is highly concentrated. We cannot allow the release of output that comes “too close” to revealing any individual firm’s

value. To make this practical, we must have criteria for deciding when release of the information in a cell as tabulated comes “too close” to disclosing confidential information. Such a cell is called *sensitive*, and we use mathematical rules for determining, which cells are sensitive. Two such rules are called the *p% rule* and the *(n,k) rule*. These rules acknowledge that the other firms contributing to the cell total (the firm’s competitors) are the ones in the best position to determine the values contributed by the other firms represented in the cell. Under the *p% rule*, a cell is sensitive if the second largest firm can determine the largest firm’s value to within *p%*. The value of the rule parameter *p* is considered confidential. The Census Bureau currently uses this rule in its publications. Under the *(n,k) rule*, which is used in Census Bureau publications before 1992, a cell is sensitive if the largest *n* members of the cell contribute more than *k%* of the cell total. The values of rule parameters *n* and *k* are also considered confidential.<sup>4</sup>

CES uses a combination of these rules; the general principle is that we use the rules in effect for the particular survey and year in question. The general rules:

**Standard Primary Disclosure Rules for Research Output based on Economic data  
(from business establishments and firms)**

- The *p%* rule applies for statistics in cells that represent
  - individual years 1992 and later or
  - multiple (pooled) years involving the year 1992 and later.
- The *(n,k)* rule applies for statistics in cells that represent
  - individual years 1991 and earlier or
  - multiple (pooled) years involving any year prior to 1992 (e.g., a table from data representing both 1991 and 1992 will use the *(n,k)* rule).

The values of *p*, *n*, and *k* are confidential. Your RDCA will give you their values as needed since they may be revealed only to individuals with Special Sworn Status who need to know them.

### Cell Suppression in Primary and Secondary Disclosure Analysis

We must ensure that *all* the individual observations in the table are protected. The traditional method of protection is *cell suppression*, where the value in the sensitive cells, as determined by either the *p%* or *(n,k)* rules, are not released. Cells suppressed in this way are called *primary* suppressions. For example, indicator variables used as controls (e.g., industry dummies) are often run in regressions but not reported in released output. They contribute to the estimates, but are often suppressed from the released output; for model output, this is usually sufficient. For tabular output, primary suppression is rarely enough to protect the values of sensitive cells in additive tables. For example, if only one cell were sensitive and suppressed, its value could easily be determined by subtraction using the marginal totals. For this reason, we must suppress other cells, which are called *secondary suppressions*. To do this properly is very complex and involves

<sup>4</sup> For the values of *p*, *n*, and *k* used at the Census Bureau, the two rules give very similar results. The *p%* rule is considered superior for reasons given in the references cited above. For the parameters the Census Bureau uses, the *(n,k)* rule suppresses a few more cells than the comparable *p%* rule; in that sense, it over suppresses. CES began using the *(n,k)* rule, and we have not changed to maintain consistency across years. However, the Census Bureau largely switched to the *p%* rule in the 1992 Economic Census so CES practice has not been entirely consistent with Census Bureau practice; we will use the rules in [Table 1](#).

linear programming techniques. It is by far the most difficult part of disclosure analysis<sup>5</sup> for tables, and is the main reason why CES does not allow complex tables.<sup>6</sup>

---

<sup>5</sup> For example, see Jewett (1992) p. 1–D–4.

<sup>6</sup> Another main reason is that it is not the mission of the RDC program to generate tables to supplement those tables released in official publications.

## **Disclosure Analysis Rules for CES and the RDCs**

The basic rule at CES and the RDCs for disclosure analysis of demographic and Decennial Census output is the threshold rule. The basic rules for disclosure analysis of economic output are a combination of the threshold and the concentration rules.

For consistency, we will ask you to produce the same supporting statistics for many types of output. This may not be strictly necessary, but the output we request should cover almost all situations, so it should save time by making things more predictable.

We provide SAS and Stata programs that you can use to carry out primary disclosure analysis. These programs are available in /public/disclosure. The programs do not contain the values of the parameters  $p$ ,  $n$ , or  $k$ . Your RDCA will give these values to you, and you must not reveal them to anyone outside the special sworn members of your project team. To do so is considered a disclosure of confidential information.

### **Important Reminder: Carry Along Identifiers in Your Data Sets**

To apply the rules stated in this section using computer programs, you need to have firm, establishment, household, and individual identifiers in your data sets. Please keep these variables in your analysis data set, even if you are not using them for any other purpose!

If we find problems or estimates that do not meet our criteria for release, we will do one of the following things:

- Ask you to *collapse*—that is, combine—certain cells. This will avoid disclosure problems at the expense of output detail, and is the preferred course of action.
- *Suppress* (e.g., replace by "D") the numbers in the affected cells. As described above, we will almost always need to carry out complementary (secondary) suppression on other cells. Complementary suppression is by far the more difficult, time-consuming part of disclosure analysis. We will only allow release of relatively simple tables (such as those generally found in journal articles) for, which secondary disclosure can be carried out simply.
- Ask you to reconsider the output, by asking yourself what you are trying to show and what other ways you can show it besides tables. For example, you may be able to summarize the information in the cells rather than showing all the cells.

Finally, we state some general (and obvious but worth emphasizing) guidelines:

- Statistics computed from large samples will require minimal review, because person or business counts are likely to be well above thresholds (and concentration of businesses will tend to be low).
- Statistics computed for detailed industry/geographic areas will require more intensive review.
- We reserve the right to take any output to the Disclosure Review Board.

## **Disclosure Analysis Rules for Tabular Output**

For our purposes, standard tabular output, often called "descriptive statistics," is summary information consisting of, for example, counts, totals, and basic distribution moments. We ask that you limit your tabular output to the minimum necessary to describe the sample used in your

models and how it might compare to the underlying population(s) of interest. These are the types of tables that usually appear in academic papers.

### Tables From Economic (Business and Establishment) Data

For economic (business and establishment) tables, our primary disclosure rules are based on the Census Bureau's practices for the various surveys and censuses.

- We follow the disclosure rules used at the time (year) of the survey.
- The variables used in primary disclosure analysis (p%, (n,k)) differ across surveys, and within the survey they sometimes change over time. For this reason, [Table 1](#) may not be directly applicable to your data set/year combination. The most important point is *we will not release tabular output for cells that the Census Bureau has suppressed*. It is your responsibility to check this; although, we can help. (For example, in the Annual Survey of Manufactures (ASM), the Census Bureau has done disclosure analysis on every variable since 2004). Please see your RDCA for guidance.
- We also require the number of firms to be at least a certain number, greater than  $n$  in the (n,k) rule.
- We will use secondary disclosure analysis as described above.

#### **Use of “Noise Infusion” in Census Bureau Programs**

The Census Bureau has begun to use a new method of confidentiality protection, which involved infusing random “noise” to individual observations before tabulating them for publication (Evans, Zayatz, and Slanta 1998). CES receives the underlying (non-noisy) data and the noise infusion “factors.” The following programs are beginning to use noise infusion, starting with the data years indicated.

- Integrated Longitudinal Database (ILBD), 2005
- Commodity Flow Survey (CFS), 2007
- Survey of Business Owners (SBO), 2007
- Economic Census of Puerto Rico and Island Areas, 2007

For these, the basic rules are:

- Use “noisy” data for tabulations.
- You may use non-noisy data for models.

In addition, the LEHD Quarterly Workforce Indicators (QWI) program uses noise infusion.

This is a preliminary discussion of confidentiality protection using these data sets. Please see your RDCA for further guidance.

**Table 1: Primary Disclosure Rules for Economic Data (Business Establishments and Firms)**

(See Notes at end of table)

Data Set	Primary Disclosure Rules	Comment
Census of Manufactures and Annual Survey of Manufactures (LRD)	Continuous variables: SR (see table notes) on value of shipments and capital investment, but see comment 3 in the adjacent column.	1. TVS is a key variable used to cover disclosure for variables considered correlated with it (i.e., most variables). Investment is a "lumpy" variable that is not considered highly correlated with shipments so disclosure is done separately for it. If TVS is not included in the study, then shipments or payroll are acceptable substitutes. 2. The ASM and other surveys use unweighted numerator, weighted denominator in calculating shares. Our rules are more conservative. But most studies using the ASM use unweighted data, so our procedures are appropriate. 3. Since 2004, the Census Bureau has performed disclosure on every continuous variable in the ASM.
Economic Census—sectors outside manufacturing	Continuous variables—SR on sales measure.	Sales is a key variable.
Pollution abatement Costs & Expenditure (PACE)	SR on every continuous variable	PACE variables not considered correlated; PACE expenditures can be lumpy, with some groups of plants known to have only certain types of expenditures.
Manuf. Energy Cons. (MECS)	Same as PACE	Similar to PACE
Survey Of Manuf. Tech (SMT)	Firm count	SMT variables not considered very sensitive by themselves; no disclosure in publications.
Research & Dev. (R&D)	SR on every variable	Similar to PACE
Quarterly Financial Reports (QFR)	SR on all variables	
National Employer Survey (NES)	Continuous variables: plant count Discrete variables: Smaller plant count	Rules formulated in first survey year (1994). Survey considered less sensitive than other economic surveys, because there are no certainty strata, geography is limited, and most variables are discrete responses to questions about types of training or hiring practices. We may increase numbers and make disclosure firm-based.
Medical Expenditure Panel Survey—Insurance Component (MEPS-IC)	SR on total employment	
Survey of Bus. Owners (SBO)	Firm count	
Longitudinal Business Database (LBD)		Payroll and employment are key variables
Integrated Longitudinal Business Database (ILBD)	SR on Receipts	The ILBD is derived from the Census Bureau's Nonemployer Statistics program. From 2005 on, this survey uses noise infusion in disclosure protection. See your RDC Administrator for guidance.

## Notes on Table 1:

- SR means "Standard Rule"—the p% or (n,k) rules as described in the section "Disclosure Analysis Rules for CES and the RDCs."
- The table does not give rule parameters—i.e., n and k in the (n,k) rule, n and p in the p% rule, or the firm counts. Your RDC Administrator will provide those you need to know. *The parameters are confidential.*
- The above rules are for primary disclosures. In addition, we will very likely need to suppress additional cells to ensure that it is not possible to calculate the values of suppressed cells using relationships derived from other cells.
- When two surveys are merged, use the more conservative rules; e.g., for the NES, use LRD rules for continuous variables when NES is merged to LRD.
- You may not remove data for cells suppressed in the publications. We may ask you to examine publications to confirm this.

**IMPORTANT NOTE:** Again, the rule parameters and the disclosure programs that contain those parameters are confidential. You may not reveal the parameters to anyone outside the group of researchers with Special Sworn Status who are working on your project. The programs may not leave the secure RDC site.

## Tables From Households and Individuals (Demographic Data)

This category includes tabulations from demographic survey or Decennial Census data. The main considerations are these.

- To the extent possible, all such tabulations must be created from a public-use version of the data file.
- For tabulations that cannot be created from public use files, disclosure analysis is based on threshold (count) rules.
  - Tabulations created from the nonpublic data files will be at the state level or above, and any cell must contain data from at least 10 unique individuals (unweighted). In these cases, the number and size of tabulations generated must be very limited.
  - The Census Bureau's Disclosure Review Board (DRB) must review any tabulations that do not meet these criteria.
- In cases where the researcher has linked third-party microdata or geographic-based aggregate data with the Title 13 confidential data and wishes to run tabulations, the RDCA will inform the Disclosure Officer. The DRB will, at its discretion, review the proposed tabulations prior to their removal from the center.

The DRB meets every Monday (with a few exceptions; e.g., holidays). The Disclosure Officer, who is a DRB member, will attempt to bring your request to the DRB within two meetings; and you will need to submit an easy-to-follow summary of what your requested output describes, how that output relates to your overall approved project, the issues in the output requiring DRB review, and why you think your output does not contain a potential disclosure problem. You should include how you will use your output to make particular points. This will help the DRB suggest ways to respecify your output (if needed) so that it does not present a disclosure risk and yet gets your intended message across. If you are using terminology or statistics that would not be familiar to those outside your field, please write a description of those terms or statistics that someone not in your field can understand. These additional documents are meant to make it easier for the DRB (composed of people knowledgeable in statistics, but from varied backgrounds) to understand your output and make decisions on it.

## Decennial Census and American Community Survey

The following special considerations apply in addition to the ones listed immediately above.

