

Số: 3788 /BT-TT-THH

Hà Nội, ngày 26 tháng 12 năm 2014

V/v Hướng dẫn liên thông, trao đổi dữ liệu có cấu trúc bằng ngôn ngữ XML giữa các hệ thống thông tin trong cơ quan nhà nước

Kính gửi:

- Các Bộ, cơ quan ngang Bộ, cơ quan thuộc Chính phủ;
- Ủy ban nhân dân các tỉnh, thành phố trực thuộc Trung ương;
- Các tập đoàn kinh tế, tổng công ty Nhà nước.

Căn cứ Luật Công nghệ thông tin ngày 29/6/2006;

Căn cứ Nghị định số 64/2007/NĐ-CP ngày 10/4/2007 của Chính phủ về ứng dụng công nghệ thông tin trong hoạt động của cơ quan nhà nước;

Căn cứ Nghị định số 132/2013/NĐ-CP ngày 16/10/2013 của Chính phủ quy định chức năng, nhiệm vụ, quyền hạn và cơ cấu tổ chức của Bộ Thông tin và Truyền thông;

Căn cứ Thông tư số 22/2013/TT-BTTTT ngày 23/12/2013 của Bộ trưởng Bộ Thông tin và Truyền thông Ban hành Danh mục tiêu chuẩn kỹ thuật về ứng dụng công nghệ thông tin trong cơ quan nhà nước;

Bộ Thông tin và Truyền thông công bố Hướng dẫn liên thông, trao đổi dữ liệu có cấu trúc bằng ngôn ngữ XML giữa các hệ thống thông tin. Hướng dẫn này được áp dụng làm cơ sở trong việc khảo sát, thiết kế các hệ thống thông tin trong hoạt động ứng dụng công nghệ thông tin của các cơ quan nhà nước.

Đề nghị Quý cơ quan tải bản điện tử tài liệu hướng dẫn từ cổng thông tin điện tử <http://www.mic.gov.vn> hoặc trang thông tin điện tử <http://www.aita.gov.vn>.

Trong quá trình thực hiện, nếu có điều gì vướng mắc, đề nghị Quý cơ quan phản ánh về Bộ Thông tin và Truyền thông để được hướng dẫn giải quyết./.

Nơi nhận:

- Như trên;
- Bộ trưởng (để b/c);
- Đơn vị chuyên trách về CNTT của các Bộ, cơ quan ngang Bộ, cơ quan thuộc Chính phủ (gửi email);
- Đơn vị chuyên trách về CNTT của Văn phòng Trung ương Đảng, Văn phòng Chủ tịch nước, Văn phòng Quốc hội, Tòa án nhân dân tối cao, Viện kiểm sát nhân dân tối cao, Kiểm toán Nhà nước;
- Đơn vị chuyên trách về CNTT của Cơ quan Trung ương của các đoàn thể;
- Sở TT&TT các tỉnh, thành phố trực thuộc Trung ương (gửi mail);
- Cổng thông tin điện tử Bộ TT&TT;
- Lưu: VT, THH (CSHTTT).

**KT. BỘ TRƯỞNG
THỨ TRƯỞNG**

(Đã Ký)

Nguyễn Minh Hồng

BỘ THÔNG TIN VÀ TRUYỀN THÔNG

HƯỚNG DẪN LIÊN THÔNG, TRAO ĐỔI DỮ LIỆU CÓ CẤU TRÚC BẰNG NGÔN NGỮ XML GIỮA CÁC HỆ THỐNG THÔNG TIN TRONG CƠ QUAN NHÀ NƯỚC

*(Kèm theo Công văn số 3788 /BT-TT-THH ngày 26 tháng 12 năm 2014
của Bộ Thông tin và Truyền thông)*

Hà Nội, 2014

I. HƯỚNG DẪN CHUNG

1. Phạm vi hướng dẫn

Tài liệu này hướng dẫn các điều kiện cơ bản, mô hình, cách thức, quy trình thực hiện, hướng dẫn áp dụng các tiêu chuẩn công nghệ thông tin trong việc triển khai kết nối liên thông, trao đổi dữ liệu có cấu trúc bằng ngôn ngữ XML giữa các hệ thống thông tin trong cơ quan nhà nước.

Cơ quan, tổ chức chịu trách nhiệm xây dựng các cơ sở dữ liệu quốc gia, cơ sở dữ liệu của các Bộ, ngành, địa phương và các cơ sở dữ liệu, hệ thống thông tin của cơ quan nhà nước áp dụng hướng dẫn này để trao đổi dữ liệu với nhau tạo điều kiện thúc đẩy khai thác, sử dụng chung dữ liệu.

Đối với các hệ thống thông tin đặc thù, ngoài việc trao đổi dữ liệu theo hướng dẫn này, các cơ quan, đơn vị triển khai bổ sung các phương thức khác để tạo điều kiện thuận lợi hơn trong việc chia sẻ, kết nối sử dụng dữ liệu của các hệ thống thông tin.

Đối với việc trao đổi văn bản điện tử, thực hiện theo văn bản của Bộ Thông tin và Truyền thông về hướng dẫn kỹ thuật liên thông giữa các hệ thống quản lý văn bản và điều hành trong cơ quan nhà nước đã được ban hành.

2. Đối tượng áp dụng

Đối tượng áp dụng của tài liệu này là các cơ quan, tổ chức, cá nhân có liên quan đến việc xây dựng, quản lý các hệ thống thông tin, các cơ sở dữ liệu trong hoạt động ứng dụng công nghệ thông tin của các cơ quan nhà nước.

3. Thuật ngữ

Trong văn bản này, những thuật ngữ được hiểu như sau:

Dữ liệu có cấu trúc: là dữ liệu có thể được mô hình hóa bởi các mô hình cấu trúc dữ liệu, dữ liệu được lưu trong các trường cố định trong một bản ghi hoặc tệp dữ liệu. Cơ sở dữ liệu quan hệ hay bảng tính là hình thức lưu trữ thông dụng của dữ liệu có cấu trúc.

Bên cung cấp: là cơ quan, đơn vị có vai trò cung cấp thông tin, dữ liệu hoặc hệ thống thông tin của cơ quan, đơn vị có vai trò cung cấp thông tin, dữ liệu.

