

**NGHIÊN CỨU ỨNG DỤNG PHẦN MỀM QUẢN LÝ DỮ LIỆU VI MÔ
(MICRODATA MANAGEMENT) VÀO VIỆC LƯU TRỮ,
KHAI THÁC DỮ LIỆU CÁC CUỘC ĐIỀU TRA**

1. Cấp đề tài : Cơ sở
2. Thời gian nghiên cứu : 2008
3. Đơn vị thực hiện : Đoàn TNCS Hồ Chí Minh (TCTK)
4. Chủ nhiệm đề tài : CN. Nguyễn Quang Phương

PHẦN I

GIỚI THIỆU PHẦN MỀM QUẢN LÝ DỮ LIỆU VI MÔ

**I. Giới thiệu sơ lược và các ưu điểm của phần mềm quản lý dữ liệu vi mô
(Microdata management)**

Phần mềm quản lý dữ liệu vi mô được thiết kế chuyên dụng cho việc lưu trữ các thông tin về các cuộc điều tra hộ gia đình. Nó không cung cấp các chức năng tổng hợp, phân tích số liệu mà chỉ đơn thuần lưu trữ thông tin.

Phần mềm này cũng cung cấp một số công cụ để người sử dụng có thể dễ dàng chiết suất dữ liệu thô được lưu trong phần mềm ra các định dạng file khác nhau để từ đó sử dụng các chương trình phân tích thích hợp, quen thuộc cho người sử dụng.

Trước đây đã có rất nhiều phần mềm được ứng dụng để lưu trữ thông tin về các cuộc điều tra ở Tổng cục Thống kê. Tuy nhiên các phần mềm này chủ yếu tập trung quản lý số liệu và phân tích số liệu mà chưa hoặc rất hạn chế trong việc lưu trữ các thông tin vi mô liên quan.

Phần mềm này có ưu điểm so với các phần mềm khác là nó cho phép chúng ta lưu trữ tất cả các thông tin liên quan đến cuộc điều tra trong cùng một hệ thống một cách cực kỳ chi tiết, giúp cho người tra cứu dễ dàng tiếp cận được các thông tin này. Ví dụ: cuộc điều tra do ai tiến hành, được thiết kế như thế nào, cỡ mẫu điều tra là bao nhiêu, bảng câu hỏi và các công cụ điều tra khác như thế nào, chỉ đạo điều tra ra sao, nhất là công cụ này mô tả chi tiết từng biến số của bộ số liệu ... giúp cho người sử dụng bộ số liệu có thể hiểu rõ về bộ số liệu, tránh mắc sai lầm khi phân tích.

Microdata Management là một bộ công cụ bao gồm 4 module:

1. Metadata Editor: Đây là module dành cho các nhà quản trị dữ liệu để xây dựng cơ sở dữ liệu vi mô cho các cuộc điều tra, các tài liệu có liên quan đầu vào sẽ được chuẩn hóa theo tiêu chuẩn quốc tế về dữ liệu vi mô (DDI và Dublin Core), các thông tin đầu vào là các thông tin đầy đủ, từ các công văn liên quan, bảng câu hỏi, các tài liệu hướng dẫn, bộ số liệu thô đến các chương trình nhập tin và các chương trình phân tích, các báo cáo kết quả đầu ra v.v. .

2. Nesstar Explorer: là module dành cho người sử dụng cơ sở dữ liệu, công cụ này dùng để đọc các file do *Metadata Editor* tạo ra, cho phép người sử dụng có thể xem được số liệu vi mô và các tài liệu liên quan đến bộ số liệu đồng thời cung cấp chức năng chiết xuất dữ liệu ra các định dạng file phổ biến như Stata, SPSS... tùy theo nhu cầu của người sử dụng.

3. CD-ROM Builder: module này cho phép ta tạo ra các sản phẩm số liệu đầu ra thân thiện với người sử dụng như là đĩa CD-Rom, đĩa DVD, ... người sử dụng có thể khai thác trực tiếp trên đĩa mà không cần cài đặt phần mềm *Metadata Editor* hoặc *Nesstar Explorer*.

4. NADA : Cũng có nhiệm vụ tạo ra sản phẩm đầu ra giống như *CD-Rom Builder*, phần mềm *NADA* là module công cụ tạo nên sản phẩm trên các trang Web. Nhà quản trị có thể lựa chọn các quyền truy cập và sử dụng số liệu.

