

GIẢI PHÁP CHUYỂN ĐỔI DỮ LIỆU TỪ PHẦN MỀM MÃ NGUỒN MỞ GREENSTONE SANG PHẦN MỀM DSPACE

Phan Ngọc Đông
Trường Đại học Đà Lạt

1. Đặt vấn đề

Hiện nay, qua quá trình hoạt động, nhiều cơ quan TT-TV trong nước đang sử dụng các phần mềm khác nhau để quản lý nguồn tài liệu số. Phổ biến hơn cả là hai phần mềm mã nguồn mở Greenstone và DSpace. Mỗi phần mềm có một lợi thế riêng, song xu hướng hiện nay đối với các đơn vị là chuyển sang sử dụng phần mềm DSpace. Tuy nhiên không ít đơn vị trước đây đã từng sử dụng phần mềm Greenstone để quản lý cá bộ sưu tập số của đơn vị. Vậy làm thế nào để có thể chuyển đổi dữ liệu cũ từ phần mềm Greenstone sang phần mềm mới DSpace để giúp cho quản lý tập trung dữ liệu. Dưới đây, xin được giới thiệu quy trình để thực hiện việc chuyển đổi này.

Quy trình chuyển đổi

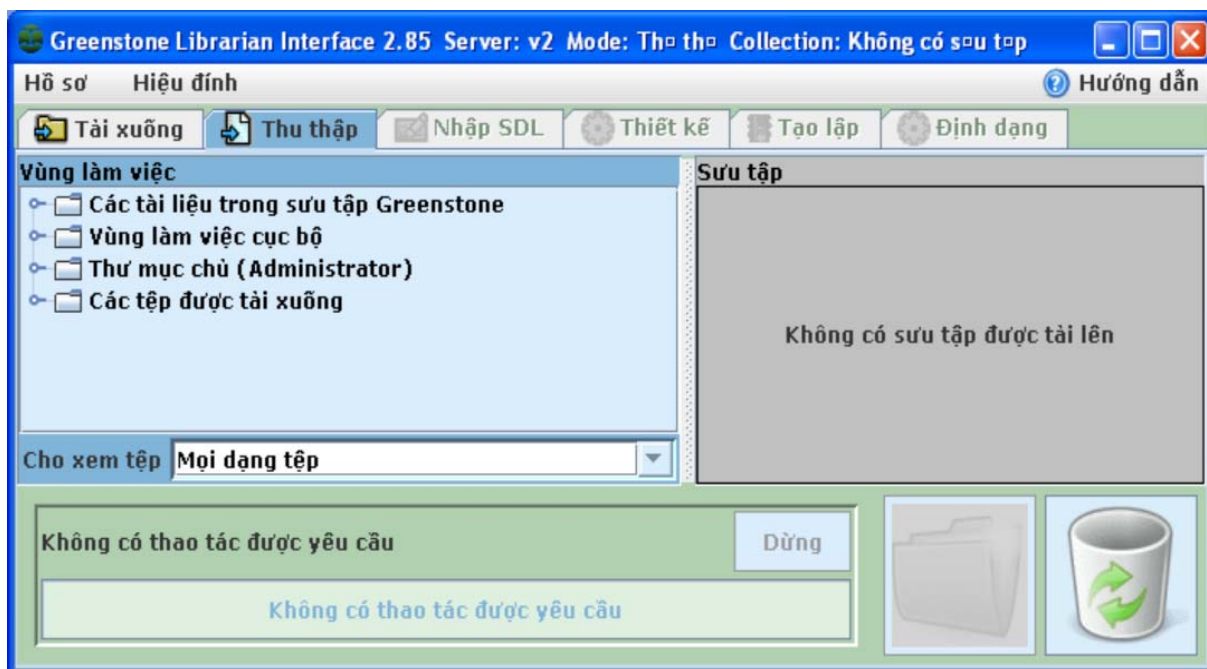
2.1. Xuất dữ liệu từ phần mềm Greenstone

Bước 1: Lựa chọn phiên bản Greenstone

Qua quá trình thử nghiệm, tác giả nhận thấy phiên bản hỗ trợ tốt nhất cho việc chuyển đổi là Greenstone – phiên bản 2.85. Đối với các đơn vị đang sử dụng phiên bản khác có thể tải bản cài đặt này tại địa chỉ <http://tinyurl.com/greenstone285> và tiến hành cài trên máy có các bộ sưu tập Greenstone.

