

TIÊU CHUẨN QUỐC GIA

TCVN 7981-6:2009

ISO/TS 17369-6:2005

TRAO ĐỔI SIÊU DỮ LIỆU VÀ DỮ LIỆU THỐNG KÊ - PHẦN 6: HƯỚNG DẪN SDMX SỬ DỤNG DỊCH VỤ WEB

Statistical data and metadata exchange - Section 6: SDMX guidelines for the use of web services

Lời nói đầu

TCVN 7981-6:2009 hoàn toàn tương đương với **ISO 17369-6:2005**

TCVN 7981-6:2009 do Ban Kỹ thuật Tiêu chuẩn quốc gia TCVN/TC 154 “Quá trình, các yếu tố dữ liệu và tài liệu trong thương mại, công nghiệp và hành chính” biên soạn, Tổng cục Tiêu chuẩn Đo lường Chất lượng đề nghị, Bộ Khoa học và Công nghệ công bố.

Bộ tiêu chuẩn TCVN 7981 (ISO 17369) gồm các phần sau:

- TCVN 7981-1:2008 (ISO/TS 17369-1:2005); Phần 1: Khung tổng quát về các tiêu chuẩn SDMX.
- TCVN 7981-2:2008 (ISO/TS 17369-2:2005); Phần 2: Mô hình thông tin: Thiết kế khái niệm UML.
- TCVN 7981-3:2009 (ISO/TS 17369-3:2005); Phần 3: Lược đồ và tài liệu SDMX-ML.
- TCVN 7981-4:2009 (ISO/TS 17369-4:2005); Phần 4: Cú pháp và tài liệu SDMX-EDI.
- TCVN 7981-5:2009 (ISO/TS 17369-5:2005); Phần 5: Hướng dẫn thực thi tiêu chuẩn định dạng SDMX.
- TCVN 7981-6:2009 (ISO/TS 17369-6:2005); Phần 6: Hướng dẫn SDMX sử dụng dịch vụ web.

TRAO ĐỔI SIÊU DỮ LIỆU VÀ DỮ LIỆU THỐNG KÊ - PHẦN 6: HƯỚNG DẪN SDMX SỬ DỤNG DỊCH VỤ WEB

Statistical data and metadata exchange - Section 6: SDMX guidelines for the use of web services

1. Giới thiệu

Các dịch vụ web tiêu biểu cho thế hệ tương lai của công nghệ Internet. Dịch vụ này cho phép các ứng dụng máy tính trao đổi dữ liệu trực tiếp qua mạng Internet. Về bản cơ bản, nó cho phép mô đun hoặc việc tính toán phân tán theo dạng linh hoạt hơn trước. Tuy nhiên, với mục đích giúp dịch vụ web thực hiện chức năng của nó, cần nhiều tiêu chuẩn để: yêu cầu và cung cấp dữ liệu; biểu diễn việc lập phong bì dữ liệu được sử dụng để đóng gói dữ liệu trao đổi; mô tả dịch vụ web cho một dịch vụ web khác, cho phép tích hợp dễ dàng vào các ứng dụng sử dụng dịch vụ web khác như nguồn dữ liệu.

SDMX tập trung vào việc trao đổi dữ liệu sử dụng các công nghệ Internet, cung cấp một số tiêu chuẩn liên quan đến dữ liệu và siêu dữ liệu thống kê. Tuy nhiên, nhiều tiêu chuẩn dịch vụ web đã có nên không cần tái tạo chúng để sử dụng riêng trong cộng đồng thống kê. Nói một cách cụ thể, SDMX sử dụng giao thức truy cập đối tượng đơn (SOAP) và ngôn ngữ mô tả dịch vụ web (WSDL) để bổ sung định dạng trao đổi dữ liệu và siêu dữ liệu đang chuẩn hóa.

Mặc dù đã có cam kết của SOAP và WSDL, nhưng trên thực tế, các quá trình thực thi khác nhau không liên tác với nhau. Đây là lý do để bắt đầu sáng kiến liên tác các dịch vụ web (WS-I); bao gồm một nhóm các nhà cung cấp thực thi các tiêu chuẩn dịch vụ web giống nhau theo cùng một cách và đã được kiểm chứng trong thực tế thông qua việc thực hiện các thử nghiệm về tính liên tác. Các nhà cung cấp này công bố các hồ sơ mô tả cách sử dụng tiêu chuẩn dịch vụ web có thể tiên tác. SDMX tác động đến công việc của WS-I một cách thích hợp để đáp ứng các nhu cầu

của cộng đồng thống kê.

