

SỐ HÓA BÁO CÁO KẾT QUẢ NGHIÊN CỨU TẠI TRUNG TÂM THÔNG TIN KHCN QUỐC GIA - NHỮNG KINH NGHIỆM THỰC TIỄN

ThS. Phan Huy Quế

Trung tâm Thông tin KHCN Quốc gia

Luận chứng về sự cần thiết số hoá các KQNC. Phân tích các đặc điểm của tài liệu KQNC và ảnh hưởng của nó tới việc số hoá. Trình bày các bước trong giải pháp kỹ thuật để số hoá KQNC tại Trung tâm thông tin KHCN Quốc gia.

I. Sự cần thiết phải số hóa kết quả nghiên cứu (KQNC) tại Trung tâm Thông tin KHCN Quốc gia

Nghị định của Chính phủ số 159/2004/NĐ-CP ngày 31/8/2004 về hoạt động thông tin khoa học và công nghệ quy định Trung tâm Thông tin Khoa học và Công nghệ Quốc gia (sau đây gọi là Trung tâm) là cơ quan Nhà nước có thẩm quyền về đăng ký và lưu giữ KQNC. Đồng thời, cũng quy định cần ứng dụng công nghệ thông tin và các công nghệ tiên tiến khác vào phát triển nguồn thông tin KHCN, trong đó có KQNC.

KQNC là một trong các dạng tài liệu được xếp vào loại "tài liệu xám", loại tài liệu có giá trị khoa học rất cao và được sử dụng lâu dài. Mặt khác, báo cáo KQNC là một trong những tài liệu phản ánh rõ nét nhất các tính chất cơ bản của hoạt động nghiên cứu khoa học và phát triển công nghệ. Bằng việc số hóa KQNC, các tính chất này có điều kiện thuận lợi để phát huy, làm tăng hiệu quả sử dụng KQNC. Cụ thể là:

- *Tính hữu dụng:* việc tạo ra các kết quả có khả năng áp dụng vào thực tế là một trong những mục tiêu chủ yếu của hoạt động nghiên cứu khoa học. Các kết quả tạo ra có thể được ứng dụng tại nhiều địa chỉ khác

nhau và trong các môi trường tự nhiên, xã hội khác nhau. Số hóa KQNC tạo điều kiện thuận lợi để nhân bản và phổ biến KQNC, qua đó phát huy tính hữu dụng của chúng;

- *Tính kế thừa:* Kế thừa các kết quả nghiên cứu giúp tiết kiệm công sức và kinh phí nghiên cứu. Một trong những điều kiện thực hiện tính kế thừa KQNC là cần lưu giữ chúng lâu dài, tin cậy, đồng thời phổ biến một cách nhanh chóng và dễ dàng tới người sử dụng. Số hóa là một trong những biện pháp hữu hiệu đáp ứng các điều kiện trên;

- *Tính trữ của KQNC:* Do một trong những phương pháp nghiên cứu là dự báo, tiên đoán hướng phát triển của đối tượng nghiên cứu, nên có những KQNC không thể phát huy tính hữu dụng ngay tại thời điểm kết thúc nghiên cứu. Vì vậy, cần số hóa KQNC để bảo đảm lưu giữ lâu dài và tin cậy, nhằm phát huy tính hữu dụng của chúng trong tương lai.

Áp dụng công nghệ thông tin trong hoạt động thông tin tư liệu là xu hướng tất yếu hiện nay. Việc này mang lại lợi ích thiết thực không chỉ cho các cơ quan thông tin trong việc thu thập, lưu giữ và phổ biến thông tin mà còn giúp người dùng tin tiếp cận và khai thác thông tin một cách nhanh chóng và thuận tiện nhất. Số hóa KQNC là

một trong các công việc cụ thể áp dụng CNTT để phục vụ một cách hiệu quả đối tượng người dùng tin là cán bộ NCPT trong phạm vi cả nước.

II. Số hóa KQNC tại Trung tâm - Thực tế và một số kinh nghiệm

- Giai đoạn từ 1986 đến 1990;

- Giai đoạn từ 1991 đến nay.

Đặc điểm hiện trạng vật lý tài liệu

Các đặc điểm hiện trạng vật lý của từng giai đoạn ảnh hưởng đến việc số hoá được trình bày trong Bảng 1 dưới đây.

