SECOND WORKSHOP ON BIG DATA FOR OFFICIAL STATISTICS

Host country: Brazil

Organizers: IBGE - Brazilian Institute of Geography and Statistics, Castlab - Computational Agricultural Statistics Laboratory of the Federal University of Pernambuco, NIC.br - Brazilian Network Information Centre and INE Paraguay.

Date: August 30th to September 1st of 2022

Time: From 8:30 a.m. to 5:30 p.m., Monday through Wednesday. Mornings will be reserved for theoretical discussion and applied demonstration and afternoons will be dedicated to practice via the assisted development of a mini project.

Format: in-person only

Context:

With the rise and fast expansion of different kinds of big data, National Statistical Offices – NSOs are being challenged to incorporate such new non-traditional data into their production system, for improvement and expansion of the statistics they provide. As in other parts of the world, most NOS in Latin America and the Caribbean are willing to incorporate the use of big data. Awareness and training initiatives are key to foster this use. The possibility of knowledge exchange in a regional network is invaluable given that realities might be very similar, posing common challenges to different countries.

The **Regional Hub for Big Data in Brazil** has one workstream particularly turned to training and promoting the use of big data in NSOs. The concern here is to provide experimentation, hands on experiences for the theme to be more easily grasped by this public, shortening the way to potential application.

The workshops:

A series of 2 workshops were planned for the period 2021-2022. The first workshop was entitled **Introduction to web scraping applied to price indexes**. The second workshop is entitled **Calculating Price Index from alternative data sources**, and it will be a sequence of the 1st event.

The workshop will take place daily, from 29 to 31 August 2022. Class will meet from 9 a.m. to 5 p.m. In the mornings, the facilitating team will discuss and illustrate some of the aspects related to the use of alternative data sources and offer demonstrations using a package in R especially created for performing the most important steps for the calculation of price indices from big data sets. The developer of the package, Professor Jacek Bialek of the University of Lods in Poland, will have a special participation in the demonstration classes.

In the afternoons, participants will develop a mini project from the facilitators'

instructions and with their close support. Countries will also be encouraged to present current experiences under development in their institutes to share practices and major challenges. This is key for mapping main problems in the region and help building a local network of collaboration. Portuguese will be the main language and Spanish and English speakers from the team will be available. A pre-requisite for participation is an intermediate knowledge level of R. English knowledge is desirable.

Target audience:

The workshop aims primarily to serve technicians and employees of NSOs in Latin America and the Caribbean, who are involved with studies or production of price statistics or big data use projects.