

**ĐỀ TÀI KHOA HỌC**  
**SỐ: 2.2.11-CS07**

**NGHIÊN CỨU VÀ THỬ NGHIỆM TIN HỌC HÓA**  
**KHO TƯ LIỆU KHOA HỌC THỐNG KÊ**

1. Cấp đề tài : Cơ sở
2. Thời gian nghiên cứu : 2007
3. Đơn vị chủ trì : Viện Khoa học Thống kê
4. Đơn vị quản lý : Viện Khoa học Thống kê
5. Chủ nhiệm đề tài : CN. Hoàng Minh Thiện
6. Những người phối hợp nghiên cứu:
  - TS. Nguyễn Thị Đông
  - CN. Nguyễn Thái Hà
  - CN. Trần Mạnh Hùng
  - Ths. Đỗ Văn Huân
  - CN. Đặng Thu Bình
7. Điểm đánh giá nghiệm thu đề tài: 9

## **ĐẶT VẤN ĐỀ**

Viện khoa học thống kê có vai trò đầu mối trong phát triển các kết quả nghiên cứu vào thực tế ứng dụng của Ngành, cho nên vai trò của công tác thông tin khoa học đang trở nên thiết thực với các hoạt động tác nghiệp cũng như quản lý, đào tạo và nghiên cứu! Rõ ràng hiện nay công tác tư liệu khoa học thống kê còn một số bất cập, tập trung vào mấy điểm chính sau: chưa đa dạng hóa các tư liệu cung cấp, chưa đưa tận tay người sử dụng, nguồn tư liệu chưa đầy đủ và hệ thống, công cụ xử lý tư liệu chưa thống nhất và hiện đại... Chính vì vậy, nhu cầu củng cố và áp dụng các kỹ thuật thông tin mới vào công tác tư liệu khoa học thống kê, cộng với nhu cầu sử dụng thông tin khoa học thống kê của người sử dụng cuối cùng đang thay đổi và tăng lên, đã đòi hỏi cần có cải tiến trong công tác này tại Viện Khoa học thống kê;

Đề tài “*Nghiên cứu và thử nghiệm tin học hóa kho tư liệu khoa học thống kê*” là một đề tài cấp cơ sở nhằm vào giải quyết các vấn đề cơ bản nhất của công tác tư liệu khoa học thống kê của ngành; trước hết đánh giá đúng đắn thực trạng kho tư liệu hiện có, từ nguồn hình thành tư liệu, khả năng tiếp tục bổ sung, hiện trạng của công tác xử lý và lưu trữ, các ấn phẩm đang có từ xử lý nguồn tư liệu, các công cụ xử lý tư liệu đang sử dụng...cho đến so sánh thực trạng trên với đòi hỏi của người dùng tin, cũng như so với tình hình trong nước và thế giới. Sau khi tham khảo các ứng dụng tương tự trong nước và thế giới, sẽ tìm kiếm những phương pháp ứng dụng mới trong tất cả các khâu của công tác tư liệu, chủ yếu là các phương pháp tin học hóa, hệ thống hóa.

Trên đây là khái quát về mục đích, yêu cầu và nội dung của đề tài, sau đây là các kết quả chính được và hệ thống lại từ các chuyên đề của các thành viên tham gia; kết quả này là nỗ lực của tất cả các thành viên tham gia đề tài, thay mặt ban chủ nhiệm đề tài chúng tôi cảm ơn sự đóng góp ý kiến và tham gia nhận xét sâu sắc của các đồng nghiệp liên quan.

### **I. THỰC TRẠNG CÔNG TÁC TƯ LIỆU KHOA HỌC THỐNG KÊ**

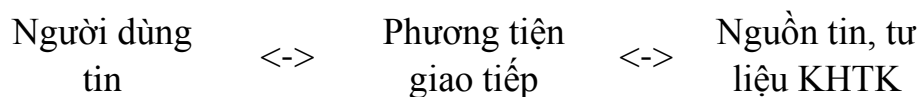
#### **1. Tổng quan về công tác tư liệu**

Như trong phần trên đã nói mục tiêu chính của đề tài đặt ra là vấn đề củng cố kho thông tin tư liệu KHTK thống kê nhằm đáp ứng nhu cầu thông tin khoa học cho người sử dụng cuối cùng là các cán bộ nghiên cứu, quản lý và tác nghiệp nên hiển nhiên là chúng ta phải đặt ra các câu hỏi sau đây:

- Người dùng tin cần những thông tin tư liệu gì?
- Chúng ta đã có những thông tin tư liệu gì?

- Và cuối cùng là bằng những cách thức nào đưa thông tin tư liệu đến tay người dùng tin?

Xuất phát từ ba câu hỏi cơ bản trên chúng ta đã có ngay một sơ đồ tổng quát như sau:



Xuất phát từ người dùng tin ta thấy các nhu cầu cơ bản như sau:

- Sách khoa học, sách tra cứu...về lĩnh vực nghiên cứu, tác nghiệp thường được đáp ứng bởi hoạt động thư viện truyền thống,

- Tư liệu về lĩnh vực nghiên cứu, tác nghiệp như những báo cáo tác nghiệp, những kết quả nghiên cứu đã đạt được, hay đã ứng dụng của đồng nghiệp, trong nước và thế giới,

- Đội ngũ các chuyên gia, các tổ chức hoạt động trong lĩnh vực nghiên cứu, tác nghiệp cần cho tìm hiểu hay hợp tác,

- Hành lang pháp lý cho hoạt động nghiên cứu, tác nghiệp đảm bảo cho hoạt động nghiên cứu, tác nghiệp không vi phạm pháp luật,

- Thông tin về tư liệu, về hoạt động thực tế của lĩnh vực để người dùng tin tự tìm kiếm hay tra cứu,

- Các số liệu cơ bản cần cho quá trình nghiên cứu, tác nghiệp,

- ...

Xuất phát từ nguồn tư liệu thì chúng ta đang có gì ?

- Hiện nay TTTT KHTK có một thư viện gần 2000 đầu sách, sách tra cứu cả bằng tiếng Việt và ngoại ngữ,

- Về tư liệu đã lưu trữ nhiều báo cáo kết quả các đề tài, chuyên đề nghiên cứu; nhiều tổng luận, kỷ yếu các hội nghị khoa học, tư liệu dịch thuật...

- TTTT KHTK có hai ấn phẩm thường kỳ là : Tờ Thông tin KHTK và phụ bản là tờ Thông tin Phục vụ quản lý thống kê. Thường niên có các thông báo thư mục, kết quả nghiên cứu,...

- Có nhiều tài liệu tác nghiệp như báo cáo hành chính các cuộc điều tra thống kê, báo cáo phân tích điều tra...

- Có nhiều báo, tạp chí trao đổi với các đơn vị khác như Bộ kế hoạch đầu tư, Tài chính, lao động...

- Tư liệu khai thác trên mạng, tài liệu hội nghị, hội thảo, khảo sát,...

- ...

Xét tới phương cách giao tiếp giữa người dùng tin và nguồn tin sẵn có thì vẫn là các giải pháp truyền thống:

- Người dùng tin đến trực tiếp sử dụng như dịch vụ của thư viện truyền thống,

- Một số ấn phẩm thông tin được gửi tới người dùng tin: hai tờ thông tin, thông báo thư mục, kỷ yếu ...

Qua phân tích cơ bản như trên có thể thấy hoạt động thông tin tư liệu KHTK hiện nay còn ở dạng giản đơn, tối thiểu và mang nặng tính bao cấp thể hiện qua những điểm sau:

- Nguồn tin không đáp ứng đủ và kịp thời cho nhu cầu người sử dụng,

- Phương cách giao tiếp thụ động,

- Chưa kể chất lượng và sự đa dạng của nguồn tin thì nhiều nguồn tư liệu đã có cũng không đến được tay người dùng tin.

Từ đây, chúng ta thấy những yêu cầu đặt ra phải giải quyết là:

- Bổ sung, đa dạng hóa và nâng cao chất lượng nguồn thông tin, tư liệu KHTK,

- Sử dụng các giải pháp của CNTT thay đổi phương cách xử lý, lưu trữ và cung cấp thông tin, tư liệu cho người dùng tin,

- Ứng dụng CNTT để đạt được hai yêu cầu trên.

Việc đánh giá thực trạng của các loại hình tư liệu dưới đây sẽ giúp chúng ta nhìn lại toàn cảnh kho tư liệu KHTK và cũng sẽ là các đánh giá rút kinh nghiệm thực tế về công tác này, cuối cùng các khảo sát này là tiền đề cho việc phát triển các ứng dụng kỹ thuật tin học mới được đưa ra trong phần tiếp theo.

## **2. Thực trạng công tác thư viện khoa học thống kê**

Công tác thư viện khoa học thống kê ở Tổng cục Thống kê do Trung tâm Thông tin Khoa học Thống kê đảm nhận. Thư viện Khoa học Thống kê được giao các chức năng, nhiệm vụ sau: Tìm kiếm, thu thập, xử lý/phân tích, lưu giữ và tổ chức phục vụ thông tin khoa học chuyên ngành Thống kê. Đối tượng phục vụ của Thư viện là đông đảo cán bộ nghiên cứu; cán bộ giảng dạy; sinh viên các trường đại học.

Từ khi ra đời đến cuối những năm 1980, Thư viện còn nhận được sự hỗ trợ có hiệu quả trong bổ sung kho tư liệu từ những nước XHCN. Vốn tài liệu của Thư viện lên tới 6000 bản, không kể báo và tạp chí.

Đến đầu những năm 1990 với sự sáp nhập của Thư viện của Trung tâm Dân số, vốn tư liệu của Thư viện đã lên tới gần 8.000 bản. Nhưng qua nhiều lần chuyển địa điểm hoạt động đã làm thất thoát và hư hỏng một số lượng lớn tài liệu của thư viện. Số tài liệu bằng tiếng Anh chỉ còn lại vài trăm bản và cũ dần qua thời gian.

Chuyển sang cơ chế thị trường, nguồn tài chính đầu tư cho Thư viện bị cắt giảm nhiều, Thư viện chỉ “tồn tại” bằng nguồn kinh phí hết sức rất hạn hẹp bằng nội tệ, việc bổ sung tài liệu ngoại văn hoàn toàn ngừng hẳn vào năm 1989. Do vậy nội dung kho tư liệu trở lên nghèo nàn. Phương tiện hỗ trợ cho công tác tin học hóa hoạt động của Thư viện ít được trang bị.

Cán bộ thư viện chưa được quan tâm đào tạo bồi dưỡng có hệ thống về nghiệp vụ thông tin-thư viện, nghiệp vụ máy tính, ngoại ngữ, ít được thâm nhập thực tế nhằm tìm hiểu nhu cầu tin của người dùng. Trong suốt một thời gian dài việc phục vụ thư viện chỉ do một cán bộ đảm nhận làm mọi khâu từ A đến Z.

