# ISSUES IN DEVELOPING A MICRODATA BASE FOR SMALL BUSINESS RESEARCH

Vito Natrella, Consultant

The Small Business Administration is required by Title II of the Small Business Act to conduct research in the field of small business. As discussed in this session, a number of approaches have been or are being developed in order to provide the data bases needed for this analysis. One of these is a microdata set representative of both small and large businesses.

Using simulation microdata models, analysis could be developed to evaluate the impacts of different policy choices affecting small business such as economic regulation, environmental and health rules, and taxation. A microdata set could also be used to study the determinants of new business ventures and the characteristics of firms going out of business. A file to meet these needs must contain a variety of information. This would include financial data, geographical information and information on employee characteristics. In addition, a basic requirement is company classification according to employment size.

A considerable amount of data exists from various sources involving these aspects. However, they need to be integrated and made internally consistent. Since no one series contains all the information needed, it is necessary to augment and merge records. Depending on the circumstances this can be achieved by means of an identical match, by a synthetic or statistical match or by a multiple regression technique. Problems associated with these techniques for inputting missing data include high cost, inconsistent definitions, sparsity of cases, poor quality, induced biases, and issues of confidentiality.

## A Proposal for a
## Small Business Microdata Base

Initial development of a small business microdata base could be in the direction of a system of three separate but related files based essentially on the IRS Statistics of Income samples: corporate, partnership, and sole proprietorship. The SOI samples have the advantage that they are stratified random samples and contain almost all the needed data. The system would encompass a Company Financial File, an Establishment File and an Employer-Employee File.

The Company Financial File would consist of the SOI sample files augmented by employment data. The SOI files contain complete income account and tax computation items for all forms of business and, in addition, balance sheet items for corporations and partnerships. Geographic designation is included but refers to the central office for large corporations filing consolidated tax returns. Since employment is to be used as the main size

classifier, this item must be added from another source. This might be done from the Standard Statistical Establishment List (SSEL) of Census or from the 941 file of IRS. Each of these has problems which will be discussed later. The Company Financial File could be made more useful by updating some industries on the basis of the Federal Trade Commission Quarterly Financial Report.

In order to permit geographical analysis an Establishment File should be set up. This can be done by obtaining data for the establishments of the firms included in SOI samples in the Company Financial File, from the SSEL of Census. Data would cover industry, employment, receipts, geographic location and employment size of the owning enterprise for each establishment.

The third file in the system of microdata files would contain information on the employees of the firms and would be designated as an Employer-Employee File. This would be developed from the Continuous Work History Sample of Social Security Administration associated with the Establishment File. The file would be arranged by employer and contain data for all its employees. This would permit analysis of the employee mix, characteristics and wages for small business.

## Access to Microdata

There are really two parts to the problem of access to microdata. One is that, in order to match data from a number of data files, identification is essential if we want to achieve the greatest accuracy. The other involves obtaining access to microdata without identification so that researchers in small business would have the opportunity and flexibility to work directly with the data.

1. Internal Revenue Service

According to Section 6103 of the Internal Revenue Code authorizations are limited, very specific as to conditions, and restricted to particular stated purposes. Paragraph (j) provides for disclosure of tax returns and return information for statistical use. Only four agencies are included in this section to receive authority for specific statistical purposes. They are the Bureau of the Census for all returns and the Bureau of Economic Analysis for corporate returns, to the extent necessary in the structuring of censuses and national economic accounts. The Federal Trade Commission has access to corporate tax returns in order to select its survey sample for the Quarterly Financial Report. The Treasury has access in order to prepare required forecasts, projections and analyses.

These data may be disclosed to others