

**Dr Jörg Drechsler**, Institute for Employment Research, Nuremberg, Germany

**Title:** Synthetic Datasets for Statistical Disclosure Control – A Personal Appraisal

**Abstract:**

Arguably, there are strong benefits when datasets can be shared between researchers and research institutions. Broader access would allow research that might otherwise never be undertaken due to time and cost constraints. Furthermore researchers might be able to link datasets from different sources to enable analyses that would not be possible based on one dataset alone. On the other hand most data are collected under the pledge of confidentiality and there are obvious legal and ethical obligations to keep this promise.

Several approaches are discussed in the literature that try to balance the trade-off between broad data access and maintaining confidentiality by altering the data or reducing the amount of information before the data are released. However, most of these approaches inevitably lead either to a substantial loss of information or distort the relationships between the variables leading to false results based on the altered data.

Recently, a promising alternative that might help overcome the shortcomings of traditional statistical disclosure control techniques has gained increasing attention: Releasing multiply imputed synthetic datasets. With this approach first proposed by Rubin in 1993 sensitive values in the dataset are replaced by synthetic values generated from a model fit to the original data.

These synthetic datasets are then released to the public. If the models for generating the synthetic data are carefully selected, most of the relationships between the variables are preserved and valid inferences can be obtained based on the released data.

In this talk, I will present the basic idea of the concept, introduce the two main approaches (fully and partially synthetic datasets) and illustrate the potentials of the method through an application to a German establishment survey, the IAB Establishment Panel.