Nguyên nhân cơ bản của các tồn tại trên đây là do chúng ta chưa ý thức được đầy đủ vị trí, vai trò và chức năng công tác thư viện để có chiến lược và kế hoạch phát triển cụ thể, đầu tư không đồng bộ, thiếu nhất quán, sử dụng bất hợp lý cán bộ nghiệp vụ.

Trong bối cảnh phát triển mạnh mẽ của công nghệ thông tin, việc tăng cường chức năng phục vụ thông tin của Thư viện Khoa học Thống kê đã trở thành một yêu cầu tất yếu khách quan trong đổi mới công tác thông tin-thư viện của Viện. Mô hình hoạt động thông tin-thư viện khoa học thống kê phải được xây dựng phù hợp với xu thế hiện nay của Việt Nam và thế giới trong lĩnh vực này.

### **3. Yêu cầu khách quan đặt ra đối với việc quản lý các tài liệu nội sinh**

Trong phần này đề cập đến một phần của việc tin học hoá kho tư liệu Khoa học Thống kê để thực hiện nhiệm vụ chức năng của Viện nói chung và phục vụ bạn đọc nói riêng ngày một tốt hơn. Đó là sự cần thiết của việc quản lý các tài liệu nội sinh theo hướng tin học hoá. Ở đây khái niệm tài liệu nội sinh được hiểu là sản phẩm do chính cơ quan tạo ra. Tài liệu nội sinh của Viện Khoa học Thống kê được nói tới là các đề tài của Viện và các Vụ nghiệp vụ của Tổng cục Thống kê thực hiện do Viện quản lý, các chuyên đề, tổng luận chủ yếu do cán bộ của Viện thực hiện.

Hàng năm có một số lượng các đề tài, chuyên đề, tổng luận, được chuyển giao cho thư viện - Trung tâm thông tin Khoa học Thống kê.... Dù là sản phẩm chính thống của Viện nhưng so với sách Việt, sách ngoại văn hiện có tại thư viện cũng chỉ là một phần nhỏ mà chưa được quan tâm quản lý theo cách truyền thống (Xử lý vật lý - logic để tìm kiếm, tra cứu, bổ sung) chứ chưa nói đến xử lý tiền máy để nhập vào Cơ sở dữ liệu (CSDL) với một phần mềm hỗ trợ. Vì thế để phát huy tác dụng của tài liệu này tốt hơn cho ứng dụng, tham khảo ...thì chúng cần được quản lý tốt hơn phù hợp với xu hướng tin học hóa.

#### **4. Thực trạng khai thác tư liệu khoa học thống kê từ mạng Internet**

Trong mấy năm qua, Trung tâm thông tin Viện Khoa học Thống kê đã thực hiện tìm kiếm, khai thác thông tin tư liệu thống kê từ mạng Internet. Việc tìm kiếm thông tin tư liệu bổ sung nguồn thông tin và tư liệu thống kê chuyên ngành chủ yếu bằng tiếng Anh là thực sự có ý nghĩa. Thông qua truy cập các website của các tổ chức thống kê quốc tế, các tổ chức thống kê khu vực và trang web của các nước có nền thống kê tiên tiến đã tìm kiếm bổ sung, cập nhật được những thông tin mới về phương pháp thống kê, số liệu thống kê và hoạt động thống kê.

Nghiên cứu quản lý tư liệu thống kê khai thác từ mạng Internet sẽ tập trung vào một số nội dung về hoạt động khai thác tư liệu, xử lý tư liệu, quản lý tư liệu và hướng tin học hóa tư liệu đã khai thác của Trung tâm thông tin Viện Khoa học thống kê.

##### *- Nguồn tư liệu khoa học thống kê từ mạng Internet*

Nguồn khai thác thông tin chủ yếu từ mạng Internet trên có sở các Website của các tổ chức thống kê quốc tế và website của các cơ quan thống kê quốc gia. Chúng tôi đơn cử đưa ra một số tên tổ chức và địa chỉ truy cập mà các cán bộ của Trung tâm đã trực tiếp truy cập, tìm kiếm và download tư liệu, cụ thể như sau:

Các bảng phân loại thống kê, các tài liệu về phương pháp luận thống kê, các tài liệu nghiệp vụ thống kê thường có nguồn gốc từ các trang web được ví dụ dưới đây:

- Website thống kê của Liên Hợp Quốc (United Nation Statistic Department - UNSD)

*<http://unstats.un.org/unsd/default.htm>*

- Tổ chức hợp tác phát triển kinh tế (Organization for Economic Co-operation and Development - OECD) cổng Statistic Portal

<http://www.oecd.org/statsporta>.

- Viện thống kê quốc tế khu vực Châu Á - Thái Bình Dương: SIAP

[http://www.unsiap.or.jp/link/stat\\_link.htm](http://www.unsiap.or.jp/link/stat_link.htm)

- Trang Web thống kê của Nhật Bản, ...

<http://www.stat.go.jp/english/index.htm>

- .....

Tư liệu khai thác từ Internet rất phong phú thường xuyên được khai thác gồm có:

- Bản tin thống kê của tổ chức các nước hợp tác và phát triển kinh tế “OECD Newsletters”

- Bản tin thống kê của tổ chức các nước khu vực Châu Á - Thái Bình Dương” ESCAP Statistics Newsletters”

- Báo cáo phát triển con người qua trang web thống kê Liên hợp quốc trong nhiều năm.

- Tài liệu hội nghị, hội thảo, đào tạo, sổ tay hướng dẫn nghiệp vụ thống kê chuyên ngành của các nước và các tổ chức quốc tế.

## **5. Tình hình quản lý tờ Thông tin KHTK và một số ấn phẩm điện tử**

Ngày 28/2/1980 Tổng cục trưởng Tổng cục Thống kê có Quyết định số 14- TCTK/QĐ cho phép Viện xuất bản tờ Thông tin Khoa học Thống kê. Từ tháng 2 năm 1980 tờ Thông tin Khoa học Thống kê đã xuất bản 2 số được in Roneo (khoảng 100 bản/số). Thời kỳ 1981-1985 tờ Thông tin Khoa học Thống kê được in ấn phát hành với số lượng từ 200 bản/số đã tăng lên 400 bản/số, đáp ứng được nhu cầu tra cứu và tham khảo của cán bộ nghiệp vụ ở trung ương cũng như địa phương.

Đặc biệt năm 1985, tờ Thông tin Khoa học Thống kê đã được Bộ Văn hoá cấp giấy phép số 586/XB-BC phát hành rộng rãi trong cả nước. Đồng thời được cấp chỉ số phân loại ISSN 0868-3689. Từ 1989 đến 1997, việc biên soạn, phát hành 1 năm khoảng 4 số Thông tin Khoa học Thống kê và được duy trì thường xuyên.

Từ năm 1997 trở lại đây, hàng năm định kỳ được phát hành 2 tháng 1 số, và từ 2-4 số chuyên san.

Như vậy tính đến năm 2007, tờ Thông tin Khoa học Thống kê xuất bản được 27 năm, với khối lượng khoảng trên 120 số định kỳ và khoảng trên 30 số chuyên san. Tuy nhiên, công tác lưu trữ không được chú trọng nên việc

tìm kiếm lại toàn bộ các bản tin gặp nhiều khó khăn, nếu có cũng chỉ được thực hiện trên bản giấy. Hơn nữa, do những năm đầu bản tin chỉ được in ấn trên giấy chất lượng kém, thời gian tính đến nay cũng đã lâu nên nhiều bản tin cũng đã rách, chữ mờ, khó có thể khôi phục được.

Việc hệ thống lại toàn bộ bản tin đã là một công việc khó khăn, nên việc lưu trữ các tệp dữ liệu trên máy tính dường như không được chú trọng. Do những nguyên nhân sau đây:

- Những năm đầu được in bằng Roneo
- Máy tính chưa được trang bị
- Bộ nhớ máy tính nhỏ nên khó lưu trữ
- Cán bộ chưa có ý thức trong việc lưu trữ trên máy.

#### **Về các tư liệu ảnh và đa phương tiện khác:**

Nguồn tư liệu ảnh và đa phương tiện kỹ thuật số tương lai sẽ rất phổ dụng, tuy nhiên hiện nay vì nhiều điều kiện sự có mặt của nó còn rất khiêm tốn trong kho tư liệu KHTK cho nên chúng tôi chưa có điều kiện để nghiên cứu và thử nghiệm kỹ, nên đề tài có điểm qua như một khả năng không thể bỏ qua trong tương lai mà thôi.

**Về tư liệu hình ảnh**, hiện nay Viện Khoa học Thống kê đã trang bị một máy ảnh kỹ thuật số bán tự động Sony Cyber-shot 10.0 Megapixels để có thể ghi lại các hình ảnh về hoạt động của Viện như: hội thảo khoa học, bảo vệ đề tài, các đoàn công tác, kỷ niệm các ngày lễ lớn của Viện/Tổng cục,... đồng thời các ảnh giấy thời gian qua đã được scan và lưu trữ, tuy nhiên nó chưa được lưu trữ khoa học mà chỉ phân chia theo các thư mục cho nên việc tìm kiếm và sử dụng còn hạn chế.

Việc cần thiết hiện nay đối với kho tư liệu ảnh của Viện Khoa học Thống kê là việc tổ chức lại kho tư liệu ảnh khoa học để dễ dàng tìm kiếm và sử dụng. Qua tìm hiểu chúng tôi đề xuất sử dụng phần mềm quản lý ảnh miễn phí của Google là Picasa (dung lượng nhỏ khoảng 5,5MB), phần mềm này có thể hiệu chỉnh, thay đổi kích thước ảnh,... xuất ảnh sang các dạng khác nhau, rất thuận tiện trong quá trình sử dụng.

**Về tư liệu âm thanh và phim** hiện nay Viện Khoa học Thống kê chưa có tư liệu dạng này vì chưa được trang bị thiết bị ghi âm, ghi hình, nhưng trong thời gian tới dạng tư liệu này nên được chú ý ví dụ như âm thanh hoặc phim về các buổi hội thảo, sinh hoạt khoa học, nói chuyện của các chuyên gia về một lĩnh vực nào đó có liên quan, hay những âm thanh, hình ảnh hoạt



động của Viện trong các ngày đặc biệt, đây là những tư liệu sống động, việc lưu trữ những tư liệu này có ý nghĩa quan trọng.