Bước 2: Tùy chỉnh thông số cho bộ sưu tập chuẩn bị xuất

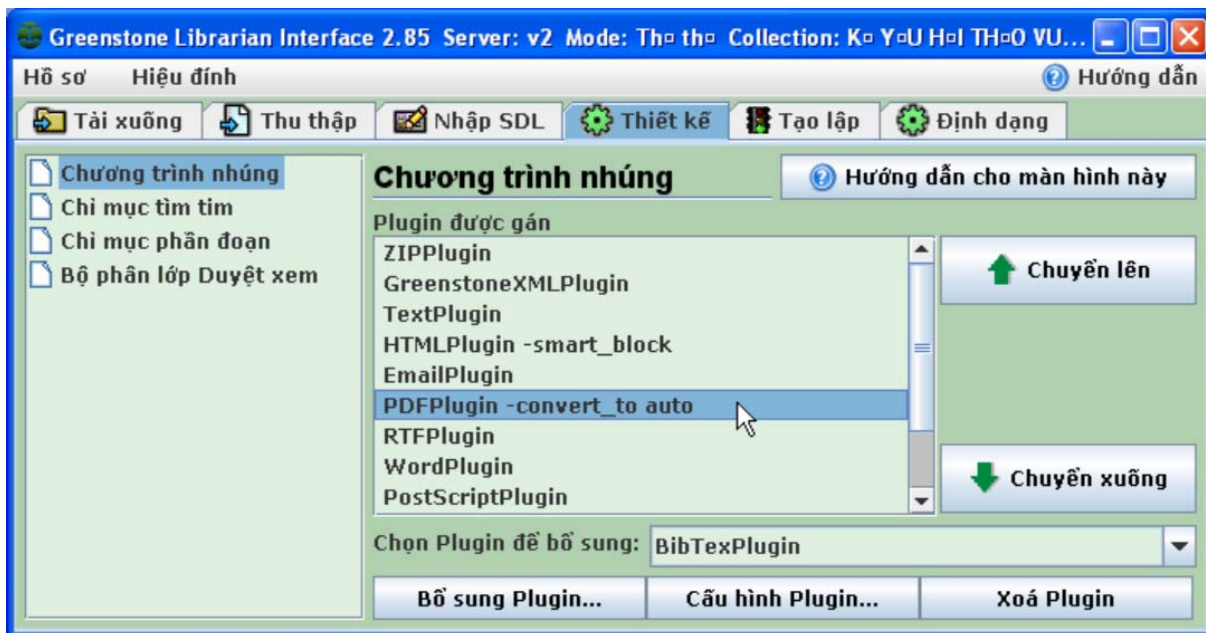
Khởi chạy chương trình Greenstone *Librarian Interface (GLI)* từ trình đơn hệ thống



Hình 1: Giao diện phần mềm Greenstone

Nghiệp vụ-Kỹ thuật

Bước 3: Tùy chỉnh thông số cho bộ sưu tập định dạng tài liệu số là PDF ta click chọn Plugin. Mở bộ sưu tập dự định chuyển đổi sang là: **PDF Plugin** -> Nhấp chọn **Cấu hình Plugin** -> Sau đó chọn thẻ **Thiết kế** -> Nếu **Plugin...**



Hình 2: Tùy chỉnh bộ sưu tập

Sau đó nhấp vào tùy chọn **noimages** để khi chuyển sang định dạng hình -> Nhấp chọn **Ok** xuất bộ sưu tập, các tập tin pdf không tự động