- For further reference on disclosure analysis rules for tables on households and individuals, please see the disclosure analysis rules for special tabulations for the 2000 Decennial and ACS data sets. These rules are posted on the Census website and are in /public/disclosure on the RDC servers.
  - 2000 Decennial: <http://www.census.gov/population/www/cen2000/sptabs/disclosure.html>
  - ACS: [http://www.census.gov/acs/www/Products/spec\\_tabs/drb\\_rules.htm](http://www.census.gov/acs/www/Products/spec_tabs/drb_rules.htm)
- For the 1970 and 1980 Decennial Censuses, we use the rules for the 1980 census. These involve cell suppression.
  - Each cell must be based on 30 or more *unweighted* people, and/or 10 or more *weighted* housing units (whether it is “and” or “or” depends on the universe of the table) in a geographic area in order to show data for that area.

- In addition, if the table shows characteristics by race/Spanish, then there must be 30 weighted people in the race group to be shown or it must be suppressed, and complimentary suppression is applied to avoid getting the suppressed number by subtraction.<sup>7</sup>
- For the Decennial Census and ACS group quarters data, results may be released for breakdowns no more detailed than institutional/noninstitutional. There are no exceptions; this rule applies to all years.

## Disclosure Analysis for Other Distributional Characteristics

Care must be taken in presenting other statistics intended to describe the shape of distributions. We prefer that you present general statistics on distributional shape (e.g., skewness, kurtosis measures) rather than showing anything closely related to individual observations (e.g., quantiles). In general, please think about what you are trying to say and what is the best way to say it before asking for anything that approximates data on individual observations.

### Quantiles

If you do require quantiles (e.g., to show journal reviewers), then the following considerations apply. Under these rules, it is difficult to release detailed percentiles of a distribution of the actual data. There are two possible approaches.

- Recommended: Estimate a continuous approximation (e.g., kernel density) and compute percentiles from the estimated distribution. For this case, disclosure rules (counts/concentration) for regression analysis apply.
- Also allowed: Use the observed data to estimate the percentiles based on a sufficiently large number of units.
  - For demographic data
    - There must be at least five observations both above and below the quantile (i.e., 10 observations “around” the quantile).
    - For multiple quantiles, these 10 observations must not overlap.
    - The quantile value shown must be rounded to two significant digits.
  - For economic data
    - There must be at least ten *firms* represented around the quantile, five on each side.
    - For multiple quantiles, these 10 observations must not overlap.
    - To calculate the quantile, average the observations of the n firms included in the quantile. This average must pass the standard tabular disclosure tests. (The number of firms used to calculate the quantile is confidential.)

In addition:

- The above rules imply that maxima and minima are not allowed.
- If your table contains summary aggregates including maxima and minima of the aggregates, the minimum and maximum must pass disclosure review. For example, if you want to show a table that shows the mean variance, maximum, and minimum of a group of county-level incomes, the maximum and minimum county-level incomes must pass disclosure review.

<sup>7</sup> For an example, see [http://www2.census.gov/prod2/decennial/documents/1980a\\_alD-01.pdf](http://www2.census.gov/prod2/decennial/documents/1980a_alD-01.pdf), p. 7

## Histograms, Kernel Densities, and Scatterplots

We generally discourage histograms of univariate distributions, which give counts of numbers of observations within certain classes. We instead encourage kernel densities, which essentially are smoothed versions of histograms.

In estimating kernel densities, we ask you to choose bandwidth values that do not obviously suggest the presence of individual observations. Unfortunately, as of now there are no fixed rules for doing this.

For multivariate distributions, we generally prohibit scatterplots of data on individual observations. For example, we do not permit scatterplots of value of shipments versus employment for individual establishments or firms. Instead, we will consider bivariate kernel densities, which do not show the individual observations.

These considerations also apply to data that are not in the scale of the observation, e.g., productivity measures, which are ratios and not tightly related to size. Researchers should think of alternative ways to show these relationships; e.g., bivariate kernel densities.

## Maps

We treat maps as tables since typical maps show the numbers of observations in categories for an area. We require that each category in each area adhere to the appropriate disclosure rules. Maps *must* go to the Disclosure Review Board (DRB) if, for demographic or Decennial data, maps show geography below the state level or if they are made using confidential demographic or Decennial Census microdata that show sub-state areas. However, we frequently take maps to the DRB that do not fall into these categories so please plan for this in timing your disclosure submission.

## **Disclosure Analysis for Model Output**

For model output, disclosure is usually not a problem if the sample contains a sufficient number of observations of establishments, firms, households, or individuals. The exceptions usually involve categorical variables, and disclosure analysis focuses on these variables. In a regression with dichotomous (0,1) variables, for example, the binary may sometimes take values of 1 for establishments belonging to a small number of observations (firm or known small group of firms). These are disclosures because they identify, which plants belong to that specific firm(s).

To summarize disclosure analysis for model output (e.g., regressions):

- We require that you make the standard disclosure analysis calculations described above (threshold or concentration) as appropriate for economic or demographic data sets for every regression coefficient “cell” for, which you request to release information (e.g. a coefficient or descriptive statistics).
- If the model has a continuous variable on the left hand side and if you are reporting indicator (binary) variables on the right hand side, then make the standard disclosure calculations for all the establishments, firms, households, or individuals in each binary category.
- If the model has a categorical (e.g., binary) variable on the left-hand side and you are reporting indicator (binary) variables on the right-hand side, then make the standard

disclosure calculations for all of the establishments, firms, households, or individuals in each binary category, crossed by the categories on the left-hand side of the model.

- If you interact binary variables, then this constructs a new set of binary variables. Make the calculations for each of the possible categories in the cross classification.

You only need to make this calculation for binary variables that you report. For example, if you use dummies as “controls” or “fixed effects” for establishments, firms, households, individuals, or states, but do not report them, you do not need to make these calculations for them.

*You need to make the above calculations **for the observations that actually appear in the model.** Sometimes estimation procedures in SAS or Stata automatically drop observations during an estimation procedure because of missing values of one of the variables. Make sure those same observations are also excluded from the sample used to calculate disclosure review statistics for that set of regression results or other statistics. Your samples may include different observations from model to model because of this. If this is the case, **we need a separate set of disclosure review statistics for each sample.** (Hint: a different N between regressions is a sign that this is occurring. Please make sure the observation count for your disclosure review statistics matches that for the output you are requesting.)*

If we find problems or estimates that do not meet our criteria for release, we will either ask you to alter your specification or not report the values of the affected coefficients. In some cases, we may ask you to report that the coefficient is in a certain range (e.g., positive and significant). We will work with you to ensure, as much as possible, that your research results remain meaningful after respecification. However, this will require more time and we may take the output to the Disclosure Review Board.

### **Disclosure Analysis for LEHD (Mixed Economic and Demographic) Data**

For the LEHD data available at the RDCs, we use a mixture of the rules for economic and demographic data. Specifically:

- The standard disclosure rules described above apply for either person-level or business-level analysis. This means you should use the threshold (count) or concentration (p% or (n,k) rules described above. All results to be disclosed must include data from **at least three states**, unless your project has obtained a specific exemption to this rule during proposal review. Models may include geographic controls for more detailed geographic levels, but the coefficients on these controls may not be reported. It is okay to note on the table of coefficients: includes controls for [insert geography].
- Some other considerations mentioned above, but worth repeating here:
  - The finest level of detail that may be shown for Group Quarters data is Institutional/ Non-institutional. There are no exceptions to this rule, which applies to all years.
  - Under these rules, it is difficult to release detailed percentiles of a distribution of the actual data.

- Under the agreements that allow the Census Bureau to use their data in the LEHD program, some states must review research output before the output is released publicly. Please contact your RDCA about these requirements and plan for any needed extra review time.

### **Preparing Files for Disclosure Avoidance Review<sup>8</sup>**

This section describes the process for preparing files for the release of project output from the RDC for projects using the Census Bureau's data. (See the section "Release of Output From Other Agencies' Data" for the procedures for obtaining release of output from data provided by AHRQ and NCHS.) The section discusses the roles of the various parties—the researcher, the RDCA (including the Lead RDC Administrator), the CES Disclosure Officer, and (sometimes) the Census Bureau's Disclosure Review Board (DRB). For the actual formal submission of your request, see the section "Requesting File Clearance using the Disclosure Avoidance Review Tracking System (DARTS) in the CES Project Management System."

There are two general types of project documents that can be released: files containing numerical research results (e.g., regression output, tables, graphs) and other files (typically programs but also other documents such as research notes). "Other" documents do not require the same sets of forms as data/output and their release typically can be expedited. In the on-line Disclosure Avoidance Review Tracking System (DARTS), there are two types of requests that you will be able to select: (1) output and (2) programs. For the purposes of the on-line system, select "programs" for the release of other documents such as research notes. Since programs and other types of qualitative information do not require the same sets of forms as output, their release typically can be expedited. The clearance process for both types involves the following general steps.

- Inform your RDCA that you are planning a request to release files and provide the approximate date you will need the released files. For output, discuss the pending request with your RDCA in light of the requirements for disclosure analysis. In particular, ask the administrator (possibly consulting with the CES Disclosure Officer) whether the Disclosure Officer or the Census Bureau's Disclosure Review Board (DRB) needs to review the output.
- Provide the materials for release, in the proper places with proper organization—including any required supporting materials. You will be asked to provide materials that document all requests for release of files. For output files, the documentation must indicate that the output was produced as part of your approved project, and it must document that published output does not disclose confidential information. To accomplish this, you must follow the process as detailed in the following sections of this chapter.
- Request clearance of your files formally, using the Disclosure Avoidance Tracking System (DARTS) by logging into your account on the CES website ([www.ces.census.gov](http://www.ces.census.gov)). See the section "Requesting File Clearance using the Disclosure Avoidance Review Tracking System (DARTS) in the CES Project Management System" for specific instructions.

### **Clearing Research Results**

Clearing research results requires preparation and interaction with your RDCA and possibly the Disclosure Officer. When you are close to having research results you wish to remove from the

---

<sup>8</sup> The glossary defines some confidentiality and disclosure related terms that may be unfamiliar to you.

RDC, you can begin the process. The following steps provide information about the general process, but the process can be iterative, so the steps may not be linear.

1. Make an appointment with your RDCA to discuss the clearance request. Your administrator may refer you to the Disclosure Officer. The appointment should be made well enough in advance to allow for a thorough analysis of the results to be removed from the RDC. As a general rule, the appointment should be made at least a month in advance of the desired release date. If DRB review is required, then an even longer lead-time is recommended. During this appointment, discuss the samples you will use in your analyses and show examples of the types of output that you will request be released. It is *not* imperative that all files are final at this point; it *is* imperative that the RDCA and/or Disclosure Officer have a good sense of what you will want released. This process may also take more than one meeting and may require you to make changes in your output to satisfy clearance requirements.
  - For a **typical** release of output, we recommend allowing 3 weeks from the date of submission to receipt of the final cleared file(s). This includes NCHS and AHRQ review of output.
  - For a **typical** release of programs, we recommend allowing 3 business days for receipt of cleared program files.
  - Review times can be longer in certain cases; e.g., if the output has to go to the Census Bureau's Disclosure Review Board (DRB) or if there is a large amount of output being requested.

#### **When does the Disclosure Review Board (DRB) Review Output?**

For output from Census Bureau *household* censuses or surveys, the DRB must approve release of any output when

- any cell (table or regression coefficient) is based on fewer than 10 unweighted, unique individuals, and/or
- the output is presented for a geographic level lower than the state level.

For large amounts of tabular output (either for households or for businesses), the CES Disclosure Officer reserves the right to take any output to the DRB based on the complexity of the request. CES will not release large tabulations that were not explicitly described in the approved proposal.

2. In the disclosure folder in your project space, create a *dated clearance request subdirectory* for this request, e.g., rdccluster2:/rdcprojects/wo1/wo00000/disclosure/YYYYMMDD. This directory should contain the following.
  - A directory called "output," containing your output files (i.e., the files you want released).
  - A directory called "support," containing the following information:
    - *Disclosure analysis files* that show the required disclosure analysis statistics. The specific statistics needed for disclosure analysis vary with the output. See the sections on models and tabulations.
    - *Program files* used to create the data set being used, to produce the output, and to produce the disclosure analysis statistics.

- The *Clearance Request Memo*.<sup>9</sup> A template is provided in Appendix D and is also available on the RDC server under /public/disclosure. The section “Preparing the Clearance Request Memo” explains the process for completing this document.
3. Submit the request online using the Disclosure Avoidance Review Tracking System (DARTS) available by logging into the CES website ([www.ces.census.gov](http://www.ces.census.gov)). See Requesting File Clearance using the Disclosure Avoidance Review Tracking System (DARTS) in the CES Project Management System for more information.

