

INTERAGENCY AGREEMENTS FOR MICRODATA ACCESS: THE ERUMS EXPERIENCE

Thomas B. Petska, Internal Revenue Service and
Lois Alexander, Social Security Administration

The Employer Reporting Unit Match Study (ERUMS) was a pilot record linkage study carried out under the auspices of the Federal Committee on Statistical Methodology of the Office of Management and Budget. The study linked records of employers and their reporting units from three agencies: the Bureau of Labor Statistics (BLS), the Social Security Administration (SSA) and the Internal Revenue Service (IRS). The primary linkages involved samples of the agencies' records for employers in the State of Texas covering their activities in 1982.

For the ERUMS Workgroup to gain access to the data sets needed for the study, arrangements had to be developed that would comply with the confidentiality provisions and statutes of the Federal and State agencies that controlled these data sets. This paper gives an overview of these arrangements and agreements. In the first section, background information on the statistical content and confidentiality provisions of each of the data sets is provided. In the second section, the actual arrangements for the release of confidential microdata are described. The last section provides a summary of what we have learned about such data sharing arrangements.

BACKGROUND INFORMATION

The goal of ERUMS was to demonstrate the feasibility of matching employer and reporting unit data from different agency record systems as a means of obtaining more precise information about the coverage and content of the data in those systems. A purpose was to examine and evaluate differences in wage and employment data at the state and county level as reported to those agencies. Despite the many difficulties encountered in establishing the data access agreements, ERUMS demonstrated that data such sharing projects can be successful under current laws.

1. Data Set

The ERUMS study was a three-way data linkage study in which individual microdata records from BLS, SSA, and IRS were matched by Employer Identification Number (EIN).

- a. BLS provided a 1982 Unemployment Insurance (UI) Address File, which, for each state, consists of data for individual employers and their reporting units, which are often equivalent to "establishments." The data for this file are submitted to BLS by the State employment security agencies that operate the Federal-State UI Program. BLS uses the data submitted by the states as a basis for statistical reports on employment and wages and uses the UI Address File as a national sampling frame for its establishment surveys.
- b. SSA provided an edited file of Form W-3 annual reports for 1982 and the Single-Unit and Multi-Unit Code Files. The Form W-3 file provided data on individual employers and, in some cases, for each of their reporting units, which are frequently equivalent to establishments. The Single-Unit Code File contains a record for most entities that have filed an application for an Employer Identification Number. The Multi-Unit Code File contains a record for each reporting unit of multi-unit employers who are participating in the Establishment Reporting Plan, a voluntary program under which employers report wage information on Form W-3 separately for each of their reporting units.
- c. IRS data used for ERUMS were from a Census-edited file based on Forms 941 and 943 for Tax Years 1981-83. These forms are used by em-

Form 943) to IRS on income taxes withheld from wages and other payments to employees and on taxes under the Federal Insurance Contributions Act (FICA) under the Social Security system. Extracts of data from these forms are provided annually by IRS to the Census Bureau for use in the latter's County Business Patterns Program and other statistical programs. The Census Bureau edits the files, particularly the industry codes, and imputes certain missing data. This file was made available to the IRS Statistics of Income (SOI) Division for use in its business employment and payroll studies and was used for ERUMS. In addition, copies of Form 940, Federal Unemployment Tax Return, were obtained for a substantial proportion of the ERUMS sample cases.

2. Data Sharing Issues

For the ERUMS Workgroup to gain access to the data sets needed for the study, it was necessary to develop working arrangements that complied with the provisions of confidentiality statutes, regulations, and policies of the Federal and State agencies that controlled these data sets.

Although interagency exchange of identifiable microdata was the key to ERUMS, such data sharing is restricted by Federal confidentiality laws which generally permit agencies to disclose statistical information only in summary or other unidentifiable form. Since ERUMS was designed to link and compare information about individual employers collected separately by the different agencies, the Workgroup had to develop and implement lawful methods of transferring data on identifiable business units among the participants. A related task was to minimize the disclosure of identifiers in making those transfers and linkages.

The Workgroup was particularly interested in the different ways an employer may report establishment or multi-unit enterprise data to various State and Federal agencies. To examine these differences, the Workgroup needed to compare employers' reports to the BLS State UI programs, the SSA FICA reporting, and the IRS employment tax returns. Members of the

Workgroup included employees of these agencies, plus employees of the Bureau of Economic Analysis, Office of Management and Budget, the Bureau of the Census, and the Committee on National Statistics of the National Academy of Sciences.

The Workgroup planned to analyze the information that corresponded to each EIN as it was reported to each agency. The analysis and findings would be entirely statistical in nature with no reference to the individual (identifiable) cases. Nevertheless, the planning, processing, and analysis phases each required access to identifiable data.

