

**STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE**

**COMMISSION OF THE  
EUROPEAN COMMUNITIES**

**CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROSTAT**

**Joint ECE/Eurostat Work Session on  
Statistical Data Confidentiality**  
(Thessaloniki, Greece, 8-10 March 1999)

Working Paper No. 14 (Summary)  
English only

Topic (i): new applications of disclosure control methods

**POSSIBILITIES OF THE CREATION OF A SCIENTIFIC-USE-FILE  
FOR THE IAB-ESTABLISHMENT PANEL**

Prepared by Ruth Brand (University of Hannover, Germany)  
Stefan Bender and Susanne Kohaut (Institute for Employment Research (IAB), Germany)

**Contributed paper**

**Summary**

1. Recently, firm data has gained much importance in applied social research because the demand side of the labour market needed to be analysed as well as the supply side. Therefore, more and more researchers ask for access to firm data existing in institutions like the Institute for Employment Research at the Federal Employment Services. That means that the data must be made available by a scientific-use-file that meets the demand of social scientists. At the same time the confidentiality of the data has to be maintained.

2. One of the most comprehensive establishment surveys in Germany is the IAB-Establishment-Panel containing information of about 9000 firms in Germany. Like other firm panels the IAB-Establishment-Panel consists of a wide range of continuous and discrete variables like sales, sectors, patterns of employment.

3. In Germany, statistical disclosure control rules exist for the creation of scientific-use-files for individual and household data. So far, the re-identification risk of German business data has not been subject of intensive research. Obviously, there are important differences between data covering personal information and business data, so that the rules for individual data cannot simply be applied on business data for the following reasons:

- a) For most of the firms, additional information is available and easily accessible for everybody.
- b) Large and medium-sized firms or enterprises can easily be identified by a few key variables.
- c) The additional information available is very precise on the enterprise level, e.g. because of the publication requirements for balance sheet informations of limited liability companies.

4. Therefore, the aim of this paper is to present the first results of research on the re-identification risk for establishments. Moreover, masking techniques to anonymise business data are applied and presented. The method chosen is the approach of Sullivan (1989, 1990), where the variables are anonymised by means of transformation to normality and addition of random noise. One advantage of this method is that for both discrete and metric variables the variance-covariance-matrix and the univariate distributions are sustained. The first results indicate that this masking technique is a suitable way to anonymise establishment data.