## **6. Tư liệu là các văn bản pháp lý và các cơ sở dữ liệu khoa học thống kê**

Các văn bản pháp lý

Trong quá trình nghiên cứu và tác nghiệp cán bộ khoa học luôn luôn tiếp cận các giới hạn đã được quy định bởi các văn bản pháp lý; vì thế nhu cầu có một kho tư liệu là các văn bản pháp lý phù hợp là tự nhiên. Khoa học thống kê là một ngành hẹp chúng ta chỉ cần những văn bản pháp lý gần gũi và tránh lặp lại các kho văn bản pháp lý đã có của nhiều cơ quan khác nên chúng ta cần chọn lựa trong hệ thống văn bản pháp lý của nhà nước để có được kho tư liệu riêng về lĩnh vực này.

Trong những năm đã qua, Trung tâm thông tin KHTK chưa có điều kiện xây dựng kho tư liệu này một cách bài bản và đặc biệt chưa đưa nó theo hướng tin học hóa. Xuất phát từ đặc trưng của tư liệu là đòi hỏi chính xác cao nên tự nhiên nó là loại cần lưu trữ toàn văn. Một đặc tính nữa của loại hình tư liệu này là không thường xuyên phải cập nhật nó nên chi phí duy trì không đáng kể; về cấu trúc của tư liệu này chúng ta có thể chia theo các đặc tính sau:

- + Chia theo mức độ hiệu lực của văn bản,
- + Chia theo vấn đề là văn bản giải quyết như: chức năng, nhiệm vụ, tài chính, kinh phí, phạm vi, đối tượng....,
- + Chia theo thời gian ban hành,
- + Chia theo cấp ban hành,
- ...

Sự phân chia này tạo thành cấu trúc lưu trữ tư liệu và giúp cho người sử dụng dễ tìm kiếm tư liệu mà mình quan tâm.

Qua những khảo sát ban đầu về loại hình tư liệu là văn bản pháp lý chúng ta thấy rằng việc xây dựng kho tư liệu này là không khó và có thể tiến hành theo các bước sau:

- + Thu thập các văn bản pháp lý,
- + Phân loại và tổ chức lưu trữ trên máy tính,
- + Xây dựng các thư mục tìm kiếm (Metadata)
- + Đưa lên trang WEB của Viện để phổ biến cho người sử dụng.

Trong thời gian nghiên cứu của đề tài, dù đặt ra vấn đề tư liệu là văn bản pháp lý như một chủ điểm, song việc tạo dựng và đưa vào khai thác là đơn giản, lại có thử nghiệm trên trang WEB đang thiết kế nên chúng tôi không đi sâu vào vấn đề này; xem như vấn đề đã được giải quyết nên các phần sau chúng tôi không nhắc đến nữa.

Tư liệu từ các CSDL trợ giúp tác nghiệp

Như trong mục tổng quan đã nhắc đến chúng ta cần có một số CSDL trợ giúp cho người nghiên cứu và hoạt động tác nghiệp khác mà hai CSDL chính yếu cần có là:

- + CSDL các nhà nghiên cứu thống kê,
- + CSDL các tổ chức hoạt động nghiên cứu thống kê,

Trong đề tài trước đây chúng tôi đã đặt ra vấn đề này và tiến hành thử nghiệm xây dựng CSDL các nhà nghiên cứu thống kê vào năm 2003, tuy nhiên do kinh phí ít nên việc tiến hành điều tra thu thập đã không bao khắp được các nhà nghiên cứu thực sự là những chuyên gia, mà chỉ thu thập phiếu về một số cán bộ nghiên cứu trẻ; thêm nữa, từ đó đến nay CSDL không được duy trì cập nhật cho nên có thể nói đó là một kinh nghiệm thất bại.

## II. CÁC GIẢI PHÁP CHO CÔNG TÁC TƯ LIỆU KHOA HỌC THỐNG KÊ THEO HƯỚNG TIN HỌC HÓA

Tin học hoá là quá trình áp dụng tổ hợp các phương tiện kỹ thuật, đảm bảo phần mềm và ngôn ngữ nhằm mục đích hình thành các hệ thống thư viện tự động hoá trong việc chọn lựa, đảm bảo quản lý, xử lý và sử dụng thông tin thuộc mọi loại hình và công dụng.

### *Các mục tiêu của tin học hoá*

- Tạo lập và quản trị mục lục tài liệu điện tử cho vốn tài liệu của thư viện mình, dần dần tiến tới tạo lập các cơ sở dữ liệu khác trong đó có CSDL toàn văn các mảng tài liệu quan trọng;
- Hiện đại hoá các quá trình bổ sung và xử lý tài liệu, tìm kiếm thông tin;
- Tạo lập hệ thống phục vụ người dùng tin mới về chất trên cơ sở các CSDL của thư viện và tiếp cận với các CSDL trong và ngoài nước, đáp ứng kịp thời các yêu cầu về tài liệu;
- Tự động các hoạt động thống kê và kiểm kê;
- Mở rộng danh mục và khối lượng các sản phẩm thông tin cũng như các dịch vụ giành cho người dùng tin;

- Tạo nên mạng thông tin - thư viện toàn quốc cho phép bảo quản và sử dụng các nguồn thông tin quốc gia, tiếp cận tự do và kịp thời, hoà nhập vào không gian thông tin thế giới thông qua nối mạng toàn cầu.

## **1. Các giải pháp xử lý tư liệu thư viện: cho sách và tạp chí**

### **a, Lựa chọn hệ quản trị dữ liệu và khổ mẫu nhập tin**

Hiện nay trên thị trường Việt Nam tồn tại khá nhiều các phần mềm quản trị tư liệu, mỗi phần mềm lại có những mặt ưu và nhược điểm khác nhau. Các cơ quan thông tin-thư viện lớn của Việt Nam thường sử dụng các phần mềm thư viện tích hợp như: Tinh Vân, Libol, Lạc Việt... để quản trị kho tư liệu của mình.

Đối với các cơ quan thông tin-thư viện vừa và nhỏ ở nước ta thường sử dụng phần mềm CDS/ISIS do UNESCO phát hành miễn phí. Phần mềm này được UNESCO xây dựng chuyên để quản trị tư liệu và cập nhật thường xuyên nhằm nâng cao tính thân thiện của nó đối với người dùng. Trên thực tế Thư viện khoa học Thống kê cũng đã và đang sử dụng phần mềm này để quản trị kho sách của mình.

Cấu trúc CSDL trong CDS/ISIS gồm 4 file quan trọng nhất cho phép quản trị dữ liệu:

- Bảng xác định trường: FDT. Bao gồm các mã, nhãn, kiểu và trạng thái của các trường nhập thông tin;
- Biểu mẫu nhập tin: FMT. Cấu trúc các trường nhập tin;
- Format in/trình bày: PFT. Cho phép định dạng thông tin đầu ra;
- Bảng chọn trường: FST. Dùng để tạo file đảo, cho phép tìm kiếm thông tin theo các trường đã lựa chọn.

Sau khi đã lựa chọn phần mềm quản trị dữ liệu, công việc tiếp theo là phải xác định khổ mẫu nhập thông tin cho CSDL. Trong bối cảnh hiện nay khổ mẫu MARC 21 được sử dụng như một sự lựa chọn chủ yếu trên thị trường quản trị thư viện và được áp dụng rộng rãi trong cộng đồng thông tin-thư viện thế giới. Hiện nay số lượng lớn các biểu ghi theo MARC 21 được lưu giữ ở Mỹ - 800 triệu biểu ghi, mạng OCLC - 50 triệu biểu ghi, Thư viện Quốc hội Mỹ - 20 triệu biểu ghi [4]. Hầu hết các thư viện ở Việt Nam cũng đang theo xu hướng chuyển đổi sang dùng MARC 21 (phiên bản MARC Việt Nam) và Thư viện Khoa học Thống kê cũng không nằm ngoài khuynh hướng đó.

### **b, Ứng dụng tin học hóa quản lý kho sách**

Việc ứng dụng tin học hóa quản lý kho sách của Thư viện được đề xuất thực hiện qua các khâu nghiệp vụ sau: Tìm kiếm, thu thập, xử lý/phân tích, lưu giữ, phổ biến và phục vụ thông tin.

**Tìm kiếm, thu thập:** Trước hết cần thống nhất xây dựng chính sách bổ sung cụ thể nhằm đa dạng hóa kho tư liệu bằng các loại hình tài liệu ở các dạng khác nhau: In truyền thống và điện tử. Để triển khai việc bổ sung tài liệu đòi hỏi phải thiết lập các qui chế, cơ chế linh hoạt về trao đổi, hợp đồng mua trực tiếp từ các nhà xuất bản hoặc qua các đại lý phát hành. Việc bổ sung kho tư liệu khoa học thống kê có thể được thực hiện bằng các phương thức sau: qua email, qua điện thoại, tải từ Internet, tải từ các mạng thông tin - thư viện khác thông qua hợp tác trao đổi.

### **Xử lý/phân tích:**

- **Xử lý vật lý:** Đóng dấu, dán nhãn tài liệu, lập chỉ số xếp giá (mã kiểm soát, chỉ số cutter), vào sổ đăng ký cá biệt. Ở khâu này có thể sử dụng công nghệ mã vạch để tự động hóa thông tin đầu vào.

- **Xử lý logic:** bao gồm xử lý hình thức và xử lý nội dung tài liệu.

+ **Xử lý hình thức:** Xử lý thư mục cấp 1 nhằm cung cấp các thông tin cần thiết về tên tác giả, tác phẩm, năm xuất bản, số trang, ngôn ngữ... Việc mô tả thư mục cần tuân thủ chặt chẽ các qui tắc mô tả thư mục chuẩn quốc tế - ISBD.

+ **Xử lý nội dung:** Bao gồm các công đoạn: Phân loại; tóm tắt nội dung; chú giải; lập từ khóa (chủ đề, địa lý).

**Phân loại:** Trước hết cần lựa chọn bảng phân loại phù hợp về đặc thù khoa học thống kê, triển vọng cập nhật của bảng phân loại, tính dễ dàng sử dụng và phát triển chỉ số phân loại cục bộ... Hiện nay trong hoạt động thông tin-thư viện của Việt Nam và thế giới tồn tại rất nhiều bảng phân loại khác nhau: khung phân loại DDC, BBK, khung đề mục quốc gia, khung thập phân, khung phân loại của Thư viện Quốc gia, khung 19 dãy...).

Gần đây với sự trợ giúp của Mỹ (Thư viện Quốc hội Mỹ) về kinh phí cho dịch và phổ biến bảng phân loại thập phân Dewey (DDC) ở Việt Nam. Tính hiện đại của bảng phân loại DDC là hiển nhiên, bởi nó được Mỹ thường xuyên cập nhật và sẵn sàng tài trợ kinh phí cho áp dụng.

\* **Tóm tắt nội dung tài liệu:** Đòi hỏi cán bộ thư viện phải nắm được ngôn ngữ xử lý thông tin. Khác với ngôn ngữ tự nhiên, ngôn ngữ nhân tạo trong xử lý thông tin nhằm loại bỏ các ngôn từ vô nghĩa đảm bảo cho chất lượng tìm và phục vụ thông tin cho người dùng.