II. Module quản trị Metadata Editor và ứng dụng

Hiện nay, Việt Nam cũng như trên thế giới công nghệ thông tin đã đi sâu vào trong đời sống nên hầu hết các công việc đã được điện tử hóa. Tại Tổng cục Thống kê các tài liệu điều tra cũng được điện tử hóa gần như toàn bộ.

Metadata Editor cho phép người quản trị dữ liệu đưa các dữ liệu ở các dạng khác nhau vào và đồng bộ hóa trong hệ thống này. Các tài liệu đó được chuẩn hóa theo một chuẩn chung Quốc tế có định dạng Data Documentation Initiative và Dublin Core – chuẩn về siêu dữ liệu hiện nay trên thế giới.

1. Đưa một bộ dữ liệu của một cuộc điều tra vào lưu trữ trong Microdata Management

Toàn bộ cơ sở dữ liệu và các siêu dữ liệu đầu vào liên quan đến cuộc điều tra sẽ được ghi lại trong một file có phần mở rộng là .nesstar. *Metadata Editor* gọi đây là một Project, hay 1 cơ sở dữ liệu.

Mỗi một Project lưu trữ thông tin của một bộ dữ liệu của một cuộc điều tra.

2. Đưa các thông tin chung về một cuộc điều tra vào cơ sở dữ liệu

Khi thiết kế một cơ sở dữ liệu về một cuộc điều tra nào đó, Metadata Editor yêu cầu khai báo các thông tin chung nhất mô tả về cuộc điều tra đó. Từ việc thiết kế điều tra, lên kế hoạch, xây dựng phương án điều tra, thiết kế bảng hỏi, điều tra thử, tập huấn, chọn mẫu, thu thập, xử lý số liệu đến khâu phân tích và ra kết quả cuối cùng.

Việc mô tả thông tin chung này, giúp cho người sử dụng dữ liệu có thể hiểu sâu hơn về cuộc điều tra, về chọn mẫu, sai số chọn mẫu, về tính đại diện để có thể sử dụng bộ số liệu một cách hiệu quả nhất.

3. Đưa các văn bản, tài liệu khác liên quan vào quản lý chung với database

Điểm khác biệt nhất so với các phần mềm lưu trữ số liệu và dữ liệu vi mô trước đây của Metadata Editor là nó cho phép chúng ta lưu trữ và quản lý các tài liệu liên quan dạng siêu dữ liệu trong cùng một thư mục và việc tra cứu được thể hiện ở dạng đường dẫn tới các file điện tử liên quan.

Các phần mềm lưu trữ dữ liệu vi mô trước đây chủ yếu quan tâm đến việc lưu trữ thông tin của bộ dữ liệu thô, chưa chú trọng đến dữ liệu vi mô.

Toàn bộ các tài liệu liên quan đến cuộc điều tra khác chưa được đưa vào lưu trữ trong phần Metadata Editor ở trên sẽ được đưa vào quản lý chung trong một thư mục với cơ sở dữ liệu chính và trong mỗi tiêu đề mục quản lý dữ liệu đều đưa đường dẫn để mở file nguồn.

III. Nesstar Explorer xem và khai thác cơ sở dữ liệu

Module Nesstar Explorer cho phép người sử dụng xem và khai thác các dữ liệu và siêu dữ liệu về định dạng thống kê chung được thiết lập từ module Metadata Editor.

Dữ liệu được đưa vào để sử dụng trong Nesstar explorer rất đa dạng, cốt lõi là bộ số liệu thô mà dựa vào các chức năng chiết suất dữ liệu trong module này người sử dụng có thể xuất dữ liệu thô của các cuộc điều tra đã được chuẩn hóa theo nesstar ra các định dạng file ưa thích để từ đó có thể khai thác sâu hơn. Ví dụ, người dùng tin muốn sử dụng chương trình SPSS để tính toán một chỉ tiêu nào đó trong Tổng điều tra Nông thôn, nông nghiệp và thủy sản, họ chỉ cần xuất số liệu thô từ Nesstar explorer ra định dạng file của SPSS.