Bên khai thác: là cơ quan, đơn vị có nhu cầu khai thác thông tin, dữ liệu hoặc hệ thống thông tin của cơ quan, đơn vị đóng vai trò tiếp nhận thông tin, dữ liệu từ hệ thống thông tin của cơ quan, đơn vị cung cấp thông tin, dữ liệu.

Phần mềm/công cụ ETL (Extract - Transfer - Load): là phần mềm có trách nhiệm trích xuất, chuyển đổi, tích hợp dữ liệu giữa các định dạng dữ liệu khác nhau, có hỗ trợ đọc định dạng dữ liệu được mô tả bằng ngôn ngữ XML.

4. Hướng dẫn chung về trao đổi dữ liệu có cấu trúc thống nhất giữa các hệ thống thông tin trong cơ quan nhà nước

4.1. Yêu cầu về trao đổi thông tin, dữ liệu

Yêu cầu trong việc tổ chức trao đổi thông tin, dữ liệu giữa các hệ thống thông tin:

- Tuân thủ các tiêu chuẩn, quy chuẩn kỹ thuật quy định bảo đảm sự tương thích, thông suốt và an toàn giữa các cơ quan nhà nước trong quá trình chia sẻ, trao đổi thông tin, dữ liệu;

- Cần lập kế hoạch trước khi tạo ra hoặc thu thập dữ liệu mới với mục đích sử dụng rõ ràng, xác định rõ phạm vi khai thác dữ liệu và vai trò của dữ liệu cung cấp cho các đơn vị, cơ quan nhà nước khác khai thác, sử dụng;

- Dữ liệu trao đổi cần phải được chuẩn hóa, có tính ổn định và không phụ thuộc vào một hệ thống thông tin cụ thể;

- Dữ liệu trao đổi phải có cấu trúc dữ liệu tham chiếu. Bên cung cấp có trách nhiệm xây dựng và ban hành cấu trúc dữ liệu;

- Công bố công khai các cấu trúc dữ liệu trao đổi, mô tả nội dung, thành phần dữ liệu, đầu mối tiếp nhận yêu cầu khai thác dữ liệu. Đối với việc trao đổi dữ liệu trực tuyến trên mạng, cần công bố rõ địa chỉ kết nối cung cấp dữ liệu;

- Lược đồ cấu trúc dữ liệu, địa chỉ khai thác dữ liệu trên mạng phục vụ trao đổi dữ liệu giữa các cơ quan nhà nước đề nghị gửi về Bộ Thông tin và Truyền thông (Cục Tin học hóa) qua đường văn bản đồng thời gửi bản điện tử tới địa chỉ thư điện tử vanthucucud@mic.gov.vn sau khi hoàn thành việc xây dựng hệ thống thông tin để Bộ Thông tin và Truyền thông theo dõi, hướng dẫn, cung cấp cho các cơ quan, đơn vị có nhu cầu.

4.2. Cách thức trao đổi

Bên cung cấp và bên khai thác có thể trao đổi qua các cách thức sau:

- Trao đổi trực tiếp: Hệ thống thông tin của bên khai thác có nhu cầu khai thác thông tin, dữ liệu kết nối trực tiếp tới hệ thống thông tin của bên cung cấp để truy cập thông tin, dữ liệu.

- Khai thác trực tuyến: Bên khai thác truy cập vào cổng thông tin điện tử cung cấp dữ liệu của bên cung cấp để chọn, trích lọc dữ liệu và tải về hoặc tải dữ liệu đã chuẩn bị sẵn để nhập vào hệ thống của bên khai thác.

Khuyến nghị triển khai cả hai cách thức trao đổi dữ liệu, ưu tiên cách thức trao đổi trực tiếp, tự động giữa hai hệ thống thông tin không cần can thiệp bởi con người.

4.3. Giao thức trao đổi

Thông nhất giao thức trao đổi dữ liệu sẽ đảm bảo dữ liệu được chuyển từ bên cung cấp tới bên khai thác trọn vẹn. Các yêu cầu sau phải được áp dụng thống nhất:

- Thông nhất sử dụng giao thức mạng TCP để trao đổi dữ liệu trong môi trường mạng;

- Đối với việc trao đổi bằng phương thức khai thác trực tuyến trên môi trường mạng, sử dụng các giao thức HTTP/FTP khi không đòi hỏi mức độ bảo mật và giao thức HTTPS/FTPS khi đòi hỏi mức độ bảo mật.

- Đối với phương thức trao đổi trực tiếp:

- + Thông nhất sử dụng dịch vụ web (webservice) trên nền tảng các giao thức HTTP hoặc HTTPS trong trao đổi dữ liệu trực tiếp giữa các hệ thống thông tin;

- + Sử dụng ngôn ngữ WSDL để mô tả dịch vụ web được cung cấp, sử dụng tiêu chuẩn SOAP để đóng gói yêu cầu truy vấn dữ liệu và dữ liệu cần trao đổi.

4.4. Ngôn ngữ, định dạng trao đổi

Đảm bảo bên cung cấp và bên khai thác có thể hiểu được nội dung thông điệp dữ liệu trao đổi. Ngôn ngữ mô tả, định dạng dữ liệu trao đổi cần áp dụng các tiêu chuẩn thống nhất:

- Sử dụng ngôn ngữ XML là ngôn ngữ mô tả dữ liệu phục vụ trao đổi. Nội dung dữ liệu được mô tả bằng ngôn ngữ XML phải chặt chẽ, tường minh; Kiểu dữ liệu được mô tả theo quy định thống nhất của lược đồ dữ liệu cơ bản XML (W3C XMLSchema);

- Thông tin được mô tả trong ngôn ngữ XML sử dụng thống nhất bảng mã UTF8;

- Cấu trúc dữ liệu được mô tả dưới dạng lược đồ cấu trúc dữ liệu. Lược đồ cấu trúc dữ liệu được xây dựng trên cơ sở tiêu chuẩn XML Schema - XSD. Dữ liệu trao đổi mô tả bằng ngôn ngữ XML phải thống nhất và tham chiếu tới các lược đồ này;

- Sử dụng tiêu chuẩn XSLT (Extensible Stylesheet Language Transformations) làm ngôn ngữ mô tả để chuyển đổi, trình diễn dữ liệu trao đổi dưới dạng thân thiện, trực quan với người sử dụng;

- Các thông tin, dữ liệu thuộc phạm vi chuyên ngành như dữ liệu địa lý, dữ liệu giao dịch điện tử,... sử dụng các ngôn ngữ mô tả mở rộng từ ngôn ngữ XML như GML, ebXML... và tuân thủ, tham chiếu tường minh các lược đồ dữ liệu tương ứng.

