

**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. 10 (Summary)
English only

Topic (ii): software and computing developments

FINGERPRINTS IN MICRODATA SETS

Submitted by Statistics Netherlands¹

Contributed paper

Summary

The paper discusses a method called fingerprinting that is used to find the “risky” records in a microdata set. The intuition on which the method rests is that records with many short unique (in the file) combinations of scores on key variables are potentially dangerous, and should be identified. Tracking the risky records in a microdata set is the aim of fingerprinting. Once the risky records with (some of) their fingerprints have been identified, appropriate data protection techniques can be applied to make the file safer by reducing the number of fingerprints. Various algorithms are discussed that identify the set of risky records in a microdata set, or an approximation thereof.

¹ Prepared by Leon Willenborg and Jan Kardaun.