Case Study

13

The handling of research data in the social sciences at University of the Andes – Data Centre (CEDE) – Colombia

Authors: Nicolás Fuertes (Data Centre Manager) & María Alejandra Galeano (Research Assistant), CEDE University of the Andes
Email: nd.fuertes1359@uniandes.edu.co / ma.galeano258@uniades.edu.co

13.1. HABEAS DATA[1] VS OPEN DATA

In Colombia, there is a regulatory framework that protects the publication of personal data such as names, municipality, identification documents, phone numbers and addresses. In addition any information regarding children's development and victims of conflict is also protected by this regulation. For this reason, Centro de Datos (CEDE) at University of the Andes has created a range methods in order to publish research data and make it as open as possible, but without revealing the elements protected by Habeas Data. These methods include: confidential agreements, users with different levels of access, and a special data processing room located in the School of Economics at the University of the Andes.

13.2. TYPES OF USERS AND ACCESS POLICY

As a dissemination platform for its information CEDE uses its web page https://datoscede.uniandes.edu.co/ Nonetheless, due to the sensitivity and the costs of the data only users with the appropriate level of granted access clearance can access this information. To this end the platform has three different categories of users: professors from the Economics department, other University of the Andes members, and finally external users. The first have free access to all of the data sets the web page offers, the main reason for this being that the information has been collected or requested by them. The second group has limited access to the information; an authorisation from a professor of the University, along with a brief summary of the investigation for which the data is needed, are necessary in order to access restricted information. Lastly, the third group can only access the public data available in the web page, and no further information is allowed due to the contracts by which the University obtains restricted information.

13.3. ANONYMISATION PROTOCOLS

In order to publish restricted data, our CEDE created an anonymisation protocol. The algorithm used consists of:

- Generating a random number in order to identify the household. This number is generated based on the interviewer identification number, order of the municipalities visited and order of the interviews conducted in the house;
- Generating a consecutive number based on the interview order and on the order of the municipalities visited.

---

[1] Habeas Data: https://en.wikipedia.org/wiki/Habeas_data; last accessed 7 February 2017.

13.4. RESTRICTED DATA AND CONFIDENTIAL AGREEMENTS

In line with the Open Data initiative, CEDE has classified its datasets into five different groups: open data, public use, licensed data, external repository and data that is not available. The data that is considered licensed is mainly information from public institutions, which hold material that is relevant for economic research, but where that information is not generally freely accessible. To access this type of data, the University subscribes to confidentiality agreements or contracts. However, due to the sensitive nature of this data, its availability is subjected to an authorization process by which professors attest to the use of the data by a student, subject to a prior discussion as to its necessity within the specific research process. The data that contains information protected by Habeas Data also has an additional step of authorization, in which the researcher has to sign a confidentiality clause where they agree not to reveal information about particular individuals, not to use the data for other purposes, and/or not to circulate the information among other people that lack the necessary permission to use it.

13.5. DATA PROCESSING ROOM

As a contingency measure to protect the information subject to Habeas Data and to guarantee access to the information, CEDE jointly with the National Department of Statistics (DANE) have made available a data processing room. In this space researchers can access information with no anonymity process in order to estimate more complete economic models. To access this room, researchers have first to request the information from DANE by way of an email in which they give details of the specific data set and give a brief introduction to the research project they are working on. After this, DANE grants access by creating a designated user file into which the information requested is placed. Finally, researchers sign a confidentiality agreement and schedule an appointment to work on the computers available in the data processing room. These computers provide no Internet connection or USB ports, in order to control the entry and extraction of data. When researchers want to enter or take information, they must first send an email to DANE concerning the material they want to enter or file, as well as the output format, to obtain the results from the calculation made in the data processing room. Under no circumstances is the raw data made available out of this room.

13.6. CONCLUSIONS

For CEDE it is important to make the data available and as open as possible. However, in this process, it is necessary to respect the law related to personal data protection in Colombia. CEDE has identified and operates some methods that are useful to open the data without revealing the protected elements within it. For this reason, the University believes that it is important to find a way under the law to publish this data and make it as open as possible.