3. Confidentiality of Federal and State Tax Records

In the ERUMS study, the Employer Identification Number (EIN) was the identifier that was common to all the reporting systems. It was used to define the sample drawn by BLS and was used as the basis for retrieving, linking and comparing records containing information from the SSA and IRS files. By law, the EIN is a tax identification number, and even when standing alone is protected by Internal Revenue Code confidentiality restrictions.

ERUMS required access to data from W-3 records which by law are Federal tax records that are processed and maintained at SSA in conjunction with the computation of Social Security retirement benefits. Since these are tax records, it was necessary to satisfy IRS that the selection by SSA of sample cases, SSA's disclosure of W-3 data to BLS, and the use of employer data by other members of the Workgroup met the requirements of the Internal Revenue Code dealing with disclosure of tax information. (See No. 4 below.)

BLS selected Texas as the State whose records it would sample, and it obtained written permission from the Texas State Employment Security Agency to use their UI records in the project. The Texas Unemployment Compensation Act requires Texas employers to maintain records and file reports to the Texas Employment Commission with detailed information about the business operations and the number and compensation of employees. Texas law prohibits disclosure except for administering the Act, and it

its disclosure except for administering the Act, and it makes improper disclosure punishable by fines or imprisonment.

4. Other Confidentiality Considerations

Since the Workgroup was composed of employees from several agencies and organizations, confidentiality laws did not apply to them uniformly. In varying degrees, certain laws, regulations, and policies affected each agency's access to identifiable records from particular sources and provided differential access to various individuals in the Workgroup. A recurring theme was the necessity at each phase of the process to identify the persons who needed to use identifiable data and to ensure that no others had access at that time.

Besides affidavits and other written procedures to protect the confidentiality of records, certain technical safeguards were adopted to minimize disclosure risk. The first of these methods was to avoid identifying sample cases by EIN to persons who performed processing in the participating agencies but were not directly associated with the Workgroup. This method was adopted to conform to the Internal Revenue Code requirements for tax information under the agreement BLS had with the State of Texas. At BLS this led to a decision not to process the data on the mainframe computer system at the Department of Labor that is operated by a private contractor. Instead, BLS used a mini-computer which was accessible only to BLS employees who were members of the Workgroup.

State agencies periodically submit to BLS UI address files that compile identification data for all reporting units at the most-detailed level that is available from employers' reports. BLS compiles these reports under a pledge of confidentiality that allows the data to be used only by authorized persons for statistical purposes.

Once BLS selected the Texas sample, it had to create a finder list so that SSA could extract corresponding records from its W-3 and related files for employers in the sample. The technical staff who performed these operations at SSA have routine access in their usual jobs to the employer records

maintained at SSA. However, they did not need to know which of the employers' records comprised the sample selected by BLS from the Texas UI file. To avoid identifying those cases that were actually in the sample, BLS furnished SSA with a listing of 7 of the 9 digits of sample EINs. SSA staff then extracted records from the W-3 and related files for all records in which these 7 digits appeared without knowing which employers were actually in the BLS sample. This procedure effectively masked the identities of sample cases derived from State UI files, and thus significantly limited the number of SSA employees who were required to sign BLS non-disclosure affidavits.

AGREEMENTS FOR INTERAGENCY DATA SHARING

Access by the Workgroup to the data sets needed for the study was accomplished through three interagency agreements plus an additional access arrangement.

The Workgroup had originally planned a tripartite arrangement through interagency agreements of SSA and BLS with IRS. However, IRS counsel raised objections that such a multi-party agreement would be unduly cumbersome, and approval would probably not be forthcoming. As an alternative, IRS proposed to contract exclusively with BLS for the performance by BLS of services that required access to tax data. SSA staff would be designated as special agents of BLS to process the data. Bilateral BLS/IRS and BLS/SSA agreements would also have to be drafted under this arrangement.

The drafting of these arrangements proved to be a delicate task. By law, the purposes of IRS participation in the project and its service contract with BLS had to be related to IRS administration of the tax laws. Section 6103(n) of the Internal Revenue Code (IRC) allows IRS to disclose tax return information to persons outside of the agency as long as it is for purposes of tax administration [1]. Specifically, this purpose is to conduct statistical studies based on return information, which Section 6108(a) of the IRC authorizes IRS to perform [2]. A case was made that the ERUMS study was one such purpose.

1. BLS and Texas Agreement

BLS has cooperative agreements with 50 State Employment Security Agencies to use employment statistics collected by the states for its laboreconomics research. The 1982 data used in the ERUMS study was furnished to BLS in its ES-202 program by the Texas State Employment Commission under a cooperative agreement. It was necessary for BLS to obtain authorization from the State Commission to use the microdata for the ERUMS study and to provide access for the Workgroup members. Under this cooperative agreement, the access and use of the data were subject to the confidentiality requirements of the Texas Employment Compensation statute as well as those set out in the BLS Commissioner's Order No. 2-80.

