

Sử dụng dữ liệu hành chính trong công tác thống kê: Những khó khăn, thách thức với ngành Thống kê

Nguồn: Cổng thông tin điện tử Chính phủ

Sử dụng dữ liệu hành chính trong công tác thống kê là xu thế tất yếu

Trong một tài liệu gần đây của Ủy ban kinh tế - xã hội châu Á-Thái Bình Dương (ESCAP) nghiên cứu về việc sử dụng dữ liệu hành chính (DLHC) trong điều tra dân số và nhà ở đã nhận định, đây là xu hướng mới nổi ở Châu Á và Thái Bình Dương. Có 47 trong số 58 quốc gia thành viên và liên kết của ESCAP đã sử dụng DLHC trong Tổng điều tra dân số chu kỳ 2020.

Hàn Quốc đến năm 2000, Tổng điều tra dân số vẫn tiến hành theo phương pháp truyền thống với việc thu thập thông tin hoàn toàn thông qua bảng hỏi. Đến năm 2015, Hàn Quốc đã tiến hành Tổng điều tra dân số với phương pháp kết hợp với sổ hộ khẩu (Resident Registration Register - RRR).

Ngoài ra, thống kê Hàn Quốc còn sử dụng sổ đăng ký thống kê dựa trên thông tin từ 25 sổ đăng ký hành chính và tạo ra một bộ dữ liệu vi mô cho tổng điều tra dân số dựa trên sổ đăng ký.

Tương tự như Hàn Quốc, năm 2010, tổng điều tra dân số của Indonesia vẫn theo cách truyền thống. Đến năm 2020, Indonesia đã sử dụng dữ liệu dân số từ hệ thống đăng ký dân số của mình để hiện đại hóa phương pháp điều tra.

Để tạo điều kiện cho cơ quan thống kê Indonesia có quyền truy cập vào nguồn DLHC phục vụ cho công tác thống kê, một số công cụ lập pháp đã được ban hành hoặc sửa đổi tập trung vào cung cấp cho Cơ quan Thống kê Indonesia những thông tin cần thiết truy cập dữ liệu từ Hệ thống đăng ký dân số.

Australia, Indonesia, Hàn Quốc và Thổ Nhĩ Kỳ là những ví dụ về các quốc gia sử dụng kết hợp DLHC trong tổng điều tra dân số.

Ở Việt Nam, hiện Tổng cục Thống kê sử dụng 3 nguồn dữ liệu chính: Dữ liệu từ điều tra thống kê, DLHC và chế độ báo cáo thống kê. Trong 3 nguồn này thì việc biên soạn các chỉ tiêu thống kê chủ yếu được lấy từ dữ liệu điều tra thống kê và dữ liệu từ các báo cáo thống kê. Riêng nguồn DLHC hiện nay chủ yếu khai thác gián tiếp thông qua báo cáo thống kê, cũng như sử dụng kết hợp với nguồn dữ liệu điều tra.

Hiện, Tổng cục Thống kê đã 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.

Các cơ sở dữ liệu quốc gia (CSDLQG) về đăng ký doanh nghiệp đã kết nối liên thông với CSDL của Tổng cục Thuế nhằm trao đổi, chia sẻ thông tin về đăng ký doanh nghiệp, tình trạng hoạt động, số liệu sản xuất kinh doanh của doanh nghiệp... theo thời gian thực (realtime).

Ngoài ra, hoạt động đăng ký hộ tịch hiện đang được thực hiện tại Việt Nam, điều này sẽ 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 mã 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.

Có thể thấy, việc khai thác và sử dụng DLHC trong sản xuất thông tin thống kê đang trở thành xu hướng mới trong công tác thống kê của nhiều nước trên thế giới nói chung và Việt Nam nói riêng.

Nguồn dữ liệu hành chính đang nằm trải rộng ở tất cả các cơ quan, bộ ngành

Dù đạt được một số kết quả đáng ghi nhận, song việc sử dụng DLHC trong công tác thống kê hiện nay vẫn gặp một số khó khăn, thách thức.

Đó là, trước hết, nguồn DLHC ở Việt Nam hiện đang nằm trải rộng ở tất cả các cơ quan, bộ, ngành gây khó khăn trong việc khai thác nếu không có chế tài cũng như mô hình/quy trình khai thác thích hợp.

Hai là, các bộ ngành chưa triển khai đồng bộ tất cả các cơ sở dữ liệu (CSDL), cũng như chưa kết nối các thông tin. Chẳng hạn Bộ Giáo dục và Đào tạo đã thực hiện được hầu hết các CSDL về cán bộ; CSDL ngành về giáo dục mầm non, phổ thông, về giáo dục đại học....;

Bộ Nội vụ đã hoàn thành việc xây dựng CSDL về công chức, viên chức, đang tập huấn cho những người làm công nghệ thông tin của các bộ, ngành. Bộ Công an quản lý dữ liệu dân cư...