(Các tiêu chuẩn liên quan đến giao thức, ngôn ngữ, dạng dữ liệu được trình bày tại Phụ lục I, khung thông điệp dữ liệu XML, khung lược đồ cấu trúc dữ liệu được trình bày trong Phụ lục II kèm theo)

4.5. Các điều kiện, yêu cầu khác thúc đẩy trao đổi dữ liệu giữa các hệ thống thông tin thống nhất

- Thống nhất, đồng bộ và nhất quán cách thức mã hóa thông tin giữa các cơ sở dữ liệu trong cơ quan nhà nước, cụ thể:

+ Các thuộc tính dữ liệu được lưu dưới dạng mã (mã đơn vị hành chính, mã quốc gia,...) cần sử dụng theo các quy định hiện hành trong cơ quan nhà nước, không tạo ra các hệ thống mã danh mục riêng đặc thù cho từng hệ thống thông tin, cơ sở dữ liệu;

+ Tham khảo và sử dụng danh mục dữ liệu dùng chung được Tổng cục Thống kê (Bộ Kế hoạch và Đầu tư) và các danh mục dữ liệu chuyên ngành do Bộ, ngành chuyên ngành ban hành đảm bảo thống nhất nội dung thông tin khi tích hợp dữ liệu.

- Một thành phần thông tin có một cấu trúc dữ liệu thuộc một lược đồ mô tả. Không xây dựng lại các cấu trúc dữ liệu của thành phần thông tin đã có, nếu cần sử dụng phải tham chiếu và sử dụng chung.

- Đối với việc đảm bảo an toàn, an ninh, các đơn vị có thể áp dụng các biện pháp: triển khai đường truyền mạng phù hợp, sử dụng giao thức mã hóa đường truyền, sử dụng chữ ký số, đăng nhập xác thực tài khoản và các biện pháp khác phù hợp với các quy định của pháp luật và mục đích, nhu cầu trong trao đổi dữ liệu trên thực tế. Phương pháp sử dụng cần công bố cụ thể cho bên khai thác sử dụng.

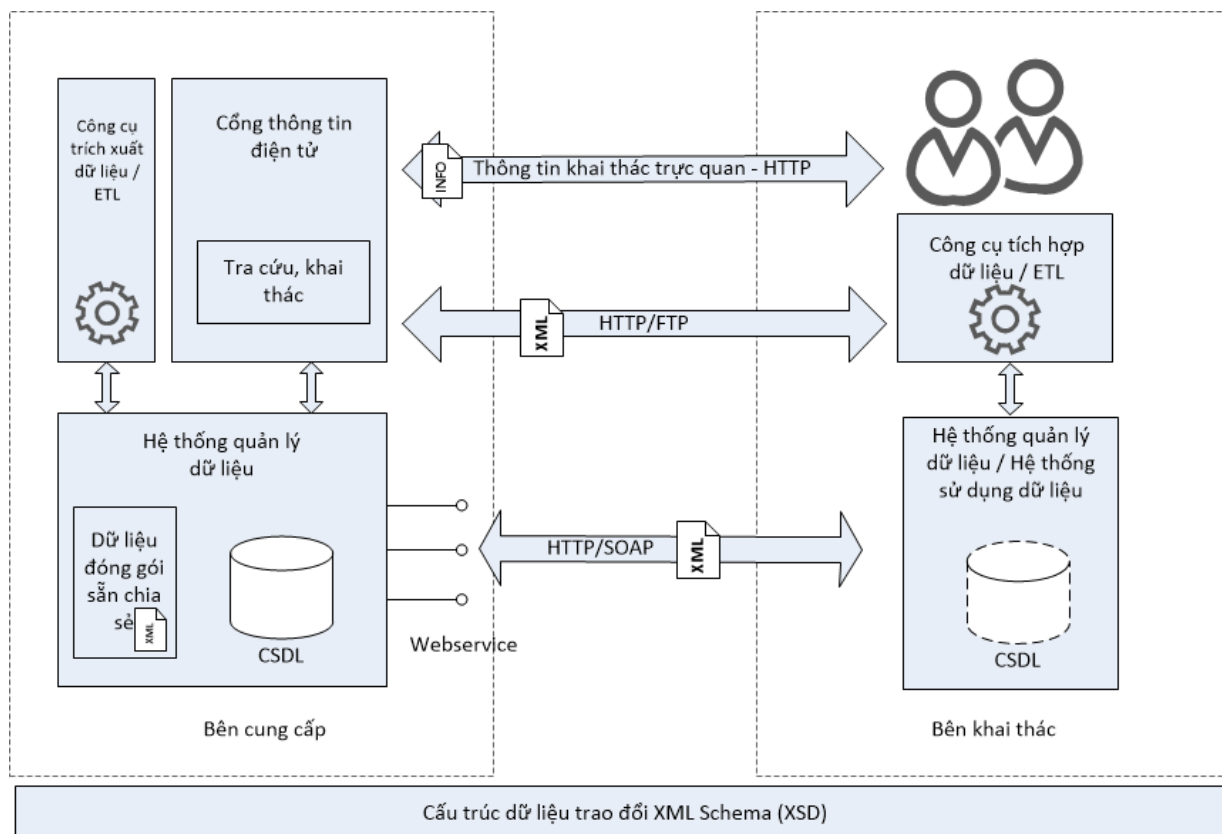
II. HƯỚNG DẪN CỤ THỂ

1. Mô hình và quy trình trao đổi dữ liệu

Mô hình trao đổi dữ liệu giữa bên cung cấp và bên khai thác được thực hiện qua hai phương thức chính là trao đổi trực tiếp và khai thác trực tuyến.