### **Clearing Other Files—Programs and Research Notes**

It is easy to write programs and compose notes that pose minimal risk; guidelines are provided below to enable you to write programs and notes that can be released. Since CES must review **every line** of the files to be released, please minimize the number of files and the size of the files (length of program or notes) in each request. Following these guidelines will reduce the review time required and ensure that your request will be released in a timely manner.

#### *Programs:*

- Do not include code that edits (or otherwise refers to) individual observations via their identifiers (as opposed to groups) or that includes hard-coded values based on unreleased information such as quantiles or means. Use separate programs or “include” statements for the same effect. Do not refer to individual identifiers or observations in your notes.
- Document the source of any number you use in your program files. If there are numbers in the program files to be released and the reviewer is uncertain of their source, then the files will not be released.
- To reduce the volume of information in your programs, use macros or include statements for other code that appears in many programs (e.g., data manipulation statements that prepare data for analysis). CES can usually release the macros, and the reviewer needs to look at this code only once. (With planning, this practice also can speed up your coding in general.)
- Do not include notes in a program that describe “sensitive” information; e.g., the relationship between an individual and unique identifiers. When in doubt, the reviewer will not approve release of the file.

#### *Notes:*

- Do not refer to individual identifiers or observations in your notes, even in general terms.
- Do not include confidential information such as notes on specific observations you may have “cleaned” or numbers that CES has not cleared for release.

We must look at all such notes before clearing them so keep them as brief as possible. See the section “

- Disclosure Avoidance Review: Rules and Considerations” for more details about information that can and cannot be released.

To submit a request for clearance of these documents, create a dated directory under your project’s disclosure directory and copy the files you want released into that directory. (You do

---

<sup>9</sup> In the on-line disclosure avoidance review tracking system, you will be given the option to upload this document (it must be .rtf or .pdf). Most researchers complete these in the RDC and leave them in the supporting documents folder in the dated disclosure directory; however, some researchers fill them in outside of the RDC and this feature allows them to upload the form.

not need to provide a Clearance Request Memo for programs or for notes.) Then, submit the request using the on-line DARTS system. See “Requesting File Clearance using the Disclosure Avoidance Review Tracking System (DARTS) in the CES Project Management System.”

## Roles and Responsibilities

The disclosure avoidance review process relies on the interaction of many parties: you, the researcher; your RDCA; the reviewer assigned to review the request (often but not always your RDC Administrator); the CES Disclosure Officer; and (sometimes) the Census Bureau's Disclosure Review Board (DRB). The responsibilities of each party are as follows.

- Researchers
  - To conduct the initial review of the output to ensure that all the parameters being requested meet the standard rules for release of the output. If the standard rules are not met, then researchers need to make a justification as to why the output is required.
  - To ensure that the documentation is complete and formatted appropriately.
- RDC Administrators
  - To assist the researcher in understanding the rules and in ensuring that the documentation is complete and ready for review.
- Disclosure Avoidance Reviewers
  - To analyze the output to be released and the supporting documentation to ensure that there are no potential risks from releasing the output.
- Disclosure Officer
  - To assist the reviewers and administrators in releasing complicated clearance requests.
  - To analyze output that does not meet the standard disclosure rules and to take the requisite requests to the Disclosure Review Board.
- Disclosure Review Board (DRB)
  - To set the standard disclosure rules followed by RDC Researchers.
  - To make final decisions on requests that do not meet the standard disclosure rules but do not pose substantial disclosure risk.

## File Preparation for Releasing Files

The disclosure avoidance review process requires a significant amount of resources from Census Bureau personnel; CES requires that researchers appropriately prepare requests for review. It is the researcher's responsibility to disclosure-proof output that is to be released. The RDCA, the reviewer, and the Disclosure Officer are there to help you through the process and to ensure that the output has been disclosure-proofed.

Your responsibilities include the following.

- You must provide properly documented programs, descriptions of data files, research output files, and disclosure analysis files. If you do not, your request will be **not approved** and you will need to resubmit a new request to have the properly formatted files released.
- The clearer and more complete your request is, the more quickly your files can be released. In addition, working interactively with the reviewer will facilitate learning on both sides, avoid costly misunderstandings, and speed up the review process.
- It is your responsibility to ensure that the files to be released are in a format that you will be able to read outside of the RDC (e.g., .rtf or .xls).

- **Output files must be as close to a final format as possible.** These files can be spreadsheets, word processing documents, or other formats as long as they conform to the formatting requirements. *We will **not** release log files unless they meet all the requirements for a properly formatted document.*

The documentation aspect of this process is important. As projects become longer and larger, it is possible that multiple reviewers will review your files. This is especially true for project reactivations, where access is granted in order for you to revise and resubmit a publication, since there is typically a long period of time between clearance requests (and even you may be fuzzy on the details of previous releases). There is also always the possibility that another reviewer will need to be assigned to review output for your project if the primary reviewer is unavailable at the time of your request. Given that new samples and new output need to be compared to old samples and previously released output, it is important that a new person could do that comparison quickly and fairly easily. Having everything documented properly means that both you and a new reviewer could gain a quick understanding of the output that has been previously released and the supporting documents that provided the basis for those releases. Though it may take you more time in the short run, it is likely to gain you time in the long run. For these reasons and simply to expedite the process, we provide the following guidance on preparing files for the disclosure avoidance review process.

### **Guidelines for preparation of files**

*For output files to be released, all extraneous information needs to be removed from the file.*

This is especially relevant for log files, which often contain information about number of observations dropped and other information that the researcher does not need. In general, if you do not need the information, then remove it from the file. The reviewer should not have to remove this information for you.

- *The files should be formatted such that they can easily be read and/or printed by the reviewer.* This includes ensuring that the permissions to the folders and files are set properly (i.e., the group should at least have read permissions to the file and rwx permissions on the directories). The reviewer should also not have to spend a lot of time interpreting variable names, expanding columns, or formatting pages for viewing/printing. Reviewers are well within their rights to ask researchers to format files to make them easier to read or even to not approve a file that is unreadable with standard software.
- *Files to be released will be reviewed as is and will be released in the format that you have specified.* It is your responsibility to ensure that the files are ready for review and that you will be able to read the file outside of the RDC (e.g., .rtf or .xls).
- *All variables should be named consistently*—e.g., identical variables (such as industry, IND) should have the same names across data sets and files. This is true both for output files and their supporting documentation.
- *The link between an output file and the supporting documentation for that file should be clear.* The reviewer should not have to guess, which disclosure analysis output (or sample) corresponds to a given table or file.
- *All tables (in both output and supporting documents) should have clear titles* indicating what the output is. Tables should also indicate whether the analysis uses weighted or unweighted data (and, when weighted, what the weight variable is).
- *Output in the files you are requesting should be releasable (based on your disclosure analysis).* This means that you should look over the disclosure analysis files, determine,

which cells do not meet disclosure rules, and deal with small cells in the output file before submitting your request for review. If the information you need does not pass the disclosure rules, then discuss possible alternatives with your RDCA. There are typically ways to provide the information that you need, but it may require some discussion with the RDCA and the Disclosure Officer.

- *The programs that produced the materials are part of the documentation.* Please provide brief comments at the top of your programs whenever possible to explain what the program does.
- *Graphs must be saved in a format that does not have the data points embedded within the document.* Formats that can be released include .jpeg and .tif files. The following file types will not be released because they contain the underlying data: Stata graphs, PDF, and post-script files.
- *CES will provide programs for Disclosure Analysis for SAS and Stata.* These programs can be found under /public/disclosure. Separate programs are provided for demographic and economic data sets.
- *Document all samples used over time in your project.* Due to secondary disclosure risk, the reviewer needs to look at all samples used over time by the project. If your sample changes (for a similar analysis), then be sure to run the disclosure analysis on the change in your sample.

The Disclosure Officer (or other designated CES employee) will download cleared files from your disclosure directory and will send them to you via e-mail. In addition to the information in DARTS, this e-mail also serves as a record of what files have been removed from the RDC.

### ***Preparing the Clearance Request Memo<sup>10</sup>***

CES wants you to fill out the form clearly enough so that given your descriptions, they could repeat what you did and replicate your results. Another important way to think of it is that you should put on this form whatever will make it easier for the Disclosure Office (or anyone else) to determine if your output is safe. Give clear and organized descriptions of your output and its disclosure safeguards.

At the top of the form, put in the project number and the date of submission.

#### 1. General Information

This section describes the disclosure request in general and how it relates to the project.

- a. Name of the subdirectory: The path to the dated disclosure directory under the project's main clearance directory. For example, /20090101/.
- b. Outputs proposed for removal: Describe the general nature of the statistical output you are requesting.
- c. Part of the approved project: Explain how these results are related to your proposal. If your proposal posed some question, for example, explain how these results answer that question. Do not just refer the reader back to the proposal (e.g., see proposal). You can quote or summarize sections/pages within your proposal, but please still give a general description.

#### 2. Descriptions of Research Samples

---

<sup>10</sup> This section is taken from *The Researcher's Practical Guide to Clearance of Output* originally written by Ethan Lewis, updated by Phillip Doctor (December 2003), and now updated again by CES staff to be included here.

This section will ask you several questions about your data samples that underlie your output. You will need to start by briefly describing each data sample that you've used. After you've done this, you must also describe any relationships between the samples along with any relationships between the samples and other data samples, which might be similar.

### 3. Output Files

This section draws a distinction between “research” output files and “disclosure” analysis files. Research output files are the output files you propose to remove from the RDC and should be the only files in the output subdirectory. The disclosure analysis files are the ones where you provide documentation that your research output passes disclosure rules and the tables that provide disclosure statistics. All of these files should be in the directory labeled support.

In this section, you will have to detail information on your research output files. Each research output file should have a supporting disclosure analysis output file containing statistics that support your request for disclosure. For each of these files you should enter:

- *File Number.* This is nothing more than an ordinal numbering of all the files (1, 2, 3, ...) you are listing.
- *File Name.* Provide the name of the file (e.g., tables.xls). This file should be in the output subdirectory for this request.
- *File Description.* Describe the sample you used (e.g., “large northeastern manufacturing firms”) and the type of output in this research file (e.g., “regressions of capital stock on employment”).
- *Research Output Program.* Provide the name of the program that produced your research output. This should be in the support subdirectory for this request.
- *Research Sample Number.* This is the sample number that underlies the research output (from #2).
- *Disclosure Analysis File Name.* This is the name of the file containing supporting statistics that allow your research output to be released. This should be in the support subdirectory for this request.
- *Disclosure Analysis Program.* This is the name of the file used to generate your disclosure analysis statistics. This should be in the support subdirectory for this request.
- *Results.* Leave this area blank, CES fills in whether or not this file was cleared.
- *Comments.* Provide a general description of how this meets disclosure requirements, (e.g., by involving a large number of firms) and any other information that would aid the reviewer in assessing the disclosure risk.

### 4. Variable Definitions

This section asks you for information regarding every variable in your research output. For each variable please list:

- *Variable Name.* This is simply asking you to indicate the name of the variable you are about to discuss.
- *Definition.* Write a brief description of the variable (e.g., variable "c0192" contains the cost of cement). If you generated this variable yourself or if this is a modified variable rather than one that was present in the original data set(s), then please

indicate the equations/methods used to generate this variable. Also include the type of variable, e.g., continuous, binary (0,1), categorical (with categories).

- *Source*. List, which data set this variable came from, or if you generated it yourself, please indicate this.

This information could be provided as part of a proc contents if the information was clearly documented in the variable label.

#### 5. Other Information or Comments

Please provide any other information that you think the reviewer should know. Once again, think of any information that you think might make it easier for the Disclosure Officer to easily determine if your output is safe for clearance.

### **Requesting File Clearance using the Disclosure Avoidance Review Tracking System (DARTS) in the CES Project Management System**

This section describes the process for the formal submission of your request for disclosure avoidance reviews and assumes that your documents have been appropriately formatted and documented in the section “Preparing Files for Disclosure Avoidance Review.”

Only one person from a project should submit the request. The person making the request should follow these steps.

