

## **5. NGHIÊN CỨU XÂY DỰNG MÔ HÌNH KHAI THÁC DỮ LIỆU HÀNH CHÍNH PHỤC VỤ BIÊN SOẠN MỘT SỐ CHỈ TIÊU PHÁT TRIỂN BỀN VỮNG Ở VIỆT NAM**

Mã đề tài: 2.1.5-B19-20

Nội dung nghiên cứu:

1. Lý luận chung dữ liệu hành chính và mô hình

a. Lý luận chung về dữ liệu hành chính

(1) Cơ sở dữ liệu thống kê:

- Khái niệm dữ liệu thống kê và cơ sở dữ liệu thống kê;
- Phân loại dữ liệu thống kê;
- Các nguồn dữ liệu thống kê.

(2) Hồ sơ hành chính:

- Khái niệm, định nghĩa hồ sơ hành chính;
- Phân loại hồ sơ hành chính.

(3) Dữ liệu hành chính:

- Hệ thống hóa lý luận về dữ liệu hành chính
  - + Khái niệm/Định nghĩa về dữ liệu hành chính của một số nước trên thế giới cũng như của một số các cơ quan, tổ chức nghiên cứu về thống kê;
  - + Đặc điểm của dữ liệu hành chính;
  - + Phân loại dữ liệu hành chính.
- Dữ liệu hành chính trong sản xuất số liệu thống kê nhà nước.

b. Ưu điểm và thách thức khi khai thác dữ liệu hành chính để sản xuất số liệu thống kê nhà nước.

(1) Những ưu điểm của việc khai thác dữ liệu hành chính để sản xuất số liệu thống kê nhà nước.

- Giảm chi phí thu thập số liệu thống kê;
- Giảm tải gánh nặng của việc điều tra thống kê;

- Tính kịp thời và mức độ thường xuyên của số liệu thống kê.

(2) Những thách thức gặp phải khi khai thác dữ liệu hành chính để sản xuất số liệu thống kê nhà nước

- Thách thức về việc kiểm soát chất lượng;

- Các thách thức đối với thống kê Việt Nam đòi hỏi phải đổi mới quy trình sản xuất, thu thập số liệu thống kê:

- + Tính công khai thông tin cũng như mức phân quyền tiếp cận vào các hồ sơ hành chính để khai thác và sử dụng các dữ liệu này trong quá trình sản xuất số liệu thống kê.

- + Tính thay đổi về mục đích của các nguồn dữ liệu hành chính: các nguồn dữ liệu hành chính thông thường được đề xuất thu thập với mục đích quản lý của nhà nước và sẽ thay đổi theo từng giai đoạn cũng như khi có các thông tư, nghị định mới ra, việc này sẽ kéo theo các khái niệm định nghĩa và nội dung của từng đơn vị trong cả một tổng thể dữ liệu hành chính sẽ thay đổi.

- + Việc đánh mã và phân cấp có sự chưa thống nhất.

- Thách thức về sai số hệ thống trong dữ liệu hành chính;

- Thách thức về công nghệ.

c. Lý luận chung về mô hình và mô hình khai thác dữ liệu

- Khái niệm, định nghĩa về mô hình; phân loại mô hình.

- Khái niệm về mô hình khai thác dữ liệu.

- Đặc điểm của các loại mô hình khai thác dữ liệu.

- Các loại mô hình khai thác dữ liệu: Tìm hiểu về 2 loại mô hình khai thác dữ liệu được biết đến là Mô hình dự đoán (Predictive) và Mô hình mô tả (Descriptive).

2. Kinh nghiệm, thực tiễn xây dựng mô hình và khai thác dữ liệu hành chính trong sản xuất số liệu thống kê của quốc tế và trong nước

(1) Kinh nghiệm xây dựng mô hình và khai thác, sử dụng dữ liệu hành chính trong sản xuất số liệu thống kê trên thế giới.

- Thực tế khai thác sử dụng dữ liệu hành chính trong sản xuất số liệu thống kê của các nước khối Scandinavia, Ca-na-đa, châu Âu và một số nước châu Á;

- Kinh nghiệm xây dựng mô hình khai thác dữ liệu hành chính trong sản xuất số liệu thống kê của các nước đã thực hiện thành công;

- Kinh nghiệm khai thác, sử dụng dữ liệu hành chính của các nước trong sản xuất số liệu thống kê;

- Những thách thức, hạn chế, vướng mắc ở mỗi quốc gia trong quá trình khai thác và sử dụng dữ liệu hành chính trong sản xuất số liệu thống kê.