1. Xem database

Phần mềm này có giao diện hiển thị dưới dạng 2 khung, khung bên trái là cây thư mục thiết kế giống như chuẩn Metadata editor sử dụng để tạo ra nesstar file, khung bên phải là phần hiển thị nội dung chi tiết các mục đã lựa chọn ở khung bên trái

2. Chiết suất/export data

Chúng ta không thể sử dụng các phần mềm phân tích thống kê thông thường như SPSS, STATA hoặc SAS để phân tích dữ liệu đang được lưu trữ dưới dạng nesstar file. Tuy nhiên Nesstar Explorer cho phép người sử dụng chiết suất ra thành các định dạng file phù hợp với SPSS, STATA hoặc SAS để sau đó sử dụng các phần mềm phân tích database đó.

3. Chiết suất/export metadata

Nesstar Explorer cho phép người sử dụng chiết suất các văn bản/tài liệu liên quan sang các định dạng siêu văn bản hoặc định dạng phù hợp để sử dụng

IV. Sử dụng CD-ROM Builder để tạo đĩa CD

CD-ROM Builder là một gói công cụ, nó xuất dữ liệu ra định dạng HTML cơ bản có thể lưu trữ trên CD-ROM, DVD, mạng intranet hay mạng internet. Mặc dù nó được thiết kế chủ yếu là công cụ để phổ biến vi dữ liệu, nó cũng còn được sử dụng lưu trữ dữ liệu và siêu dữ liệu.

Việc tạo đĩa CD hầu như được tiến hành 1 cách tự động. Đầu vào duy nhất cần thiết là bộ dữ liệu ở định dạng của Nessta và lựa chọn tên chủ đề. Các tùy chọn được cung cấp để chỉnh sửa nội dung và mẫu thiết kế của sản phẩm đầu ra.

Một file Nessta là file kết quả lưu trữ bộ dữ liệu sử dụng Metadata Editor. Nhãn hiệu bao gồm những tựa đề lớn được thiết kế tùy biến sẽ xuất hiện trên đầu mỗi trang của CD-ROM. Nhãn hiệu có thể bao gồm logo, tên và các phần tử đặc trưng khác mà nó đại diện.

PHẦN II

THỰC TRẠNG CÔNG TÁC QUẢN LÝ DỮ LIỆU VI MÔ CỦA MỘT SỐ CUỘC ĐIỀU TRA TRONG TỔNG CỤC THỐNG KÊ

Ngành Thống kê có hệ thống cơ sở hạ tầng về CNTT rất tốt. Các máy tính có cấu hình mạnh, đã có mạng LAN tại TCTK và mạng WAN toàn ngành. Phần mềm hệ thống được sử dụng của Microsoft.

Đường truyền Internet tốc độ cao: Tổng cục Thống kê có 2 đường truyền internet tốc độ cao đó là: đường cáp quang và đường SHSL 2M được thuê từ nhà cung cấp dịch vụ Bưu điện Trung ương.

Phần mềm hệ thống: Máy chủ : Microsoft Windows Server 2003 64bit, máy trạm: Microsoft Windows XP và Microsoft Windows Vista, phần mềm văn phòng: Microsoft Office 2003 và Microsoft Office XP, phần mềm diệt virus: Symantec Antivirus 10.1 Net, hệ thống thư tín điện tử (E-Mail): Mdaemon, hệ thống truyền và nhận file (Ftp): ServerU và Ftp của Microsoft

I. Thực trạng ứng dụng CNTT trong ngành Thống kê

1. Ứng dụng công nghệ thông tin trong việc truyền số liệu

Việc tin học hoá công tác xử lý và truyền đưa số liệu của một số lĩnh vực công tác thống kê được duy trì, hoàn thiện và phát triển đã góp phần nâng cao hiệu quả công tác, chất lượng số liệu và rút ngắn thời gian báo cáo. Mạng GSO phát huy tác dụng trong việc truyền đưa số liệu, phổ biến nhanh các văn bản, tài liệu hướng dẫn của Tổng cục đến các địa phương. Thư tín điện tử đã được sử dụng, tạo điều kiện để giải quyết nhanh và hiệu quả các công việc. Trang WEB thống kê được duy trì trên mạng intranet, tiến hành công tác quản trị mạng và cập nhật thông tin trên trang WEB.