*Step 0: To begin, log into the management system and go to the disclosure module for the project.*

- Log into the management system in your web browser by going to [www.ces.census.gov](http://www.ces.census.gov) and clicking on “sign in.” Then enter your user name (which is your e-mail address) and your password. If you are unsure about your user name, contact your RDCA.
  - The system requires you to change your password every 45 days. If you need your password reset, there is a password reset button on the sign in screen (<http://www.ces.census.gov/index.php/ces/formlogin>).
  - In addition, your account will be locked if inactive for more than 45 days. If you have not logged into the on-line system for 45 days, send an e-mail to Brian Holly ([brian.p.holly@census.gov](mailto:brian.p.holly@census.gov)) and Lynn Riggs ([tammy.l.riggs@census.gov](mailto:tammy.l.riggs@census.gov)) in order to have your account unlocked.
- Once you are logged into the system, click on the + next to My Account (under Site Map on the left side of the screen).
- Click on the tab called Clearance Request. This brings up the Clearance Request Screen; all of the active projects on which you are working should now be visible.
- Click on the + next to the project number under, which you are requesting clearance. Click on the + next to the relevant folder(s) in order to get to the project’s *disclosure directory* and then to the dated directory that you created specifically for this disclosure. (There is no content contained in this information.)

**Since the project management system captures the information over night, the folder names and files may not be visible until the day after you create them.**

*Step 1: Select the files for clearance*

- All the files you wish to clear (and only those files) should be in the output subdirectory if you have organized the files properly in the output and support subdirectories as directed. Click on the + in front of this subdirectory to see all the files within that folder. Each file and subdirectory will have an open box, and subdirectories have a + to expand. Checking the box next to a folder will select all the files within that folder for clearance. To select individual files, simply check the box in front of that folder. Continue until all of the files that you want released have been selected.
- Once all of the files (and directories) have been selected, click on the Select Files button. This will make the file names appear in the box under Step 2. If a file was forgotten, then go back to Step 1, check the box in front of the file, and hit the Select Files button again. The additional file (or files) will now be shown under Step 2.

*Step 2: Create the Clearance Request*

- Check the box in front of any file that needs to go to the Disclosure Review Board (DRB) for review. If you do not know whether or not the file needs to go to DRB, then leave the box unchecked. The reviewer assigned to the project will determine whether DRB review is required and will be able to request it at that time.
- Select the review type: Programs or Output. Please make separate requests for different types of files. The type Program should be selected for qualitative research output (e.g., research notes).
- Put in the “Deadline.” This is the date by, which the materials are needed.

Note: CES does not guarantee that this deadline will be met. When entering the date remember, also, that CES typically requires 3 weeks for disclosure requests for output and 3 business days for programs (see “Clearing Research Results”). These estimates are for **typical** requests—requests for large numbers of files, for large amounts of output, or requests that need to go to DRB may take longer. Moreover, you do not have to enter a deadline in order to submit the request.

- When you have entered all the information, click on the Create Request button. This will move the request into Step 3.

*Step 3: Submit the Clearance Request*

Here, the request on which you are currently working should be open to show the full detail of the request. This step provides the opportunity to review the request information before formally submitting it for review.

- If something was entered incorrectly, simply click the Withdraw Request button and go back to Step 1.
- Here, you can upload of the Clearance Request Memo (it must be .rtf or .pdf). This is an option for users who fill in the Clearance Request Memo outside of the RDC. If the file was completed in the RDC, then no file needs to be uploaded since the file should be in the disclosure directory for the request.
- To formally submit the request, click on the Submit Clearance Request button.

- The system will notify you, the Disclosure Officer, RDCA, and the Lead RDC Administrator via e-mail of the clearance request. The request will then be assigned to a reviewer.

Under Step 3 of the Clearance Request Screen, users will also be able to see the status of all previous disclosure requests for that project. The initial view will show the clearance request record, and clicking on the  next to the clearance request number will show the individual file information. Files that have been approved for release will be sent to the researcher via e-mail by the CES Disclosure Officer.

## **Decisions and Release of Output**

When the reviewer decides whether a file (or group of files) is approved for clearance, the management system will send you (the originator of the disclosure request) an e-mail. The Disclosure Officer or the reviewer will send you the output files as an attachment (a zipped folder) in a separate e-mail.

If you disagree with decisions on clearance of research output, then you may discuss these decisions with the reviewer and (if appropriate) the CES Disclosure Officer, who may take the research output to the DRB. **The DRB's decisions are final.**

## **Release of Output From Other Agencies' Data**

Some projects use data from other agencies—sometimes in conjunction with Census data and sometimes using just the other agency's data. Primarily, these projects use data provided by AHRQ or NCHS. The Census Bureau has to adhere to the disclosure rules and policies prescribed by these other agencies. If your project uses other agency data, please check with your RDCA to see if there are special output review procedures required for your project.

## **Release of Output Using AHRQ or NCHS Data**

There are several types of potential projects that could be covered under these procedures:

- Projects using AHRQ or NCHS data exclusively
- Projects using AHRQ and NCHS data
- Projects using AHRQ and Census data
- Projects using NCHS and Census data
- Projects using AHRQ, NCHS, and Census data

## **Projects using AHRQ or NCHS data exclusively**

Projects using AHRQ or NCHS data exclusively should speak with the data provider about the specific disclosure statistics required and any other rules and requirements that need to be met. For the request, researchers should put the requisite files for that request into their project's disclosure directory under a dated subdirectory (e.g., /rdcprojects/wo1/wo00000/disclosure/YYYYMMDD/). The researcher should then submit the request using the steps outlined in the section "Requesting File Clearance using the Disclosure Avoidance Review Tracking System (DARTS) in the CES Project Management System." The CES Disclosure Officer will be notified of the request and will send the files to the appropriate agency for review. As with requests for Census Review of output, please plan for at least 3 weeks for receipt of the output.

## **Projects using Combinations of Census, AHRQ and/or NCHS data**

For these multi-agency data projects, each agency will appoint one person during the proposal review process to take the lead on disclosure review for the project for that agency. When a request for clearance is received the Disclosure Officer will immediately contact the appropriate personnel and start the process. A typical review of output takes approximately 3 weeks; however, the process will likely require additional time due to additional review required by each agency.

If Census data are involved in the project, Census will conduct the initial review and will not send the files to the other agencies unless the files are considered releasable based on the Title 13 (Bureau of the Census) data protections. Researchers need to complete the documentation under the normal rules and procedures outlined in this chapter. For requests involving other agency data in the creation of the output, personnel from the other agencies will have to review the output and the researcher will need to submit the required information according to each agency's specification for disclosure review.<sup>11</sup> These processes for disclosure avoidance review should be documented in the Review Synopsis at the time of project approval. The researcher should contact the other agencies for more details on the requirements for their disclosure avoidance review process.

Regardless of, which agency is reviewing the files, the researcher should submit the request using the steps outlined in the section "Requesting File Clearance using the Disclosure Avoidance Review Tracking System (DARTS) in the CES Project Management System."

---

<sup>11</sup> The NCHS/AHRQ annual agreement states that the agencies will jointly review proposals and conduct disclosure review on projects that will use NHIS/MEPS data. NCHS will have the primary responsibility for disclosure risk and will need to keep AHRQ informed.

## Bibliography

The following is a short annotated bibliography giving references on many of the topics discussed in this manual, as well as certain other topics related to privacy, confidentiality, and data access. At the end we list some useful related web sites.

Doyle, P., J. Lane, J. Theeuwes, and L. Zayatz (eds.). *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*. Amsterdam: Elsevier Science B.V., 2001.

Duncan, G.T., T. B. Jabine, and V.A. deWolf (eds). *Private Lives and Public Policies*. Washington, D.C., National Academy Press, 1993. An excellent reference on privacy, confidentiality, and data access issues related to the government statistical system.

Evans, T., L. Zayatz, and J. Slanta, “Using Noise for Disclosure Limitation of Establishment Tabular Data.” *Journal of Official Statistics*, Vol. 14, No. 4, 1998, pp. 537–551. Available at <http://www.jos.nu/Articles/abstract.asp?article=144537>.

Federal Committee on Statistical Methodology (FCSM), “Report on Statistical Disclosure Limitation Methodology.” Statistical Policy Working Paper 22. Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget, May 1994 (Second Version 2005). Available at <http://www.fcsm.gov/working-papers/spwp22.html>.

Hundepool, A., J. Domingo-Ferrer, L. Franconi, S. Giessing, R. Lenz, J. Longhurst, E.S. Nordholt, G. Seri, and P. De Wolf. “*Handbook on Statistical Disclosure Control*” Version 1.0.1, March 2007. Available at [http://neon.vb.cbs.nl/cenex/CENEX-SDC\\_Handbook.pdf](http://neon.vb.cbs.nl/cenex/CENEX-SDC_Handbook.pdf)

Interagency Confidentiality and Data Access Group. “Checklist on Disclosure Potential of Proposed Data Releases.” Interagency Confidentiality and Data Access Group, an interest group of the Federal Committee on Statistical Methodology, 1999. Available on FCSM web site below.

Jabine, T. “Procedures for Restricted Access.” *Journal of Official Statistics*, 9(2). 537–590, 1993. An excellent and oft-cited background paper used in preparing Duncan, Jabine and deWolf .

Jewett, R. “Disclosure Analysis for the 1992 Economic Census.” A manual describing the disclosure programs used for the 1992 Economic Census. The manual and a version of the programs are available on the FCSM web site cited below.

Willenborg, L., and de Wall, T. *Statistical Disclosure Control in Practice*. Lecture Notes in Statistics 111. New York: Springer-Verlag, 1996.

———. *Elements of Statistical Disclosure Control*. Lecture Notes in Statistics 155. New York: Springer-Verlag, 2001. The two books by Willenbourg and de Wall are general references, perhaps a bit more technical than the Eurostat manual.

**Web sites of interest:**

Federal Committee on Statistical Methodology: <http://www.fcsm.gov>. Contains information on FCSM activities, and links to information on confidentiality and disclosure analysis, including conferences and working groups.

American Statistical Association Committee on Privacy and confidentiality home page: <http://www.amstat.org/comm/cmtepc/> Contains information on committee activities, and links to information on confidentiality and disclosure analysis, including conferences and working groups.

Searchable U.S. Code website: <http://uscode.house.gov/lawrevisioncounsel.shtml>.

## Appendix A: Glossary

This appendix is a glossary of terms that can be divided into two general groups: terms related to confidentiality and terms related to the specific procedures for clearing research output at Census Bureau Research Data Centers. Please note many of the terms in the first group have multiple definitions; we have given only one here. For good discussions, see Duncan, Jabine, and deWolf (1993) pp. 22–24, and the web site of the American Statistical Association Committee on Privacy and Confidentiality.)

*Administrative records*—systems of records collected and/or maintained by government agencies for the purpose of administering programs—i.e., for *administrative purposes*. Administrative records are distinct from the systems of information collected exclusively for *statistical purposes*, such as those that the Census Bureau produces under authority of Title 13 U.S. Code.

*Business establishment*—a location at, which business is carried out. Usually has a permanent plant number (PPN) associated with it.

*Business firm*—one or more establishments under common ownership. In our files, usually identified by the "ID" or "CFN" (Census file number).

*Clearance request memorandum*—a memo, contained in the clearance request subdirectory, describing a request to clear (release) research output.

*Clearing of research output*—the approval to remove research output from the secure RDC facility. Only the RDC Administrator or other authorized Census Bureau employees may clear research output.

*Confidentiality*—the legal status that indicates that some body of data or information is sensitive and must be protected against loss, misuse, destruction, or unintentional change.

*Dated clearance request subdirectory*—a directory, under the project clearance directory, in, which you place the material for each clearance request: the clearance request memorandum, the research output (and the programs that produced it), and the disclosure output (and the programs that produced it).

*Disclosure*—the inappropriate attribution of information to a data subject, whether an individual or an organization. Disclosure occurs when a data subject is identified from a released file (*identity disclosure*), sensitive information about a data subject is revealed through the released file (*attribute disclosure*), or the released data make it possible to determine the value of the characteristic of an individual more accurately than otherwise would have been possible (*inferential disclosure*). Comments: (1) Disclosure of sensitive data does not have to be exact; disclosure is considered to occur if a characteristic of an individual (e.g., income) is estimated sufficiently closely. (2) Revealing the presence or absence of a unit in a data collection typically is considered to be a disclosure.

*Disclosure analysis*—the act of ensuring that research output does not disclose confidential information about individual business establishments, business firms, households, or individuals.

*Disclosure output (disclosure analysis output)*—certain output, in addition to the research output, that provides the information we need to ensure that your research output does not disclose

confidential information.

*Disclosure output files (disclosure files)*—computer files containing disclosure output.