(2) Thực trạng khai thác sử dụng dữ liệu hành chính trong sản xuất số liệu thống kê tại Việt Nam và biên soạn các chỉ tiêu VSDGs

- Đánh giá chung về cơ sở dữ liệu hành chính ở Việt Nam;

- Kinh nghiệm xây dựng mô hình khai thác dữ liệu hành chính phục vụ công tác biên soạn các chỉ tiêu thống kê của Việt Nam;

- Đánh giá thực trạng thu thập, khai thác và sử dụng dữ liệu hành chính để sản xuất số liệu thống kê tại Việt Nam và biên soạn các chỉ tiêu VSDGs.

Từ tháng 8 năm 2016 đến tháng 4 năm 2017, Tổng cục Thống kê đã nhận được sự hỗ trợ từ UN-ESCAP giúp xây dựng phần mềm chiết xuất số liệu thống kê kinh tế từ dữ liệu thuế, dựa trên Biên bản ghi nhớ giữa Tổng cục Thống kê và Tổng cục Thuế (ban hành ngày 10 tháng 11 năm 2015). Tổng cục Thống kê cũng đang xúc tiến hợp tác với Tổng cục Hải quan trong việc xây dựng quy trình, biểu mẫu và thử nghiệm sử dụng dữ liệu hải quan để sản xuất số liệu thống kê xuất nhập khẩu. Liên quan tới việc sử dụng số liệu hành chính trong thống kê giáo dục, cơ quan Thống kê Đan Mạch đã thực hiện đợt khảo sát thí điểm một số trường tiểu học, trung học cơ sở và Đại học tại tỉnh Bắc Ninh để xem xét, đánh giá việc sử dụng dữ liệu hành chính trong sản xuất số liệu thống kê cũng như luồng dữ liệu từ các trường đến cơ quan chủ quản giáo dục đào tạo, từ đó nghiên cứu xây dựng báo cáo nghiên cứu khả thi làm cơ sở để tìm các nhà tài trợ thực hiện dự án cho Tổng cục Thống kê.

Ngoài ra, hoạt động đăng ký hộ tịch hiện đang được thiết kế và thực hiện tại Việt Nam đã tạo điều kiện thuận lợi giúp chuyển dịch việc sản xuất một số chỉ tiêu thống kê nhất định từ điều tra thống kê sang sử dụng hồ sơ hành chính,

đặc biệt việc sử dụng duy nhất một định danh cá nhân sẽ giúp Tổng cục Thống kê xây dựng kho dữ liệu chuỗi thời gian theo chiều dọc (longitudinal time series) trong lĩnh vực thống kê xã hội.

- Thực trạng của việc thu thập biên soạn các chỉ tiêu VSDGs tại Việt Nam thông qua khai thác, sử dụng dữ liệu hành chính.

(3) Thuận lợi và thách thức khi khai thác và sử dụng dữ liệu hành chính trong sản xuất số liệu thống kê tại Việt Nam và phục vụ biên soạn các chỉ tiêu VSDGs

3. Đề xuất mô hình khai thác dữ liệu hành chính để sản xuất thông tin thống kê nhà nước

- Các loại mô hình chung (mô hình phân cấp - Hierarchical model, mô hình mạng – network model hay mô hình dữ liệu quan hệ - relational model..v.v) và các mô hình khai thác dữ liệu (mô hình dự đoán và mô hình mô tả).

- Dựa trên các cách thức xây dựng một mô hình của các chuyên gia trên thế giới để đề xuất ra cách thức xây dựng mô hình khai thác dữ liệu hành chính.

4. Đề xuất các bộ công cụ hỗ trợ cũng như yêu cầu về mặt khoa học công nghệ cho quá trình thực hiện mô hình

- Đề xuất các công cụ hỗ trợ về kỹ thuật và khung pháp lý

- Các yêu cầu ứng dụng khoa học công nghệ

5. Ứng dụng mô hình khai thác dữ liệu hành chính để biên soạn một số chỉ tiêu VSDGs

(1) Nội dung các chỉ tiêu trong VSDGs và đề xuất các chỉ tiêu có thể được tính toán nhờ khai thác dữ liệu hành chính sau khi đã có mô hình chung.

(2) Ứng dụng mô hình khai thác dữ liệu phục vụ biên soạn những chỉ tiêu VSDGs được lựa chọn.

(3) Hoàn thiện mô hình đồng thời đưa ra các kiến nghị cần thiết để thực hiện mô hình.

Chủ nhiệm đề tài: ThS. Nguyễn Thanh Ngọc

Thời gian thực hiện: tháng 7/2019 đến tháng 12/2020

Email: ntngoc@gso.gov.vn