

SDMX - XU HƯỚNG TRAO ĐỔI THÔNG TIN CỦA THỐNG KÊ HIỆN ĐẠI

Tổng quan về SDMX

SDMX là gì? SDMX - Statistical Data and Metadata Exchange - Chuẩn trao đổi dữ liệu và dữ liệu đặc tả thống kê, là xu hướng trao đổi dữ liệu trên thế giới hiện nay và nó càng ngày càng mở rộng tạo điều kiện để sự trao đổi dữ liệu cả về số lượng và chất lượng đáp ứng tính sẵn sàng, kịp thời và tiện ích. Từ trước đến nay, các tổ chức hầu như không sử dụng bất kỳ tiêu chuẩn hoặc cấu trúc dữ liệu thống kê thống nhất để trao đổi. Trong trường hợp tốt nhất mà họ sử dụng Excel hoặc CSV file. SDMX là bước tiến quan trọng trong việc chia sẻ, trao đổi dữ liệu và dữ liệu đặc tả thống kê, nó cung cấp một mô hình tuyệt vời để giải quyết vấn đề thống kê chung chung và nó cung cấp một giao diện chuẩn hóa ra bên ngoài thông qua các dịch vụ web (webservices).

SDMX là một sáng kiến chung của các nhà lãnh đạo của 7 tổ chức thống kê quốc tế bao gồm Ngân hàng Thanh toán quốc tế (BIS), Ngân hàng Trung ương Châu Âu (ECB), Thống kê Châu Âu (Eurostat), Quỹ Tiền tệ quốc tế (IMF), Tổ chức Hợp tác và Phát triển kinh tế (OECD), Thống kê Liên hợp quốc (UNSD) và Ngân hàng Thế giới (WB). Sáng kiến này xuất phát từ thực tế: Việc thu thập, xử lý và trao đổi số liệu thống kê tốn nhiều thời gian và nhân lực; mỗi tổ chức, cơ quan thống kê lại có một phương pháp trao đổi và xử lý số liệu riêng và cùng với đó là nhu cầu áp dụng các công nghệ mới trong thống kê ngày càng tăng. Sáng kiến về SDMX ra đời giải quyết những vấn đề trên bằng việc phát triển một quy trình hiệu quả hơn để trao đổi và chia sẻ số liệu (data) và dữ liệu đặc tả (metadata) bằng cách

sử dụng công nghệ hiện đại. Đây là tiêu chuẩn mã hóa điện tử (e-standard) cho số liệu và dữ liệu đặc tả mà các tổ chức áp dụng chung, áp dụng công nghệ dựa trên nền web (web-based technology). SDMX xây dựng các định nghĩa cấu trúc dữ liệu (data structure definition - DSD) và định nghĩa cấu trúc dữ liệu đặc tả (metadata structure definition - MSD) dưới dạng mã hóa, tiêu chuẩn này sẽ giống như một ngôn ngữ chung để các tổ chức dễ dàng trao đổi dữ liệu và dữ liệu đặc tả chỉ thông qua máy tính mà không mất nhiều công sức của con người. Đối với người dùng tin, sau khi đăng ký trên các trang web của các tổ chức thống kê, sẽ dễ dàng thu thập được thông tin và sẽ được tự động cập nhật thông tin mới nhất khi có thay đổi.

Như vậy, SDMX là bộ tiêu chuẩn kỹ thuật và hướng dẫn của các đối tượng nội dung tạo thuận lợi cho trao đổi dữ liệu và dữ liệu đặc tả thống kê sử dụng trong công nghệ thông tin hiện đại với trọng tâm là dữ liệu tổng hợp.

SDMX từ khi ra đời năm 2001 đã có rất nhiều tiến bộ. Cho đến nay, SDMX đã được nhiều tổ chức thống kê sử dụng và áp dụng trong rất nhiều lĩnh vực thống kê như: tài khoản quốc gia, cán cân thanh toán, đầu tư nước ngoài, MDG, hệ thống phổ biến số liệu GDDS, v.v... và tiếp tục được mở rộng cho các lĩnh vực khác. Những tiến bộ chính trong việc phát triển chuẩn SDMX, các công cụ và chỉ dẫn cho người sử dụng, xây dựng các phần mềm dựa trên chuẩn SDMX đang tiếp tục được hoàn thiện.