Hình 3: Tùy chỉnh cấu hình đối số cho PDF Plugin

Nghiệp vụ-Kỹ thuật

Bước 4: Xuất bộ sưu tập

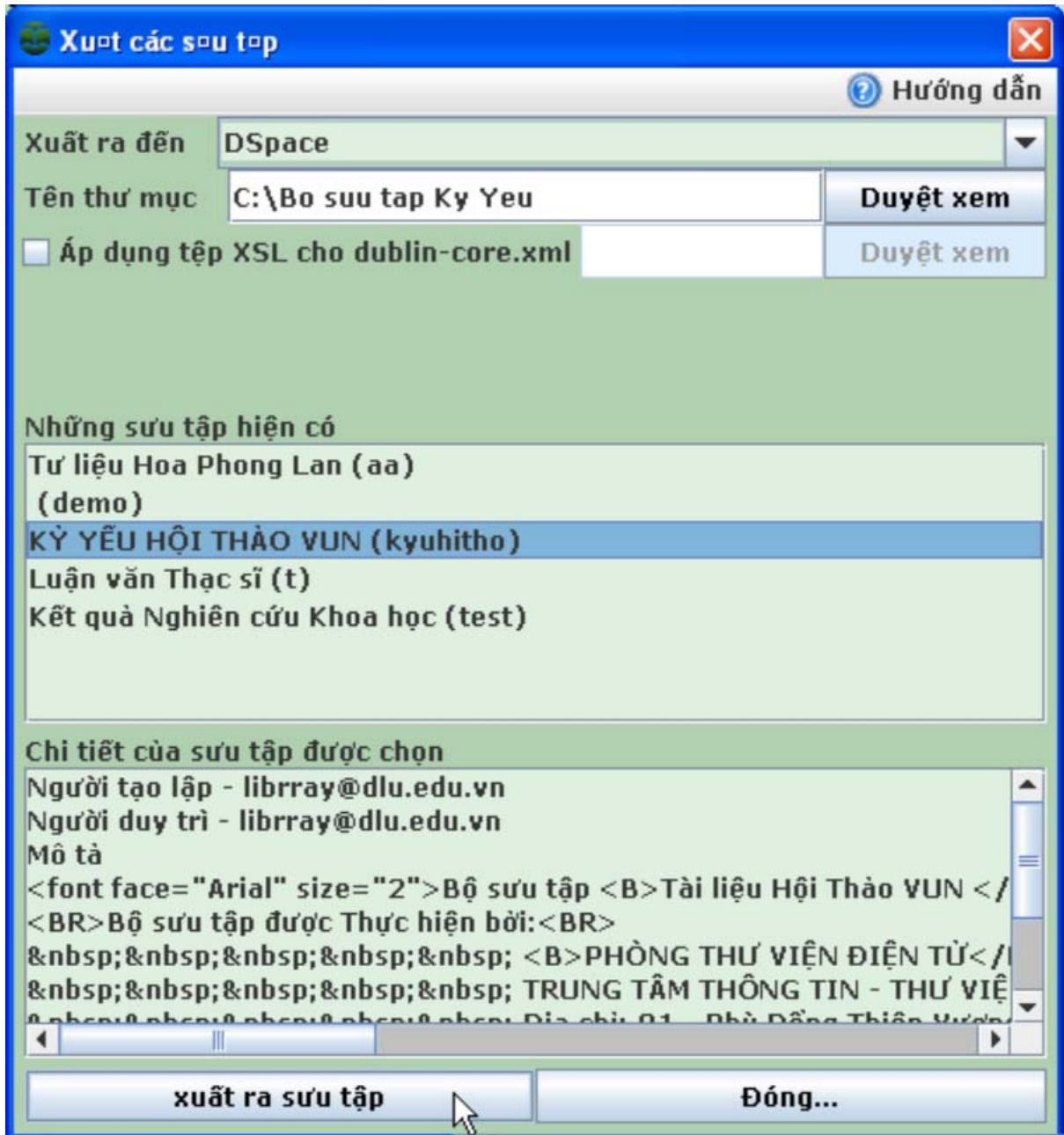
Vào thực đơn **Hồ sơ** -> Chọn **Xuất ra**, sau đó khai báo các tùy chọn trong hộp thoại hiện ra

- **Xuất ra đến:** chọn **DSpace**;

- **Tên thư mục:** Chọn thư mục sẽ lưu dữ liệu sau khi xuất ra;

- **Những sưu tập hiện có:** Nhấp chọn bộ sưu tập muốn xuất;

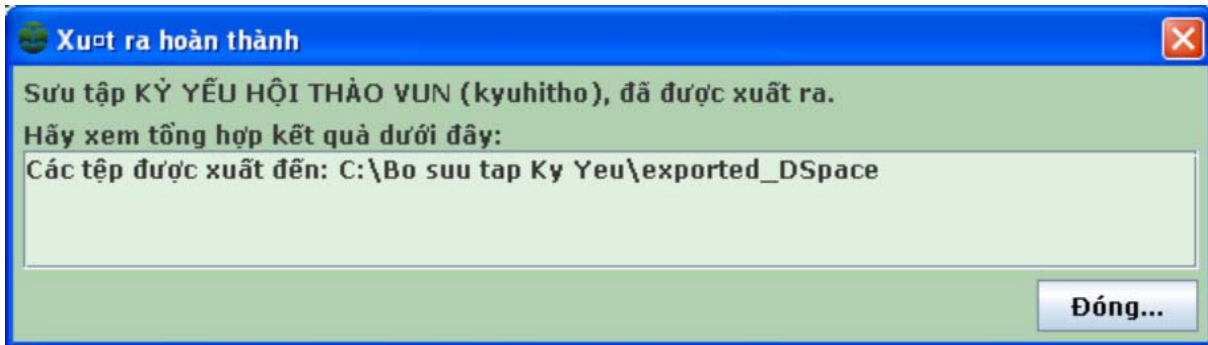
- Sau đó nhấp nút **xuất ra sưu tập**.