Tiêu chuẩn này mang tính tham khảo và đưa ra quy phạm thực hành hiệu quả khi sử dụng các tài liệu SDMX-ML và tiêu chuẩn dịch vụ web cho việc trao đổi dữ liệu và siêu dữ liệu thống kê. Trong tương lai, nhóm sáng kiến SDMX đề xuất các tiêu chuẩn mang tính quy định để sử dụng các công nghệ dịch vụ web dựa trên các hướng dẫn được cung cấp ở đây.

2. Dịch vụ web và SDMX-ML

Theo truyền thống, các ứng dụng và dịch vụ quy ước thể hiện chức năng của chúng thông qua các giao diện lập trình ứng dụng (API). Các dịch vụ web cũng tương tự - chúng cung cấp phiên bản công khai các cuộc gọi chức năng có thể truy cập vào trang web có sử dụng giao thức dịch vụ web. Để tạo ra tập các dịch vụ web liên tác, cần có một khái niệm hoặc mô hình chuẩn là cơ sở cho các chức năng công khai này. Các lợi ích của SDMX từ việc có một mô hình thông tin chung, sử dụng mô hình thông tin SDMX này làm cơ sở cho các cuộc gọi chức năng dịch vụ web chuẩn là cách mở rộng tự nhiên.

Các dịch vụ web trao đổi dữ liệu theo một định dạng XML: đây là cách truyền dữ liệu giữa các dịch vụ web được định dạng. SDMX-ML là một tiêu chuẩn XML cho việc trao đổi dữ liệu và siêu dữ liệu cấu trúc trong lĩnh vực thống kê, SDMX-ML cung cấp một định dạng XML hữu ích để xuất bản dữ liệu dịch vụ web theo từng kỳ. Trong khi có một số kỹ thuật cho các trao đổi dữ liệu dịch vụ web đơn giản - thường sử dụng các phép gọi thủ tục từ xa (RPCs), việc sử dụng một tập các trao đổi XML dựa trên một mô hình thông tin chung được xem như phương pháp khả thi hơn để đạt được tính liên tác.

Có một số kiểu tài liệu khác nhau có sẵn trong SDMX-ML rất quan trọng đối với người tạo và người sử dụng các dịch vụ web SDMX.

1. Thông điệp “Envelope” (phong bì): Thông điệp này sử dụng trong các ứng dụng không phải dịch vụ web, nó dư thừa một phần với SOAP. Có thể sử dụng tất cả các thông điệp SDMX mà không có trình bao này.

2. Thông điệp “Structure” (cấu trúc): Thông điệp này mô tả các khái niệm, tập khóa và danh sách mã mà xác định cấu trúc của dữ liệu thống kê. Mỗi tập dữ liệu theo SDMX phải có một cấu trúc tập khóa mô tả nó. Dịch vụ web SDMX phải luôn sẵn có mô tả XML này khi được yêu cầu.

3. Thông điệp dữ liệu “Generic” (chung): Đây là cách đánh dấu dữ liệu SDMX “chung”. Lược đồ này mô tả định dạng phi tập khóa để trao đổi dữ liệu SDMX, đó là yêu cầu mà mỗi dịch vụ web SDMX tạo cho các dữ liệu của nó sẵn có ít nhất trong dạng này. (Thông thường, các dịch vụ tương đương cũng hỗ trợ các dạng XML theo tập khóa khác).

4. Thông điệp dữ liệu “Compact” (cô đọng): Đây là định dạng lược đồ chuẩn được tạo từ mô tả cấu trúc có sử dụng các thẻ và một ánh xạ chuẩn. Nó đặc trưng cho cấu trúc của tập khóa cụ thể, vì vậy, mỗi tập khóa đều có lược đồ “Compact” (cô đọng) riêng. Thông điệp này được thiết kế để cho phép truyền các tập dữ liệu lớn và cho phép các cập nhật tăng. Đây là một định dạng dữ liệu mà dịch vụ web cung cấp phụ thuộc vào các yêu cầu dữ liệu mà chúng trao đổi.

