

NGHIÊN CỨU ÁP DỤNG CHUẨN TRAO ĐỔI DỮ LIỆU VÀ DỮ LIỆU ĐẶC TẢ THỐNG KÊ TRONG CÔNG TÁC PHỔ BIẾN SỐ LIỆU THỐNG KÊ NƯỚC NGOÀI TẠI VIỆT NAM

Bùi Ngọc Tân*

Tóm tắt:

Với sự phát triển của Internet và mạng toàn cầu, việc trao đổi và chia sẻ dữ liệu điện tử đã trở nên dễ dàng và phổ biến hơn. Khi trao đổi dữ liệu thống kê được thực hiện thường xuyên, liên tục, việc áp dụng các phương pháp và tiêu chuẩn chung mang lại lợi ích rất lớn cho cả người cung cấp dữ liệu và người dùng dữ liệu. Sáng kiến trao đổi dữ liệu thống kê và dữ liệu đặc tả thống kê (SDMX) được bảy tổ chức quốc tế gồm: Ngân hàng thanh toán quốc tế, Ngân hàng trung ương châu Âu, Thống kê châu Âu, Quỹ tiền tệ quốc tế, Tổ chức hợp tác và phát triển kinh tế, Cơ quan Thống kê Liên hợp quốc và Ngân hàng Thế giới thông qua năm 2001. Tổng cục Thống kê đã nhận được hỗ trợ kỹ thuật từ nhiều tổ chức quốc tế và quốc gia nhằm sử dụng các khung có sẵn của các tổ chức quốc tế (IMF, Liên hợp quốc...) để phổ biến số liệu theo SDMX. Tuy nhiên, vẫn chưa có nghiên cứu chính thức nào để tự xây dựng một khung SDMX giúp phổ biến số liệu thống kê của Việt Nam. Do đó, cần có một nghiên cứu để xây dựng và áp dụng khung SDMX phục vụ việc phổ biến số liệu thống kê nước ngoài tại Việt Nam. Dựa trên kinh nghiệm xây dựng khung SDMX này, TCTK có thể áp dụng để xây dựng các khung SDMX khác dùng để phổ biến số liệu trong các lĩnh vực khác nhau của Việt Nam, hướng tới áp dụng rộng rãi SDMX trong việc phổ biến số liệu thống kê của Việt Nam.

1. Tổng quan về Chuẩn trao đổi dữ liệu và dữ liệu đặc tả thống kê

Với sự phát triển của Internet và mạng toàn cầu, việc trao đổi và chia sẻ dữ liệu điện tử đã trở nên dễ dàng và phổ biến hơn. Tuy nhiên, việc trao đổi thông tin và dữ liệu thường được thực hiện sử dụng nhiều định dạng và khái niệm không theo bất kỳ tiêu chuẩn nào. Khi trao đổi dữ liệu thống kê được thực hiện thường xuyên và liên tục, việc áp dụng các phương pháp và tiêu chuẩn chung mang lại lợi ích rất lớn cho cả người cung cấp dữ liệu và người dùng dữ liệu.

Sáng kiến trao đổi dữ liệu và dữ liệu đặc tả thống kê (SDMX) được phát triển nhằm mục đích đảm bảo dữ liệu đặc tả luôn đi kèm với dữ liệu thống kê, làm cho thông tin trở nên dễ hiểu và hữu ích. Do đó, các tiêu chuẩn và nguyên tắc SDMX xử lý cả dữ liệu và dữ liệu đặc tả. SDMX phát triển và sử dụng các quy trình hiệu quả hơn để trao đổi và chia sẻ dữ liệu thống kê và dữ liệu đặc tả giữa các tổ chức quốc tế và các quốc gia thành viên. Để đạt được mục tiêu này, SDMX cung cấp các định dạng chuẩn cho dữ liệu và dữ liệu đặc tả, cùng với hướng dẫn nội dung và kiến trúc công nghệ thông tin (CNTT) để trao đổi dữ liệu và dữ liệu đặc tả. Các tổ chức được tự do sử

* Vụ Thống kê nước ngoài và Hợp tác quốc tế

➤➤➤ THỐNG KÊ VÀ CUỘC SỐNG

dụng bất kỳ thành phần nào của SDMX phù hợp nhất trong một trường hợp nhất định.