Hình 4: Khai báo thông số để xuất bộ sưu tập

Khi đó Greenstone sẽ tiến hành xuất dữ liệu ra hoàn tất, chương trình sẽ thông báo thư mục chúng ta vừa khai báo. Khi quá trình xuất

Nghiệp vụ-Kỹ thuật



Hình 5: Thông báo hoàn tất quá trình xuất bộ sưu tập

Kết quả của bước này, chúng ta sẽ được một thư mục chứa dữ liệu của bộ sưu tập (**C:\Bo sưu tap Ky Yeu\exported_DSpace**). Khi mở thư mục này, ta sẽ thấy có các thư mục con và các tập tin. Đối với các tập tin (*archiveinf-src.gdb, archiveinf-doc.gdb, OIDcount, earliestDatestamp*), ta sẽ xóa đi. Chúng ta chỉ quan tâm đến các thư mục con bên trong thư mục *exported_DSpace*. Mỗi thư mục con (*thường có tập tin toàn văn, tập tin dublin_core.xml và tập tin contents*) chứa thông tin của một tài liệu. Đây chính là dữ liệu của bộ sưu tập chuẩn bị cho quá trình chuyển sang phần mềm DSpace.

2.2. Nhập dữ liệu vào phần mềm DSpace

Bước 1: Chuẩn bị dữ liệu nhập

Chuyển thư mục dữ liệu đã xuất ra được từ

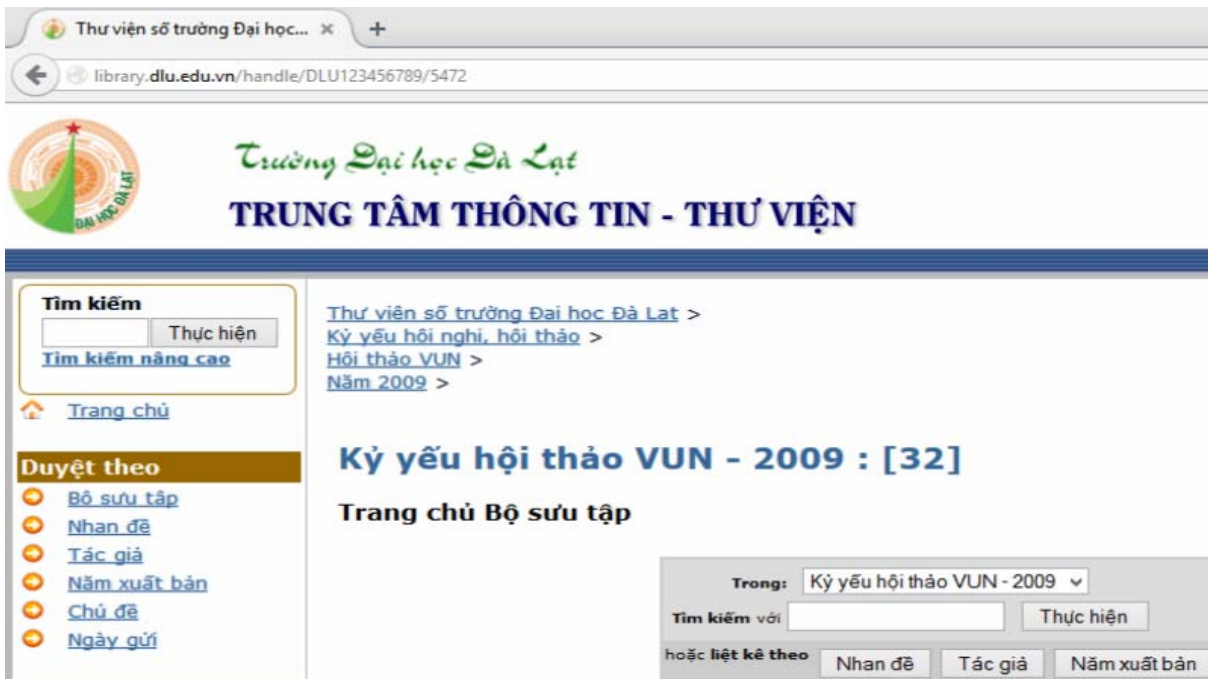
phần mềm Greenstone sang máy cài DSpace (*thư mục exported_DSpace ở bước trước*). Ví dụ ta đưa vào ổ đĩa C, khi đó ổ đĩa C sẽ có thư mục **C:\exported_DSpace**