Tuy nhiên, việc chia sẻ với các bộ, ngành, kết nối chia sẻ trên NGSP (Hệ thống kết nối, liên thông các hệ thống thông tin ở Trung ương và địa phương) hoặc cung cấp cho người dân/doanh nghiệp thì chưa được thực hiện hoặc thực hiện lẻ tẻ...

Ba là, các CSDL quốc gia và các CSDL chuyên ngành hiện nay đều chưa có một quy định, hay cơ chế chia sẻ cụ thể, chi tiết và đầy đủ đối với cơ quan thống kê.

Ngành Thống kê đang nghiên cứu thay thế các quy trình sản xuất số liệu thống kê tốn kém và công kênh

Có thể thấy, sử dụng DLHC cho mục đích thống kê đang trở thành xu thế chung trên toàn thế giới cũng như ở Việt Nam. Nền thống kê nước ta hiện nay sẽ bị hạn chế và không bắt kịp xu thế nếu chỉ dựa trên hệ thống sản xuất số liệu thống kê hiện có.

Ngoài việc cần phải thích ứng liên tục với các yêu cầu mới, để giảm gánh nặng thống kê nhà nước, hiện ngành Thống kê đang nghiên cứu thay thế các quy trình sản xuất số liệu thống kê tốn kém và công kênh bằng các quy trình sản xuất tích hợp giúp tiết kiệm chi phí, thời gian.

Kết hợp với việc sử dụng mới và mở rộng các nguồn dữ liệu hiện có, dựa nhiều hơn vào các DLHC sẵn có từ các cơ quan Chính phủ. Để làm được điều này, chỉ riêng ngành Thống kê là không đủ, cần sự chung tay từ tất cả các cấp chính quyền trung ương, địa phương, các bộ ngành./.

Ưu điểm khi sử dụng DLHC

Giảm chi phí : Nguồn DLHC là nguồn dữ liệu lớn, đa dạng và sẵn có vì vậy nếu các cơ quan thống kê khi khai thác sử dụng nguồn dữ liệu này để sản xuất số liệu thống kê nhà nước thì sẽ tiết kiệm được chi phí với việc thu thập dữ liệu thông kê thông qua các cuộc điều tra bởi sẽ không mất thêm các khoản chi phí khác ngoại trừ các khoản phụ phí hoặc chi phí liên quan đến làm sạch dữ liệu.

Giảm tải gánh nặng của thu thập thông tin thống kê : Khi khai thác nguồn DLHC sẵn có sẽ giúp giảm tải gánh nặng đáng so với quy trình khai thác sản xuất số liệu thống kê truyền thống như: giảm gánh nặng về nguồn nhân lực và thủ tục hành chính; giảm gánh nặng đối với người được phỏng vấn trong các cuộc điều tra.

Tính kịp thời và mức độ thường xuyên, liên tục của số liệu: Thông tin thu thập từ các cuộc điều tra chuyên môn luôn có độ trễ nhất định do để triển khai một cuộc điều tra thống kê cần phải được tiến hành đúng theo trình tự, đảm bảo đúng và đầy đủ tất cả các khâu. Trái lại, với nguồn DLHC thì các thông tin đã sẵn có không cần phải tiến hành các khâu: chuẩn bị, tập huấn và triển khai thu thập thông tin.

DLHC là dữ liệu được thu thập có tính liên tục, trực tiếp. Các nguồn DLHC luôn được cập nhật thường xuyên hàng năm, hàng quý, hàng tháng và thậm chí hàng ngày, hàng giờ tùy thuộc vào nhu cầu về nguồn dữ liệu theo quy định của pháp luật.

Có độ bao phủ rộng, thông tin đa dạng và phân tổ được theo nhiều tiêu thức: DLHC được thu thập dựa trên quy định của pháp luật phục vụ công tác quản lý của các cơ quan hành chính của tất cả từ các cơ quan, tổ chức đến các cá nhân chính vì thế, DLHC sẽ có độ bao phủ rộng. Ở nhiều quốc gia trên thế giới hay ở một số lĩnh vực cụ thể thì DLHC có tính bao phủ gần như 100% dân số giúp có thể phân tổ cũng như đảm bảo độ tin cậy ở cấp nhỏ.

Giảm sai số trong điều tra thống kê: Trong nghiên cứu thống kê có hai loại sai số là “sai số phi chọn mẫu” và “sai số chọn mẫu”. Cả hai loại sai số này sẽ được khắc phục bằng việc thay thế bằng việc khai thác nguồn DLHC sẵn có