Trong kết luận tại Phiên họp thứ 39 của Ủy ban Thống kê Liên hợp quốc diễn ra tại New York vào tháng 2/2008, SDMX đã được công nhận là “tiêu chuẩn ưu tiên để trao đổi và chia sẻ dữ liệu và dữ liệu đặc tả trong cộng đồng thống kê toàn cầu”. Việc chấp nhận SDMX ở cấp Liên hợp quốc này là một bước tiến quan trọng hướng tới việc sử dụng SDMX rộng rãi hơn trên toàn thế giới.

Việt Nam đã xây dựng Tiêu chuẩn Việt Nam, TCVN 7981 “Trao đổi siêu dữ liệu và dữ liệu thống kê” do Ban Kỹ thuật Tiêu chuẩn quốc gia TCVN/TC 154 biên soạn, Bộ Khoa học và Công nghệ công bố năm 2008. Bộ tiêu chuẩn này gồm các phần: (1) TCVN 7981-1:2008 (ISO/TS 17369-1:2005): Khung tổng quát về các tiêu chuẩn SDMX; (2) TCVN 7981-2:2008 (ISO/TS 17369 - 2:2005): Mô hình thông tin: Thiết kế khái niệm UML; (3) ISO/TS 17369:2005: Tài liệu và lược đồ SDMX-ML; (4) ISO/TS 17369:2005: Tài liệu và cú pháp SDMX-EDI; ISO/TS 17369:2005: Hướng dẫn người thực thi SDMX; (6) ISO/TS 17369:2005: Hướng dẫn sử dụng các dịch vụ web trong SDMX.

Việc áp dụng SDMX vẫn còn tương đối mới đối với Việt Nam khi mới chỉ ứng dụng để phổ biến các chỉ tiêu phát triển quốc gia được phổ biến trên trang Cơ sở dữ liệu quốc gia (UNData) của Liên hợp quốc. Năm 2015, trong khuôn khổ dự án “Tăng cường công tác phổ biến số liệu Thống kê quốc gia của Việt Nam” do Ngân hàng Thế giới tài trợ, Tổng cục Thống kê (TCTK) đã tổ chức Hội thảo về SDMX. Với sự hỗ trợ từ IMF, TCTK phối hợp với Bộ Tài chính và Ngân hàng Nhà nước Việt Nam tiếp tục hoàn thiện và phổ biến thông tin về tài chính quốc gia theo định dạng SDMX.

Ngoài ra, TCTK cũng tiếp nhận các hỗ trợ kỹ thuật về SDMX từ Liên hợp quốc và các dự án hợp tác song phương liên quan đến phổ biến các chỉ tiêu phát triển bền vững, hiện đại hóa hoạt động thống kê.

Trong các hoạt động trên, TCTK được hướng dẫn sử dụng các khung có sẵn của các tổ chức quốc tế (IMF, Liên hợp quốc...) để phổ biến số liệu theo SDMX. Tuy nhiên, vẫn chưa có nghiên cứu chính thức nào để tự xây dựng một khung SDMX giúp phổ biến số liệu thống kê của Việt Nam. Do đó, cần có một nghiên cứu để xây dựng và áp dụng khung SDMX phục vụ việc phổ biến số liệu thống kê nước ngoài tại Việt Nam. Dựa trên kinh nghiệm xây dựng khung SDMX này, TCTK có thể áp dụng để xây dựng các khung SDMX khác dùng để phổ biến số liệu trong các lĩnh vực khác nhau của Việt Nam, hướng tới áp dụng rộng rãi SDMX trong việc phổ biến số liệu thống kê của Việt Nam.