*Disclosure statistics (disclosure analysis/review statistics)*—certain statistics that enable us to determine whether or not the research output can be cleared. The statistics include information on the numbers of observations (establishments, firms, households, or individuals) in all cells of a table or all categories in models; in some cases they include measures of the possible dominance of a cell total by a small number of observations (usually firms).

*Household*—all persons who occupy a particular housing unit as their usual residence; or who live there at the time of the interview and have no usual residence. The usual residence is the place where the person lives and sleeps most of the time. This place is not necessarily the same as a legal residence, voting residence, or domicile.

*Primary Disclosure Rules for Establishment and Firm Data*—rules for determining when a cell discloses confidential information. The two main rules: (1)  $(n,k)$  rule—a cell discloses confidential information if its  $n$  largest respondents contribute more than  $k$  percent of the cell value. (3)  $p$ -percent rule—a cell discloses confidential information if the second largest respondent can use the cell value to estimate the largest respondent's contribution to within  $p$  percent.

*Project*—the research specified in an approved research proposal.

*Project clearance directory*—a directory (one for each project) on your computer in that contains overall information about clearance activity for the project.

*Research output*—output you would like to remove from the Research Data Center. This will consist of output from statistical models (e.g., regression and other econometric models), together with a limited amount of supporting descriptive statistics (e.g., means and variances of variables in your research sample). Note: Output may not be released that shows data for individual observations. This currently includes residual plots; e.g., plots of residuals against the dependent variable. (This policy regarding plots is under review and may be changed.)

*Research output files*—computer files containing research output, usually generated with statistical packages such as SAS or Stata. The output often includes tables that summarize the coefficient estimates from groups of models, and supporting tables of sample means and variances.

*Statistical purpose*—the motivation for the collection of information to generate an aggregate description of a group of persons or businesses. Data collected for statistical purposes may not be used to make decisions regarding a specific individual or business, although the information may result in policy changes that could result in benefits or costs to persons or businesses. Examples: investigating geographical factors related to crime, the relation between welfare benefits and educational levels of welfare recipients. Stands in contrast to data collected for *administrative purposes*, which may be used to make decisions regarding individual persons or businesses—e.g., to determine eligibility for welfare benefits.

**Appendix B: ADRF**  
**CENTER FOR ECONOMIC STUDIES**  
**ACCOUNT AND DATA REQUEST FORM**  
**(ADRF)**

**I. Description of Work**Researcher Name: **NAME**Project Number: **CES PROJECT NUMBER**Title: **TITLE**Principal Researcher: **NAME OF PI**Starting Date: **DATE**Ending Date: **DATE**Duration of Work: **DURATION IN MONTHS** months

Location of Research

Primary RDC: **NAME AND ADDRESS OF PROJECT'S RDC**Researcher RDC Location: **NAME OF RDC THAT RESEARCHER WILL USE**

Computer Login Information:

SERVER:

USERID:

PRIMARY GROUP:

ACCOUNT EXPIRATION DATE:

DATE REQUEST FILLED:

***Detailed Description of Work***

The U.S. Census Bureau's Center for Economic Studies (CES) and its network of Census Research Data Centers (RDCs) support and encourage research activity using Census Bureau microdata to improve Census Bureau programs and inform public decision makers. As an integral part of its ongoing research program to improve Census Bureau programs, CES shall provide support for the research project described in this document and its attachments.

The CES research program is carried out with both CES staff and researchers from outside the Census Bureau (called research associates). It is centered on empirical analysis of confidential data collected in the Bureau's regular survey and census programs. The research associates carry out research projects at secure CES facilities, Census RDCs.

Through the network of RDCs, including the facility at CES headquarters in Suitland, MD, the Census Bureau grants researchers with approved research projects and Special Sworn Status restricted access to non-publicly available data files in secure environments. These secure environments ensure that the confidentiality of the data are maintained, but at the same time, provide limited access to data that are not a regular part of the Census Bureau data publication program. These research projects benefit the Census Bureau's programs. Access to non-publicly available data files is a privilege, not an entitlement. This privileged access carries with it specific obligations and responsibilities as specified in this document, the ADRF.

This activity is authorized under Title 13 U.S.C. Sections 6, 8, and 23c.

Attachment A: Project Description (see Proposal on CES Management System)

Attachment B: Benefits to the Census Bureau (see Predominant Purpose Statement on CES Management System)

## **II. Center for Economic Studies Furnished Equipment, Software, and Services**

The Center for Economic Studies (CES) shall provide the following facilities and support towards the execution of this work during the specified period:

1. A physical workspace that is suitable to perform research-oriented tasks.
2. Access to a thin-client computing device that is used to communicate with the CES computer server in a graphical interface mode. This will provide remote access (only from within the RDC) to the CES data specified in this document and its attachments.
3. Usual and reasonable computer and data support. Researcher may consult with CES staff on computer-related and/or data topics to better understand the environment in, which this work is conducted. CES shall make every reasonable effort to provide researcher with documentation for each data set to, which they are provided access.
4. The Linux operating system, with all usual system tools, and utilities.
5. A graphical user environment, such as CDE, GNOME, or KDE, configured to support the requirements of the researcher.
6. A browser installed and configured to view HTML documentation.
7. The OpenOffice office software suite.
8. Statistical and/or econometric software packages, such as SAS and Stata.

Please refer to the *Thin Client User Guide* for a full itemization of currently available analytic applications. Availability of any specific application can be revoked at any time if it is determined that the application does not meet Census security requirements. Census alone determines an application's security compliance.

If researcher wants to use a software package that is not currently offered in the RDC, then the researcher must document why existing packages do not meet the requirements of the project as outlined in the approved proposal. If this argument is valid (as determined exclusively by CES) and there are no security risks, then the researcher may be responsible for any licensing fees before the requested software is uploaded to the CES servers.

9. Allocated disk storage based on the estimated data and computing requirements to complete the work as defined in this document and its attachments. Requests for additional workspace shall be submitted in writing to the RDC Administrator for forwarding to CES Headquarters for approval. There may be an additional fee for extra disk space.
10. Archival storage of selected files. Due to limited disk resources, completed RDC research project directories are archived after completion of the project. Before the end date of this project, RDC research teams are expected to copy files (e.g., log files, data files, programs, user provided data) they wish to have archived into the /archive directory. The archive directory is created by the research team and it should be located at the top of the project directory tree (along side /data, /programs and /disclosure). For details of the current archive policy, please refer to *The Researcher Handbook*.

## **III. Project Data**

The Center for Economic Studies shall provide to the researcher direct access to the following analytical data sets:

### LIST OF CENSUS PROVIDED DATA

These data shall be provided in a SAS data set format. Any conversion of these SAS data sets to other formats shall be carried out by the researcher, not by CES staff.

The researcher is granted access only to the data specified in this document and its attachments. Researcher agrees to access only those data specified in this document and its attachments. Requests for new Census Bureau data will require the researcher to submit a new research proposal to CES for review under standard project review procedures.

The researcher acknowledges and agrees that Census Bureau data are collected for the purpose of producing published aggregate statistics. Use of these data for other purposes is of variable reliability. All survey and census data contain measurement error and may contain sampling error. Available information about these potential errors are provided or referenced in relevant publications. The Census Bureau recommends that data users incorporate this information into their analyses as these errors could affect inferences. Researchers analyzing the data in order to create their own estimates are responsible for their validity and shall not cite the Census Bureau as the source of these derived estimates, but only as the source of the underlying data.

The researcher shall obtain any necessary approval(s) for the use of non-Census data in the Census RDCs from the data producer(s) as appropriate. The researcher must provide these written approvals to designated CES staff before non-Census Bureau data are loaded onto CES computer systems. As specified in the *User Provided Data Memo* (UPDM), different categories of data have different requirements. A completely filled-out UPDM must accompany the request to upload each non-Census data set into the CES computer system. Once the non-Census data set is provided to CES by the researcher with the attendant documentation cited in the UPDM, the data will be uploaded to the project directory. The UPDM and its attendant documentation become Attachment C of this document, Data Permissions for Researcher Provided Data.

### *Data Access Policies and Confidentiality*

1. Researcher agrees to access only those data specified in this document and its attachments.
2. Researcher shall not ask other researchers at the RDC for access to their data, and researcher shall not give other researchers access to their data.
3. Researcher shall perform only analyses that support the work specified in this document and its attachments. The researcher agrees not to carry out analyses for any colleagues or other researchers who are not assigned to this research project.
4. Researcher acknowledges and agrees that research data files shall not contain sensitive identifiers such as name, address, or social security number. If manipulation of the data based on sensitive identifiers is necessary, then this is conducted by internal Census Bureau staff, requires approval and requisite fees.
5. Researcher acknowledges and agrees that absolutely no data and/or research results shall leave the facilities without being examined and approved by a designated CES employee. Designated CES staff must examine all materials that the researcher wishes to remove from the RDC to prevent disclosure of confidential information. The researcher agrees that he/she shall not remove research notes, printouts, or research output before obtaining explicit approval from a designated CES staff member. The CES Disclosure Officer or other designee shall release

electronic files only, usually via e-mail.

6. Researcher acknowledges and agrees that the Census Bureau considers it important not only to avoid disclosure of confidential information, but also to avoid the perception of disclosure. The researcher shall avoid any perception of disclosing sensitive information concerning the data they were provided.

7. Researcher shall not discuss any information about individual survey respondents in any manner in research papers, in e-mail, on the telephone, in a fax, or in conversation with persons outside Census Bureau facilities. In addition, researcher shall not discuss individual respondents' data with any other researcher not affiliated with their project.

8. Researcher shall use the current standard disclaimer on all research papers and reports, which can be found in *The Researcher Handbook*.

9. Researcher acknowledges and agrees that any research data sets produced as part of this project can be made available to CES or to research associates on other approved projects at the discretion of the Census Bureau.

10. Researcher acknowledges and agrees that all notes, printed material, and other media containing data that have not been cleared for release must contain the following label(s):

"Disclosure Prohibited-Title 13 U.S.C."

"Disclosure Prohibited-Title 13 and Title 26 U.S.C." (if using FTI data)

#### **IV. Protection of Confidential Information Provided to the Researcher**

All researchers who participate in this project shall abide by all the following conditions and rules of behavior in order to guarantee the confidentiality of any and all data provided to them during the execution of the work defined within this document and its attachments. These rules clearly delineate the responsibilities for all individuals with access to confidential CES data. These rules shall be enforced through sanctions commensurate with the severity of the infraction; this includes the possibility of criminal prosecution and termination of the project at the discretion of CES.

All researchers participating in this project who will have access to any confidential data or any undisclosed output must obtain Special Sworn Status (SSS) from the Census Bureau. No researcher shall have access to any data or undisclosed output prior to obtaining SSS and completing the training described below. Researchers who have not obtained SSS shall not be allowed in the RDC. The Census Bureau handles the administrative details of granting and maintaining SSS for the researcher. As individuals with SSS, the project researcher agrees to conform to all applicable confidentiality and security provisions of the Census Bureau and the Department of Commerce **forever**. In particular, these provisions state that no one shall disclose any information that would allow identification of any individual respondent (business firm, business establishment, household, or individual) of a Census Bureau census or survey. Additionally, the data supplied may be used for statistical purposes only. Disclosure of confidential information is prohibited by law (Title 13 U.S.C. section 214) and is punishable by a fine of up to \$250,000 (Title 18 U.S.C. section 3559 and 3571), a prison term of up to five years, or both. (Title 13, U.S.C., Section 214, as amended; Title 18, U.S.C., Sections 3559 and 3571)

Part of obtaining Special Sworn Status is the requirement of a researcher to take the Title 13 and Title 26 (FTI) Awareness Trainings and to be certified for Title 13 and Title 26 data access. Researcher must also complete Information Technology Security Awareness training. After obtaining SSS, all certification processes must be repeated every fiscal year for the duration of the project.

Researcher is granted the use of a security badge, bearing their photograph, which allows

access to the RDC facility specified in section I. Researcher must keep this badge secure at all times. This badge is the property of the Census Bureau and as such must be relinquished to Census upon project completion or upon demand, whichever occurs first.

All researchers shall be provided with a copy of *The Researcher Handbook*. This handbook contains regulations and policies that researchers shall follow. Researcher acknowledges and agrees that violation of any rules or regulations contained in this handbook shall be enforced as would any other provision in this document, including possible termination of the project at the discretion of CES.

If this project uses data that contains Federal Tax Information (FTI), then access to these data is subject to prior approval of the Internal Revenue Service. Title 26, U.S.C. Sections 7213 and 7431 provide penalties for unauthorized use or disclosure of FTI.