1.1. Mô hình trao đổi dữ liệu

Mô hình trao đổi dữ liệu tổng thể giữa các hệ thống thông tin như sau:



Trong đó:

- Ngôn ngữ mô tả dữ liệu được sử dụng là ngôn ngữ XML đã được quy định bắt buộc sử dụng tại Thông tư số 22/2013/TT-BTTTT.

- Lược đồ cấu trúc dữ liệu trao đổi XMLSchema (XSD) được bên cung cấp chuẩn hóa và cung cấp rộng rãi cho các bên khai thác phục vụ mục đích thống nhất cấu trúc dữ liệu được mô tả bằng XML. Lược đồ cấu trúc dữ liệu được đăng tải trên đường dẫn mạng phù hợp với không gian tên của lược đồ.

- Công cụ trích xuất/tích hợp dữ liệu là một thành phần của hệ thống quản lý dữ liệu được xây dựng để trích xuất/tích hợp dữ liệu từ cơ sở dữ liệu sang định dạng dữ liệu trao đổi XML hoặc ngược lại. Công cụ trích xuất/tích hợp có thể là một phần mềm thương mại ETL hoặc phần mềm nội bộ được xây dựng theo nhu cầu.

- Tùy theo mục đích khai thác, các hệ thống của bên khai thác có thể sử dụng dữ liệu trực tiếp hoặc lưu trữ nội bộ để sử dụng.

- Dữ liệu đóng gói sẵn được trích xuất từ cơ sở dữ liệu của bên cung cấp bằng công cụ phần mềm trích xuất hoặc phần mềm thương mại ETL theo mục đích sử dụng thông dụng nhất và lưu trữ trên hệ thống quản lý phục vụ khai thác của bên khai thác. Vai trò của dữ liệu đóng gói sẵn giúp giảm tải quá trình xử lý dữ liệu của bên cung cấp và đáp ứng nhanh nhu cầu sử dụng.

1.2. Trao đổi trực tiếp

Hệ thống thông tin của bên khai thác thông tin kết nối trực tiếp tới hệ thống thông tin của bên cung cấp để truy cập dữ liệu qua dịch vụ web (webservice) trực tuyến trên mạng. Dịch vụ web được mô tả chi tiết bằng tiêu chuẩn ngôn ngữ WSDL đi kèm, sử dụng giao thức đóng gói yêu cầu truy vấn dữ liệu và dữ liệu trao đổi bằng tiêu chuẩn SOAP (ưu tiên). Dữ liệu trao đổi được đóng gói bằng ngôn ngữ mô tả dữ liệu XML. Cách thức này hoàn toàn tự động và không cần can thiệp bởi con người.

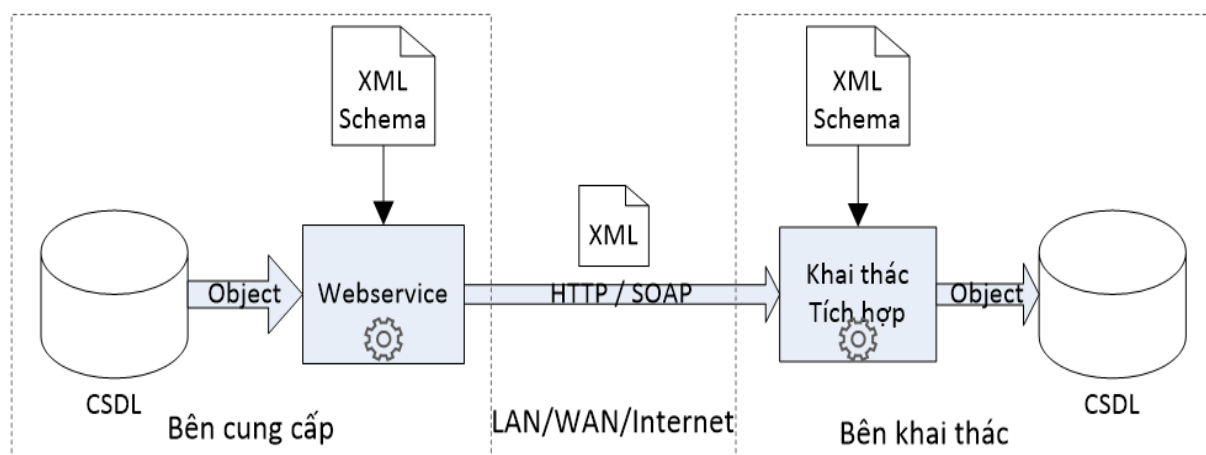
Quy trình trao đổi:

- Bên khai thác đóng gói yêu cầu dưới dạng SOAP (ưu tiên) hoặc sử dụng giao thức dưới dạng HTTP/Post, kết nối đến dịch vụ web của bên cung cấp để gửi yêu cầu;

- Bên cung cấp tiếp nhận yêu cầu, truy vấn dữ liệu và đóng gói dữ liệu dưới dạng ngôn ngữ mô tả XML theo cấu trúc XSD (tiếp đó, đóng gói dưới dạng thông điệp SOAP nếu sử dụng giao thức SOAP) và gửi về cho bên khai thác;

- Bên khai thác tiếp nhận dữ liệu mô tả bằng ngôn ngữ XML, theo chỉ dẫn cấu trúc trong XSD để đọc và phân tích dữ liệu;

- Bên khai thác tích hợp dữ liệu vào cơ sở dữ liệu nội bộ hoặc sử dụng trực tiếp dữ liệu được khai thác bởi các chức năng, nghiệp vụ của hệ thống.



Đặc điểm:

- Kết nối hoàn toàn tự động giữa hai hệ thống thông tin;

- Bên khai thác cần xây dựng module phần mềm kết nối đến các dịch vụ web theo nội dung dịch vụ web được mô tả qua ngôn ngữ WSDL;
- Phần mềm bên khai thác xử lý dữ liệu XML trên cơ sở lược đồ dữ liệu trao đổi XSD;
- Trong trường hợp số lượng các dịch vụ lớn, bên cung cấp có thể tổ chức thành thư mục hỗ trợ đăng ký, tìm kiếm và khai thác các dịch vụ web sử dụng tiêu chuẩn UDDI;
- Có thể mở rộng triển khai hệ thống trung gian như trực tích hợp dịch vụ (ESB) đóng vai trò quản lý và trung chuyển các dịch vụ giữa bên cung cấp và bên khai thác tạo thuận tiện cho việc tìm kiếm, kết nối và sử dụng các dịch vụ.