2. Đề xuất các thành phần Chuẩn trao đổi dữ liệu và dữ liệu đặc tả thống kê phục vụ phổ biến số liệu thống kê nước ngoài tại Việt Nam

Trong SDMX, Định nghĩa cấu trúc dữ liệu (DSD) là thành phần quan trọng nhất, gồm một tập hợp các chú thích về cách thức xây dựng và cách diễn giải tập dữ liệu. Các chú thích này được xây dựng dựa trên các khái niệm thống kê. DSD được hình thành theo ba bộ khái niệm: Chiều (Dimension); Đo lường (Measure); và Thuộc tính (Attribute).

Thứ nhất, Chiều (Dimension) là những khái niệm mô tả dữ liệu thống kê và tạo thành mã định danh (Khóa) cho dữ liệu liên quan. Do đó, chiều được nhóm thành Khóa, cho phép xác định một bộ dữ liệu cụ thể.

Thứ hai, đo lường (Measure) bao gồm dữ liệu chuỗi thời gian một thước đo - thước đo

chính, áp dụng cho cả dữ liệu chuỗi thời gian và dữ liệu chéo.

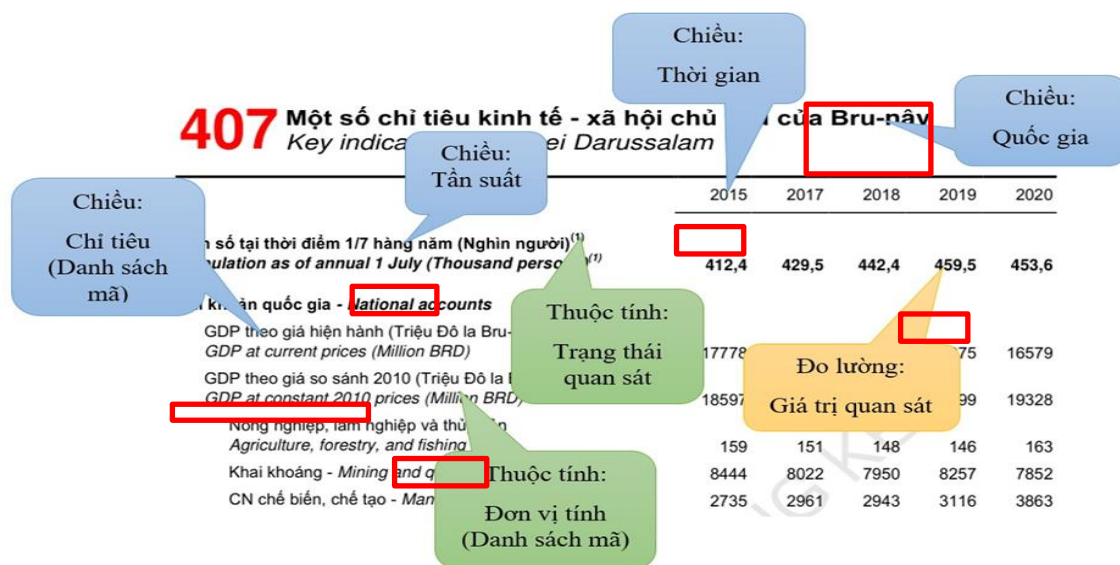
Thứ ba, thuộc tính (Attribute) là các khái niệm được sử dụng để cung cấp thêm thông tin về một số phần của bộ dữ liệu. Mỗi thuộc tính trong DSD phải được gán cho một phần định danh của bộ dữ liệu (trong mô hình được gọi là 'cấp độ đính kèm' hoặc 'cấp nhóm').