Đặc điểm vật lý	Giai đoạn		
	Từ 1985 trở về trước	1986-1990	Từ 1991 đến nay
Giấy	Phần lớn là giấy mỏng (ơ-luya), màu xám đen, dính nhau, dễ rách	Giấy cứng nhiều hơn, nhưng đa phần màu giấy còn xỉn, giòn, dễ rách	Giấy tương đối đảm bảo tiêu chuẩn, trắng, đúng khổ A4. Tuy nhiên trọng lượng (g/m) khác nhau
Chữ viết	Chủ yếu là đánh máy chữ, bản giấy than hoặc đen, hoặc xanh, chữ mờ, khó đọc	Đánh máy chữ + máy tính, chủ yếu là in kim, chữ mờ, khó đọc	
Đóng quyển	Không có bìa hoặc giấy bìa chất lượng kém. Khổ báo cáo dài hơn A4 khoảng 3 cm	Phần lớn là bìa mềm, còn nhiều khổ khác nhau	Chủ yếu là bìa cứng, khổ tương đối thống nhất (A4)
Khối lượng báo cáo	Ít trang, thông thường khoảng 25-45 trang, cá biệt từ 5-10 trang	Số lượng trang nhiều hơn, trung bình khoảng 50-100 trang	Số lượng trang trung bình 100-125 tr/1 báo cáo. Cá biệt có báo cáo > 800 trang.

1. Những đặc điểm của KQNC ảnh hưởng đến việc số hóa

Các đặc điểm ảnh hưởng đến việc số hóa chủ yếu tập trung ở 2 khía cạnh. Đó là đặc điểm hiện trạng vật lý và đặc điểm trình bày thông tin của các báo cáo KQNC. Tuy nhiên, mức độ ảnh hưởng của các đặc điểm này phụ thuộc vào từng giai đoạn thời gian hình thành tài liệu. Qua thực tế xử lý KQNC để số hóa, chúng tôi tạm chia 3 giai đoạn hình thành KQNC như sau:

- Giai đoạn từ 1985 trở về trước;

Đặc điểm trình bày thông tin

Có thể kể ra một vài đặc điểm trình bày thông tin của các báo cáo KQNC theo từng giai đoạn thời gian ảnh hưởng đến quá trình số hoá như sau:

- Từ 1985 trở về trước: Chủ yếu là văn bản (text) và biểu bảng. Đôi khi có hình vẽ, bản đồ nhưng thường là vẽ tay bằng bút mực, bút chì hoặc bút bi. Cá biệt có báo cáo xen lẫn ảnh minh họa đen trắng, ảnh được chụp riêng và dán vào báo cáo.

- Từ 1986-1990: Các biểu bảng và hình

vẽ nhiều hơn, chủ yếu là vẽ bằng máy tính. Nhiều báo cáo có phụ lục là các bản thiết kế khổ A3 đóng quyển trực tiếp cùng báo cáo.

- Từ 1991 đến nay: Báo cáo có xen kẽ các ảnh màu minh họa và được "trộn" cùng văn bản. Đa phần ảnh màu thường là biểu đồ, điều kiện tự nhiên, tài nguyên khoáng sản, diễn biến tính chất của đối tượng nghiên cứu, v.v..., cần dùng màu sắc để phân biệt.

Tất cả những đặc điểm hiện trạng vật lý và trình bày thông tin kể trên có ảnh hưởng rất lớn đến quá trình và chất lượng số hóa, đòi hỏi người tổ chức và thực hiện số hoá phải xác định các yêu cầu và các giải pháp kỹ thuật phù hợp để bảo đảm hiệu quả số hóa.

2. Một vài kinh nghiệm về số hóa KQNC tại Trung tâm

Xác định các yêu cầu số hóa

Xuất phát từ đặc điểm hình thành tài liệu và giá trị của loại hình tài liệu KQNC cũng như các yêu cầu đối với việc tổ chức lưu giữ và sử dụng loại tài liệu này, chúng tôi đã xác định một số yêu cầu cơ bản về số hoá KQNC như sau:

- Bản số hoá phải bảo đảm chính xác tuyệt đối so với nội dung và hình thức trình bày của tài liệu gốc;

- Phải bảo đảm người quản trị và người sử dụng không thể sửa chữa, thêm bớt hoặc cắt dán các thông tin từ bản số hóa KQNC;

- Kho số hóa phải phản ánh trung thành kho tài liệu gốc để dễ đối chiếu, so sánh và kiểm kê khi cần thiết.

- Báo cáo KQNC số hóa phải đọc được dễ dàng, dung lượng nhỏ, thuận tiện cho khai thác và chuyển giao.

