

## Anonymity and risk of re-identification of health data

### Topic 4 – Getting the statistics out

Keywords: Health data, anonymity, re-identification

#### Introduction

Since 1990 France has set up a national systematic collection of information regarding health data of inpatients: at the end of each hospital stay, the diagnoses presented by the inpatient and the surgical procedures performed during his stay are collected, coded and recorded according to a standardized process, in the same way in all public and private health institutions.

By this way, each health institution maintains its own file containing medical and economic data. With a unique software, distributed by the government, each institution must transform that file into an anonymized file and send it to a central organization. This one has the responsibility of determining the price of each stay and the amount to be paid to the institution and it also has the responsibility of producing national health economic statistics.

Despite all the precautions for anonymity, it turns out that in 89% of cases, one can find back the anonymous stays for inpatients for which are known only some features (and even in 100% of cases if he was hospitalized at least twice a year): one just have to know the age, the length of stay, the place of residence and the hospital which relate to this patient. Under these circumstances, it raises privacy concerns to provide researchers with the national health database for statistical purposes. That is why France has decided to settle a specific legislation for these kind of files.