

1. NGHIÊN CỨU KHAI THÁC NGUỒN DỮ LIỆU LỚN (BIG DATA) PHỤC VỤ SẢN XUẤT CHỈ SỐ GIÁ TIÊU DÙNG

Mã đề tài: 2.1.1-B18-19

Nội dung nghiên cứu:

1. Tổng quan về nguồn dữ liệu lớn: Khái niệm, đặc trưng, nội dung, nguồn hình thành, các loại dữ liệu lớn, các lĩnh vực ứng dụng; thuận lợi, thách thức trong khai thác, sử dụng dữ liệu lớn...;

2. Nghiên cứu các khuyến nghị của Cơ quan thống kê Liên hợp quốc về khai thác, sử dụng dữ liệu lớn để sản xuất số liệu thống kê chính thức;

3. Kinh nghiệm của một số cơ quan thống kê quốc gia về khai thác, sử dụng dữ liệu về giá bán lẻ hàng hóa, dịch vụ tiêu dùng để sản xuất CPI và bài học kinh nghiệm cho Thống kê Việt Nam. Đặc biệt, bài học kinh nghiệm về xử lý dữ liệu đối với các trường hợp: Mặt hàng xuất hiện không liên tục giữa các kỳ trích xuất dữ liệu; xác định các mặt hàng cá thể vào nhóm hàng theo rổ hàng hóa truyền thống; phạm vi về mặt hàng so với rổ hàng hóa truyền thống; phạm vi theo vùng, địa phương; kỳ trích xuất dữ liệu; phương pháp tính giá bình quân, phương pháp tính CPI...

4. Các đề xuất về khai thác, sử dụng dữ liệu giá bán lẻ hàng hóa và dịch vụ tiêu dùng trên các trang web để tính CPI ở Việt Nam, gồm: Đề xuất về nguồn dữ liệu lớn về giá bán lẻ hàng hóa và dịch vụ; đề xuất về công nghệ trích xuất, xử lý, lưu trữ dữ liệu; đề xuất về nghiệp vụ thống kê liên quan đến tính CPI từ dữ liệu lớn;

5. Tiến thành thử nghiệm khai thác, sử dụng dữ liệu về giá bán lẻ hàng hóa và dịch vụ tiêu dùng từ các trang web để tính CPI, gồm: Chuẩn bị thu thập dữ liệu; thu thập dữ liệu; xử lý dữ liệu; phân tích và đánh giá chất lượng dữ liệu; tính giá bình quân mặt hàng; tính CPI; lưu trữ dữ liệu;

6. Báo cáo toàn bộ quá trình tiến hành thử nghiệm, gồm: Nhật ký quá trình thử nghiệm; bài học kinh nghiệm; các nội dung/vấn đề cần nghiên cứu hoàn thiện.

Chủ nhiệm đề tài: ThS. Nguyễn Văn Thụy

Thời gian thực hiện: tháng 6/2018 đến tháng 12/2019

Email: nvthuy@gso.gov.vn /nv.thuy.gso@gmail.com