

Working Paper No. 36 (Summary)
ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

Joint ECE/Eurostat work session on statistical data confidentiality
(Luxembourg, 7-9 April 2003)

Topic (vi): Software tools for statistical disclosure control

**COORDINATION OF CELL SUPPRESSIONS:
STRATEGIES FOR USE OF GHMITER**

Contributed paper

Submitted by the Federal Statistical Office of Germany¹

¹ Prepared by Sarah Giessing (sarah.giessing@statistik-bund.de).

Co-ordination of Cell Suppressions: strategies for use of GHMITER

Sarah GIESSING,
Federal Statistical Office of Germany
 65180 Wiesbaden
 E-mail: sarah.giessing@statistik-bund.de

Extended abstract of supporting paper

1. Introduction

GHMITER is a software package for tabular data protection, offering to protect multiple hierarchical tables by secondary cell suppression using the GHQUAR hypercube method. For description of the GHQUAR method see [2], or [1]. An input data set to be processed by GHMITER typically contains “primary suppressions”: cells, which must not be published, because otherwise the published data of the table might be used to disclose individual respondent data contributing to those cells. In order to prevent these primary suppressions from exact, or interval disclosure from the linear relationship between the cells of the table, additional cells (so called “secondary” or “complementary” suppressions) must be suppressed. This second step is called “secondary cell suppression”.

The problem of finding an optimum set of suppressions is known as the ‘secondary cell suppression problem’. It is computationally extremely hard to find exact, or close-to-optimum solutions for the secondary cell suppression problem for multiple, large hierarchical tables. The GHMITER algorithm developed at the Statistical office of Northrhine-Westphalia/Germany offers a quick heuristic solution. Lately, the GHMITER program has become more easily accessible and usable: It was one of the tasks of the EPROS² project CASC (Computational Aspects of Statistical Confidentiality) to integrate GHMITER into the software ? -ARGUS for tabular data protection. And while ? -ARGUS is a control centre for a variety of alternative algorithms for secondary cell suppression, on behalf of Eurostat a specialised user interface for GHMITER, the CIF package, has been developed recently.

The projected paper will suggest strategies for application of the GHMITER algorithm. Some of those strategies are implemented in ? -ARGUS. In addition, the paper will hint on some advanced strategies, suitable to improve results of the method when applied to sets of linked tables. The paper will focus particularly on the problem of co-ordinating suppression patterns between linked tables provided by different agencies. It will discuss models for such co-operation, and present first empirical results.

2. Background: Information loss and data security concept of GHMITER

The paper will briefly introduce into the method of the GHMITER algorithm, focussing particularly on aspects of data security, and the information loss concept used. It will explain in which way weighting options allow the user to influence the selection of secondary suppressions. The paper will introduce weighting options that are consistent with the general information loss concept of GHMITER.

3. Co-ordination of cell suppression between tables

Due to technological advance in the information age, it is much easier nowadays for users of statistical data to compare and analyze suppression patterns in multiple tables. This increases the disclosure risk for linked tables immensely. Using proper procedures for co-ordination of suppression patterns in multiple linked tables is therefore becoming an issue of growing importance. This does impose not only methodological and technical, but also organizational problems in a situation where those tables are provided by different statistical agencies. The paper will present and discuss ideas to co-ordinate suppressions between tables on the background of facilities provided by GHMITER. We will draft models for co-operation between agencies to co-ordinate suppression patterns in multiple tables. The underlying

² European Plan for Research in Official Statistics

methodology has been tested empirically, using tabular data on German national and sub-national level. Test scenarios will be described.

4. Empirical Results

We will present and comment on results of the empirical tests mentioned above.

5. Summary and Final Conclusions

Based on the results of the empirical tests, we will give recommendations for use of GHMTER and suggest similar experimentation with alternative cell suppression algorithms such as HiTaS.

References

1. Giessing, S., Repsilber, D. (2002), 'Tools and Strategies to Protect Multiple Tables with the GHQUAR Cell Suppression Engine', In: *Inference Control in Statistical Databases* Domingo-Ferrer (Ed.), Springer (Lecture notes in computer science; Vol. 2316)
2. Repsilber, D. (2000), 'Sicherung persönlicher Angaben in Tabellendaten' - in *Statistische Analysen und Studien Nordrhein-Westfalen, Landesamt für Datenverarbeitung und Statistik NRW*, Ausgabe 1/2002 (in German)