Số liệu thống kê nước ngoài trong Niên giám Thống kê của TCTK bao gồm:

Phần thứ nhất bao gồm 18 biểu tổng hợp các quốc gia trên toàn thế giới, mỗi biểu gắn với một hoặc một vài chỉ tiêu cụ thể. Dữ liệu trong các biểu này được xem là dữ liệu chéo.

Phần thứ hai bao gồm 13 biểu trình bày số liệu của các quốc gia trong ASEAN và các đối tác quan trọng trong khu vực châu Á của Việt Nam. Mỗi biểu sẽ thể hiện các chỉ tiêu về kinh tế, xã hội chủ yếu của một số quốc gia theo thời gian. Dữ liệu trong bảng này được xem như dữ liệu chuỗi thời gian.

Hình 1. Phân tích các thành phần trong định nghĩa cấu trúc dữ liệu SDMX sử dụng để phổ biến số liệu thống kê nước ngoài.



Nguồn: TCTK

Những phân tích ở trên sẽ đưa ra các thành phần chính trong DSD SDMX phục vụ phổ biến số liệu thống kê nước ngoài theo mô tả dưới đây.

Biểu 1. Các thành phần SDMX sử dụng để phổ biến số liệu thống kê nước ngoài

Khóa	Định danh khái niệm	Tên khái niệm	Cấp độ đính kèm	Trạng thái sử dụng	Định danh Danh sách mã	Mô tả Danh sách mã
CHIỀU						
1	FREQ	Tần suất		Bắt buộc	CL_FREQ	Danh sách mã tần suất
2	COUNTRY	Quốc gia		Bắt buộc	CL_AREA	Danh sách mã quốc gia
3	FOREIGN_IND ICAT	Chỉ tiêu thống kê nước ngoài		Bắt buộc	CL_FOREIGN_IND ICAT	Danh sách mã Chỉ tiêu thống kê nước ngoài
	TIME_PERIOD	Khoảng thời gian		Bắt buộc		

ĐO LƯỜNG						
	OBS_VALUE	Giá trị quan sát		Có điều kiện		
THUỘC TÍNH						
	OBS_STATUS	Trạng thái của quan sát	Quan sát	Có điều kiện	CL_OBS_STAT US	Danh sách mã trạng thái quan sát
	UNIT	Đơn vị	Chuỗi thời gian	Bắt buộc	CL_UNITU	Danh sách mã đơn vị
	TIME_FORMAT	Định dạng thời gian	Chuỗi thời gia	Bắt buộc	CL_TIME_FOR MAT	Danh sách mã định dạng thời gian

Nguồn: Tác giả đề xuất

Để bổ sung cho các thành phần ở trên, cần sử dụng các danh sách mã (Code list) theo đúng tiêu chuẩn SDMX. Danh sách mã chỉ đơn giản là một tập hợp các giá trị được sử dụng để thể hiện một Khái niệm (Chiều hoặc Thuộc tính) trong định nghĩa cấu trúc dữ liệu SDMX. Mỗi mã được xác định duy nhất bằng một giá trị và mô tả có thể được cung cấp bằng nhiều ngôn ngữ. Mô hình này cho phép một danh sách mã có hệ thống phân cấp mã đơn giản. Danh sách mã được đề xuất gồm:

Thứ nhất, danh sách mã đơn vị tính (CL_UNIT_MULT). Danh sách mã này cung cấp các giá trị mã để chỉ ra độ lớn theo đơn vị đo lường;

Thứ hai, danh sách mã quốc gia (CL_AREA) cung cấp các khu vực tham chiếu thống kê sử dụng theo ISO3166 phù hợp với bảng mã quốc gia mà các tổ chức quốc tế (Ngân hàng Thế giới, UNICEF,...);

Thứ ba, danh sách mã tần suất (CL_FREQ) tuân thủ theo tiêu chuẩn SDMX-ML. Danh sách mã này cung cấp một tập hợp các giá trị biểu thị "tần suất" của dữ liệu (ví dụ: hàng tuần, hàng tháng, hàng quý). Khái niệm "tần suất" có thể đề cập đến nhiều giai đoạn khác nhau trong quy trình sản xuất, ví dụ: thu thập dữ liệu hoặc phổ biến dữ liệu;