5. Thông điệp dữ liệu “Utility” (tiện ích): Thông điệp này ít được sử dụng để cung cấp các dịch vụ web SDMX nhưng hữu ích trong một số lĩnh vực. Giống như thông điệp dữ liệu “Compact”, thông điệp dữ liệu “Utility” đặc trưng cho tập khóa của dữ liệu được sử dụng để đánh dấu. Thông điệp này được tạo theo các ánh xạ chuẩn từ mô tả tập khóa. Nó được thiết kế để cung cấp lược đồ XML điển hình cho kiểu dữ liệu thống kê riêng khi được sử dụng bởi công cụ trình diễn và biên tập XML. Không giống với thông điệp Compact, dữ liệu này khá dài dòng và yêu cầu một tập dữ liệu đầy đủ. Do vậy, nó không được sử dụng cho các cập nhật tăng.

6. Thông điệp “Cross-SectionalData” (dữ liệu phần giao): Thông điệp này cho phép nhiều quan sát đơn được cung cấp cùng một giá trị thời gian quan sát cho trước, hơn nữa nó cho phép một số giá trị của khóa ở mức quan sát đó (thay vì ở mức chuỗi hoặc mức cao hơn, giống như chuỗi thời gian liên quan đến định dạng dữ liệu SDMX). Điều này đặc biệt hữu ích đối với một số tập dữ liệu thống kê. Cũng giống như thông điệp Compact (cô đọng) và thông điệp Utility (tiện ích), thông điệp này được tạo từ việc mô tả cấu trúc theo ánh xạ chuẩn.

7. Thông điệp “Query” (truy vấn): Đây là thông điệp được sử dụng để gọi dịch vụ web SDMX. Đó là đặc điểm chung của tất cả các tập khóa, tạo các truy vấn dưới dạng các giá trị được quy định cho khái niệm của một cấu trúc cụ thể (như được quy định trong mô tả cấu trúc). Nó cho phép người sử dụng truy vấn dữ liệu, khái niệm, danh sách mã và tập khóa - do đó, tất cả các chức năng này được hỗ trợ bởi dịch vụ web SDMX.

Chú ý rằng đối với mỗi thông điệp dữ liệu, phần tử toàn cục luôn sẵn có để sử dụng cùng với các phong bì SOAP. Dịch vụ web SDMX không nên sử dụng phần tử <wsdl : types>, mà sử dụng phần tử <wsdl : import> để quy định các lược đồ liên quan.

Chú ý rằng tất cả các dịch vụ web SDMX được yêu cầu để hỗ trợ các trao đổi, các trao đổi này cho phép truy vấn đến các tập khóa, danh sách mã và khái niệm. Khuyến cáo rằng định dạng dữ liệu chung được hỗ trợ tối thiểu. Điều này đảm bảo có ít nhất một định dạng dữ liệu tồn tại chung giữa người công bố dữ liệu và người sử dụng dịch vụ web. Trong nhiều trường hợp, các định dạng dữ liệu tối ưu hơn được yêu cầu và sử dụng chung khi chúng được tối ưu hóa để sử dụng với các quá trình liên kết chung với dữ liệu đó. Tuy nhiên, việc đảm bảo định dạng dữ liệu chung và đơn lẻ là cơ sở để phổ biến tính liên tác được xây dựng để sử dụng dữ liệu đó trong thời gian tiếp theo.

3. Mẫu trao đổi cho các dịch vụ web SDMX

Do SDMX yêu cầu một số định dạng dữ liệu (mặc dù chỉ yêu cầu một định dạng) và do nó liên quan đến cả dữ liệu và siêu dữ liệu cấu trúc cần thiết để hiểu và xử lý dữ liệu đó, dịch vụ web SDMX bao gồm một tập các trao đổi dữ liệu. Vì vậy, dịch vụ web SDMX thực thi “mẫu trao đổi đa thông điệp” (theo thuật ngữ WSDL). Các trao đổi này được liệt kê dưới đây, cùng với chỉ dẫn dịch vụ web SDMX đó có được yêu cầu để hỗ trợ chúng hay không:

1. Thu được tập khóa: Đây là trao đổi được gọi bởi thông điệp Query (truy vấn), trong đó thông điệp trả về là các mô tả hay mô tả tập khóa được biểu diễn như thông điệp Structure (cấu trúc). Khuyến cáo nên hỗ trợ.

2. Thu được danh sách mã: Đây là trao đổi được gọi bởi thông điệp Query (truy vấn), trong đó thông điệp trả về là một hay nhiều danh sách mã được biểu diễn như thông điệp Structure (cấu trúc). Khuyến cáo nên hỗ trợ.