Mục đích của SDMX là giúp quy trình sản xuất thống kê để thực hiện các bước một cách

chính xác phù hợp với các thủ tục chuẩn; giúp người dùng tìm kiếm thông tin cần thiết: Mô tả tổng thể điều tra; Dữ liệu nào sẵn có; Tìm kiếm bằng từ khóa; Tìm kiếm bằng các trình đơn (menu) tìm kiếm một cách logic; Giao diện giữa các phần khác nhau trong hệ thống thông tin thống kê. SDMX - công nghệ được sử dụng để tự động và chuẩn hóa dữ liệu và dữ liệu đặc tả nên dữ liệu sẽ được cập nhật dễ dàng và có được sớm hơn. Với SDMX, không những các cơ sở dữ liệu (CSDL) quốc gia được lưu trữ, trao đổi dễ dàng trên phạm vi toàn cầu, mà các CSDL khác của các cơ quan thống kê, ban bộ ngành khác, các cấp quản lý hành chính khác cũng có thể phổ biến và trao đổi rộng rãi trong một quốc gia, vv...

Thành phần SDMX: (1) Mô hình thông tin SDMX (Khái niệm); (2) Hướng dẫn nội dung theo định hướng (Khái niệm và danh sách mã được chia sẻ); (3) Cơ sở hạ tầng công nghệ thông tin để trao đổi và chia sẻ (Nền tảng công nghệ thông tin).

Ứng dụng của SDMX ngày càng trở nên rộng rãi và phổ biến trên toàn thế giới. SDMX là cách thức trao đổi dữ liệu tốt nhất hiện có, cạnh tranh và hơn nữa có các tiêu chuẩn bổ sung là Cơ sở dữ liệu trao đổi DDI và Các bộ dữ liệu xuất bản ngôn ngữ DSPL. SDMX giúp hài hòa hóa các phân loại; hài hòa hóa các cấu trúc (định nghĩa cấu trúc dữ liệu); tổ chức của dữ liệu đặc tả; trực quan; phổ biến thông tin và tạo thành một hệ thống dữ liệu như một bộ sưu tập. Trao đổi dữ liệu thống kê và dữ liệu đặc tả sử dụng nhiều quy trình tự động khác nhau, nhưng có một số điểm chung là sử dụng công nghệ thông tin là gần như phổ biến trong các khâu trong hệ thống. Vì vậy SDMX nhằm mục đích cung cấp các tiêu chuẩn có ích nhất cho các quy trình tự động và công nghệ. Chúng có thể được mô tả như sau:

1. Trao đổi hàng loạt dữ liệu và siêu dữ liệu: Việc truyền tải toàn bộ hoặc một phần cơ sở dữ liệu giữa các đối tác, bao gồm cập

nhật thường xuyên hay đột xuất với tần suất gia tăng.

2. Cung cấp các dữ liệu một cách dễ dàng và có thể xử lý dữ liệu đặc tả trên Internet: sử dụng Internet công nghệ là rất phổ biến. Công nghệ cao đáp ứng dịch vụ XML và web như cơ chế chính để tự động hoá dữ liệu và cung cấp dữ liệu đặc tả, cũng như HTML tĩnh truyền thống và xuất bản cơ sở dữ liệu có cấu trúc.

3. Quy trình chung: Trong khi nhiều ứng dụng và quy trình cụ thể đối với một số tập hợp các dữ liệu và dữ liệu đặc tả, các loại dịch vụ tự động đều được thiết kế để xử lý bất kỳ loại dữ liệu và dữ liệu đặc tả thống kê gì. Điều này đặc biệt đúng trong trường hợp các trang web và cổng thông tin điện tử cung cấp nguồn dữ liệu được thực hiện có sẵn trên Internet.

4. Trình bày và chuyển đổi dữ liệu: Để làm cho dữ liệu và dữ liệu đặc tả hữu ích cho người tiêu dùng, nhà cung cấp phải hỗ trợ các quy trình tự động chuyển đổi chúng thành các định dạng ứng dụng cụ thể để xử lý, định dạng tiêu chuẩn, và định dạng phần mô tả rõ ràng nếu việc trao đổi thông tin giữa các đối tác được hỗ trợ.

Các tiêu chuẩn SDMX quy định ở đây được thiết kế để hỗ trợ các yêu cầu của tất cả các quá trình tự động hóa và công nghệ.

Các công cụ hỗ trợ ứng dụng SDMX: SDMX Reference Infrastructure (SDMX-RI) là cơ sở hạ tầng dịch vụ tổng quát cho phép tái sử dụng một phần hoặc toàn bộ các công cụ/module của SDMX-RI để trao đổi dữ liệu. Có thể sử dụng trực tiếp SDMX-RI hoặc có thể bổ sung thêm các module mới, hoặc có thể sửa đổi các module hoặc có thể tích hợp thêm một số chức năng vào môi trường phổ biến số liệu hiện tại. Các module và các công cụ hỗ trợ chủ yếu của SDMX-RI bao gồm: SDMX Query Parser, Data Retriever, Structure Retriever, SDMX Data Generator, Web Client, Web

Service Provider, Mapping Assistant. Trong đó, sử dụng công cụ Mapping Assistant để báo cáo và phổ biến thông tin trong cơ sở dữ liệu hiện có theo chuẩn SDMX là chủ yếu.