If this project uses an Agency for Health Research and Quality extract, then access to these data is subject to prior approval of AHRQ. This agency requires researchers to sign a separate confidentiality agreement, *AHRQ Nondisclosure Affidavit for Data Center Users*. Refer to this document regarding penalties for disclosing AHRQ confidential information.

If this project uses a National Center for Health Statistics extract, then access to these data is subject to prior approval of NCHS. This agency requires researchers to sign a separate confidentiality agreement, *NCHS Designated Agent—NCHS Research Data Center (RDC)*. Refer to this document regarding penalties for disclosing NCHS confidential information.

## **V. Supremacy of Census Bureau and Data Supplier Administrative Requirements**

This document does not supersede the administrative procedures and requirements of the Census Bureau and other data providing agencies. During the course of this project, changes may occur in administrative procedures and/or requirements of the Census Bureau and/or data providing agencies governing the use of the data provided to the researcher. In addition, changes may occur in the availability of data under both the governing Memorandum of Understanding (MOU) in existence as of the date of this document, and related MOU that come into existence during the course of the project. In such cases, relevant portions of this document may no longer be valid and shall be superseded by new language in order to accommodate administrative requirements of the Census Bureau and/or data providing agencies.

## **VI. Office Security Policies**

1. Researcher shall keep security clearance and entrance information (identification, key code, security badge, lock combination, and the like) current and secure.
2. The researcher shall not share with nor lend to anyone their security clearance information, entrance information, and/or security badge.
3. Researcher acknowledges that security surveillance cameras are used in the RDC facilities.
4. Researcher agrees to keep the RDC door closed and to ensure that it is locked whenever they enter or exit the RDC.
5. Security policies specific to individual RDCs may require the researcher to arm/disarm the RDC alarm if the researcher is the last person to leave or the first person to arrive at the RDC.
6. Unless otherwise approved by the RDC Administrator, the researcher shall schedule all visits to the RDC. The researcher acknowledges and agrees that access to the RDC may be denied if a reservation has not been made.
7. Researcher agrees to sign in every day that they work on the project at the RDC.

8. Researcher shall not allow **anyone** into the RDC.
9. Researchers shall not remove printouts from the RDC. Additional information regarding the release of output is discussed in section IX of this document and in *The Researcher Handbook*.

## **VII. Computer Security Policies**

1. Researcher acknowledges and agrees that all their computing activities shall be monitored by CES staff and that information from the monitoring programs shall be accumulated and saved in a database. This information shall be used to identify security violations, and to assess system use.
2. Researcher shall not include passwords in any computer files, such as the body of an e-mail or in a login script.
3. Researcher shall not write their computer password(s) down and must not give their password(s) to anyone.
4. If the researcher suspects that their password has been compromised, the researcher shall change the affected password immediately and notify the RDC Administrator.
5. CES Computer Staff personnel may employ password-checking software to determine the "guessability" of user passwords. If the researcher is informed by a CES staff member that the researcher's password is too easily "guessable," the researcher shall change it immediately.
6. Researchers are prohibited from loading software, "home-grown" programs, or data on any computer(s) in the RDC. Requests for uploading materials must be submitted to your RDC Administrator or other designee.
7. Portable data storage devices are prohibited in the RDC.
8. Laptop computers and any other portable computing devices that include an internal or external modem (e.g., Palm Pilot) are prohibited in the RDC. The researcher shall never bring a laptop or other portable computing device that incorporates a network data connection into the RDC.

## **VIII. Printing Security Policies**

1. The local RDC printer is located in the RDC Administrator's office or in a locked container. Researchers may send print jobs to this printer at any time, but researchers may retrieve their printouts only when the RDC Administrator is present. The RDC Administrator's office or printer-container is locked and inaccessible to researchers at all other times.
2. All such printouts shall be considered research output. Researchers may not remove these printouts from the RDC. The CES Disclosure Officer or other designee shall release electronic files only, usually via e-mail. The Disclosure Officer will review any such output in accordance with current disclosure avoidance analysis policy.
3. Researcher shall log all printouts that contain Title 26 (FTI) data.
4. Researcher shall place all output containing confidential information that they no longer need only into designated receptacles. Logged printouts shall be recorded in the disposal log.

## **IX. Disclosure Avoidance Analysis**

Disclosure avoidance analysis is a process by which CES staff ensures that no confidential information concerning census and/or survey respondents are included in any information that is allowed to leave CES facilities. Designated CES staff members review all research output for inadvertent disclosure of confidential information prior to release of the output. To aid this disclosure process, researchers shall prepare a disclosure memorandum that documents the files requested for release, their preparation, and shall prepare supporting statistics about

underlying data samples that were used to construct estimates.

The process of disclosure avoidance analysis takes time and advance planning. Upon beginning the project, the researcher shall meet with at least one designated CES staff member to discuss the specific output the researcher wishes to remove from the Census Bureau and to discuss the process of disclosure avoidance analysis. More information about this process is in *The Researcher Handbook*.

1. The researcher acknowledges and agrees that release of some analytical results specified in this document or its attachments may not be feasible due to problems identified during disclosure avoidance analysis. In particular, the researcher acknowledges and agrees that it may not be feasible to release any tabular descriptive statistics. The researcher acknowledges and agrees that CES decisions are final regarding the form and content of any research results released to the public.
2. The researcher shall submit supplementary information, in addition to the research output they wish to remove, to the CES staff members who carry out the disclosure avoidance analysis. The researcher shall supply all such supplementary output in the form requested.
3. The researcher shall allow sufficient lead-time for CES staff to perform disclosure avoidance analysis. The amount of time needed will depend on the specific results to be removed and the form in, which the required information is submitted to CES staff. To aid CES staff members in carrying out disclosure avoidance analysis, the researcher shall meet with these staff members at one or more agreed-upon times, specified well in advance.
4. Due to the complexity of disclosure avoidance analysis, any delays in providing the required information or in meeting with CES staff until the project is completed or nearly completed may greatly delay or even prevent the researcher from receiving any output. The researcher acknowledges and agrees that the timing of the release of any research outputs is determined by CES staff.
5. For purposes of this document, the parameters for the rules used in disclosure avoidance analysis are themselves confidential under Title 13, and to reveal them is considered a disclosure. Researcher acknowledges and agrees that the disclosure rules shall not be discussed by the researcher with any persons other than CES staff members.
6. Before the first release of research output, the researcher shall provide a statement of benefits to Census Bureau programs resulting from the work as described in the approved project proposal and Predominant Purpose Statement. CES will determine whether the statement is sufficient to release the research output.
7. Before the final release of research output, the researcher shall provide a Post Project Certification (PPC) and a Technical Memorandum. These (and other required documents) are described in section X. At its discretion, CES may require further statements of benefits before other releases of research output.
8. In the event of litigation based on this research, Census Bureau and/or U.S. Government confidentiality and disclosure avoidance rules cannot be waived. Moreover, Census Bureau personnel would not be made available to testify regarding the data and analysis the researcher develops.

## **X. Project Results**

The researcher agrees to provide work products at the completion of and throughout the duration of this project. More specifically, the researcher agrees to provide to CES the following documents based on project results.

1. Yearly progress reports. These reports will document the activities accomplished to date and

provide the designated CES staff member access to all project documentation.

2. A Post Project Certification (PPC) statement. This document states the key findings pertinent to Census Bureau programs, particularly, Chapter 5, Title 13 programs, including whether and how the accomplished benefits differed from those proposed in the Predominant Purpose Statement (PPS) and describing their potential application. In the case that a proposed benefit was not achieved, the PPC states why the project was unable to produce the benefit. This document should *not* include any undisclosed information; instead this document can make reference to such information by referring the reader an undisclosed technical note in the project directory on the CES server.

3. A Technical Memorandum (not to be confused with the technical note mentioned above). This document is a highly detailed account of how the research progressed. It should include information about how data sets were linked, the variables involved, any difficulties that were encountered and so forth.

For the following two items, note that any publications or presentation materials must include a disclaimer as described in *The Researcher Handbook*.

4. At least one research paper for inclusion in the CES Discussion Paper Series (or other appropriate Census Bureau series as designated by CES staff). These research papers may also appear in other paper series (e.g., university or NBER working paper series). The researcher agrees to provide CES with a copy of the paper as it appears in other paper series.

5. Copies of any published reports, journal articles, book chapters, slides from conference presentations, and the like.

Researcher acknowledges and agrees that failure to provide the documents stated above shall result in the termination of all data access, and that CES will reject all subsequent project proposals until the researcher provides these documents to CES. In particular, researcher acknowledges and agrees that CES, at its discretion, will not release research output until the researcher provides required documentation of benefits to Census Bureau programs.

## **XI. Signature Page**

Project **CES PROJECT NUMBER** titled "**PROJECT TITLE**" will begin **DATE START** and end **DATE END** for a total of **NUMBER OF MONTHS** months.

I, **NAME OF RESEARCHER**, certify that I have read this computer account and data request form, understand my responsibilities, and agree to the terms specified herein.

\_\_\_\_\_  
Signature of Researcher

Date: \_\_\_\_\_

This page of the ADRF should be faxed, mailed, or hand-delivered to your RDC Administrator. **ADMINISTRATOR CONTACT INFORMATION FOR RDC WHERE RESEARCHER WILL WORK**

## Appendix C: Data Breach Memorandum

May 1, 2007

### U.S. CENSUS BUREAU CENTER FOR ECONOMIC STUDIES (CES) PROCEDURES FOR DEALING WITH A SUSPECTED DATA SECURITY BREACH BY CENSUS RESEARCH DATA CENTER (RDC) RESEARCHERS

*Definition:* A data breach is defined as the unauthorized **access to, inspection of, or release of** information required by law to be kept confidential under Title 5 (personal information), Title 13 (Census Bureau information), Title 26 (Internal Revenue Service information), the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2002 or other relevant U.S. statute (e.g., for data supplied by another federal government agency); or for data supplied by any government agency (e.g., a State) required to be kept confidential.

#### Determining whether a data breach has occurred.

##### A. Personally Identifiable Information

1. As soon as a data breach involving Personally Identifiable Information (PII) is suspected, the individual identifying the suspected breach is required to inform the Census Bureau's Computer Incident Response Team (<BOC.CIRT@census.gov>; 1.877.343.2010 or 1.301.763.5141) within one hour.<sup>1</sup> (BOC CIRT will inform appropriate management officials.) All actions with regard to PII must be consistent with procedures specified by Census Bureau Data Stewardship DS-022 "Data Breach Policy".
2. The BOC CIRT Data Breach Team (DBT, as described in DS-022) will undertake an investigation of all suspected data breaches of PII, in cooperation with any security office of the partner institution hosting the RDC. The researcher, the CES Disclosure Officer, the CES Security Officer, and the RDC Administrator will be asked to provide information regarding the incident, and the RDC Executive Director will also be consulted. The DBT will attempt to determine if the security breach occurred; and if so, if it involved PII, Title 13, Title 26, or other data; if the disclosure was inadvertent or deliberate; and if the disclosure was reported by the researcher or identified by someone else. All individuals involved in the investigation will be requested to protect the researcher's identity until a breach has been verified. To the extent necessary, other federal agencies may be involved in the investigation.
3. If the BOC CIRT determines that a data security breach has occurred, the researcher will be given an opportunity to submit a statement for the record.

##### B. Non-PII Title 13, Title 26, or CIPSEA/other government data

1. As soon as a data breach involving non-PII Title 13, Title 26, or CIPSEA/other government data is suspected, the researcher must inform a CES management official (the CES Division Chief, the Assistant Division Chief for Research, the Assistant Division Chief for Research

---

<sup>1</sup> Memorandum M-06-19 (July 12, 2006) from the Office of Management and Budget requires the reporting of all *suspected* as well as confirmed incidents involving personally identifiable information (PII). According to the memo, PII is defined as "any information about an individual maintained by an agency, including, but not limited to, education, financial transactions, medical history, and criminal or employment history and ... name, social security number, date and place of birth, mother's maiden name, biometric records, etc." See <<http://www.whitehouse.gov/omb/memoranda/fy2006/m-06-19.pdf>>. The definition of PII is evolving so any release of information that might be construed as PII must be reported. The Census Bureau will determine the classification of all information.

Support, or the Lead RDC Administrator) within one business day. (If no senior CES management official is available, the RDC Administrator for the location that the suspected breach occurred must be informed.)