\* **Lập từ khóa (chủ đề, địa lý):** Đây là một trong các điểm truy cập thông tin quan trọng trong tìm kiếm được tài liệu. Có thể chia ra từ khóa chủ đề và từ khóa địa lý. ..

**Xử lý tiền máy:** Điền các thông tin xử lý hình thức và nội dung tài liệu vào các trường của biểu ghi thư mục trước khi nhập vào CSDL. Dưới đây là thiết kế mẫu “Phiếu nhập dữ liệu” của CSDL. Các trường nhập tin trong khổ mẫu này được xây dựng theo qui định của phần mềm CDS/ISIS. Cấu trúc dữ liệu của các trường như sau:

### PHIẾU NHẬP DỮ LIỆU CỦA CSDL KHTK

<i>Số TT</i>	<i>Mã tr-ờng</i>	<i>Tr-ờng dữ liệu</i>	<i>Tr-ờng con</i>
1	111	MÃ SỐ TÀI LIỆU	
2	113	DẠNG TÀI LIỆU	
3	120	NGÔN NGỮ	
4	210	TÁC GIẢ (CÁ NHÂN)	ab
5	215	TÁC GIẢ (TẬP THỂ)	ab
6	220	TÊN TÀI LIỆU (TIẾNG VIỆT)	
7	221	TÊN TÀI LIỆU (NGOẠI NGỮ)	
8	230	NHAN ĐỀ NGUỒN TRÍCH LÀ SÁCH	
9	232	NHAN ĐỀ NGUỒN TRÍCH TỪNG THƯ	
10	234	NHAN ĐỀ NGUỒN TRÍCH LÀ ÁN PHẨM	
11	236	TÊN HỘI NGHỊ HỘI THẢO	
12	238	TÊN XERI	
13	239	SỐ XERI	
14	250	NHÀ XUẤT BẢN	
15	251	NƠI XUẤT BẢN	ab
16	252	LẦN XUẤT BẢN	
17	253	SỐ ISBN	
18	254	TẬP	
19	255	SỐ TẠP CHÍ/BÁO	
20	256	SỐ ISSN	
21	260	NĂM XUẤT BẢN	
22	278	TRANG	
23	310	CHI SỐ ĐỀ MỤC	
24	312	TỪ CHUẨN	
25	314	TỪ CHUẨN ĐỊA LÝ	
26	320	TÓM TẮT	
27	440	GIÁ TIỀN	
28	445	GHI CHÚ	

Dưới đây là thiết kế biểu mẫu nhập dữ liệu cho CSDL tài liệu khoa học thống kê:

### BIỂU MẪU NHẬP DỮ LIỆU

<i>Số TT</i>	<i>Mã trường</i>	<i>Trường dữ liệu</i>	<i>Trường con</i>
1	001	Mã số kiểm soát	
2	003	Cơ quan gán số kiểm soát	VN- TK
3	020	Số ISBN	^a ^cĐiều kiện thu thập:
4	024	Các số nhận dạng chuẩn khác	
5	040	Cơ quan tạo biểu ghi BM gốc	^aVN- TK^bvie^e
6	041	Mã ngôn ngữ	^aChính văn: ^hNguyên bản/bản dịch trung gian:
7	044	Mã nước xuất bản	^a
8	082	Ký hiệu phân loại DDC	^aKý hiệu DDC: ^bSố thứ tự (Cutter): ^2Lần XB:
9	084	Ký hiệu phân loại khác	^aKý hiệu ^bSố thứ tự (Cutter): ^2Nguồn:
10	100	Tiêu đề chính - Tên cá nhân	^aHọ, tên: ^bThông tin trách nhiệm:
11	101	Tiêu đề chính - Tên tập thể	^aTên cơ quan: ^bTên đơn vị trực thuộc: ^eThuật ngữ xđịnh trách nhiệm liên quan: ^uCơ quan chủ quản:
12	111	Tiêu đề chính -	^aTên hội nghị:
13	242	Nhan đề dịch	^aNhan đề dịch: ^bPhần còn lại của nhan đề: ^eNgôn ngữ nhan đề dịch:
14	245	Nhan đề chính	^aNhan đề chính: ^nSố của phần tập: ^Nhan đề của phần tập: ^bPhần còn lại của nhan đề: ^cThông tin trách nhiệm:
15	246	Các dạng khác của nhan đề	^aNhan đề: ^bPhần còn lại của nhan đề:
16	250	Lần XB	^a
17	260	Địa chỉ xuất bản/phát hành	^aNơi XB: ^bNhà XB: ^cNăm XB:
18	300	Mô tả vật lý	^aKhối lượng, số trang: tr. ^bCác đặc điểm vật lý khác: ^eTài liệu kèm theo:

<i>Số TT</i>	<i>Mã trường</i>	<i>Trường dữ liệu</i>	<i>Trường con</i>
19	490	Tùng thư	^aNhan đề tùng thư:
20	500	Phụ chú chung	^aPhụ chú:
21	504	Phụ chú thư mục	^aThư mục: ^b tr.
22	520	Tóm tắt/chú giải	^aNội dung tóm tắt/chú giải:
23	650	Đề mục chủ đề (từ khoá)	^aĐề mục chủ đề:
24	651	Tiêu đề bổ sung	^a Đề mục con địa lý:
25	653	Thuật ngữ cần không kiểm soát	^aThuật ngữ:
26	700	Tiêu đề bổ sung- tên cá nhân	^aTên cá nhân: ^eThông tin trách nhiệm:
27	710	Tiêu đề bổ sung- tên tập thể	^aTên cơ quan/địa điểm: ^bTên đơn vị trực thuộc: ^eThuật ngữ xđịnh trách nhiệm liên quan: ^uCơ quan chủ quản/địa chỉ:
28	765	Bản ngôn ngữ gốc	^tNhan đề:
29	767	Bản dịch	^tNhan đề:
30	852	Nơi lưu trữ	^aNơi lưu trữ:VN-TTTC ^hKý hiệu phân loại: ^iChi số xếp giá-Cutter: ^jSố kiểm soát:
31	856	Địa chỉ và truy cập điện tử	^aTên máy chủ: ^dĐường dẫn: ^fTên điện tử:

**Lưu giữ thông tin:** Hiện tại sách của Thư viện Khoa học Thống kê được lưu giữ trong cơ sở dữ liệu (CSDL) sách quốc văn (KHTK) và CSDL sách ngoại văn (PDIC). Hiện tại tổng số biểu ghi đã cập nhật vào CSDL KHTK là 1283 biểu và CSDL PDIC là 1.800 biểu.

**Phổ biến thông tin:** Trên cơ sở thông tin đã cập nhật trong CSDL có thể: Ra thông báo thư mục sách mới; thư mục chuyên đề; sao chụp cung cấp thông tin theo yêu cầu; cung cấp nội dung thông tin tư liệu khoa học thống kê lên trang Web của Viện.

Việc sử dụng các thông tin được lưu giữ trong CSDL để phổ biến cho người dùng theo yêu cầu được thực hiện bằng sự hỗ trợ của format cho thông tin đầu ra trong Winisis như sau:

*Các lệnh chế độ dữ liệu bao gồm:*

- mdl: chế độ dữ liệu thay trường con bằng dấu (,) hoặc (.), cuối trường luôn có dấu (.);

- mhl: không hiển thị trường con;
- mpl: chế độ kiểm định dữ liệu (nhập thế nào ra như vậy);
- if ... then.

*Các lệnh trong format bao gồm:*

- v: Lệnh chọn trường - v(n);
- m: Lệnh căn lề - m(x,y);
- tab: độ nhảy dữ liệu (tính từ vị trí dữ liệu);
- f: chọn font (thực hiện trong file hệ thống theo các tham số qui định);
- cl: chọn màu;
- % và / : lệnh xuống dòng và triệt tiêu dòng trống;
- #: dấu xuống dòng và để lại dòng trống.

*Các hằng trong format bao gồm:*

- Hằng có điều kiện "... " (dấu nháy kép);
- Hằng vô điều kiện '...' (dấu nháy đơn);
- Hằng lặp |...| (dấu sỏ);
- Hằng tiền tố và hậu tố

**Ví dụ, format hiện toàn bộ thông tin về một tài liệu có dạng sau:**

```
mhl,/{cl6,(v1^*l. l, v852^i)},{cl5,b,c35,v84}/{cl13,b,v100^a+;
l/v110^a+; l/v111^a,v111^a,v111^d,v111^c}/l
lv245^a,v245^b,v245^a,v245^p,v245^c,
l.- lv250^a,l.- lv260^a,l: l v260^b, l, lv260^c,l.-
lv300^a,v300^b,v300^c,v300^e,("
- "(lv440^a,l ISSN lv440^x,v440^v,d440)" ),(" - "(lv49
0^a,l ISSN lv490^x,l lv490^v,d490" )/l lv500^a/(l lv502^a,l
lv502^b,v502^c)/(l ISBN lv20^a),
(l lv711^a l,v711^nl l,v711^dl l,v711^cl l)/{cl5,i,("Từ khóa: "v650^a,l-
lv650^z, d650l; l)/
("Từ khóa: "v653^a+; l)}/#{cl20,"Tóm tắt: "v520/}/c35,'BOOK-',mf(5)
```

Nội dung dữ liệu đầu ra sẽ có dạng sau:

MFN: 1026

1: N266-4B5



3: TTKHTK

40: ^aTTKHTK^bvie^eaacr

41: ^aViệt

41: ^aAnh

44: ^aViệt nam

84: ^aZ24.h5(2B-4B5)

110: ^aCục Thống kê Bắc Ninh

242: ^aBac Ninh statistical yearbook 2005

245: ^aNiên giám thống kê tỉnh Bắc Ninh 2005

260: ^aHà nội^bThống kê^c2006

300: ^a383 tr.

520: ^aBao gồm những số liệu cơ bản phản ánh khái quát động thái và thực trạng kinh tế xã hội của tỉnh Bắc Ninh từ năm 2001 đến năm 2005 (số liệu năm 2005 là số ước tính). Ngoài ra trong nội dung của niên giám còn có số liệu tổng quát của các cuộc điều tra trong năm; các chỉ tiêu liên quan đến phát triển con người (HDI) và một số chỉ tiêu kinh tế xã hội chủ yếu của các tỉnh thuộc khu vực Đồng bằng sông Hồng.

651: ^aBắc Ninh

651: ^aĐồng bằng sông Hồng

653: ^asố liệu

653: ^athống kê

653: ^aking tế xã hội.