Ứng dụng SDMX trong việc phổ biến thông tin của Thống kê Việt Nam

Việc áp dụng SDMX vẫn còn mới đối với thống kê khu vực nói chung và Việt Nam nói riêng. Tuy nhiên, để bắt kịp với sự phát triển về công nghệ của thống kê thế giới, thống kê Việt Nam đã ứng dụng SDMX để phổ biến các chỉ tiêu phát triển quốc gia trên Trang Cơ sở dữ liệu quốc gia (Country Data) của Liên hợp quốc và phổ biến Niên giám Thống kê trên website của TCTK để phổ biến tới người dùng tin.

(1) Phổ biến các chỉ tiêu phát triển quốc gia

Để có nhiều thông tin hơn về các chính sách và xu hướng phát triển của một số quốc gia, nhu cầu cấp bách đặt ra là phải thúc đẩy sự sẵn có của số liệu thống kê, đồng thời tạo điều kiện thuận lợi hơn cho người sử dụng trong việc truy cập các chỉ tiêu thống kê quốc gia nhằm cải thiện sự liên kết, tính có sẵn và tính so sánh được của thông tin thống kê thông qua việc ứng dụng chuẩn SDMX

Bộ chỉ tiêu phát triển quốc gia bao gồm các chỉ tiêu thuộc các lĩnh vực kinh tế - xã hội, môi trường phản ánh động thái và sự phát triển của quốc gia. Mỗi chỉ tiêu được gắn với dữ liệu đặc tả gồm: Thông tin về khái niệm, định nghĩa, phương pháp tính, nguồn số liệu, phân tổ chủ yếu và các hạn chế khuyến nghị nếu có. Các chỉ tiêu này đều định dạng Định nghĩa cấu trúc dữ liệu và Định nghĩa cấu trúc dữ liệu theo chuẩn của SDMX. Việt Nam đã chọn Danh mục gồm 60 chỉ tiêu phát triển để trao đổi và chia sẻ với mục đích:

- Nâng cao sự hiểu biết lớn hơn về các chính sách và xu hướng phát triển quốc gia, từ đó cải tiến việc phổ biến số liệu cấp quốc gia và cấp toàn cầu theo phương thức gắn kết thông qua việc xác lập một cơ sở dữ liệu quốc

gia, cổng phổ biến thông tin một cửa toàn cầu.

- Thực hiện việc một hệ thống so sánh và phổ biến thông tin cho một số quốc gia đang phát triển được lựa chọn. Xây dựng hoàn chỉnh một ứng dụng các chỉ tiêu phát triển quốc gia khác trên trang web để tạo ra sự truy cập lớn hơn và dễ dàng hơn tới các số liệu chính thống của quốc gia thông qua Cơ sở dữ liệu của Liên hợp quốc UNdata,

- Việc truy cập được cải thiện và dễ dàng hơn tới các số liệu thống kê sẵn có sẽ giúp cho các nhà hoạch định chính sách và đưa ra quyết định có căn cứ và giám sát được tính hiệu quả của các quyết định đó đồng thời giảm sự thiếu nhất quán giữa cơ sở dữ liệu quốc gia và quốc tế.

- Cải thiện sự liên kết giữa các chỉ tiêu phát triển do Tổng cục Thống kê sản xuất; thúc đẩy việc sử dụng các định nghĩa theo tiêu chuẩn, các phương pháp luận, các mô hình trao đổi dữ liệu và chuyển giao dữ liệu.

- Cải thiện tính có sẵn của số liệu thống kê chính thống của Việt Nam tại cấp quốc tế: Xây dựng cơ sở dữ liệu đối với các chỉ tiêu phát triển ở cấp quốc gia; Kết nối cơ sở dữ liệu này với Cơ sở dữ liệu của Liên hợp quốc UNdata.

- Tăng cường kiến thức trong hệ thống thống kê chính thức của Việt Nam thông qua đào tạo nhận thức thống kê và công nghệ hiện đại, liên quan đến các chủ đề như: Các chỉ tiêu phát triển; quy trình lưu giữ số liệu; trao đổi, phổ biến số liệu; trình bày và công bố số liệu.

(2) Phổ biến Niên giám Thống kê trên website của TCTK

Niên giám thống kê là ấn phẩm thống kê được Tổng cục Thống kê xuất bản hàng năm, là sản phẩm thống kê quan trọng của ngành Thống kê, bao gồm: Những số liệu thống kê cơ bản phản ánh khái quát động thái và thực trạng kinh tế - xã hội của cả nước, của các vùng và các địa phương. Việc lựa chọn Niên giám thống kê để phổ biến trên web và nhằm mục đích trao đổi thông tin qua lựa chọn ứng

dụng SDMX là cần thiết theo xu hướng chung của thế giới và của phương thức trao đổi dữ liệu và dữ liệu đặc tả thống kê hiện đại.