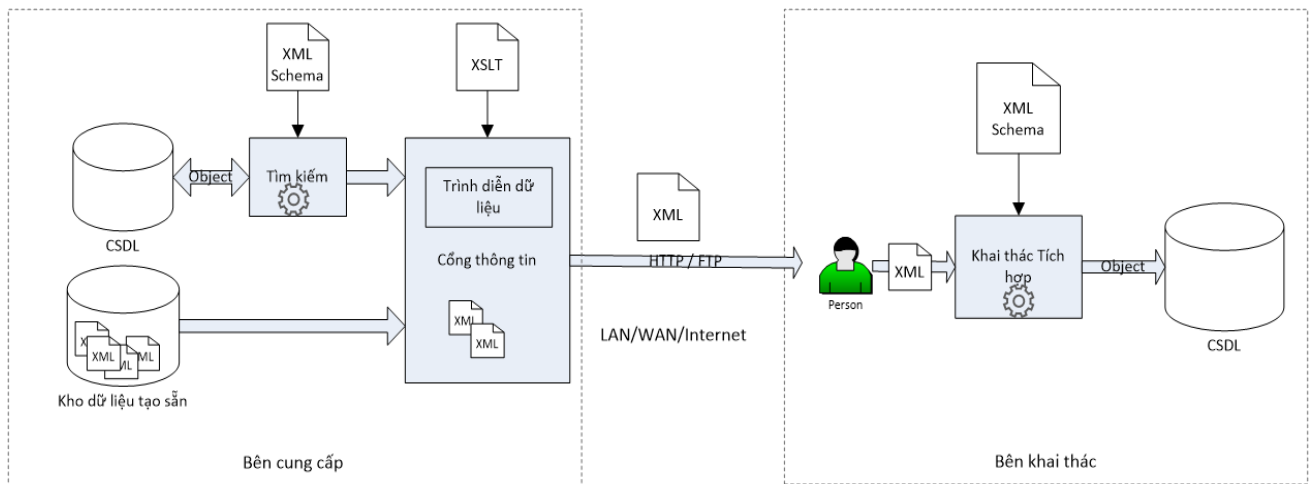
1.3. Khai thác trực tuyến

Bên khai thác truy cập vào cổng thông tin điện tử (hoặc trang thông tin điện tử) cung cấp dữ liệu của bên cung cấp để tải về dưới định dạng trao đổi XML. Dữ liệu khai thác có thể dưới hai hình thức.

- Trích lọc theo điều kiện của người sử dụng tương tác với cổng thông tin điện tử, người sử dụng có thể xem dữ liệu trước khi tải về bằng công nghệ chuyển đổi XSLT được công bố và chỉ dẫn ngay trong thông điệp dữ liệu mô tả XML.
- Dữ liệu đã được chuẩn bị sẵn (dưới dạng hàng hóa trao đổi) theo nhu cầu thông dụng và cho phép người sử dụng chọn từ danh sách và tải về.

Quy trình trao đổi:

- Bên khai thác truy cập vào cổng thông tin điện tử trao đổi dữ liệu của bên cung cấp để tìm kiếm, tải dữ liệu;
- Người sử dụng sử dụng các ứng dụng trên trang web của bên cung cấp để tìm kiếm dữ liệu theo điều kiện hoặc chọn dữ liệu đã được bên cung cấp chuẩn bị sẵn theo từng mục đích phù hợp. Dữ liệu có thể được trình diễn theo định dạng HTML (đã được chuyển hóa bằng XSLT) thân thiện với người sử dụng để xem trước dữ liệu (yêu cầu dữ liệu XML cần liên kết với tệp dữ liệu trình diễn XSLT);
- Bên khai thác tải tệp dữ liệu đã mô tả bằng ngôn ngữ XML về và tiến hành tích hợp vào hệ thống nội bộ.



Đặc điểm:

- Đây là hình thức tự động bên cung cấp nhưng bán tự động bên khai thác;
- Thường sử dụng cho các trường hợp công bố dữ liệu chia sẻ rộng rãi, sử dụng dữ liệu tạo sẵn giúp giảm tải xử lý của bên cung cấp;
- Dữ liệu tìm kiếm động theo nhu cầu và tải về thường sử dụng trong trường hợp phạm vi dữ liệu nhỏ;
- Dữ liệu chuẩn bị sẵn được trích xuất trên cơ sở nhu cầu sử dụng dữ liệu thông dụng giúp giảm tải xử lý của hệ thống bên cung cấp. Dữ liệu chuẩn bị sẵn cần được sắp xếp thành nhóm phù hợp với mục đích sử dụng dữ liệu hoặc chủ đề dữ liệu tạo thuận lợi cho việc tìm kiếm và sử dụng. Dữ liệu chuẩn bị sẵn có thể nén giảm kích cỡ sử dụng tiêu chuẩn nén liệt kê tại Phụ lục 1;
- Quá trình tạo ra các sản phẩm dữ liệu chuẩn bị sẵn của bên cung cấp hoặc quá trình tích hợp dữ liệu trao đổi vào hệ thống thông tin của bên khai thác có thể sử dụng các phần mềm thương mại ETL sẽ không cần phải xây dựng các phần mềm nội bộ đặc thù.

2. Trình tự tổ chức thực hiện trao đổi dữ liệu

2.1. Bên cung cấp

- Xác định nhu cầu khai thác, sử dụng thông tin (của các hệ thống thông tin ngoài phạm vi hệ thống thông tin được xây dựng) của các cơ quan, đơn vị bên ngoài trên cơ sở:
 - + Các hệ thống thông tin hiện hành có nội dung thông tin quản lý liên quan;
 - + Kế hoạch ứng dụng công nghệ thông tin 05 năm, hàng năm của Bộ, ngành, địa phương, của cơ quan, đơn vị mình;
 - + Nhu cầu khai thác thông tin thực tế trong quy trình, nghiệp vụ quản lý giữa các cơ quan, đơn vị.