Thứ tư, danh sách mã trạng thái quan sát (CL_OBS_STATUS). Danh sách mã này cung cấp thông tin được mã hóa về (i) Chất lượng của một giá trị; hoặc (ii) Các giá trị bất thường hoặc bị thiếu;

Thứ năm, danh sách mã định dạng thời gian (CL_TIME_FORMAT). Danh sách mã này cung cấp thông tin được mã hóa (dựa trên tiêu chuẩn ISO 8601) cho biết loại tham chiếu thời gian được sử dụng trong dữ liệu;

Thứ sáu, danh sách mã giới tính (CL_SEX). Danh sách mã này cung cấp thông tin được mã hóa về trạng thái nam hay nữ, được áp dụng nếu dữ liệu cần được phân loại theo giới tính;

Thứ bảy, danh sách mã hiển thị số thập phân (CL_DECIMALS). Danh sách mã này cung cấp danh sách các giá trị hiển thị số chữ số thập phân được sử dụng trong dữ liệu;

Thứ tám, danh sách mã về bảo mật thông tin (CL_CONF_STATUS). Danh sách mã này cung cấp thông tin được mã hóa về trạng thái nhạy cảm và bảo mật của dữ liệu;

Thứ chín, danh sách mã tuổi. Danh sách mã này cung cấp một tập hợp các mã cơ sở được sử dụng để tạo ra các mã định danh đơn giản hoặc phức tạp liên quan đến khái niệm "tuổi";

Thứ mười, danh sách mã tiền tệ quốc gia. Danh sách mã này cung cấp các giá trị được mã hóa cho các loại tiền tệ trên thế giới. Đối với phần số liệu thống kê nước ngoài, sẽ sử dụng mã 3 ký tự ISO 4217;

Thứ mười một, danh sách mã các chỉ tiêu. Danh sách này do tác giả xây dựng dựa trên tên các chỉ tiêu phần thống kê nước ngoài trong Niên giám Thống kê Việt Nam.

Ngoài những mã trên, có thể sử dụng bổ sung các danh sách mã có thể sử dụng chung cho các lĩnh vực thống kê khác nhau đã được Ban thư ký SDMX chính thức thông qua dựa trên các đề xuất từ Nhóm công tác thống kê của Liên hợp quốc và các tổ chức quốc tế như hệ thống ngành kinh tế, danh mục nghề...

3. Thử nghiệm phổ biến một số chỉ tiêu thống kê nước ngoài theo Chuẩn trao đổi dữ liệu và dữ liệu đặc tả thống kê

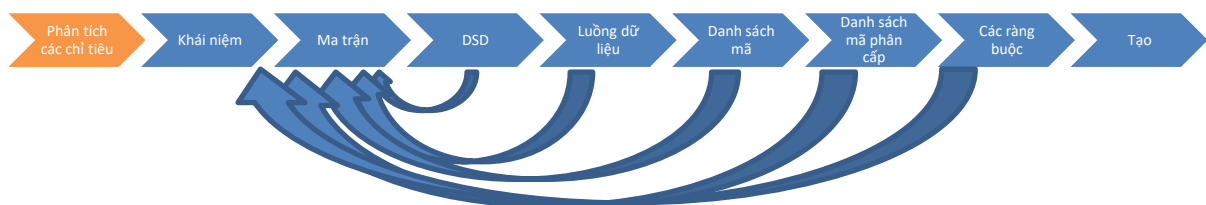
Hiện có rất nhiều công cụ CNTT được các tổ chức quốc tế xây dựng để tạo thuận lợi cho việc sử dụng rộng rãi SDMX trong thống kê. Hiện tại trên trang thông tin điện tử SDMX.ORG liệt kê khoảng 28 công cụ phần mềm dùng để thực hiện và phát triển SDMX.