Với sự trợ giúp của bảng format chọn trường trong CSDL ta có thể tìm kiếm tài liệu theo nhiều yếu tố khác nhau: số thứ tự biểu ghi (MFN), tác giả cá nhân, tác giả tập thể, nhan đề, từng thư, nơi xuất bản, nhà xuất bản, năm xuất bản, chỉ số đề mục, ký hiệu xếp giá, từ khóa (chủ đề và địa lý). Các yếu tố tìm tin này có thể dùng chế độ dữ liệu để thực hiện tìm cả trường hoặc từng từ, từng cụm từ. Trong thuật ngữ tìm tin thường sử dụng các toán tử sau:

- AND: để thu hẹp kết quả tìm;
- OR: mở rộng kết quả tìm;
- NOT: loại trừ giá trị không cần tìm.
- \$: dấu chắt chốt phải trong trường hợp chỉ cần giá trị tìm bên trái.

Trong phần thử nghiệm chúng tôi trình bày một số kết quả tìm kiếm thông tin theo các yêu cầu khác nhau như: Theo từng chuyên đề, theo tên tác giả, theo tên tác phẩm, theo năm xuất bản... Riêng thử nghiệm cho ra thông báo thư mục đã được thực hiện thành công qua một số ấn phẩm “Thông báo thư mục sách mới” của Trung tâm.

**Phục vụ thông tin:** Việc phục vụ thông tin có thể thực hiện theo hai hình thức sau: Phục vụ tài liệu (tài liệu in truyền thống) và phục vụ theo nhu cầu tin (tài liệu điện tử). Đối người dùng tin từ xa có thể phục vụ yêu cầu tin thông qua mạng Internet. Hiện nay, theo Pháp lệnh Thư viện, việc đáp ứng nhu cầu tin từ xa được coi là các dịch vụ thông tin có thu.

*(xem kết quả thử nghiệm ở Phụ lục1).*

### c, Ứng dụng tin học hóa quản lý kho tạp chí

Một vài năm gần đây việc sử dụng một số phần mềm thư viện điện tử đã giúp các thư viện dễ dàng hơn trong việc tự động hóa công tác quản lý tạp chí. Hiện nay, để giúp các thư viện có thể quản lý tốt loại hình tài liệu này, Thư viện Quốc hội Mỹ đã xây dựng khổ mẫu lưu giữ tư liệu (Holding Data Format - HDF) thuộc MARC21. Việc áp dụng HDF cho phép tự động hóa công tác quản lý tạp chí trong các phần mềm thư viện điện tử.

Khổ mẫu HDF được thiết kế nhằm chứa các thông tin liên quan đến từng tư liệu cá lẻ cho các loại tài liệu mang tính định kỳ xuất bản mang tính đặc thù của từng thư viện cụ thể như: chuyên khảo một tập, nhiều tập, báo/tạp chí.

Biểu ghi về vốn tư liệu (HDF)		
001		88888
003		Tên thư viện
004		99999
852	##	\$anoi lưu giữ\$CChỉ số lưu giữ
853	20	\$aSố\$inăm\$wq\$zd
863	40	\$a1-3\$1977
863	40	\$a4-5\$1978
863	40	\$a6-7\$1979
863	40	\$a1-4\$1980-1982
866	40	\$aSố1-4(1977-1982)\$z những năm thiếu số

## 2. Giải pháp cho tư liệu nội sinh: cho đề tài và chuyên đề nghiên cứu

Khổ mẫu được khuyến nghị sử dụng trong bối cảnh hiện nay là: Marc21 cho dữ liệu thư mục (TCVN 7539: 2005 Thông tin Tư liệu) được cộng

đồng thông tin thư viện sử dụng rộng rãi, đang trở thành một tiêu chuẩn quốc tế; áp dụng nó, Thư viện - Trung tâm thông tin Khoa học Thống kê sẽ có điều kiện trao đổi dữ liệu biên mục với các Thư viện, Trung tâm Thông tin khác.

Năm 2006, Viện Khoa học Thống kê (VKHTK) đã cùng các chuyên gia thông tin của Trung tâm Thông tin Khoa học Công nghệ Quốc gia xây dựng một CSDL quản lý tài liệu nội sinh.

Mọi tài liệu khi nhập vào thư viện, trong đó có tài liệu nội sinh đều cần được xử lý /phân tích, bao gồm xử lý về mặt vật lý, logic để có những thông tin cho xử lý tiền máy. (xem thêm phần thư viện)

Hiện nay, việc phân loại các tài liệu nội sinh, chủ yếu là các đề tài, Thư viện – TTTTKHTK kiến nghị tạm thời sử dụng theo khung Đề mục quốc gia vì Trung tâm Thông tin Khoa học Công nghệ Quốc gia, cơ quan TTKHCN đầu ngành của cả nước đang sử dụng khung này để phân loại các kết quả nghiên cứu khoa học (rất rộng) trong đó có kết quả nghiên cứu khoa học của ngành Thống kê và mục 83 là Thống kê từ trang 171 đến trang 173 (trong Khung Đề mục Hệ thống thông tin Khoa học và Kỹ thuật quốc gia. - Tập 1. Bảng chính. - Hà Nội: 1987), ngoài ra có thể tham khảo các mục khác về các ngành kinh tế quốc dân, toán học có liên quan đến khoa học Thống kê.

Tham khảo **Bảng liệt kê các trường và trường con** thông thường có trong cơ sở dữ liệu ở một phần mềm quản lý thư viện theo khổ mẫu chuẩn Marc21 rút gọn cho dữ liệu thư mục

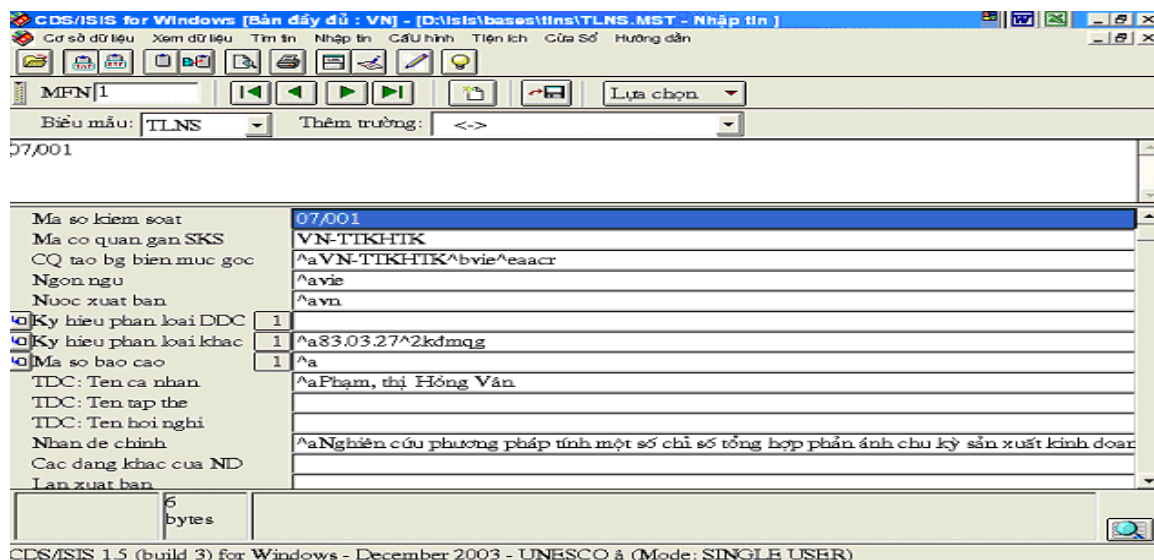
<b><i>Trường dữ liệu:</i></b>	<b><i>Trường con:</i></b>
001 Mã số kiểm soát	
003 Cơ quan gán số kiểm soát	VN- TK
020 Số ISBN	^a ^cĐiều kiện thu thập:
024 Các số nhận dạng chuẩn khác	
040 Cơ quan tạo biểu ghi biên mục gốc	^aVN- TK^bvie^e
041 Mã ngôn ngữ	^aChính văn: ^hNguyên bản/bản dịch trung gian:
044 Mã nước xuất bản	^a
082 Ký hiệu phân loại DDC	^aKý hiệu DDC: ^bSố thứ tự (Cutter): ^2Lần XB:
084 Ký hiệu xếp giá theo khung phân loại khác	^aKý hiệu ^bSố thứ tự (Cutter): ^2Nguồn:
100 Tên tiêu đề chính - Tên cá nhân	^aHọ, tên: ^bThông tin trách nhiệm:
101 Tên tiêu đề chính - Tên tập thể	^aTên cơ quan: ^bTên đơn vị trực thuộc: ^eThuật ngữ xác định trách nhiệm liên quan:

<b>Trường dữ liệu:</b>	<b>Trường con:</b>
	^uCơ quan chủ quản:
111 Tiêu đề chính Tên hội nghị	^aTên hội nghị:
242 Nhan đề dịch	^aNhan đề dịch: ^bPhần còn lại của nhan đề: ^eNgôn ngữ nhan đề dịch:
245 nhan đề chính	^aNhan đề chính: ^n Số của phần tập ^p Nhan đề của phần tập ^b Phần còn lại của nhan đề: ^cThông tin trách nhiệm:
246 Các dạng khác của nhan đề	^aNhan đề: ^bPhần còn lại của nhan đề:
250 Lần XB	^a
260 Địa chỉ xuất bản/phát hành	^aNơi XB: ^bNhà XB: ^cNăm XB:
300 Mô tả vật lý	^aKhối lượng, số trang: tr. ^bCác đặc điểm vật lý khác: ^eTài liệu kèm theo:
490 Tàng thư	^aNhan đề tàng thư:
500 Phụ chú chung	^aPhụ chú:
504 Phụ chú thư mục	^aThư mục: ^b tr.
520 Tóm tắt/chú giải	^aNội dung tóm tắt/chú giải:
650 Đề mục chủ đề (từ khoá)	^aĐề mục chủ đề:
651 Tiêu đề bổ sung	^a Đề mục con địa lý:
653 Thuật ngữ chủ đề không kiểm soát	^aThuật ngữ:
700 Tiêu đề bổ sung- tên cá nhân	^aTên cá nhân: ^eThông tin trách nhiệm:
710 Tiêu đề bổ sung- tên tập thể	^aTên cơ quan/địa điểm: ^bTên đơn vị trực thuộc: ^eThuật ngữ xác định trách nhiệm liên quan: ^uCơ quan chủ quản/địa chỉ:
765 Bản ngôn ngữ gốc	^tNhan đề:
767 Bản dịch	^tNhan đề:
852 Nơi lưu trữ	^aNơi lưu trữ:VN-TTTK ^hKý hiệu phân loại: ^iChỉ số xếp giá-Cutter: ^jSố kiểm soát:
856 Địa chỉ và truy cập điện tử	^aTên máy chủ: ^dĐường dẫn: ^fTên điện tử:

Xem ví dụ cụ thể về biểu ghi thư mục tài liệu nội sinh, các thông tin ở đây đã được giản lược phù hợp với đặc tính tài liệu và cấu trúc CSDL được chọn theo chuẩn Marc 21 (**Phụ lục 2**).