2. The CES Security Officer or the CES Disclosure Officer will lead an investigation of all suspected data breaches, possibly in cooperation with any security office of the partner institution hosting the RDC and any Census Bureau office that could provide assistance. The researcher, the CES Disclosure Officer, and the RDC Administrator will be asked to provide information regarding the incident, and the RDC Executive Director will also be consulted. CES will attempt to determine if the security breach occurred; and if so, if it involved Title 13, Title 26, or other data; if the disclosure was inadvertent or deliberate; if the disclosure was reported by the researcher or identified by someone else; and whether the disclosure would have been approved if a request had been made to do so. All individuals involved in the investigation will be directed to protect the researcher's identity until a breach has been verified.

3. In cases where the researcher does not admit that there has been an unapproved disclosure, the CES Division Chief will consult with at least one RDC Executive Director not involved in the incident about how to interpret the information that has been gathered. If the CES Division Chief does determine that a data security breach has occurred, the researcher will be given an opportunity to submit a rebuttal for the record.

### **CES policy establishing procedures for responding to a verified data security breach.**

*A. For disclosure of data or estimates without prior approval by CES Disclosure Officer or other CES official, but not in violation of Title 13 or Title 26 confidentiality requirements or those pertaining to other data accessible at the RDCs (that is, the information would have been approved for release had the proper procedures been followed).*

#### 1. Title 13 or Title 26 data

*a. First offense:* The CES Division Chief issues a formal warning letter to the individual, with a copy to the RDC Executive Director and the individual's direct supervisor (e.g., university department chair or dean). If the disclosure was inadvertent and especially if the disclosure was revealed to CES by the researcher, the CES Division Chief may, at his or her sole discretion, determine that a letter not be sent to the individual's direct supervisor.

*b. Second offense:* One-year ban on applying for any new project at an RDC effective when the current project expires. The Census Bureau's Associate Director for Economic Programs (ADEP) issues a violation letter to the offender, with a copy to the RDC Executive Director, and the head of the individual's institution (e.g., the university president). If the disclosure was inadvertent and especially if the disclosure was revealed to CES by the researcher, the ADEP may, at his or her sole discretion, determine that a letter be sent to the individual's direct supervisor instead of the head of the institution.

*c. Third offense:* Immediate suspension of the offender's Special Sworn Status (SSS) and a lifetime ban on applying to the RDC program. The Census Bureau's Deputy Director issues a notification letter to the individual, with a copy to the RDC Executive Director and the head of the individual's institution (e.g., the university president).

#### 2. CIPSEA or other government data

*a. First offense:* CES Division Chief issues a formal warning letter to the offender with a copy to the Census Bureau's Data Custodian, the RDC Executive Director, and the individual's direct

supervisor. The Data Custodian will inform the data provider if so indicated in the Memorandum of Agreement (MOA) with the provider. If the disclosure has no practical impact, the Data Custodian may determine that notification of the data provider is not appropriate. Any requirements or penalties in the MOA will apply.

*b. Second offense:* One-year ban on applying for any new project at an RDC effective when the current project expires. The Census Bureau's Associate Director for Economic Programs (ADEP) issues a formal warning letter to the offender with a copy to the Census Bureau's Data Custodian, the RDC Executive Director, and the head of the individual's institution. The Data Custodian will inform the data provider if indicated in the MOA with the provider. Any requirements or penalties in the MOA will apply.

*c. Third offense:* Immediate suspension of SSS and a lifetime ban on applying to the RDC program. The Census Bureau's Deputy Director issues a notification letter to the offender with a copy to the Census Bureau's Data Custodian, the RDC Executive Director, and the head of the individual's institution. The Data Custodian will inform the data provider if indicated in the MOA with the provider. Any requirements or penalties in the Agreement will apply.

*B. For disclosure of data without approval by the CES Disclosure Officer or other CES official that is determined to be in violation of Title 13, Title 26, or other statutory, administrative, or consensual (e.g., MOA) restriction, including disclosure of PII.*

1. PII, Title 13, or Title 26 data

a. The CES Division Chief will inform the IT Security Officer, the appropriate Associate Director(s), the Deputy Director, the Internal Revenue Service (IRS, if Title 26 data are involved), and the RDC Executive Director, and immediately suspend the researcher's SSS.

b. The corrective measures to be taken will be determined by the Deputy Director (and if appropriate, the IRS), which may include a lifetime ban on working in the RDC program and may include turning the case over to the Department of Justice for prosecution. Note that violation of Title 13 could result in a fine up to \$250,000 and/or imprisonment for up to 5 years.

c. At the appropriate time, the Census Bureau will inform the head of the individual's institution of the individual's conduct and assessed penalty.

2. CIPSEA or other government data

a. The CES Division Chief will inform the Census Bureau's Data Custodian, the IT Security Officer, the ADEP, the Deputy Director, and the RDC Executive Director and immediately suspend SSS. The Data Custodian will inform the data provider as indicated in the MOA with the provider.

b. The corrective measures to be taken will be determined by the data provider and the Census Bureau, may include a lifetime ban on applying to the RDC program, and may include turning the case over to the Department of Justice or other relevant entity for prosecution.

c. At the appropriate time, the Census Bureau will inform the head of the individual's institution of the individual's conduct and assessed penalty.

*C. When a data breach has been confirmed, the RDC involved will review its procedures for maintaining confidentiality of the data and output and take corrective action if appropriate. Two or more breaches in a 12-month period may result in temporary or permanent closure of the facility.*

### Appendix D: Clearance Request Memo

#### REQUEST FOR CLEARANCE OF RESEARCH OUTPUT

Center for Economic Studies and Research Data Centers

\*\*\*\*\*

\* Project #:

\* Submitted by:

\*

\* For CES Reviewer to complete:

\* Cleared for release:

\* Cleared by:

\*\*\*\*\*

#### 1. GENERAL INFORMATION

a. Name of this request's subdirectory under the project's main clearance directory:

b. Please provide a general description of the outputs you wish to clear:

c. Please state how the outputs are part of the research project as approved (You may summarize or copy descriptions from your proposal, with page references.)

#### 2A. DESCRIPTIONS OF RESEARCH SAMPLES:

Describe your Research sample(s) or "cuts" of data used in research output. For each sample, please describe your selection criteria and how the research sample differs from the samples underlying survey publications or other samples you have used. Take as much space as you need for each; add samples as needed.

SAMPLE 1:

SAMPLE 2:

SAMPLE 3:

#### 2B. RELATIONSHIP BETWEEN SAMPLES

Describe how your samples relate to each other (e.g., if you have two samples, is one a subsample of another?) In the cases of samples and subsamples, there is an implicit third sample, the difference between the two. Please describe this sample above. We probably will need to examine any implicit samples as well.

#### 2C. RELATIONSHIP TO OTHER PUBLICATIONS

Describe how your samples may relate to similar samples from other projects or from survey publications. (e.g., how your sample of an industry in the LRD differs from the Census of Manufactures or Annual Survey of Manufactures files in the LRD).

3. OUTPUT FILES

For each research output file to be removed, please enter the following information.

- a. File name (e.g., output.rtf)
- b. Description of file (e.g., tables relating to \_\_ or models of \_\_)
- c. Program that produced the file (e.g., output.sas or model.do)
- d. The Research Sample Number that underlies the file
- e. The disclosure analysis file's name—the file that contains supporting statistics
- f. The name of the program that generated the disclosure analysis file (e.g., output\_disc.sas). Note, if the disclosure information is produced in the same program as the research output to be removed, please cut and paste this into a new file.
- g. Results (we enter this)—whether the file was cleared or not.
- h. Comments—any relevant information you or the person who clears the file may wish to add.

FILE NUMBER: 1

FILE NAME:

FILE DESCRIPTION:

RESEARCH OUTPUT PROGRAM:

RESEARCH SAMPLE NUMBER:

DISCLOSURE ANALYSIS FILE NAME:

DISCLOSURE ANALYSIS PROGRAM:

RESULTS:

COMMENT:

FILE NUMBER: 2

FILE NAME:

FILE DESCRIPTION:

RESEARCH OUTPUT PROGRAM:

RESEARCH SAMPLE NUMBER:

DISCLOSURE ANALYSIS FILE NAME:

DISCLOSURE ANALYSIS PROGRAM:

RESULTS:

COMMENT:

4. VARIABLE DEFINITIONS

Please include all variables in the research output you wish to remove and in supporting disclosure output.

VARIABLE NAME:

DEFINITION (include type, e.g., continuous, binary (0,1), other discrete):

SOURCE:

COMMENTS:

VARIABLE NAME:

DEFINITION: (include type, e.g., continuous, binary (0,1), other discrete):

SOURCE:

VARIABLE NAME:

DEFINITION: (include type, e.g., continuous, binary (0,1), other discrete):

SOURCE:

VARIABLE NAME:

DEFINITION: (include type, e.g., continuous, binary (0,1), other discrete):

SOURCE:

VARIABLE NAME:

DEFINITION: (include type, e.g., continuous, binary (0,1), other discrete):

SOURCE:

5. OTHER INFORMATION OR COMMENTS:

Please enter any further information you feel is relevant below.

## Appendix E: User Provided Data Memo

### User Provided Data Memo

Date:

Project ID:

Researcher Name:

**Each data set that you would like uploaded requires a separate User Provided Data Memo.**

#### File information:

For SAS data set, paste “PROC CONTENTS” here AND fill in #2 below<sup>1</sup>:

For Stata data set, paste “DESCRIBE” here:

**For any other data set**, fill in all of the following (for SAS data fill in #2):

1. File name:
2. File size (KB):
3. File type:
4. Number of records:
5. List of variables (record layout is sufficient):

#### Destination Location for Files:

RDC server path (e.g. rdcluster2:rdcprojects/br1/br00999/data):

#### Permissions for the Files:

The default settings will set the owner of the file as the researcher submitting the request, the group will be your project group, and the files will have read-only access for the owner and the group. Fill in the following *ONLY IF* you want these changed.

1. Owner:
2. Group:
3. Permissions:

#### Permissions (mark all that apply):

- The data are public use and...
- downloaded from the internet. The **exact** URL(s) is(are):
- procured other than from the internet. Attached is evidence of public use.
- The data are purchased. Attached is a receipt of purchase or documentation from my institution that it subscribes to this data.
- The data are proprietary. Attached is written permission from the data custodian or vendor expressing consent to use the data on the RDC computer system and specifying any restrictions on the data's use.

#### Description (1-2 sentence description of the data):

---

<sup>1</sup> Earlier versions of SAS data sets **must** be in Unix SAS format. PC SAS data before version 8 are not compatible.

## Appendix F: Application to Add Personnel to an Approved Project

### Application to Add Personnel to an Approved Project

**Date:**

**Name:**

**Affiliation:**

**Project number:**                      **RDC:**

**Researcher Status** (check one):

| | Co-Principal Investigator/Senior Researcher

| | Research Assistant

| | Other (specify):

**Project Role**—Which aspects of the project will this person work on? What specific activities will they perform in support of the project?

**Data sets**—Identify the approved data sets that this person will be actively using. If all data, please so state.

**Record Linkages**—Which if any data sets will be subject to linkages performed by this person, and at what level (individual, household, establishment, firm, industry, geography, other)?

**External Data**—If this person's activity involves the addition of external data sets other than those already approved for the project, then please list them. (A User Provided Data Memo will be required for any such data.)

**Benefits**—To, which approved benefits to the Census Bureau will this person's activity be directed? How will those benefits be fulfilled or enhanced by this person's participation?

**Disclosure**—What types of computer output will be submitted for disclosure as a result of this person's participation?

**Products**—Will this person's contribution be that of a sole author or co-author of any documents resulting from this project such as Post Project Certification, Technical Memorandum, Research Paper (including working papers), or doctoral dissertation? If so, please specify.

## Appendix G: Legal Information

This appendix provides information (summaries and quotes) on the laws underlying the policies spelled out in this researcher handbook. The most important law is Title 13 U.S. Code for the Census Bureau. Titles 18 and 26 for the Internal Revenue Service (IRS) is also extremely important to the Census Bureau because IRS federal tax data are used in many Census Bureau data programs—chiefly in constructing and maintaining sampling frames. The IRS legislation imposes certain restrictions on this use.

### **Census Bureau Legislation—Title 13 U.S. Code**

#### **General Information**

Title 13 authorizes the Census Bureau to carry out specific data collection programs. It also requires the Census Bureau to maintain the confidentiality of the data the Census Bureau collects, and it authorizes the Bureau to use "temporary staff" to "assist the Bureau in carrying out the work authorized by this title."