Công cụ SDMX Matrix Generator đã được lựa chọn để xây dựng định nghĩa cấu trúc dữ

liệu SDMX để phổ biến thí điểm một số chỉ tiêu thống kê nước ngoài trong niên giám thống kê. Công cụ này giúp thiết kế các thành phần SDMX dễ dàng và nhanh chóng, tập trung vào các kịch bản thiết kế định nghĩa cấu trúc dữ liệu, luồng dữ liệu và danh sách mã phổ biến nhất. Ngoài ra, công cụ được xây dựng dựa trên phần mềm văn phòng Excel nên cũng giúp việc xây dựng định nghĩa cấu trúc dữ liệu SDMX trở nên quen thuộc và dễ sử dụng hơn.

Quy trình xây dựng các thành phần SDMX như Lược đồ khái niệm, Định nghĩa cấu trúc dữ liệu, Luồng dữ liệu, Danh sách mã, Ràng buộc và Phân loại theo thứ tự được trình bày trong hình dưới đây. Việc phân tích là một bước tùy chọn giúp “phân tách” thông tin trong các chỉ tiêu thành các khái niệm SDMX cụ thể, tạo thuận lợi cho xây dựng lược đồ khái niệm SDMX. Trong quá trình xây dựng các thành phần SDMX, thường có sự lặp lại giữa các bước trong khi thử nghiệm nhóm các luồng dữ liệu thành định nghĩa cấu trúc dữ liệu phù hợp, xây dựng các danh sách mã và các ràng buộc cần thiết trong thẩm định và kiểm tra dữ liệu. Các bước trên ứng với từng trang tính (worksheet) cụ thể trong công cụ SDMX Matrix Generator.

Hình 2. Quy trình xây dựng các thành phần SDMX



Nguồn: SDMX Matrix Generator

Sau khi xây dựng được các thành phần SDMX cần thiết, có thể điền dữ liệu trực tiếp từ công cụ SDMX Matrix Generator hoặc có thể trích xuất định nghĩa cấu trúc dữ liệu thành một bảng câu hỏi Excel để gửi tới người cung cấp dữ liệu

Sau khi đã xác định cấu trúc và điền dữ liệu, bước cuối cùng là chuyển đổi các kết quả ở trên thành định dạng SDMX-ML. Như đã nêu, SDMX-ML là định dạng định nghĩa dữ liệu để truyền dữ liệu thống kê một cách điện tử được thiết lập để hiện đại hóa việc thu

➤➤➤ THỐNG KÊ VÀ CUỘC SỐNG

thập và phổ biến dữ liệu thống kê trên toàn thế giới. Định dạng này mang lại những lợi ích lớn trong việc chuẩn bị, phân tích và truyền thông tin thống kê thông qua tiết kiệm chi phí

và cải thiện độ chính xác và độ tin cậy cho tất cả những người tham gia cung cấp hoặc sử dụng dữ liệu thống kê.

Hình 3. Dữ liệu và dữ liệu đặc tả của một số chỉ tiêu thống kê nước ngoài được chuyển đổi sang định dạng SDMX-ML