3. Thu được khái niệm: Đây là trao đổi được gọi bởi thông điệp Query (truy vấn), trong đó thông điệp phản hồi là một hoặc nhiều khái niệm được biểu diễn như thông điệp Structure (cấu trúc). Khuyến cáo nên hỗ trợ.

4. Thu được dữ liệu chung: Đây là trao đổi được gọi bởi thông điệp Query (truy vấn), trong đó thông điệp phản hồi là dữ liệu được đánh dấu theo thông điệp Generic Data (dữ liệu chung). Khuyến cáo nên hỗ trợ.

5. Thu được dữ liệu cô đọng: Đây là trao đổi được gọi bởi thông điệp Query (truy vấn), trong đó thông điệp phản hồi là dữ liệu được đánh dấu theo thông điệp Compact Data (dữ liệu cô đọng).

6. Thu được dữ liệu tiện ích: Đây là trao đổi được gọi bởi thông điệp Query (truy vấn), trong đó thông điệp phản hồi là dữ liệu được đánh dấu theo thông điệp Utility Data (dữ liệu tiện ích).

7. Thu được dữ liệu phần giao: Đây là trao đổi được gọi bởi thông điệp Query (truy vấn), trong đó thông điệp phản hồi là dữ liệu được đánh dấu theo thông điệp Sectional Data (dữ liệu phần giao).

Tất cả các dịch vụ web SDMX nên được mô tả bằng cách sử dụng các trường hợp WSDL cụ thể, sử dụng WSDL để quy định các khía cạnh của trao đổi đa thông điệp mà chúng hỗ trợ. Phần tử toàn cục cho mỗi định dạng dữ liệu XML trong SDMX nên được quy định đúng như nội dung các phản hồi cho mỗi trao đổi.

4. Phù hợp với WS-I

Để đảm bảo tính liên tác giữa các dịch vụ web SDMX, tất cả dịch vụ web SDMX nên phù hợp với các phần tài liệu WS-I 1.1 được khuyến cáo. Tài liệu này có thể được tìm thấy tại địa chỉ

<http://www.ws-i.org/Profiles/BasicProfile-1.1-2004-08-24.html>. Các phần được khuyến cáo liên quan đến việc sử dụng SOAP, WSDL. UDDI rất hữu ích cho việc thông báo sự tồn tại của dịch vụ web SDMX nhưng không tập trung vào tính liên tác SDMX. (Trong tương lai, SDMX sẽ chuẩn hóa mô hình sổ đăng ký được hỗ trợ để phù hợp với đặc tả sổ đăng ký SDMX có khả năng được thực hiện trên UDDI hoặc trong một số hình thức khác. Hiện tại phần này không được áp dụng)

5. Tập dữ liệu lớn và truy vấn

Bởi vì một vài truy vấn có thể tạo ra số lượng lớn các điểm dữ liệu như một phản hồi, nên khuyến cáo rằng dịch vụ web SDMX hỗ trợ việc sử dụng trường "DefaultLimit" (giới hạn mặc định) trong thông điệp QuerySDMX (truy vấn SDMX). Nếu một phản hồi lớn hơn giới hạn mặc định được đề nghị trong truy vấn, thì phản hồi đó sẽ bị cắt bớt. Phản hồi cắt bớt là phản hồi từng phần nhưng vẫn là tài liệu SDMX ML hợp lệ. Thực tế việc cắt bớt nên được chỉ ra với trường "Truncated" (cắt bớt) trong phần tử Header (tiêu đề) của thông điệp phản hồi đó.

Chú ý rằng giới hạn mặc định được giải thích theo nghĩa rộng, chứ không phải theo nghĩa hẹp, không dễ dự đoán chính xác tác dụng của việc cắt bớt khi các dịch vụ web vẫn phải tạo ra trường hợp SDMX-ML hợp lệ.

Đây là trách nhiệm của dịch vụ truy vấn nhằm điều chỉnh truy vấn và gửi lại truy vấn để tạo ra phản hồi không bị cắt bớt, nếu cần.

MỤC LỤC

Lời nói đầu

1. Giới thiệu
2. Dịch vụ web và SDMX-ML
3. Mẫu trao đổi cho các dịch vụ web SDMX
4. Phù hợp với WS-I
5. Tập dữ liệu lớn và truy vấn