2. Ứng dụng công nghệ thông tin trong xử lý số liệu điều tra

- Xử lý số liệu Tổng điều tra nông thôn, nông nghiệp với 12,4 triệu phiếu điều tra từ 12,4 triệu hộ gia đình .
- Xử lý số liệu Tổng điều tra các cơ sở kinh tế - hành chính sự nghiệp cho 2,3 triệu đơn vị kinh tế - hành chính, sự nghiệp.
- Xử lý số liệu chọn mẫu nhân khẩu học nhiều vòng với trên 1,5 triệu nhân khẩu và 200 ngàn hộ gia đình. Xử lý số liệu điều tra chọn mẫu hộ gia đình đa mục tiêu với 40 ngàn hộ gia đình được chọn mẫu cho mỗi năm.
- Xử lý số liệu điều tra Lao động và Việc làm với khoảng 120 ngàn hộ gia đình chọn mẫu cho mỗi năm.
- Xử lý số liệu Tổng điều tra Dân số và Nhà ở năm 1999 với gần 17 triệu hộ gia đình và trên 76 triệu nhân khẩu
- Và xử lý số liệu rất nhiều cuộc điều tra khác...

3. Ứng dụng công nghệ thông tin trong xây dựng cơ sở dữ liệu và phổ biến số liệu

Ứng dụng công nghệ thông tin trong xử lý và lưu giữ số liệu các cuộc điều tra: đã thí điểm xây dựng CSDL cho 4 phân hệ sau: thống kê giá cả, thống kê công nghiệp, thống kê thương mại, thống kê nông nghiệp. Trong năm 2000, các CSDL vi mô TĐTDS và Nhà ở cũng đã được xây dựng để lưu giữ dữ liệu gốc 3%, lưu giữ dữ liệu gốc 100% và khai thác qua hệ quản trị CSDL quan hệ SQL Server. Hoàn thành việc xử lý và xây dựng cơ sở dữ liệu Tổng điều tra nông thôn, nông nghiệp và thủy sản. Cùng với báo cáo tổng hợp được xuất bản dưới dạng ấn phẩm, đã sản xuất các đĩa CD-ROM về kết quả tổng điều tra của cả nước và của các địa phương.

a. Số liệu Tổng điều tra Dân số và Nhà ở năm 1999

Tổng điều tra Dân số và nhà ở (TĐTDS&NO) 1-4-1999 thu thập số liệu của gần 17 triệu hộ với trên 76 triệu nhân khẩu. Phiếu điều tra gồm 26 câu hỏi cho phiếu dài (điều tra mẫu) và 18 câu hỏi cho phiếu ngắn (điều tra toàn bộ).

Việc xử lý số liệu được thực hiện trên các hệ thống máy tính PC nối mạng nội bộ. Chương trình nhập, kiểm tra và sửa dữ liệu được phát triển bằng phần mềm ISSA, chương trình hiệu đính và tổng hợp phát triển trong phần mềm IMPS (CONCOR và CENTS). Các chương trình này được tích hợp chung trong một hệ thống thống nhất quản lý toàn bộ luồng dữ liệu, công việc, người dùng. Dữ liệu vi mô là các tệp ký tự ASCII với nhiều loại bản ghi trong cùng một tệp. Hiện tại, các tệp dữ liệu này vẫn được tiếp tục lưu giữ và sử dụng để tổng hợp các biểu, chỉ tiêu đặc biệt khi có các yêu cầu đột xuất của ngành. Dung lượng tệp dữ liệu vi mô dạng ký tự ASCII của điều tra mẫu 3% là 162 MB, còn điều tra toàn bộ là 4860 MB.

Dữ liệu vi mô toàn quốc cũng được chuyển đổi và quản lý trong cơ sở dữ liệu mô hình quan hệ trong hệ quản trị CSDL SQL Server, gồm một CSDL vi mô mẫu 3% và một CSDL vi mô toàn bộ 100%. Cơ sở dữ liệu bao gồm dữ liệu và siêu dữ liệu về cấu trúc, tên biến, các danh mục liên quan cùng với những hướng dẫn về việc chiết xuất, khai thác dữ liệu. CSDL hiện đang lưu giữ tại Trung tâm tích hợp dữ liệu TCTK và cho phép truy cập trong mạng nội bộ của TCTK.

b. Số liệu Tổng điều tra Nông thôn, Nông nghiệp và thủy sản năm 2001

TĐTNTNN&TS có thể coi là một nguồn dữ liệu phong phú và chi tiết cho thống kê xã hội với những thông tin về cơ sở hạ tầng kinh tế xã hội khu vực nông thôn, điều kiện môi trường sống, mức sống, lao động nông nghiệp.