Giới thiệu những giao diện cơ bản có ở phần mềm quản lý tài liệu nội sinh

*\* Đây là giao diện trực tiếp của một biểu mẫu nhập dữ liệu nội sinh*



Nội dung dữ liệu ứng với các trường có dạng sau:

MFN: 1

1: 07/001

3: VN-TTKHTK

40: ^aVN-TTKHTK^bvie^eaacr

41: ^avie

44: ^avn

84: ^a83.03.27^2kđmqg

88: ^a

100: ^aPhạm, thị Hồng Vân

245: ^aNghiên cứu phương pháp tính một số chỉ số tổng hợp phản ánh chu kỳ sản xuất kinh doanh và khả năng ứng dụng ở Việt Nam /^cPhạm thị Hồng Vân (Chủ nhiệm đề tài); Nguyễn thị Việt Hồng (thư ký đề tài)... [et al.].

260: ^aHà nội, ^c2005

300: ^a73 tr.

490: ^aĐề tài cấp: Tổng cục

500: ^aThời gian thực hiện: 2004-2005.

500: ^aCơ quan chủ trì: Viện Khoa học Thống kê.

500: ^aCơ quan quản lý KHCN: Viện Khoa học Thống kê

500: ^aCơ quan chủ quản: Tổng cục Thống kê

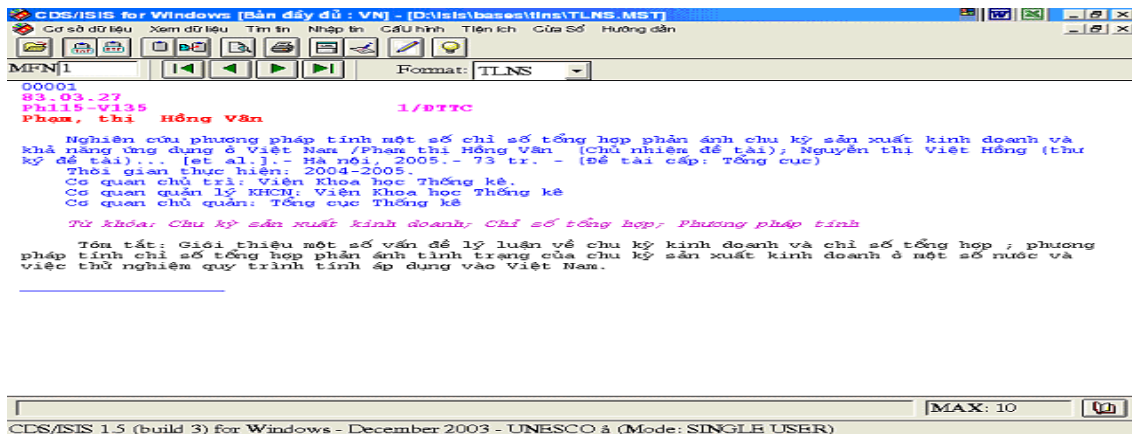
520: ^aGiới thiệu một số vấn đề lý luận về chu kỳ kinh doanh và chỉ số tổng hợp ; phương pháp tính chỉ số tổng hợp phản ánh tình trạng của chu kỳ sản xuất kinh doanh ở một số nước và việc thử nghiệm quy trình tính áp dụng vào Việt Nam.

653: ^aChu kỳ sản xuất kinh doanh  
 653: ^aChỉ số tổng hợp  
 653: ^aPhương pháp tính  
 700: ^aNguyễn, Thị Việt Hồng^e (Cán bộ phối hợp)  
 852: ^aVN-TTKHTK^h83.03.27^iPh115-V135^j1/ĐTTC  
 900: ^aĐ.T. Bình^bĐ.T. Bình^cT.M. Hùng  
 910: ^aNhà đăng ký kqnc^bSố đăng ký

*Format hiện hình đầy đủ của tài liệu nội sinh*

```
mhl,mfn(5)/{cl2,b,v852^h/v852^i,c35,v852^j}/{cl13,b,v100^a+}; |v110^a+;
|v111^a,v111^b,v111^c,v111^d,v111^e| |v245^a,v245^b,v245^c,v245^d,v245^e,
|.- |v250^a,|- |v260^a,v260^b,|- |v270^a,v270^b,v270^c,v270^d,|-
- "|(|v440^a,| ISSN |v440^x,v440^y,d440")" ),(" - "|(|v49
0^a,| ISSN |v490^x,| |v490^y,d490")")(| |v500^a/),( |v502^a,| |v502^b,v502^c/)( |ISBN |v20^a/),
(| |v711^a| |v711^b| |v711^c| |v711^d| |v711^e| )/{cl5,i,(" Từ khóa: "v650^a,|-|v650^z, d650); |)/
(" Từ khóa: "v653^a+; |); }/# {cl20," Tóm tắt: "v520/}
/'
```

*Format này cho kết quả thể hiện trên giao diện như sau:*



*Dữ liệu được tìm kiếm theo bản chọn trường sau:*

```
2 0 f(mfn,1,0)
100 0 (v100^a/)
110 4 v110^a/v110^b
111 4 v111^a
245 4 v245^a/v245^b
246 4 v246^a
260 0 (v260^a,v260^b,v260^c/)
490 4 (v440^a/) (v490^a/)
500 4 v500^a
650 0 (v650^a/v650^x/v650^y/v650^z/) (v653^a/)
650 4 (v650^a/v650^x/v650^y/v650^z/) (v653^a/)
700 0 (v700^a/)
710 4 (v710^a/v710^b/)
```

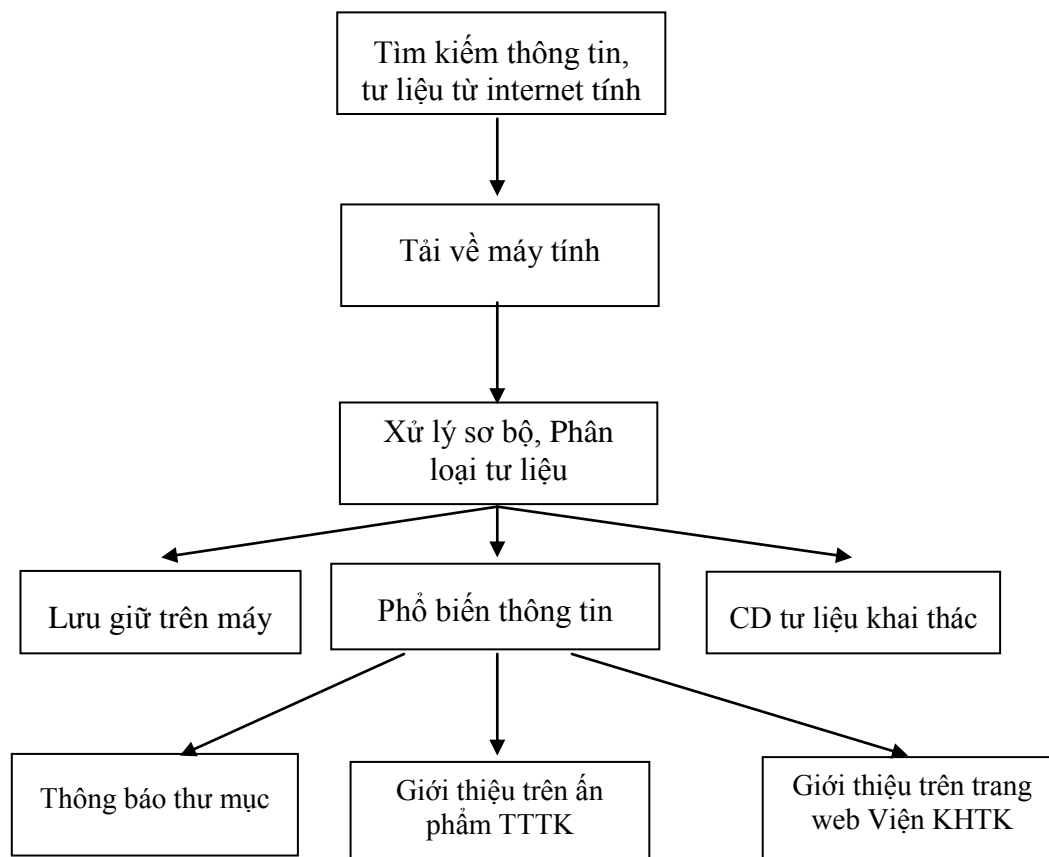
852 0 (v852^j/)  
653 0 mh1, (v653^a/)  
653 4 mh1, (v653^a/)  
852 1 mh1, v852^j  
900 1 mh1, (v900^a/v900^b/v900^c)

Ta có thể tìm kiếm theo các yếu tố Tác giả (Chủ nhiệm đề tài), nhan đề, năm xuất bản (năm hoàn thành đề tài), phụ chú cho biết được Cơ quan chủ trì, cơ quan chủ quản, từ khoá (thuật ngữ không kiểm soát), ký hiệu xếp giá... các thông tin đã được cập nhật trong CSDL có thể sử dụng để phổ biến theo các hình thức khác nhau như: ra thông báo thư mục kết quả nghiên cứu theo chuyên đề, cung cấp nội dung thông tin tư liệu KHTK lên trang Web của Viện, tra cứu sao chụp thông tin theo yêu cầu...

*(xem kết quả thử nghiệm ở Phụ lục 2).*

### 3. Các giải pháp cho tư liệu khai thác từ mạng internet

Trong mấy năm gần đây Trung tâm đã thử nghiệm thực hiện khai thác thông tin từ mạng internet, kinh nghiệm về quy trình xử lý được thể hiện dưới đây:



Thông tin, tư liệu sau khi khai thác được được phân thành các loại sau:

- bản tin/tạp chí

- tài liệu hội nghị/hội thảo,
- tài liệu đào tạo/hướng dẫn chuyên môn nghiệp vụ
- sổ tay
- văn bản pháp quy....

### ***3.1 Các sản phẩm sau xử lý***

Thông tin tư liệu đã khai thác sẽ được xử lý dưới dạng các sản phẩm phục vụ cho người sử dụng như sau:

- *Tài liệu in*
- *Thông báo*
- *Lưu giữ trên máy tính*
- *Lưu giữ trên đĩa CD*
- *Thông tin tư liệu mới khai thác trên mạng nội bộ.*

Khi trang Web của Viện Khoa học thống kê chính thức hoạt động thì việc giới thiệu các tư liệu, tài liệu thống kê mới cần được giới thiệu cho các cán bộ của Ngành Thống kê.