Căn cứ vào những yêu cầu cơ bản nêu trên, chúng tôi đã lựa chọn, áp dụng các giải pháp kỹ thuật cần thiết để số hóa KQNC.

Lựa chọn và áp dụng các giải pháp kỹ thuật

b.1. Lựa chọn khổ mẫu lưu giữ KQNC dạng số hóa

Trung tâm lựa chọn Adobe Acrobat để quản trị các báo cáo KQNC số hóa. Lý do: Acrobat là phần mềm xử lý ảnh quét nên những thông tin được số hoá sẽ đảm bảo tuyệt đối trung thành với bản gốc, phù hợp với yêu cầu số hóa KQNC. Hơn nữa, so với các phần mềm cùng chức năng xử lý ảnh quét khác như Photoshop, PaintshopPro,

FineReader, Fireworks... thì các tệp Acrobat (phần mở rộng là.pdf) có dung lượng nhỏ hơn, thao tác xử lý nhanh hơn. Đồng thời, Acrobat cũng thích hợp với nhiều loại máy quét hơn.

Trung tâm cũng lựa chọn chế độ quét đen trắng là chủ yếu để bảo đảm yêu cầu hạn chế dung lượng tệp quét, thuận tiện cho khai thác và chuyển giao.

b.2. Đặt tên tệp số hóa

Mỗi báo cáo KQNC được quét thành một tệp và được đặt tên theo số đăng ký của KQNC trong sổ đăng ký cá biệt, đồng thời cũng là ký hiệu (mã số) xếp giá của KQNC trong kho tài liệu gốc. Việc này nhằm bảo đảm yêu cầu kho số hóa luôn phản ánh trung thành kho tài liệu gốc, tạo thuận lợi cho việc đối chiếu, so sánh và kiểm kê khi cần thiết.

b.3. Lựa chọn cách quét KQNC

Đặc điểm các báo cáo KQNC, nhất là các báo cáo hình thành giai đoạn từ 1991 trở lại đây là đa phần có khối lượng lớn, đóng quyển bìa cứng. Vì vậy, chúng tôi đã chọn cách quét là quét từng tờ của báo cáo. Các quyển báo cáo sẽ được tháo rời một

cách cẩn thận. Sau khi số hóa sẽ được mang đến các cơ sở đóng quyển chuyên nghiệp để đóng lại như cũ.

b.4. Lựa chọn chế độ quét

Đây là vấn đề phức tạp nhất trong quá trình số hóa KQNC, bởi lẽ không thể áp dụng thống nhất độ phân giải, độ nét cho tất cả các báo cáo KQNC. Đồng thời cũng không thể chỉ áp dụng chế độ quét đen trắng đối với các KQNC có ảnh màu mà màu sắc trong đó giữ vai trò quan trọng trong việc phân biệt thông tin. Để bảo đảm các yêu cầu số hóa, chúng tôi giải quyết như sau:

- Áp dụng một chế độ quét thống nhất đối với các báo cáo in lade (nguyên bản hoặc bản phôtô có chất lượng). Với các báo cáo khác, phải quét thử một số trang để chọn độ nét, độ phân giải thích hợp. Thực tế, đối với loại tài liệu này, người quét phải thường xuyên thay đổi chế độ quét.

- Đối với các báo cáo có hình vẽ, bản đồ hoặc ảnh màu, nếu màu sắc giữ vai trò quyết định trong việc phân ánh thông tin thì phải quét màu các hình vẽ, bản đồ hoặc ảnh này (thường là quét riêng), sau đó chèn vào đúng vị trí của chúng trong

báo cáo.

- Đối với các hình vẽ, bản đồ kích thước vượt khổ A4 (khổ giấy máy quét cho phép), cần chia ra quét thành nhiều phần hợp với khổ giấy cho phép của máy quét. Các phần được đánh số thứ tự để người sử dụng dễ theo dõi.

b.5. Làm mục lục đọc nhanh (bookmarks)