Bước 2: Lấy định danh (ID) của bộ sưu tập cần nhập trên phần mềm DSpace

Để lấy định danh của bộ sưu tập chuẩn bị nhập trên DSpace, ta truy cập vào trang chủ của bộ sưu tập đó. Khi đó trên thanh địa chỉ của trình duyệt sẽ chứa thông tin định danh của bộ sưu tập.

Ví dụ: Trên thanh địa chỉ của bộ sưu tập có địa chỉ là: <http://library.dlu.edu.vn/handle/DLU123456789/5472>

thì định danh của bộ sưu tập là **DLU123456789/5472**



Hình 6: Trang chủ một bộ sưu tập trên phần mềm DSpace

Nghiệp vụ-Kỹ thuật

Bước 3: Tạo câu lệnh để nhập tài liệu vào bộ sưu tập được chọn

Ta khai báo các thông tin vào câu lệnh sau:

```
DSpace import --add --eperson=  
--collection= --source= --mapfile=
```

Trong đó:

- **eperson:** Địa chỉ email quản trị DSpace
- **collection:** Định danh bộ sưu tập sẽ nhập tài liệu
- **source:** Thư mục chứa tài liệu đã xuất từ Greenstone
- **mapfile:** Tên tập tin ánh xạ sẽ tạo ra khi nhập (Lưu ý: tên này không được trùng nhau giữa các lần nhập)

Kết quả khai báo ta sẽ được câu lệnh

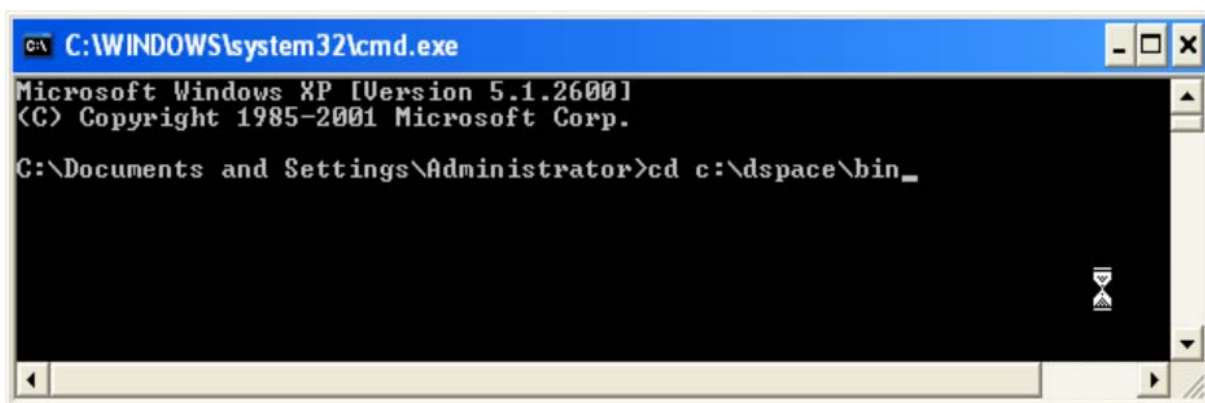
```
DSpace import --add --  
eperson=dongpn@dlu.edu.vn --collection  
=DLU123456789/5472 --source= C:\exported_  
DSpace --mapfile=KYHT
```

Bước 4: Thực thi câu lệnh để tiến hành nhập dữ liệu vào DSpace

→ Trong ví dụ này, giả sử phần mềm DSpace được cài đặt trong ổ đĩa C, có thư mục tên **DSpace**