### ***3.2 Quản lý tư liệu khoa học thống kê từ mạng Internet***

Nhóm cán bộ tìm kiếm thông tin tư liệu thống kê thường xuyên có sự trao đổi về những thông tin tư liệu mới, lựa chọn và tải xuống máy tính. Sau khi đã download được các tài liệu phù hợp về máy tính, đã trao đổi, thống nhất sơ bộ phân loại các tài liệu in và đóng quyển, chuyển sang thư viện phục vụ độc giả. Các tài liệu khác được sắp xếp, lưu giữ trong máy tính cá nhân và có những phương pháp xử lý phù hợp với yêu cầu lưu giữ và phổ biến thông tin.

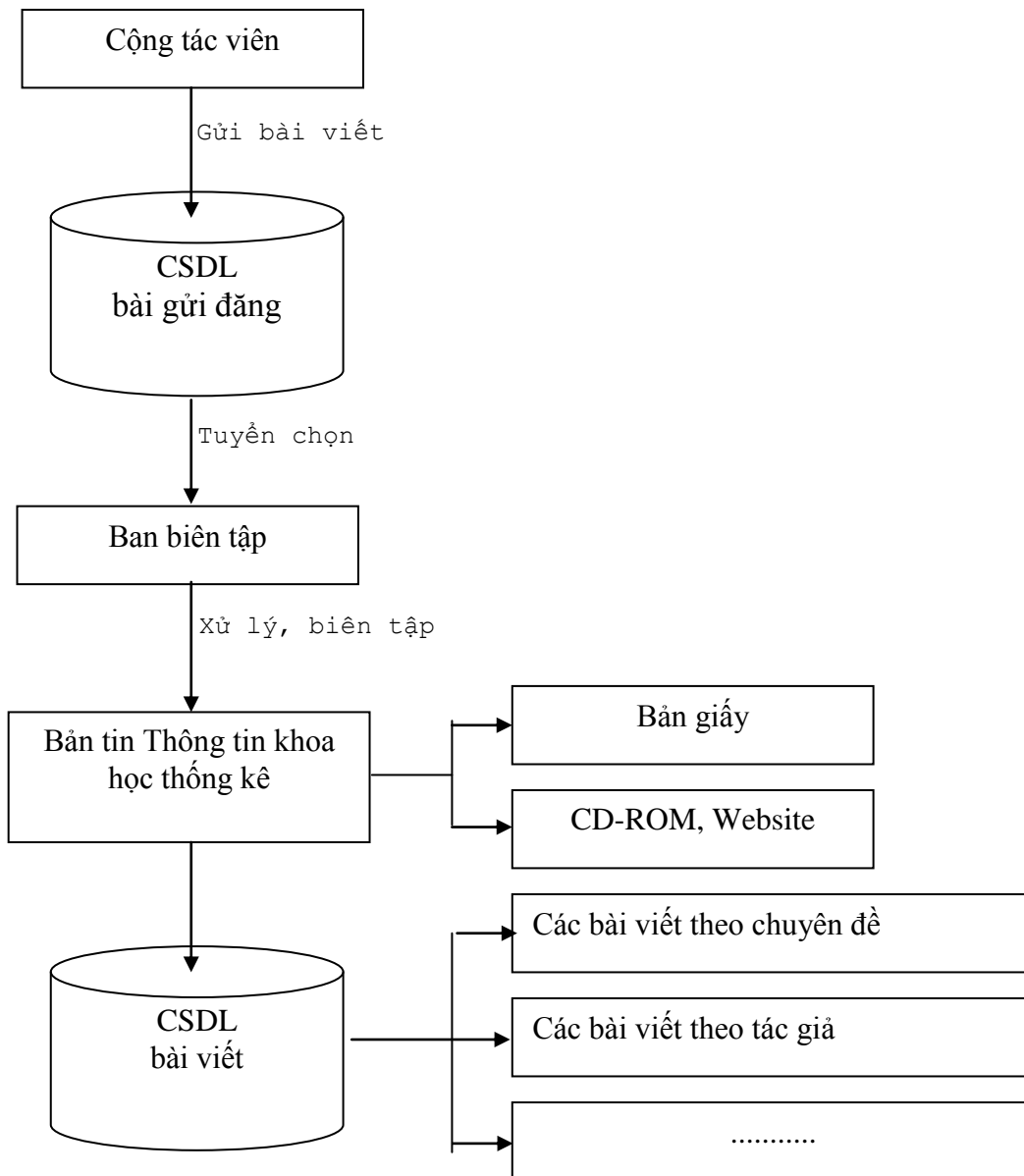
### ***3.3 Hướng tin học hoá tư liệu khai thác trên internet***

Để tin học hoá được tư liệu khai thác cần phải tổ chức sắp xếp theo hệ thống. Có thể sắp xếp tư liệu khai thác được phân loại theo chủ đề, theo thể loại, theo thời gian. Lưu giữ các địa chỉ truy cập (địa chỉ kết nối) đến từng tài liệu cụ thể để thuận tiện cho việc tra cứu. Đối với các tài liệu có tính chất cảm nang cần tải về máy và lưu giữ và phổ biến. Truy cập để tìm kiếm thông tin từ các cơ sở dữ liệu trực tuyến. Trình bày dưới dạng các thư mục.

## **4. Các giải pháp cho tờ Thông tin KHTK và các ấn phẩm đa phương tiện**

Để quản lý có hiệu quả nguồn thông tin khoa học thống kê chi tiết đến các bài viết chúng tôi xin mô tả qua sơ đồ sau:





### ***a. Đối với các bản tin đã phát hành***

Cần sưu tập, hệ thống lại toàn bộ, tiếp đó nếu có điều kiện thì tiến hành nhập lại nội dung thông tin của các bài viết đã đăng, hoặc tiến hành scan dưới dạng ảnh hoặc file dạng đuôi \*.pdf để lưu vào máy tính.

Bản tin TTKTTK kể từ năm 1990 trở lại đây được lưu trữ dưới dạng \*.pdf và xuất bản 3 lần thông qua đĩa CD-ROM phát hành tới các đối tượng sử dụng. Tuy nhiên chưa được lưu trữ chi tiết đến các bài bằng cơ sở dữ liệu nên còn gặp khó khăn trong quá trình tra cứu khi cần thiết. Vì thế sau khi xây dựng được cơ sở dữ liệu các bài viết đăng trên thông tin khoa học thống kê thì việc cần thiết là nhập các bản tin đã scan vào cơ sở dữ liệu để thống nhất quản lý và tra cứu.

### ***b. Đối với các bản tin sẽ phát hành***

Để tin học hóa toàn bộ quá trình đòi hỏi các cộng tác viên khi gửi các bài viết phải kèm theo bản softcopy (file dạng \*.doc, \*.txt,...) khi đó người nhận bài sẽ nhập một số thông tin như ngày nhận, tác giả, tên bài, thể loại, chủ đề,... tạo thành cơ sở dữ liệu các bài gửi đăng để việc tuyển chọn các bài viết được thuận tiện, nhanh chóng.

Sau quy trình quản lý bằng tin học thì kết quả là Tờ Thông tin Khoa học Thống kê bản giấy và được lưu trữ các bản tin này chi tiết (các bài viết) bằng việc thiết lập một cơ sở dữ liệu về các bài viết. Từ cơ sở dữ liệu các bài viết này, chúng ta có thể có các ấn phẩm sau: - Xuất bản đĩa CD-ROM Thông tin Khoa học Thống kê (theo năm, theo chủ đề,...) cung cấp cho bạn đọc và đưa lên Website Viện Khoa học Thống kê.

### **III. MÔ HÌNH HOẠT ĐỘNG KHO THÔNG TIN KHTK MỚI VÀ NHỮNG SẢN PHẨM CỦA NÓ**

Nền tảng cơ sở để tiến hành hoạt động thông tin tư liệu khoa học thống kê chính là kho tư liệu khoa học. Trước khi thực hiện tin học hóa kho tư liệu khoa học thống kê cần tìm hiểu khái niệm và nội hàm của kho tư liệu thống kê.

Kho tư liệu khoa học là bộ tài liệu khoa học được tập hợp theo những tính chất đặc trưng nhất định và nhằm một mục tiêu nhất định. Bộ tài liệu này được tổ chức, sắp xếp và lưu giữ theo một chuẩn thông tin - thư viện nhất định để thuận lợi cho việc tìm kiếm và phục vụ người dùng. Như vậy, kho tư liệu khoa học thống kê là tập hợp các tài liệu khoa học chuyên ngành nhằm mục đích phục vụ cho hoạt động của ngành Thống kê.

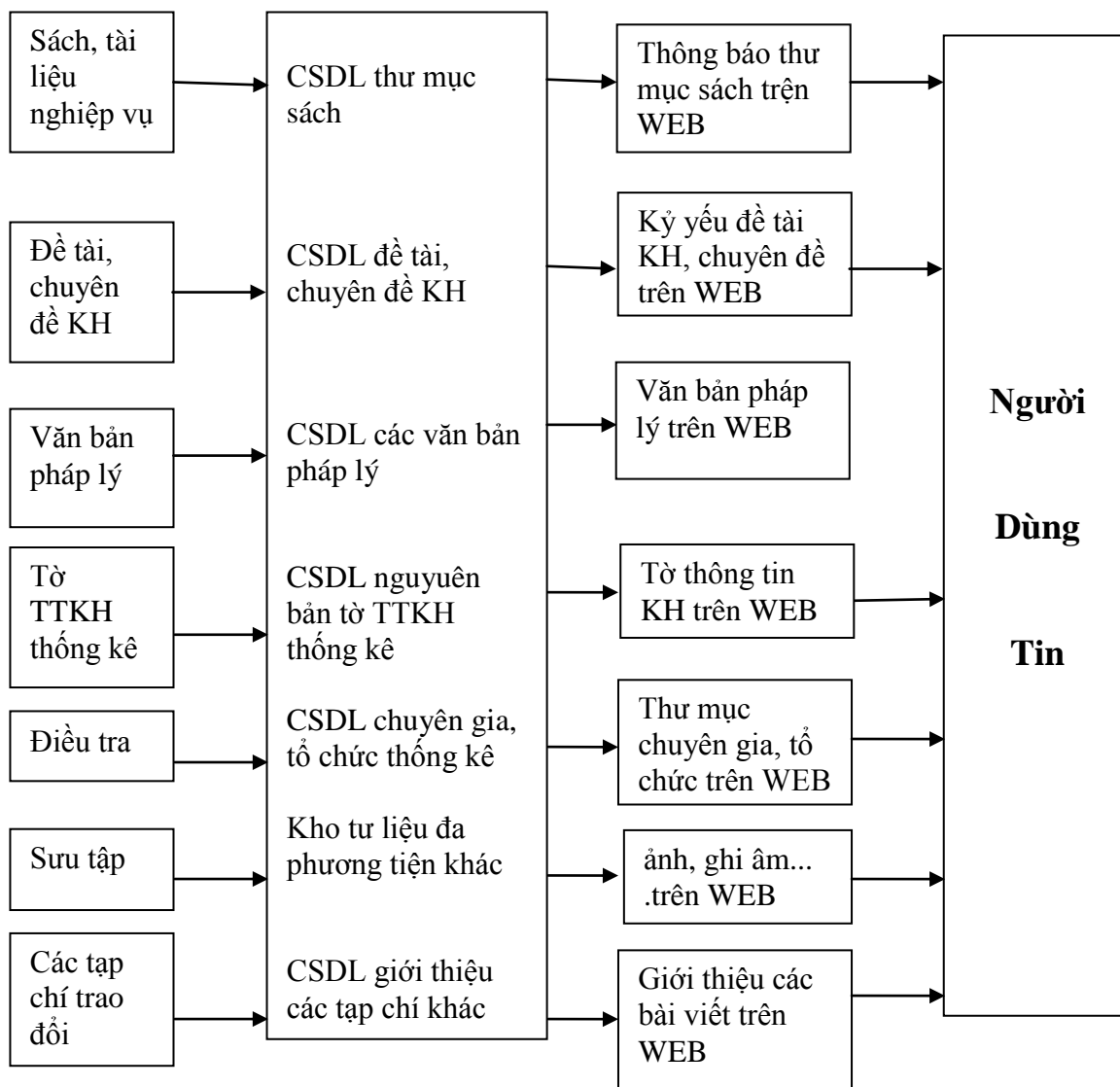
Định nghĩa “Tài liệu” (Document) theo tiêu chuẩn Việt Nam (TCVN 5453-1991) về lĩnh vực hoạt động thông tin khoa học và tư liệu - Đó là vật mang tin, trên đó cố định thông tin và là đối tượng xử lý trong quá trình xử lý thông tin và tư liệu. Như vậy, vật mang tin chính là đối tượng vật chất dùng để ghi và lưu giữ nội dung thông tin.