Do đa phần các báo cáo KQNC đều có khối lượng lớn, nên người sử dụng sẽ rất khó tìm thông tin khi phải lần lượt lật từng trang của báo cáo trên màn hình. Để khắc phục tình trạng này, chúng tôi đã sử dụng chức năng làm mục lục đọc nhanh (bookmarks) của Acrobat. Mục lục đọc nhanh về bản chất là danh mục các chương, phần, đề mục... của tài liệu nhằm giúp người đọc nhanh chóng tìm đến chỗ có nội dung cần đọc, rất tiện lợi đối với các tài liệu có khối lượng lớn như báo cáo KQNC. Khi làm mục lục đọc nhanh KQNC, Acrobat sẽ chia màn hình thành 2 phần: phần bên phải màn hình là toàn văn KQNC còn phần bên trái là mục lục đọc nhanh. Mục lục đọc nhanh thường được sắp xếp theo cây thư mục, tương ứng với chương, mục, đề mục báo

cáo KQNC. Người đọc chỉ cần nháy đúp chuột vào chương, mục cần đọc ở mục lục đọc nhanh là sẽ tìm được ngay chương, mục đó trong toàn văn báo cáo KQNC ở bên phải màn hình.

b.6. Ghi đĩa CD

Các tệp số hoá KQNC được ghi đĩa CD sau khi đã qua kiểm tra về số thứ tự trang quét, chất lượng tệp quét và làm mục lục đọc nhanh. Qua thực tế ghi và sử dụng tệp KQNC trên đĩa CD, chúng tôi rút ra 3 kinh nghiệm là:

- Không nên quá tận dụng dung lượng của đĩa. Mức ghi an toàn theo chúng tôi không nên vượt quá 600 MB đối với đĩa có dung lượng 650 MB;

- Không nên quá tận dụng tốc độ của đầu ghi. Tốc độ ghi an toàn theo chúng tôi không nên vượt quá 36 Mbs;

- Cần lựa chọn chế độ xác nhận lại dữ liệu sau khi ghi (chế độ Verify data) có trong các chương trình ghi đĩa. Việc này tuy có thể làm tăng thời gian ghi đĩa lên gấp đôi, nhưng bù lại, độ an toàn và chắc chắn của dữ liệu được tăng lên nhiều lần.

III. Thay lời kết

Số hoá tài liệu đang là một trong những hoạt động

"thời sự" của các cơ quan TT-TV hiện nay, song, đây là công việc tốn kém thời gian và tiền bạc. Một trong những biện pháp có thể hạn chế những tổn kém nói trên là phải lựa chọn chiến lược số hoá phù hợp, trên cơ sở đó đề ra các nguyên tắc số hoá đối với từng loại hình tài

liệu đồng thời với việc lựa chọn các giải pháp kỹ thuật tối ưu trong quá trình số hoá.

Tài liệu tham khảo

1. Luật Khoa học và Công nghệ
2. Nghị định số 159/2004/NĐ-CP ngày 31/8/2004 về hoạt động thông tin khoa học và công nghệ
3. Kỹ yếu Hội nghị chuyên đề

ngành về đăng ký, giao nộp và sử dụng KQNC theo tinh thần Luật Khoa học và Công nghệ / Trung tâm TTKHCNQG/ H., 2001

4. Tài liệu Hội thảo chuyên đề về quản trị và chia sẻ nguồn tin số hóa/ Cơ quan tổ chức: Trung tâm Thông tin KHCN Quốc gia , Quảng Bình, 27-28/7/2005

5. <http://www.adobe.com>

THẺ LỆ VIẾT BÀI CHO TẠP CHÍ THÔNG TIN & TƯ LIỆU

1. Bài gửi đăng trên tạp chí Thông tin & Tư liệu phải là bài chưa được đăng trên bất cứ tạp chí nào và nếu bài được đăng ở tạp chí Thông tin & Tư liệu thì tác giả cũng không được gửi đăng tiếp trên các tạp chí khác.

2. Tác giả viết bài nên thảo luận trước với bộ phận thường trực của Tạp chí về đề cương bài viết cũng như thời hạn cung cấp bài.

3. Bài viết phải sạch sẽ, rõ ràng, không quá 5000 chữ, mỗi bài viết nên có tóm tắt ngắn gọn bằng tiếng Việt.

4. Cuối bài viết tác giả cần ghi rõ họ tên, học vị, chức vụ, tên cơ quan làm việc. Ngoài ra, cần ghi địa chỉ để Tạp chí liên hệ về công việc và chi trả nhuận bút.

5. Tạp chí không trả lại bản thảo nếu không có yêu cầu đặc biệt của tác giả.

6. Thư từ, bài vở xin gửi về địa chỉ : **Tạp chí Thông tin & Tư liệu**

TRUNG TÂM THÔNG TIN KHCN QUỐC GIA

24 Lý Thường Kiệt, Hà Nội

Tel : 8257039; E-mail : lehoa70@vista.gov.vn

Tạp chí nhiệt liệt hoan nghênh sự cộng tác đông đảo bạn đọc !