Dữ liệu vi mô và vĩ mô của hai cuộc TĐTNTNN&TS các năm 2001 và 2006 được lưu giữ đầy đủ với nhiều dạng khuôn mẫu, vật lưu tin khác nhau.

Dữ liệu TĐTNTNN&TS năm 2006 cũng được lưu giữ, phân phát dưới các dạng tương tự như năm 2001 như; các tệp dữ liệu gốc FoxPro, các tệp dữ liệu ký tự ASCII, các loại đĩa CD-ROM dữ liệu và kết quả, CSDL dạng SQL Server quản lý tại TTTHDL TCTK. Khác biệt chính giữa dữ liệu của hai cuộc tổng điều tra này là: nếu năm 2001 các sản phẩm CD-ROM và CSDL vi mô chỉ được phát triển cho hai loại phiếu hộ và xã thì năm 2006, dữ liệu vi mô của tất cả các loại phiếu điều tra đều được xây dựng CSDL SQL Server và đĩa CD-ROM. Ngoài ra, trong TĐTNTNN&TS năm 2006, công cụ khai thác dữ liệu trong môi trường web còn được phát triển tạo điều kiện thuận tiện cho người dùng tạo lập các thống kê từ nguồn dữ liệu này.

c. Số liệu Tổng điều tra Cơ sở kinh tế - hành chính sự nghiệp năm 2002

Cuộc Tổng điều tra các cơ sở kinh tế, hành chính, sự nghiệp 2002 đã được Tổng cục Thống kê tiến hành trên phạm vi cả nước. Đối tượng của Tổng điều tra là toàn bộ các cơ sở đang hoạt động tại thời điểm 1/7/2002 ở cả hai lĩnh vực: sản xuất kinh doanh và hành chính sự nghiệp. Kết quả điều tra có 10.793.376 lao động làm việc trong 2.880.004 cơ sở.

d. Số liệu điều tra Biến động dân số và Kế hoạch hóa gia đình hàng năm

Các chương trình nhập tin và kiểm tra, hiệu đính số liệu được thiết kế bằng phần mềm ISSA (năm 2000 và 2001) và sau đó chuyển sang phần mềm CSPro, còn phần mềm IMPS (CENTS) được sử dụng để lập trình tính toán các biểu tổng hợp. Dữ liệu vi mô là các tệp ký tự ASCII “không phẳng”, nghĩa là một tệp dữ liệu chứa tất cả các loại bản ghi (tương tự như cấu trúc của dữ liệu vi mô TĐTDS 1989), được lưu trữ theo từng năm, mỗi năm một tệp gồm nhiều loại bản ghi. Dung lượng dữ liệu mỗi năm trên dưới 100 MB. Loại dữ liệu vẫn lưu trữ tại TTTH Thống kê và tiếp tục phục vụ cho các nhu cầu lập biểu tổng hợp đặc biệt khi có yêu cầu phát sinh.

Hàng năm, sau khi xử lý xong điều tra BDDS&KHHGD, dữ liệu vi mô đều được chuyển đổi sang dạng SPSS và gửi đến Vụ TK DSLD và các CTK tỉnh, thành phố. Số liệu vĩ mô với các biểu tổng hợp các cấp dưới dạng EXCEL cũng được lưu giữ tại TTHTK đồng thời gửi cho Vụ TK DS LĐ và các CTK. Loại dữ liệu này đáp ứng được yêu cầu kiểm tra số liệu và kết quả, tổng hợp, phân tích theo các nhu cầu đặc biệt của các cán bộ thống kê tại Vụ DS-LĐ và các Cục.

Ngoài các dạng dữ liệu trên, riêng các năm 2005 và 2006, số liệu vi mô và vĩ mô điều tra BDDS&KHHGD còn được thiết kế trong CSDL mô hình

quan hệ (SQL Server). Ngoài các dữ liệu vi mô và các biểu kết quả tổng hợp, CSDL bao gồm cả siêu dữ liệu như: mô tả dữ liệu, cấu trúc dữ liệu, sơ đồ quan hệ, tất cả các danh mục, phân loại, các siêu dữ liệu tham chiếu (tài liệu liên quan: sổ tay điều tra viên, phương án, mẫu phiếu,...). Các công cụ khai thác dữ liệu hướng người sử dụng cũng được thiết kế dưới dạng khai thác trong giao diện Web cung cấp các chức năng tìm kiếm dữ liệu, trích xuất dữ liệu sang các dạng khác nhau, lập biểu phân tổ tần xuất, lập bảng tổng hợp chéo tùy chọn (cross-tabulation), các thuộc tính áp dụng hệ số gia quyền/suy rộng hoặc không gia quyền, v.v... Đây có thể coi như những thử nghiệm tốt để xây dựng các CSDL tích hợp chung phục vụ nhu cầu khai thác của người dùng tin trên bước đường đi tới xây dựng kho dữ liệu của TCTK.