Đặc trưng nổi bật của “Tài liệu” là sự hiện hữu về các chất liệu khác nhau của vật mang tin truyền thống (giấy, băng từ, microfilm ...) và phi truyền thống (dạng điện tử), mà trên đó nội dung thông tin được cố định theo các hình thức khác nhau: Văn bản, hình ảnh, âm thanh,... Sự khác biệt của thông tin khoa học tư liệu so với thông tin truyền thống được thể hiện ở chỗ thông tin tư liệu khoa học được cố định trên vật mang tin cụ thể nhằm mục đích để lưu giữ, phổ biến và sử dụng trong không gian và thời gian.

Khác với thông tin đại chúng, thông tin khoa học được thu nhận trong quá trình hoạt động khoa học, kỹ thuật, sản xuất và phản ánh nội dung, kết

quả của các hoạt động đó. Như vậy, thông tin khoa học tư liệu thống kê chính là kết quả của quá trình hoạt động kinh tế nói chung và hoạt động thống kê, cũng như nghiên cứu khoa học thống kê nói riêng.

Từ những nghiên cứu cụ thể về kho tư liệu theo hướng tin học hóa trên, kết hợp với trang WEB của Viện KHTK đã được xây dựng chúng ta sẽ có một mô thức hoạt động thông tin tư liệu KHTK mới đáp ứng các nhu cầu đặt ra như mô tả trong sơ đồ sau:



**Nguồn tin                      Kho dữ liệu                      Sản phẩm                      Người dùng tin**

Sơ đồ nêu trên thể hiện được các yêu cầu cơ bản xây dựng nguồn thông tin tư liệu KHTK :

- + Các sản phẩm của nó trong các dạng mang tin khác nhau,
- + Phương cách giao tiếp mới tôn trọng người dùng tin.

## IV. CÁC KIẾN NGHỊ

### 1. Yêu cầu về nhân lực và trang thiết bị kỹ thuật

**Nhân lực thư viện:** Cán bộ thư viện phải là người nắm chắc nghiệp vụ thông tin-thư viện và công nghệ thông tin. Trong số các kiến thức mới mà cán bộ thông tin-thư viện cần phải có là:

- Quản trị thư viện điện tử;
- Tổ chức thu thập và lưu trữ thông tin số;
- Xử lý, số hóa và bảo quản thông tin số;
- Tìm kiếm và phổ biến thông tin số cho người dùng tin;
- Phục vụ các dịch vụ tra cứu thông tin số;
- Trợ giúp và đào tạo người dùng tin trong nghiệp vụ tìm kiếm, hệ thống hóa, phân tích và sử dụng thông tin số.

Các kỹ năng mới cần có của cán bộ thư viện:

- Sử dụng các công nghệ đa phương tiện;
- Thu thập và xử lý tài liệu trực tuyến;
- Tìm kiếm, khai thác dữ liệu và tri thức từ nhiều nguồn thông tin khác nhau;
- Biên tập và tạo lập sản phẩm thông tin theo nhu cầu tin.

Như vậy, với các kỹ năng và kiến thức mới như trên công tác thông tin-thư viện phải được chuyên môn hóa.

**Trang bị máy móc kỹ thuật:** Để đảm bảo thực hiện thành công việc tự động hóa quản lý kho tư liệu khoa học thống kê, ngoài máy tính cho cán bộ thư viện, cần trang bị cho Thư viện ít nhất 1 máy tính đủ mạnh về tốc độ, đảm bảo dung lượng bộ nhớ cao cho lưu giữ CSDL và phục vụ việc tìm kiếm thông tin. Ngoài ra cần trang bị máy in; máy quét; máy copy và các thiết bị bảo quản các CSDL trên đĩa quang. Việc trang bị đồng bộ các trang thiết bị kỹ thuật sẽ đảm bảo chất lượng phục vụ thông tin cho người dùng một cách thiết thực nhất.

### 2. Yêu cầu về quy trình, quy định kỹ thuật

Như trên đã phân tích để cho kho tư liệu khoa học thống kê làm việc hiệu quả chúng ta cần tuân thủ các yêu cầu kỹ thuật của cả quy trình xử lý, lưu trữ theo đòi hỏi của kỹ thuật tin học vì thế chúng tôi kiến nghị:

+ Soạn thảo và áp dụng một quy định chi tiết cho quá trình hoạt động thông tin theo mô hình kiến nghị,

+ Quy chuẩn các kỹ thuật áp dụng cho mỗi loại hình tư liệu như là một yêu cầu bắt buộc để thông tin thống nhất,

+ Xây dựng quy chế lưu trữ áp dụng cho các công đoạn của quá trình xử lý thông tin,

+ Xây dựng chế độ an toàn cho kho tư liệu trong các hoạt động cung cấp thông tin.

### **3. Những việc nên ứng dụng ngay**

Từ những kết quả nghiên cứu chúng tôi kiến nghị nên có những ứng dụng ngay với những mảng việc sau:

- Ứng dụng ngay hệ quản trị CDS/ISIS mới cho quản trị kho sách, tạp chí, các đề tài khoa học do viện quản lý.

- Lưu trữ tiếp tục các số của tờ Thông tin khoa học thống kê và thông tin phục vụ quản lý thống kê để có hệ thống.

- Điều tra lại để xây dựng các CSDL về các nhà nghiên cứu thống kê, các tổ chức nghiên cứu thống kê.

- Tiếp tục tổ chức lại việc khai thác các tư liệu trên mạng Internet, tìm ra cách thức khai thác hiệu quả các tư liệu này.

- Bắt đầu xây dựng kho tư liệu ảnh và đa phương tiện khác.

- Bắt đầu sưu tầm và xây dựng kho tư liệu pháp lý thống kê.

### **VÀI LỜI KẾT**

Tin học hóa và tự động hóa các khâu công tác thông tin tư liệu theo chuẩn quốc gia và quốc tế là cơ sở để tiến tới việc trao đổi và sử dụng chung nguồn lực thông tin trong phạm vi quốc gia và quốc tế, nâng cao hiệu quả và chất lượng hoạt động thông tin khoa học thống kê, đáp ứng kịp thời nhu cầu tin của người dùng tin. Vì thế đề tài hy vọng các kết quả sẽ được ứng dụng trong thực tiễn.

Cuối cùng, chúng tôi xin thay mặt toàn thể anh em tham gia đề tài chân thành cảm ơn các phản biện, các cộng tác viên đã nhiệt tình tham gia góp ý để đề tài được hoàn chỉnh về nội dung và văn phạm.

### **TÀI LIỆU THAM KHẢO**

1. Holdings Data Format. MARC21 format. <http://www.loc.gov/marc/holdings>.
2. Alan Hopkín. Standards for Serial Holding and for Serial Data in the Serials Analytic Record / Alan Hopkín.- American: Middlesex University, 2000.

3. Hướng dẫn xây dựng và khai thác CSDL chạy dưới Winisis.- Hà Nội, 1998.
4. Khổ mẫu MARC21 cho dữ liệu thư mục: Hướng dẫn áp dụng định danh nội dung.- Tập 1 và tập 2.- Hà Nội: Trung tâm Thông tin Khoa học và Công nghệ Quốc gia, 2004.
5. Mô tả các trường và hướng dẫn điền phiếu nhập tin cho các cơ sở dữ liệu tư liệu sử dụng chương trình CDS/ISIS / Trung tâm Thông tin tư liệu Khoa học và Công nghệ quốc gia.- Hà Nội, tháng 1/1995.
6. Nguyễn Đức Bình. Tập bài giảng khổ mẫu biên mục MARC21.- 2005.
7. Nguyễn Đức Bình. Ứng dụng khổ mẫu vốn tư liệu lưu giữ trong quản lý báo/tạp chí // Thư viện Việt Nam.- 2007.- Số 1.- Tr.30-36.
8. Nguyễn Thị Đông. Hoạt động thông tin tư liệu trong kỷ nguyên thông tin // Viện khoa học Thống kê: Sinh hoạt khoa học.- 2006.
9. Nguyễn Thị Đông. MARC và hoạt động thông tin-thư viện // Viện khoa học Thống kê: Sinh hoạt khoa học.- 2002.
10. Nguyễn Thị Đông. Tin học hóa công tác thông tin tư liệu khoa học thống kê // Viện khoa học Thống kê: Sinh hoạt khoa học.- 2007.
11. Phạm Sơn, Nguyễn Thị Đông. Tổ chức lại công tác thông tin-thư viện phục vụ cho hoạt động khoa học của ngành Thống kê // Thông tin Khoa học Thống kê, 2007.- Số 1.
12. Tiêu chuẩn Việt Nam. Hoạt động thông tin khoa học và tư liệu: Thuật ngữ và khái niệm cơ bản.
13. Hoàng Minh Thiện. Củng cố kho thông tin khoa học thống kê kết hợp tin học hóa//Thông tin Khoa học Thống kê 2007- số 4.