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<mes:Structure xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:mes="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/message"
  xmlns:str="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/structure" xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common"
  xsi:schemaLocation="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/message https://registry.sdmx.org/schemas/v2_1/SDP00message.xsd">
  <mes:Header>
    <mes:ID IDREF="9918"/>
    <mes:Test false/>
    <mes:Prepared 2024-10-24T10:55:47/>
    <mes:Sender id="Undefined"/>
    <mes:Receiver id="Undefined"/>
  </mes:Header>
  <mes:Structures>
    <str:Dataflows>
      <str:Dataflow xmlns:str="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/structure"
        xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common" id="DSD_TEST@DF_TEST"
        urn="urn:sdmx:org.sdmx.infomodel.datastructure.Dataflow=AGENCY:DSD_TEST@DF_TEST(1.0)" isExternalReference="false" agencyID="AGENCY"
        isFinal="false" version="1.0">
        <com:Annotations>
          <com:Annotation xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common" id="@SDMX">
            <com:AnnotationTitle>REF_AREA</com:AnnotationTitle>
            <com:AnnotationType>LAYOUT_ROW</com:AnnotationType>
          </com:Annotation>
          <com:Annotation xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common" id="@SDMX">
            <com:AnnotationTitle>TIME_PERIOD</com:AnnotationTitle>
            <com:AnnotationType>LAYOUT_COLUMN</com:AnnotationType>
          </com:Annotation>
          <com:Annotation xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common" id="@SDMX">
            <com:AnnotationTitle>MEASURE</com:AnnotationTitle>
            <com:AnnotationType>LAYOUT_ROW_SECTION</com:AnnotationType>
          </com:Annotation>
          <com:Annotation xmlns:com="http://www.sdmx.org/resources/sdmxml/schemas/v2_1/common">
            <com:AnnotationTitle>COMBINED_UNIT_MEASURE:UNIT_MEASURE;COMBINED_MEASURE:MEASURE,AGE</com:AnnotationTitle>
            <com:AnnotationType>COMBINED_CONCEPTS</com:AnnotationType>
          </com:Annotation>
        </com:Annotations>
      </str:Dataflow>
    </str:Dataflows>
  </mes:Structures>
</mes:Structure>
```

Nguồn: Tác giả đề xuất

4. Kết luận và kiến nghị

SDMX từ khi ra đời năm 2001 đã có rất nhiều tiến bộ. SDMX bắt đầu từ phiên bản 1.0 đến nay SDMX đã phát triển đến phiên bản 3.1 cho thấy sự phát triển nhanh chóng cũng như tầm quan trọng của việc áp dụng SDMX trong công tác trao đổi và phổ biến thông tin.

Cho đến nay, SDMX đã được nhiều tổ chức thống kê sử dụng và áp dụng trong 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.

Đề tài "Nghiên cứu áp dụng Chuẩn trao đổi dữ liệu và dữ liệu đặc tả thống kê trong công tác phổ biến số liệu thống kê nước ngoài tại Việt Nam" đã đưa ra cơ sở lý luận về SDMX như bối cảnh ra đời, các thành phần trong khung SDMX, định nghĩa cấu trúc dữ liệu SDMX và kinh nghiệm của một số quốc

gia trên thế giới trong việc áp dụng SDMX để phổ biến số liệu thống kê chính thức.

Sau khi đã hoàn thiện cơ sở lý luận và các thành phần SDMX phục vụ phổ biến số liệu thống kê nước ngoài tại Việt Nam, đề tài đã thử nghiệm xây dựng các tệp biểu mẫu SDMX sử dụng công cụ SDMX Matrix Generator. Từ những biểu mẫu này, sẽ tiến hành thử nhập số liệu cho một số chỉ tiêu thống kê nước ngoài trong Niên giám Thống kê. Sau khi được chuyển đổi, tệp dữ liệu sẽ sẵn sàng công bố dưới dạng SDMX.

Kinh nghiệm trong việc áp dụng SDMX để phổ biến số liệu thống kê nước ngoài là tiền đề để đưa ra các kiến nghị cho TCTK hướng tới sử dụng rộng rãi SDMX trong việc phổ biến số liệu thống kê chính thức tại Việt Nam. Đặc biệt, dựa trên kinh nghiệm xây dựng khung SDMX này, TCTK có thể áp dụng để xây dựng các khung SDMX khác dùng để phổ biến số liệu trong các lĩnh vực khác nhau của Việt Nam, hướng tới áp dụng rộng rãi SDMX trong việc phổ biến số liệu thống kê chính thức của Việt Nam./.