

Jean-Pierre Le Gléau (Former head of the coordination department of the National Institute of Statistics and Economic Studies)

Anonymity and re-identification risks: an overview

Topic 4 – Getting the statistics out

Keywords: Anonymity, Access to data, Statistical disclosure

Introduction

The production of public statistics still includes aggregated information where respect for anonymity is more often related to the number of individuals included in an elementary box of the tables. The question is of a different nature, when it comes to providing individual data files. One can imagine to remove in these files the name, surname, address and other identifying variables directly to reach a total anonymity. But this is not enough.

There is always the risk that a combination of variable of interest corresponds to a single individual. If these variables can be known to a third party, there is the risk of indirect identification of that individual. At what point do we consider that anonymity is broken? Can these files, however, continue to be disseminated? If so, how to continue to protect the anonymity of people?