Kamel Gadouche (Secure Data Access Centre)

Widening access to confidential microdata in France

Topic 4 – Getting the statistics out

Introduction

Providing access to confidential microdata to researchers is one of the priorities of national statistical institutes (NSI) insofar as working toward a wider access will eventually increase the social utility of data collected and treated by statisticians. It may also improve the quality of microdata by involving external users in such aspects: it requires a better documentation that can be published, it also gives the opportunity to get feedback from researchers about methodologies, questions and variables, etc.

The national statistical law and the deeply-rooted principle of confidentiality may however limit or re-strict this access.

This presentation will describe the process set in France for providing confidential microdata access to researchers and its underlying principles. It will also present how researchers concretly work on confidential microdata and some new features such as bigdata tools.

Methods / Problem statement

Before getting access, researchers must apply to the French Statistical Confidentiality Committee, which will assess their research project. This committee is composed of lawyers, senior researchers, business and unions representatives, members of the Data protection Authority and statisticians. They will notably check whether the applicant pursues legitimate research purposes, which data is really need…

Overall, this accreditation procedure lasts between 2 and 6 months. Once researchers are accredited, they contact the French research data centre, called CASD, to regis-ter for a training session (also called enrolment session). After this session, they receive a specific device - the SD-Box - which they can use to work remotely on confidential microdata from their insti-tution.

This SD-Box provides a biometric authentication through encrypted connections to a remote secure server hosting confidential microdata and a set of scientific software tools.

Results / Proposed solution

From the re-searcher's perspective, everything is at hand: data, software, storage, computing power… they can see the microdata but they cannot download the corresponding datafiles. They can store all outputs while working.

When they need to get the results of their analyses for publication, they must request an "output extraction" via a specific procedure during which CASD's statisticians check the output against a number of confidentiality rules. Initially set for the purpose of providing access to the French NSI's microdata, the scope of this ac-creditation procedure and access-providing infrastructure has been broadened to encompass other data sources, notably administrative data; which has allowed other data producers to give access to their confidential microdata: tax data producers; health data; Ministries of Education, Justice, Agriculture etc.

Conclusions

CASD provides secure data access to over 400 academic research projects carried out by about 1000 researchers from more than 200 institutions in France (Amiens, Lyon, Marseille, Dijon, Paris, Nantes, etc.) and abroad (UK, Germany, Italy, Austria, Luxembourg, Spain, etc.). It allows researchers to work together from different countries on the same environment with confidential microdata. Because of the size of microdata, especially because it's possible to combine datasources, CASD provides bigdata clusters to compute confidential data.