e. Số liệu Khảo sát Mức sống hộ gia đình

Điều tra năm 2002 được xử lý bằng phần mềm FoxPro và do vậy, dữ liệu vi mô cũng dưới dạng CSDL FoxPro. Từ năm 2004 (2004, 2006 và 2008) phần mềm xử lý được chọn là CPro và dữ liệu là các tệp ký tự ASCII. Các tệp ASCII này là tệp ký tự “không phẳng”, có cấu trúc gồm rất nhiều bản ghi khác nhau.


Dữ liệu vi mô của mỗi điều tra đều được chuyển đổi dưới dạng Stata phục vụ cho việc kiểm tra, tổng hợp chi tiết của vụ Thống kê Xã hội Môi trường và các Cục Thống kê. Các biểu tổng hợp các cấp dưới dạng EXCEL cũng được cung cấp cho những người dùng khác nhau và lưu giữ tại Trung tâm Tin học Thống kê và vụ Thống kê Xã hội Môi trường.

PHẦN III

THỬ NGHIỆM PHẦN MỀM QUẢN LÝ DỮ LIỆU VI MÔ VÀO LƯU TRỮ, QUẢN LÝ DỮ LIỆU CỦA KHẢO SÁT MỨC SỐNG HỘ GIA ĐÌNH NĂM 2006 VÀ TỔNG ĐIỀU TRA NÔNG THÔN, NÔNG NGHIỆP VÀ THỦY SẢN NĂM 2006

I. Tổng điều tra Nông thôn, Nông nghiệp và Thủy sản năm 2006

Ứng dụng: add group để tạo ra một nhóm với tên Agricultural data, dưới nhóm ta tạo ra một nghiên cứu với tên AC2006_V1.0 bằng cách kích chuột vào biểu tượng add new study. Trong nghiên cứu AC2006_V1.0 sẽ bao gồm các thông tin về mô tả dữ liệu, mô tả nghiên cứu, đưa số liệu vào lưu trữ, nhóm các biến (các chỉ tiêu đã thu thập), và các nguồn tài liệu có liên quan khác. Nghiên cứu này sẽ lưu trữ tất cả các thông tin liên quan đến cuộc tổng điều tra Nông thôn, nông nghiệp và thủy sản 2006.

Đưa số liệu vào lưu trữ: Để đưa số liệu tổng điều tra vào lưu trữ bằng phần mềm quản lý cơ sở dữ liệu vi mô Import dataset  Trong phần áp dụng này vì cơ sở dữ liệu của phiếu số 1 (phiếu điều tra hộ) rất lớn nên chúng tôi chỉ đưa 33% dữ liệu gốc vào lưu trữ bằng công cụ này.

II. Khảo sát Mức sống hộ gia đình Việt Nam năm 2006

Sử dụng Metadata Editor để tạo lập và quản lý cơ sở dữ liệu vi mô Khảo sát mức sống 2006.

Tương tự như cách tạo lập nhóm của Tổng điều tra Nông thôn, Nông nghiệp và Thủy sản 2006, chúng ta tạo ra cơ sở dữ liệu cho Khảo sát Mức sống hộ gia đình Việt Nam 2006.

Khai báo về cuộc điều tra:

Trong mỗi Project/nghiên cứu, đều có chuẩn các mục cần khai báo cho mỗi cuộc điều tra, phần này sẽ giúp người dùng hiểu được nội dung cuộc điều tra. Để khai báo các tài liệu của cuộc KSMS 2006, trước hết ta phải chuẩn bị hết các tài liệu có liên quan đến cuộc điều tra như: bộ số liệu, quyết định, phương án, biểu mẫu, phiếu điều tra, sổ tay hướng dẫn điều tra viên, các chỉ tiêu điều tra, qui trình thực hiện điều tra và các tài liệu khác có liên quan. Khi đã có đầy đủ các thông tin liên quan đến cuộc điều tra ta sẽ thực hiện các bước khai báo theo chuẩn của phần mềm theo các chủ đề với nội dung như sau:

Đưa các văn bản, tài liệu liên quan vào quản lý chung với database:

Metadata Editor cho phép quản lý các tài liệu liên quan (dưới dạng siêu dữ liệu như: Quyết định, phương án, bảng hỏi, phần mềm nhập tin, biểu đầu ra, sách công bố kết quả) chung với database.