Ứng dụng SDMX chia sẻ và phổ biến Niên giám thống kê trên trang web, là thích hợp và cần thiết khi nhìn từ 3 góc độ:

Xuất bản: Nhà sản xuất thông tin cung cấp luồng dữ liệu SDMX có sẵn như một sản phẩm thông tin điện tử trung gian và tăng giá trị cho người sử dụng cuối cùng.

Thu thập: Tổ chức SDMX lưu chứa dữ liệu thống kê trong một cách tiêu chuẩn hóa, vì vậy nó có thể được sử dụng để duy trì và tái sử dụng lại các hệ thống/cơ sở dữ liệu của nó.

Phổ biến: Mục tiêu chính SDMX là việc trao đổi số liệu thống kê và dữ liệu đặc tả, nó cung cấp một cách tương thích thông tin giữa các nền tảng và độc lập từ các thiết bị mang tin. Một số tổ chức Thống kê đang phát triển "công cụ phần mềm" và sử dụng SDMX như dữ liệu đầu vào để hiển thị ký tự, bản đồ...

Ứng dụng SDMX để trao đổi và phổ biến Niên giám thống kê trên trang web với mục đích:

- Tuân thủ các nguyên tắc chính thức cho việc định dạng dữ liệu và dữ liệu đặc tả, từ đó các dữ liệu có thể được trao đổi, đọc và xử lý mà không cần đến sự can thiệp của con người.

- Thiết lập các tiêu chuẩn kỹ thuật và các hướng dẫn theo định hướng nội dung, tạo thuận lợi cho việc trao đổi dữ liệu và dữ liệu đặc tả thông qua việc sử dụng các công nghệ thông tin hiện đại, tập trung vào dữ liệu tổng hợp và dữ liệu chuỗi thời gian.

- Nhu cầu chuẩn hoá thông tin thống kê

Những tiêu chuẩn của SDMX có thể được sử dụng trong một hệ thống quốc gia hay các nhà cung cấp dữ liệu tư nhân cho việc truyền tải hay chia sẻ dữ liệu và dữ liệu đặc tả thống kê. Điều này có thể mở ra sự phổ biến dựa trên cơ sở SDMX; Đáp ứng tốt những nhu cầu của người sử dụng về siêu dữ liệu và dữ liệu có

cấu trúc tốt ở dạng thức có thể tái sử dụng, và được xem như một sự lựa chọn cho các quốc gia cũng như các tổ chức quốc tế; Quản lý tổng thể cơ sở dữ liệu qua thời gian; Theo dõi cập nhật sự thay đổi cơ sở dữ liệu qua thời gian.

Tuy nhiên, việc áp dụng SDMX trong trao đổi thông tin thống kê chúng ta còn một số khó khăn sau:

- SDMX là khái niệm khá mới và chưa được phổ biến rộng rãi, đặc biệt là thống kê Bộ, ngành dường như chưa biết đến cách thức trao đổi này;

- Dữ liệu đặc tả - Metadata của một số chỉ tiêu được xây dựng mới, không có sẵn, khó khăn trong việc xây dựng theo chuẩn dữ liệu đặc tả (MSD);

- Việc chuẩn hóa metadata của một số chỉ tiêu mang tính chất đặc thù của Việt Nam để thuận tiện cho việc so sánh các khác biệt với quốc tế trong quá trình sử dụng thông tin đòi hỏi nhiều thời gian thực hiện;

- Thông tin của nhiều chỉ tiêu thống kê Việt Nam sẵn có nhưng không có trong danh mục để chia sẻ với quốc tế;

- Các cán bộ thực hiện không những đòi hỏi yêu cầu về nghiệp vụ thống kê mà phải có trình độ tốt về tin học;

Để ứng dụng SDMX thành công và có hiệu quả, cần thiết phải có sự phối hợp chặt chẽ giữa các đơn vị trong Tổng cục Thống kê cũng như giữa Tổng cục Thống kê với các thống kê Bộ, ngành để có đầy đủ số liệu và dữ liệu đặc tả cho từng chỉ tiêu; cần cử cán bộ tham dự các khóa đào tạo quốc tế về SDMX để cập nhật kiến thức mới và những cán bộ này phải có trình độ về tin học; tổ chức đào tạo về SDMX cho cán bộ thống kê và đảm bảo cơ sở hạ tầng công nghệ thông tin đáp ứng yêu cầu.

Thu Trang (lược dịch và tổng hợp)

Nguồn: <http://sdmx.org>