Each UI program is operated under state law that must conform to certain minimum federal standards, with reports that enable BLS to monitor state compliance. Under the Texas program, each employing unit is required to file (and update periodically) a status report with the Texas Employment Commission, describing the type of ownership, location, and nature of business. On a quarterly basis, employers are required to file detailed reports on wages and contributions. Multi-unit employers are asked to file a voluntary statistical supplement that provides detailed employment, wage, and contribution reports for each establishment. The ES-202 reports are compiled by BLS and form the basis for the UI Address file that BLS maintains. This is a micro-level employer file that contains first quarter information for each reporting unit, and the 1982 file provided the Texas sampling frame for the ERUMS sample.

The confidentiality of statistical data collected under the cooperative agreement is protected by interrelated state and federal procedures. At the state level, these UI reports are collected under the Texas Unemployment Compensation Act which limits the availability of its UI reports to public employees in the performance of public duties, except as the Employment Commission may find necessary in its administration of Texas law. At the federal level, BLS receives and maintains these confidential reports under the authority of the BLS

Commissioner's Order that pledges confidentiality and prohibits disclosure except to authorized persons for statistical purposes. This Order precludes any use of identifiable information for non-statistical purposes, such as investigation or enforcement.

Under this cooperative agreement with the State of Texas, it was necessary for BLS to obtain permission from the Texas Commissioner to select employer sample cases and to make information about them available to BLS and SSA employees in the ERUMS Workgroup and later to others in the Microdata Access Group. In addition, BLS procedures establish the confidentiality of the identities and all information pertaining to employers in the sample. Members of the Workgroup who were not BLS employees were appointed as BLS agents pursuant to another interagency agreement with BLS. Like BLS employees, other Workgroup members were required to sign a Non-Disclosure Affidavit before they would be given access to the microdata.

2. IRS and BLS Agreement

The initial draft of the statement of purpose by IRS representatives was acceptable to IRS counsel since its justification for sharing of confidential tax information was defined as for purposes of tax administration, which is permissible under section 6103(n) of the Internal Revenue Code [1]. However, the case that was made for IRS tax administration purposes was not acceptable to other Workgroup participants because they felt that this did not clearly describe the purposes of the ERUMS project in general or SSA's role in particular. In the subsequent draft, care was taken to define contractual purposes in language that covered the statistical purposes of the several participating agencies and that provided for the exchange of records to create a common pool of data for a variety of analytical purposes, including those related to tax administration.

In this agreement, IRS contracted with BLS for the performance of those parts of the ERUMS project that required access to tax data, including the wage report information that was to be provided by SSA. Under this agreement, SSA staff could be designated as special agents of BLS to carry out their part of the linkage and analysis operations. By law, the pur-

poses of IRS participation in the project and its service contract with BLS had to be related to IRS administration of the tax laws.

The terms of a contract between IRS and BLS which needed to be acceptable to SSA enabled BLS to receive tapes containing tax information from IRS and SSA and to combine them with records in the UI Address File maintained by BLS. It imposed strict safeguard procedures and required BLS to provide IRS with a list of all persons permitted to see confidential tax return data. This list included SSA employees who were required to sign affidavits as agents of BLS.

3. BLS and SSA Agreement

The third agreement was a Conditions of Use agreement between BLS and SSA which enabled SSA to release data from its employer files to BLS and authorized BLS to link data from these files to data in the UI Address File and data to be furnished by IRS. Like the IRS/BLS agreement, it limited access at each stage of the project to those persons who needed to use identifiable data, kept the number of such persons to a minimum, and required adequate physical security procedures. This agreement, which needed to be acceptable to IRS, enabled BLS to use SSA files for the ERUMS project. Under this agreement, SSA would furnish BLS with SSA's Single Unit Code File, Multi Unit Code File, and Employer Report (W-3) Record. The agreement authorized BLS to link data from these statistical files with data in the BLS Unemployment Insurance Address File and with data to be furnished by IRS, and prohibited any other linkage.

4. Microdata Access Group

In the planning and matching stages of the project, the persons who needed to have access to microdata were those members of the Workgroup who were performing the record matching and verification. At Workgroup meetings, members generally reviewed data in the form of frequencies and other summaries to track the progress of the matching operations and to plan future steps. Occasionally, discrepancies appeared or questions arose concerning classification of a particular employer or possible mis-match of data. Those matters were usually referred to par-

ticular members to resolve, with access to microdata as needed on an ad hoc basis.