III. Các khó khăn gặp khi ứng dụng phần mềm microdata management

Phần mềm này đã được xây dựng và ứng dụng ở một số nước trên thế giới để lưu trữ và quản lý dữ liệu của các cuộc điều tra hộ gia đình. Tuy nhiên, ở Việt Nam chúng ta đang ở giai đoạn bắt đầu ứng dụng. Vì vậy việc hỗ trợ phiên bản tiếng Việt chưa được thực hiện nên sẽ khó khăn cho người sử dụng nếu không biết tiếng Anh.

Cũng vì lý do đó khả năng phổ biến rộng rãi trong nước qua website của Tổng cục Thống kê đối với đông đảo người dùng tin chắc chắn cũng sẽ giảm đi rất nhiều.

Do công tác lưu trữ thông tin của một số cuộc điều tra trước đây không được đầy đủ nên có thể việc ứng dụng phần mềm vào lưu trữ và quản lý thông tin vi mô của các cuộc điều tra đó ít nhiều cũng sẽ bị ảnh hưởng bởi

tính đầy đủ của các thông tin liên quan đến cuộc điều tra. Có thể một số cuộc điều tra trong quá khứ của Tổng cục Thống kê chỉ còn lưu trữ thông tin dạng bản in nên sẽ rất tốn thời gian để có thể chuyển sang file điện tử để lưu trữ.

KẾT LUẬN VÀ KIẾN NGHỊ

Qua cách tiếp cận phần mềm quản lý dữ liệu vi mô vào việc lưu trữ số liệu của các cuộc điều tra, chúng ta đã thấy được tính ưu việt của phần mềm này. Theo đánh giá của chúng tôi, đây là phần mềm rất hữu ích giúp hệ thống hoá, chuẩn hóa các tài liệu liên quan theo tiêu chuẩn quốc tế về dữ liệu vi mô và lưu trữ được tất cả các thông tin liên quan đến các cuộc điều tra, tổng điều tra. Đồng thời với công cụ này có thể tạo ra được những báo cáo phân tích đầy đủ về nội dung đã được khai báo. Với cách lưu trữ bằng phần mềm này, chúng ta có thể giảm được nhiều chi phí trong việc lưu trữ và bảo quản tài liệu của các cuộc điều tra, tổng điều tra.

Qua việc thử nghiệm phần mềm quản lý dữ liệu vi mô vào việc lưu trữ và quản lý dữ liệu của Tổng điều tra Nông thôn, Nông nghiệp và Thủy sản năm 2006, và tham khảo hiệu quả thử nghiệm phần mềm này vào việc lưu trữ các cuộc điều tra của các Vụ khác như Vụ Xã hội và Môi trường, Vụ Thống kê Dân số và Lao động chúng tôi xin có một số những đánh giá của phần mềm này vào việc lưu trữ thông tin các cuộc điều tra của Tổng cục Thống kê như sau:

- Hiện tại phần mềm chưa có phiên bản tiếng Việt, các thông tin khai báo cũng không chấp nhận tiếng Việt nên rất khó cho những người không biết tiếng Anh khi sử dụng phần mềm, mặt khác do sử dụng tiếng Anh để khai báo nên sản phẩm cũng chỉ phù hợp cho các tổ chức quốc tế, đơn vị nghiên cứu trong và ngoài nước có sử dụng tiếng Anh. Tuy nhiên, đầu tháng 12 vừa qua ESCAP đã cử một chuyên gia sang công tác tại Việt Nam để tiếp tục hỗ trợ kỹ thuật trong việc tự động hoá các cuộc điều tra và tổng điều tra và đã cam kết phần mềm phiên bản tiếng Việt sẽ đưa vào sử dụng được trong 6 tháng tới.