- Xác định phạm vi nội dung dữ liệu sẽ được cung cấp và đối tượng sẽ khai thác dữ liệu. Trên cơ sở dữ liệu sẽ được xây dựng trong dự án, dữ liệu phục vụ nội bộ, cần phân hoạch rõ giới hạn phạm vi dữ liệu sẽ được cung cấp ra bên ngoài đảm bảo dữ liệu cung cấp phù hợp với các mục đích sử dụng của bên khai thác.

- Xây dựng lược đồ cấu trúc dữ liệu trao đổi XSD trên cơ sở dữ liệu sẽ được cung cấp cho các bên khai thác. Lược đồ cấu trúc dữ liệu cần được công bố và có thể tiếp cận bởi bên khai thác trên trang thông tin điện tử, cổng thông tin điện tử chính thức của cơ quan, đơn vị.

- Xây dựng trình diễn dữ liệu XSLT phù hợp với cấu trúc lược đồ dữ liệu trao đổi. trình diễn dữ liệu cũng cần công bố kèm theo dữ liệu trao đổi.

- Xác định các phương thức cung cấp dữ liệu hệ thống sẽ hỗ trợ.

- Thực hiện tổ chức cung cấp dữ liệu bao gồm:

+ Đối với phương thức trao đổi trực tiếp: bên cung cấp cần triển khai xây dựng và công bố các dịch vụ web trên môi trường mạng;

+ Đối với phương thức khai thác trực tuyến: bên cung cấp cần xác định phạm vi dữ liệu cho phép tìm kiếm, phạm vi dữ liệu tạo sẵn và trích xuất dữ liệu tạo sẵn theo mục đích thông dụng, triển khai cung cấp qua trang thông tin điện tử, cổng thông tin điện tử.

2.2. Bên khai thác

- Xác định nhu cầu và nguồn cung cấp dữ liệu phù hợp với dữ liệu cần khai thác. Để xác định nguồn cung cấp dữ liệu, bên khai thác có thể liên hệ với cơ quan nhà nước thuộc lĩnh vực quản lý, hoặc tra cứu, tìm kiếm trên trang thông tin điện tử, cổng thông tin điện tử của cơ quan nhà nước liên quan;

- Thu nhận lược đồ cấu trúc dữ liệu phục vụ phân tích và xử lý dữ liệu trao đổi. Lược đồ có thể thu được qua một số cách sau:

+ Tại đường dẫn trên mạng tương ứng với không gian tên. Không gian tên có thể được chỉ rõ trong phần đầu thông điệp dữ liệu trao đổi đã mô tả bằng XML;

+ Các văn bản, quy chế liên quan được bên cung cấp ban hành;

+ Được cung cấp trên các trang thông tin điện tử, cổng thông tin điện tử;

- Xác định phương thức khai thác dữ liệu. Tùy thuộc vào khả năng, phương thức cung cấp dữ liệu của bên cung cấp, môi trường sử dụng, bên khai thác có thể sử dụng các phương thức khai thác phù hợp.

- Thực hiện việc khai thác dữ liệu:

+ Đối với trao đổi trực tiếp: bên khai thác cần xây dựng modul khai thác dữ liệu qua dịch vụ web trên cơ sở cấu trúc dữ liệu được cung cấp và các đặc thù dịch vụ bổ sung của bên khai thác được công bố;

+ Trao đổi trực tuyến: bên khai thác tải dữ liệu, sử dụng công cụ tích hợp dữ liệu và lược đồ mô tả bằng cấu trúc dữ liệu để tích hợp dữ liệu vào hệ thống.

3. Áp dụng trong công tác khảo sát và thiết kế các hệ thống thông tin trong cơ quan nhà nước

a) Các hạng mục xác định nhu cầu khai thác của bên cung cấp và xác định nhu cầu và nguồn cung cấp dữ liệu của bên khai thác được trình bày tại Mục 2 Phần II của hướng dẫn này cần thực hiện trong quá trình khảo sát để làm căn cứ thiết kế các hệ thống thông tin trong cơ quan nhà nước.

b) Các nội dung về việc áp dụng quy chuẩn, tiêu chuẩn được giới thiệu tại Phụ lục 1 và các nội dung liên quan tại Mục 4.3, 4.4 Phần I của hướng dẫn này cần được thuyết minh cụ thể áp dụng tại Thiết kế sơ bộ, mục “Danh mục quy chuẩn, tiêu chuẩn kỹ thuật được áp dụng”.

c) Các nội dung được hướng dẫn về mô hình trao đổi dữ liệu có cấu trúc được giới thiệu tại Phần II, Mục 1 tại hướng dẫn này xem xét thiết kế cụ thể tại Thiết kế sơ bộ, mục “Phương án kỹ thuật và công nghệ, kết nối ra bên ngoài”.

d) Bên cung cấp hoặc bên khai thác đối với một hệ thống thông tin là một tác nhân (actor) của hệ thống khi thiết kế và xác định giá trị phần mềm nội bộ.

đ) Các nội dung, yêu cầu liên quan đến trao đổi dữ liệu có cấu trúc trong hướng dẫn này làm cơ sở trong việc bổ sung các trường hợp sử dụng (usecases) trong thiết kế sơ bộ của phần mềm nội bộ.

e) Đặc tả cụ thể các dịch vụ web trao đổi dữ liệu, thiết kế lược đồ dữ liệu trao đổi (XSD) cần thực hiện trong công tác lập thiết kế thi công.

g) Các nội dung thực hiện được hướng dẫn Mục 2 Phần 2 cần thể hiện rõ kết quả trong thiết kế sơ bộ và thiết kế thi công của hệ thống.

4. Một số quy tắc thống nhất đối với ngôn ngữ và lược đồ dữ liệu về trao đổi dữ liệu thống nhất trong cơ quan nhà nước

Để tăng tính sử dụng chung, dễ dàng trong việc tích hợp, sử dụng dữ liệu, trao đổi dữ liệu có cấu trúc thống nhất áp dụng chung các quy tắc sau liên quan đến ngôn ngữ XML và lược đồ cấu trúc dữ liệu sau:

4.1. Ngôn ngữ trao đổi dữ liệu XML

QT1. Chỉ thị phiên bản ngôn ngữ XML, bảng mã hóa ký tự UTF-8 trong thông điệp dữ liệu XML:

Ví dụ: `<?xml version="1.0" encoding="utf-8"?>`

QT2. Tất cả các thẻ được sử dụng trong thông điệp dữ liệu XML phải được khai báo tường minh và tham chiếu chặt chẽ tại một trong các lược đồ đã được sử dụng, tham chiếu.