Several chapters of Title 13 authorize the Census Bureau to collect data. This is important because Under IRS law (Title 26, section 6103), Federal tax data may be used only in programs authorized under Chapter 5 and to carry out "related statistical activities authorized by law." The Decennial Census and the Economic Census and related survey programs are authorized under Title 13 Chapter 5. An important exception is the Quarterly Financial Reports (QFR), which is authorized under Title 13 Chapter 3, but is specifically authorized to use Federal tax data under Title 26 section 6103 as a "related statistical activity authorized by law."

The following sections provide relevant sections from Title 13, along with brief summaries.

#### **Protection of Confidential Information—Sections 9 and 214**

##### *Summary:*

With the exception of data from censuses or interim surveys of governments, information collected and/or compiled under authority of Title 13, is confidential. The following restrictions apply:

- Data may be used only for statistical purposes
- No publication may permit identification of data for a particular business or individual
- Only sworn employees may examine individual reports
- Temporary staff (i.e., persons with Census Bureau Special sworn Status) may assist in performing work authorized by Title 13 but only if those persons are sworn to observe the confidentiality provisions in Section 9 of the statute.
- Conclusion: All Researchers at RDCs must assist the Census Bureau in performing work authorized by Title 13, and all must be sworn to observe the confidentiality provisions of Title 13, Section 9.

##### *Law:*

CHAPTER 1—ADMINISTRATION  
SUBCHAPTER I— GENERAL PROVISIONS

**Sec. 9. Information as confidential; exception**

- (a) Neither the Secretary, nor any other officer or employee of the Department of Commerce or bureau or agency thereof, or local government census liaison may, except as provided in section 8 or 16 or chapter 10 of this title or section 210 of the Departments of Commerce, Justice, and State, the Judiciary, and Related Agencies Appropriations Act, 1998.<sup>(1)</sup>
- (1) use the information furnished under the provisions of this title for any purpose other than the statistical purposes for, which it is supplied; or
  - (2) make any publication whereby the data furnished by any particular establishment or individual under this title can be identified; or
  - (3) permit anyone other than the sworn officers and employees of the Department or bureau or agency thereof to examine the individual reports. No department, bureau, agency, officer, or employee of the Government, except the Secretary in carrying out the purposes of this title, shall require, for any reason, copies of census reports, which have been retained by any such establishment or individual. Copies of census reports, which have been so retained shall be immune from legal process, and shall not, without the consent of the individual or establishment concerned, be admitted as evidence or used for any purpose in any action, suit, or other judicial or administrative proceeding.
- (b) The provisions of subsection (a) of this section relating to the confidential treatment of data for particular individuals and establishments, shall not apply to the censuses of governments provided for by subchapter III of chapter 5 of this title, nor to interim current data provided for by subchapter IV of chapter 5 of this title as to the subjects covered by censuses of governments, with respect to any information obtained therefor that is compiled from, or customarily provided in, public records

**Sec. 214. Wrongful disclosure of information**

Whoever, being or having been an employee or staff member referred to in subchapter II of chapter 1 of this title, having taken and subscribed the oath of office, or having sworn to observe the limitations imposed by section 9 of this title, or whoever, being or having been a census liaison within the meaning of section 16(2) of this title, publishes or communicates any information, the disclosure of, which is prohibited under the provisions of section 9 of this title, and, which comes into his possession by reason of his being employed (or otherwise providing services) under the provisions of this title, shall be fined not more than \$5,000<sup>1</sup> or imprisoned not more than 5 years, or both.

**Authority for Temporary Staff (Special Sworn Status)—Section 23(c)**

*Law:*

CHAPTER 1—ADMINISTRATION  
SUBCHAPTER I— OFFICERS AND EMPLOYEES

**Section 23. Additional officers and employees (excerpts)**

- (b) The Secretary may utilize temporary staff, including employees of Federal, State, or local agencies or instrumentalities, and employees of private organizations to assist the Bureau in performing the work authorized by this title, but only if such temporary staff is sworn to observe the limitations imposed by section 9 of this title. Authorizations For Data Collection Programs—Chapter 5

**Data Programs Authorized by Title 13 Chapter 5**

*Summary:*

The following data collection programs are authorized by Title 13 Chapter 5.

- Quinquennial Economic (Business) Censuses (Sec. 131)
- Decennial Census of Population and Housing (Sec. 141)
- Quinquennial Census of Governments (Sec. 161)
- Intercensal Population Estimates (Sec. 181)
- Current Survey Programs—on the subjects covered by Title 13 Census programs (Sec. 182)

<sup>1</sup> Note: The fine has been increased to \$250,000, as stated elsewhere in this handbook.

Law:

*Economic (Business) Censuses:*

**CHAPTER 5**

**SUBCHAPTER I—MANUFACTURES, MINERAL INDUSTRIES, AND OTHER BUSINESSES**

**Sec. 131. Collection And Publication; Five—Year Periods**

The Secretary shall take, compile, and publish censuses of manufactures, of mineral industries, and of other businesses, including the distributive trades, service establishments, and transportation (exclusive of means of transportation for, which statistics are required by law to be filed with, and are compiled and published by, a designated regulatory body), in the year 1964, then in the year 1968, and every fifth year thereafter, and each such census shall relate to the year immediately preceding the taking thereof.(9)

*Decennial Census Of Population And Housing:*

**CHAPTER 5**

**SUBCHAPTER II—POPULATION, HOUSING, AGRICULTURE, IRRIGATION, AND UNEMPLOYMENT**

**Sec. 141. Population And Other Census Information.**

- (a) The Secretary shall, in the year 1980 and every 10 years thereafter, take a decennial census of population as of the first day of April of such year, which date shall be known as the "decennial census date", in such form and content as he may determine, including the use of sampling procedures and special surveys. In connection with any such census, the Secretary is authorized to obtain such other census information as necessary.

*Census of Governments*

**CHAPTER 5**

**SUBCHAPTER III—GOVERNMENTS**

**Sec. 161. Quinquennial Censuses; Inclusion Of Certain Data**

The Secretary shall take, compile, and publish for the year 1957 and for every fifth year thereafter a census of governments. Each such census shall include, but shall not be limited to, data on taxes and tax valuations, governmental receipts, expenditures, indebtedness, and employees of States, counties, cities, and other governmental units.

**Sec. 163. Authority Of Other Agencies**

This subchapter does not revoke or impair the authority of any other Federal agency with respect to the collection or release of information.

**CHAPTER 5**

**SUBCHAPTER IV—INTERIM CURRENT DATA**

*Intercensal Population Estimates*

**Sec. 181. Population**

- (a) During the intervals between each census of population required under section 141 of this title, the Secretary, to the extent feasible, shall annually produce and publish for each State, county, and local unit of general purpose government, which has a population of fifty thousand or more, current data on total population and population characteristics and, to the extent feasible, shall biennially produce and publish for other local units of general purpose government current data on total population. Such data shall be produced and published for each State, county, and other local unit of general purpose government for, which data is compiled in the most recent census of population taken under section 141 of this title. Such data may be produced by means of sampling or other methods, which the Secretary determines will produce current, comprehensive, and reliable data.

*Survey Programs*

**Sec. 182. Surveys**

The Secretary may make surveys deemed necessary to furnish annual and other interim current data on the subjects covered by the censuses provided for in this title.

**Other Title 13 Data Programs**

*Summary:*

The following data programs are authorized by other parts of Title 13.

- Quarterly Financial Reports (Chapter 3)

- Foreign Commerce and Trade Statistics (Chapter 9)

*Law:*

**TITLE 13**

**Chapter 3**

**Subchapter IV—Quarterly Financial Statistics**

**Sec. 91. Collection and publication**

- (a) The Secretary shall collect and publish quarterly financial statistics of business operations, organization, practices, management, and relation to other businesses, including data on sales, expenses, profits, assets, liabilities, stockholders' equity, and related accounts generally used by businesses in income statements, balance sheets, and other measures of financial condition.
- (b) Except to the extent determined otherwise by the Secretary on the basis of changed circumstances, the nature of statistics collected and published under this section, and the manner of the collection and publication of such statistics, shall conform to the quarterly financial reporting program carried out by the Federal Trade Commission before the effective date of this section under section 6(b) of the Federal Trade Commission Act.
- (c) For purposes of section 6103(j)(1) of the Internal Revenue Code of 1986, the conducting of the quarterly financial report program under this section shall be considered as the conducting of a related statistical activity authorized by law.

**TITLE 13**

**Chapter 9—COLLECTION AND PUBLICATION OF FOREIGN COMMERCE AND TRADE STATISTICS**

**Sec. 301. Collection and publication**

- (a) The Secretary is authorized to collect information from all persons exporting from, or importing into, the United States and the noncontiguous areas over, which the United States exercises sovereignty, jurisdiction, or control, and from all persons engaged in trade between the United States and such noncontiguous areas and between those areas, or from the owners, or operators of carriers engaged in such foreign commerce or trade, and shall compile and publish such information pertaining to exports, imports, trade, and transportation relating thereto, as he deems necessary or appropriate to enable him to foster, promote, develop, and further the commerce, domestic and foreign, of the United States and for other lawful purposes.(18)

## ***Protecting Federal Tax Information (FTI)—Titles 26 and 18, U.S. Code***

This section contains a brochure (reformatted for inclusion in this handbook) that describes the need to safeguard and protect FTI. It summarizes the policies, the underlying laws, and punishments for violations.

Before gaining access to any data set containing FTI, you will be required to undergo awareness training and sign a statement indicating your understanding of and willingness to follow the policies that govern your restricted access to FTI.

## QUICK REFERENCE GUIDE

# SAFEGUARDING & PROTECTING FEDERAL TAX RETURNS & RETURN INFORMATION

The Census Bureau obtains Federal Tax Returns and Return Information (FTI) and related “Official Use Only” data from the Internal Revenue Service (IRS) for statistical purposes. As a recipient of these data, the Census Bureau must ensure that it establishes and maintains to the satisfaction of the IRS certain safeguards designed to prevent unauthorized access to and use of FTI, as well as to protect the confidentiality of that information. Failure to adequately protect these data can result in severe penalties, including fines, imprisonment, civil suits,

## Definitions

All FTI and “Official Use Only” data should be treated as confidential, which means disclosure is restricted. **Disclosure** is the making known of FTI or “Official Use Only” data in any manner to anyone. There are three types of disclosures:

- **Authorized disclosure**—disclosing confidential data to other employees with a work-related “need to know.” This is legal.
- **Willful unauthorized disclosure**—knowingly providing confidential data to anyone who is not entitled to have them. This is illegal.
- **Inadvertent unauthorized disclosure**—mistakenly giving confidential data to a person who shouldn’t have it. Penalties can apply under Title 13.

**Willful unauthorized access** or inspection of FTI is simply looking at, or browsing through, data that are not needed to complete required tasks. This is illegal.

An **authorized use of FTI** at the Census Bureau is one that both Census and IRS officials have approved through the formal review and approval process.

## IMPORTANT FACTS

- ✓ All tax data, including names and addresses, are equally sensitive and are subject to IRS safeguard requirements and penalties.
- ✓ Files that contain both Federal Tax Returns and Return Information (FTI) and non-FTI data are referred to as “commingled” data sets and are subject to the same IRS safeguard requirements as FTI, no matter how limited the use of FTI.
- ✓ Each individual is bound by these Federal laws for life, regardless of his or her employment status.
- ✓ FTI data used at the Census Bureau are subject to the privacy and confidentiality provision of Title 26 Internal Revenue Code, Title 13 Census, and Title 18 Crimes and Criminal Procedure.
- ✓ Unauthorized disclosure of FTI is illegal. Only provide FTI to those with an approved, work-related “need to know.”
- ✓ Casual browsing of FTI for, which you do not have an authorized “need to know” is illegal. For example, unless you have a “need to know,” you cannot inspect tax records of your neighbor or favorite TV personality.
- ✓ FTI must be secured at all times.
- ✓ No matter the media, FTI must be labeled as such.
- ✓ Listings and computer screens displaying FTI must not be visible to visitors or other non-authorized users.
- ✓ FTI and any material generated therefrom, such as extra copies, photo impressions, computer printouts, carbon paper, notes, and work papers should be destroyed by burning,

## **Additional Information**

For more detailed information, visit:

### **IRS Publication 1075**

<http://ftp.fedworld.gov/pub/irs-pdf/p1075.pdf>

### **Administrative Records Handbook**

<http://cww.census.gov/po/>

### ***IT Security Office***

<http://cww.census.gov/itso/>