- Phần mềm phù hợp với những cuộc điều tra mẫu hơn, còn đối với các cuộc tổng điều tra lớn, ví dụ các cuộc tổng điều tra với dung lượng file dữ liệu quá lớn thì phải chia nhỏ theo từng tỉnh hoặc theo mục trong điều tra. Việc chia nhỏ này sẽ phức tạp hơn một chút trong việc quản lý dữ liệu. Tuy nhiên, các cuộc điều tra này việc tổng hợp và lưu trữ số liệu được thực hiện bởi Trung Tâm tin học Thống kê, Bởi vậy, những khó khăn phức tạp không phải là vấn đề lớn.

- Phần mềm này có thể áp dụng được rộng rãi với những cuộc điều tra mà thông tin không cần phải bảo mật. Với việc lưu trữ bằng công cụ quản lý dữ liệu vi mô sẽ dần hình thành nên chính sách phổ biến thông tin thân thiện hơn với các đối tượng dùng tin. Nhu cầu sử dụng thông tin ngày càng phát triển, ngày càng nhiều cá nhân, tổ chức có nhu cầu sử dụng số liệu thống kê phục vụ cho công tác nghiên cứu, lập kế hoạch... Việc dễ dàng tiếp cận được với dữ liệu gốc của các cuộc điều tra sẽ giúp người dùng tin chủ động đưa ra các bảng biểu số liệu theo yêu cầu riêng mà việc này rất khó làm với các phần mềm lưu trữ khác.

- Khi có phiên bản tiếng Việt, việc áp dụng phần mềm cho các cuộc điều tra mẫu là rất cần thiết. Hàng năm, Tổng Cục Thống kê có rất nhiều các cuộc điều tra mẫu, riêng vụ Thống kê Nông, Lâm nghiệp và Thủy Sản có đến hàng chục cuộc điều tra mẫu hàng năm. Số liệu thô của các cuộc điều tra này không được lưu trữ một cách bài bản, có hệ thống nên bị mất, lưu lạc rất nhiều dẫn đến không có căn cứ để tham khảo, để so sánh giữa cuộc điều tra. Nếu phần mềm này được triển khai và áp dụng cho tất cả các cuộc điều tra mẫu sẽ rất có lợi trong việc lưu trữ, lập báo cáo phân tích và đặc biệt, các địa phương có thể học hỏi được nhau thông qua việc truy cập vào xem kết quả dữ liệu thô của các địa phương khác.

Việc áp dụng phần mềm này vào việc lưu trữ và quản lý dữ liệu của các cuộc điều tra và tổng điều tra khá thuận lợi do có sự hỗ trợ kỹ thuật của dự án ESCAP. Để triển khai áp dụng phần mềm này rộng rãi trong ngành thống kê cần thiết phải có sự ủng hộ của các đồng chí lãnh đạo Tổng Cục Thống kê cũng như các đồng chí lãnh đạo các Vụ nghiệp vụ trong việc triển khai dự án ESCAP để việc tiếp nhận sự hỗ trợ về mặt kỹ thuật trong việc tư liệu hoá số liệu các cuộc điều tra diễn ra thuận lợi để việc đào tạo đội ngũ làm công tác lưu trữ dữ liệu diễn ra nhanh chóng và hiệu quả. Như vậy việc áp dụng phần mềm này sẽ nhanh chóng được phổ biến rộng rãi và áp dụng trong toàn ngành.

TÀI LIỆU THAM KHẢO

Trong quá trình nghiên cứu ứng dụng phần mềm Microdata Management vào quản lý dữ liệu vi mô của 2 cuộc điều tra: Tổng điều tra Nông thôn, Nông nghiệp và Thủy sản; Khảo sát Mức sống hộ gia đình Việt Nam trong năm 2006 chúng tôi có sử dụng các tài liệu sau:

1. Sổ tay hướng dẫn điều tra viên Khảo sát Mức sống hộ gia đình 2006
2. Sổ tay hướng dẫn Tổng điều tra Nông thôn, Nông nghiệp và thủy sản 2006.

3. Kết quả Khảo sát Mức sống hộ gia đình 2006.
4. Báo cáo kết quả Tổng điều tra Nông thôn, Nông nghiệp và thủy sản 2006.
5. Phương án điều tra Khảo sát Mức sống hộ gia đình và Tổng điều tra Nông thôn, Nông nghiệp và thủy sản 2006.
6. Bảng câu hỏi Khảo sát Mức sống hộ gia đình và Tổng điều tra Nông thôn, Nông nghiệp và thủy sản 2006.
7. Getting_Started_v3.5sp1
8. Metadata Editor UsersGuide_EN
9. Nesstar_server_3_50_administrator_guide