QT3. Phải có cấu trúc dữ liệu đi kèm, lược đồ dữ liệu phải được khai báo trong thông điệp dữ liệu.

Ví dụ: `xmlns="http://aita.gov.vn/DataExchange/CommonType.xsd"`

QT4. Tuân thủ và tham chiếu lược đồ tiêu chuẩn W3C XMLSchema, và sử dụng định dạng thuộc tính dữ liệu của lược đồ này. Không gian tên của lược đồ tiêu chuẩn được khai báo tường minh trong thông điệp dữ liệu XML.

Ví dụ: `xmlns:xs="http://www.w3.org/2001/XMLSchema"`

QT5. Sử dụng định dạng thuộc tính dữ liệu tiêu chuẩn của W3C XMLSchema: kiểu ngày `xs:date` có định dạng: YYYY-MM-DD (năm-tháng-ngày); kiểu ngày giờ là `xs:dateTime`; kiểu giờ là `xs:time`

Ví dụ: `<NgaySinh>2000-12-31</NgaySinh>` //năm-tháng-ngày
`<GioSinh>23:30:00</GioSinh>` //giờ:phút:giây
`<BatDau>2000-12-31T23:20:0</BatDau>`
//năm-tháng-ngày{T}giờ:phút:giây

QT6. Sử dụng dấu thập phân là dấu chấm (.) trong dữ liệu kiểu số.

Ví dụ: `<TyLeTangDanSo>10.5</TyLeTangDanSo>` //10,5(phần trăm)

4.2. Cấu trúc dữ liệu trao đổi XML Schema

QT7. Sử dụng lược đồ cơ bản W3C XMLSchema làm cơ sở để thiết kế lược đồ dữ liệu. Lược đồ W3C XMLSchema đã được áp dụng trong hầu hết các hệ thống thông tin và là lược đồ cơ bản của ngôn ngữ XML, thống nhất sử dụng W3C XMLSchema sẽ giúp các hệ thống thông tin hiểu thống nhất được kiểu và cấu trúc các đối tượng dữ liệu cơ bản.

QT8. Xác định rõ không gian tên (namespace) của lược đồ cấu trúc dữ liệu cần xây dựng để khai báo cho mỗi thành phần nội dung thông tin trao đổi. Không gian tên được đặt theo địa chỉ cổng thông tin/trang thông tin điện tử chính thức của cơ quan, đơn vị.

Ví dụ:

`targetNamespace="http://aita.gov.vn/DataExchange/CommonType.xsd"`

QT9. Tất cả các phần tử dữ liệu khai báo cần chỉ rõ số lượng:

Ví dụ: đặt `elementFormDefault="qualified"` trong khai báo lược đồ và chỉ rõ số lượng trong khai báo phần tử: `minOccurs="1" maxOccurs="1"`

QT10. Các kiểu dữ liệu cần khai báo tổng thể và sử dụng chung, không khai báo riêng lẻ trong các đối tượng khác, đảm bảo tính tái sử dụng cao, cấu trúc rõ ràng.

QT11. Thuộc tính của đối tượng hoặc đối tượng trong lược đồ không đặt giá trị ngầm định hoặc cố định.

Ví dụ không khai báo:

```
<xs:attribute name="QueQuan" type="xs:string" default="Hà Nội"/>
```

QT12. Không sử dụng giá trị trống (`xs:nil`) cho đối tượng.

Ví dụ không khai báo:

```
<xs:element name="element_1" type="xs:string" nillable="true"/>
```

Và cần chỉ rõ giá trị trong thông điệp dữ liệu `<element_1>{giá trị}</element_1>`

QT13. Các thuộc tính dữ liệu thuộc danh mục có giá trị giới hạn cần được khai báo thành kiểu dưới dạng liệt kê (enumeration).

Ví dụ:

```
<xs:simpleType name="Quocgia_Type">
  <xs:restriction base="xs:string">
    <xs:enumeration value="VN"/>          <!-- Việt Nam-->
    <xs:enumeration value="US"/>          <!-- Mỹ -->
    <xs:enumeration value="FR"/>          <!-- Pháp -->
    ...
  </xs:restriction>
</xs:simpleType>
```

QT14. Đối với kiểu dữ liệu nhị phân, sử dụng đường dẫn hoặc kiểu dữ liệu `base64Binary` tiêu chuẩn. Ví dụ:

```
<element name="Image" type="xs:base64Binary"/>
```

Phụ lục I

HƯỚNG DẪN VỀ MỤC ĐÍCH, PHẠM VI SỬ DỤNG MỘT SỐ TIÊU CHUẨN CƠ BẢN PHỤC VỤ TRAO ĐỔI DỮ LIỆU CÓ CẤU TRÚC

STT	Phạm vi, mục đích sử dụng	Tiêu chuẩn sử dụng	Ghi chú
I. Đóng gói dữ liệu trao đổi			
1.	Trình diễn bộ ký tự, bảng mã ký tự dùng để mô tả dữ liệu trong trao đổi dưới định dạng ngôn ngữ mô tả XML	UTF-8	Bảng mã ký tự tiếng Việt
2.	Ngôn ngữ phục vụ mô tả dữ liệu trao đổi	XML	Định dạng dữ liệu được trao đổi. Dữ liệu có cấu trúc sẽ được mô tả dưới dạng ngôn ngữ này.
3.	Nén tệp dữ liệu XML khi trao đổi qua mạng	ZIP, GZ	Trong trường hợp dữ liệu tệp XML lớn, có thể sử dụng các tiêu chuẩn này để nén trước khi truyền qua mạng
4.	Mở rộng của XML để đóng gói dữ liệu địa lý	GML	Cấu trúc XML phải tuân theo lược đồ có không gian tên http://www.opengis.net/gml
II. Cấu trúc dữ liệu đóng gói XML			
5.	Lược đồ mô tả cấu trúc dữ liệu đã mô tả trong XML	XML Schema (XSD)	Giúp các hệ thống thông tin hiểu cấu trúc của dữ liệu khi trao đổi
6.	Định dạng tệp mô tả lược đồ cấu trúc dữ liệu XSD	XML	Nội dung của tệp lược đồ mô tả dữ liệu cũng được mô tả bằng ngôn ngữ XML
III. Trình diễn dữ liệu đã đóng gói cho người sử dụng			
7.	Trình diễn xem trước dữ liệu	XSL (XSLT)	Biến đổi dữ liệu XML thành HTML để trình diễn dữ liệu đã mô tả dưới ngôn ngữ XML thành giao diện trực quan thân thiện người sử dụng Tiêu chuẩn XSLT được áp dụng cho các hệ thống thông tin cho phép xem trước, dữ liệu hoặc tệp XSLT được phân phối kèm theo dữ liệu XML phục vụ hiển thị dữ liệu cho người sử dụng
8.	Hiển thị xem trước dữ liệu trên web	HTML	Trình diễn nội dung dữ liệu trao đổi khi sử dụng tiêu chuẩn biến đổi dữ liệu XSLT trên thông tin dữ liệu trao đổi XML