When the matching steps were completed and time came to plan the analysis, new arrangements were needed to enable a different group of persons to examine identifiable microdata. The Microdata Access Group (MAG) was formed for this purpose. At this point, IRS agreed that its contractor, BLS, would be permitted to make Workgroup members its agents as needed for the analysis stage. This enabled the Workgroup members who were employees of BEA and the Committee on National Statistics to become sworn agents who, like the employees of BLS and SSA, would be permitted to examine and analyze microdata. Thus, of the three agencies sharing microdata (BLS, SSA, and IRS), IRS was the only one that did not have access to the matched microdata file. This group met periodically to plan and perform the analysis, prepare findings, and to report its findings back to the full Workgroup.

Once the terms of all contracts were agreed upon, the contracts and the conditions of use agreement were signed by officials of the participating agencies, and the way was cleared for the data transfers.

SUMMARY AND CONCLUSIONS

To say that the process of discussion and negotiation leading to the signing of the ERUMS access agreements was painstaking, sensitive, and costly in terms of staff time and delay in the study's completion is an understatement. The disclosure aspects of the study severely tested the will and resolve of the affected agencies.

In retrospect, the signing of interagency agreements between IRS and BLS and between SSA and BLS documented a process of negotiation by which the study plan was adapted to the requirements of the various confidentiality laws that impinged on it. In addition, it summarized a process in which a combination of technical and procedural safeguards were fitted to meet the requirements of the Federal and State agencies that were involved in the data sharing.

While the participants in the ERUMS study all feel a certain degree of accomplishment due to their collective persistence, none are quite so upbeat about

the long duration of the study. Clearly, the long incubation period for the interagency data sharing agreements was a major contributor. However, it is important to recognize that the prolonged negotiation for interagency agreements did not result from lack of cooperation among the participants. On the contrary, it reflected the complex mosaic of legal restrictions on use and interagency dissemination of records.

Once it became evident that a single multi-party agreement would be unworkable for the overall project, the plan was broken down into component steps of disclosure, record linkage, and analysis. Each failure to reach an agreement required a step back to re-examine the study imperatives and to adapt the procedures to the practical and legal necessities at each stage.

In addition to adding to the overall time and resources consumed by the project, these delays further contributed to supplemental delays, including:

1. Personnel turnover among the project participants due to the extended length of the project's schedule necessitated slower progress on the technical issues.
2. The acquisition of IRS Form 940 data was adversely impacted since these have a 5 year retention and were scheduled for destruction by the time the sample EIN's were determined.

On the positive side, however, ERUMS demonstrated that such data sharing projects can be successful under current laws if there is creativity, flexibility, and most of all, persistence.

NOTES AND REFERENCES

[1] Section 6103(n) of the Internal Revenue Code (IRC) allows for the provision of confidential tax return information for purposes of tax administration. Specifically, it reads:

"Certain Other Persons.--Pursuant to regulations prescribed by the Secretary, returns and return information may be disclosed to any person, including any person described in Section 7513(a), to the extent necessary in connection with the processing, storage, transmission, and reproduction of such returns and return information, and the programming, maintenance, repair, testing, and procurement of equipment, for purposes of tax administration."

[2] Section 6108 of the IRC has three parts which call for the publication of statistical compilation of tax return information at regular intervals, but, unlike Section 6103(n), such information cannot identify a particular taxpayer. This Section is the primary "mandate" for IRS' Statistics of Income (SOI) program.

(a) **Publication or Other Disclosure of Statistics of Income.--**The Secretary shall prepare and publish not less than annually statistics reasonably available with respect to the operations of the internal revenue laws, including classifications of taxpayers and of income, the amounts claimed or allowed as deductions, exemptions, and credits, and any other facts deemed pertinent and valuable.

(b) **Special Statistical Studies.--**The Secretary may, upon written request by any party or parties, make special statistical studies and compilations involving return information (as defined in section 6103(b)(2)) and furnish to such party or parties transcripts of any such special statistical study or compilation. A reasonable fee may be prescribed for the cost of the work or services performed for such party or parties.

(c) **Anonymous Form.--**No publication or other disclosure of statistics or other in-

formation required or authorized by subsection (a) or special statistical study authorized by subsection (b) shall in any manner permit the statistics, study, or any information so published, furnished, or otherwise disclosed to be associated with, or otherwise, identify, directly or indirectly, a particular taxpayer.

Section 6108(a) has been interpreted as a tax administration purpose for the Statistics of Income (SOI) Program (unlike 6108(b) and 6108(c)); hence, if a 6108(a) study requires the use of "outsiders," then a 6103(n) contract can be initiated as was done for the ERUMS study.