9.	Giao diện người sử dụng trên web hỗ trợ tìm kiếm, khai thác dữ liệu cho người sử dụng	HTML	Tiêu chuẩn trình bày giao diện cung cấp chức năng tra cứu, tìm kiếm dữ liệu, liệt kê dữ liệu có sẵn có thể tải về
IV. Truyền tệp tin dữ liệu đã đóng gói qua mạng			
10.	Truyền tệp tin dữ liệu đã đóng gói bằng ngôn ngữ XML qua mạng	HTTP FTP	Giao thức truyền tệp tin dữ liệu đã được đóng gói bằng ngôn ngữ XML qua mạng được sử dụng trong trường hợp trao đổi trực tuyến, Bên khai thác tải về qua mạng từ trang thông tin điện tử, cổng thông tin điện tử của bên cung cấp
11.	Truyền tệp tin dữ liệu đã đóng gói bằng ngôn ngữ XML qua mạng yêu cầu bảo mật	HTTPS FTPS	
V. Kết nối trao đổi trực tiếp giữa các hệ thống thông tin			
12.	Giao thức nền cho dịch vụ web thực hiện trao đổi dữ liệu	HTTP	
13.	Ngôn ngữ đóng gói truy vấn và dữ liệu trong dịch vụ web	SOAP	Chuẩn đóng gói câu truy vấn dữ liệu và dữ liệu được sử dụng rộng rãi trong nhiều công nghệ hiện nay
14.	Ngôn ngữ mô tả chi tiết dịch vụ web trao đổi dữ liệu	WSDL	Các dịch vụ web cung cấp dữ liệu cần mô tả kèm theo, các mô tả bằng ngôn ngữ này
15.	Dịch vụ web để trao đổi dữ liệu địa lý.	WFS	Chỉ sử dụng đối với dữ liệu địa lý. Khi trao đổi dữ liệu đối tượng địa lý cần xây dựng các dịch vụ web này
16.	Phục vụ xây dựng các danh bạ dịch vụ hỗ trợ mô tả, tìm kiếm, tích hợp dịch vụ.=	UDDI	Nền tảng độc lập để mô tả dịch vụ, hỗ trợ tìm kiếm tích hợp dịch vụ trên mạng Sử dụng khi dùng

Phụ lục II

KHUNG LƯỢC ĐỒ XSD VÀ THÔNG ĐIỆP ĐÓNG GÓI DỮ LIỆU XML

1. Khung lược đồ dữ liệu trao đổi

```
<?xml version="1.0" encoding="utf-8"?>
<!-- Khai báo không gian tên cho cấu trúc dữ liệu cần định nghĩa-->
<xs:schema targetNamespace="{không gian tên lược đồ}"
  elementFormDefault="qualified"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns="{không gian tên lược đồ}" vc:minVersion="1.1"
  xmlns:vc="http://www.w3.org/2007/XMLSchema-versioning">
  <!-- Khai báo nhập không gian tên khác để sử dụng nếu cần-->
  <xs:import namespace="{không gian tên cần nhập}"/>
  <!-- Khai báo nhập thành phần lược đồ từ tệp khác có cùng không gian
  tên nếu có-->
  <xs:include schemaLocation="{đường dẫn}/{tên tệp lược đồ}.xsd"/>
  <xs:annotation>
    <xs:documentation xml:lang="vi">
      {ghi chú về nội dung lược đồ định nghĩa}
    </xs:documentation>
  </xs:annotation>
  {Định nghĩa các element, các kiểu dữ liệu mới được khai báo ở đây}
</xs:schema>
```

2. Khung thông điệp dữ liệu được mô tả bằng XML

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Chỉ thị trình diễn dữ liệu-->
<?xml-stylesheet type="text/xsl" href="{tên tệp trình diễn}.xslt"?>
<!-- Tất cả các đối tượng dữ liệu trong đối tượng gốc và phần tử con
của nó được khai báo cấu trúc trong không gian tên -->
<{Đối tượng gốc} xmlns="{không gian tên lược đồ}"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="{không gian tên lược đồ} {tên tệp}.xsd"
  xmlns:xs="http://www.w3.org/2001/XMLSchema" >
  {các đối tượng dữ liệu được liệt kê tại đây}
</{Đối tượng gốc}>
```

Đối với dữ liệu XML được đóng gói SOAP. Nội dung dữ liệu được đưa vào phần Body của thông điệp SOAP

```
<?xml version="1.0" encoding="UTF-8"?>
<soap:Envelope xmlns:soap="http://www.w3.org/2003/05/soap-envelope">
  <soap:Header>
  </soap:Header>
  <soap:Body>
    <{Đối tượng gốc} xmlns="{không gian tên lược đồ}"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="{không gian tên lược đồ} {tên tệp}.xsd"
      xmlns:xs="http://www.w3.org/2001/XMLSchema" >
      {các đối tượng dữ liệu được liệt kê tại đây}
    </{Đối tượng gốc}>
  </soap:Body>
</soap:Envelope>
```

Ghi chú: phần in nghiêng trong ngoặc kép cần thay thế phù hợp