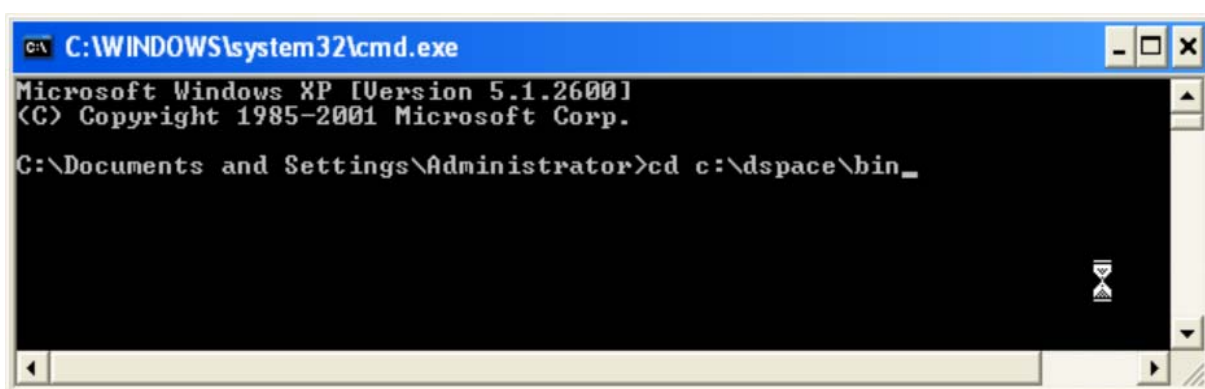
→ Nhấp Start -> chọn Run -> gõ lệnh CMD -> nhấp OK. Khi đó hệ thống sẽ mở một cửa sổ cho phép ta gõ các lệnh để thực thi.

→ Gõ lệnh **Cd c:\DSpace\bin** -> Nhấn Enter để về thư mục cài đặt DSpace



Hình 7: Màn hình thực thi lệnh

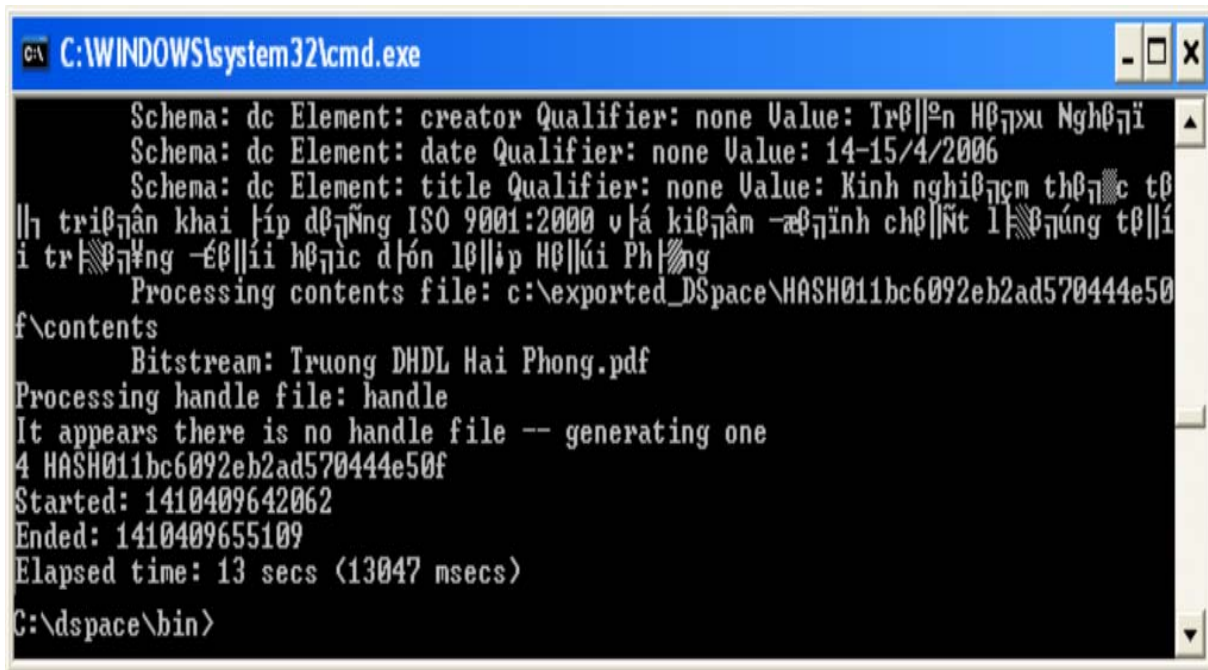
→ Gõ lệnh đã tạo được ở bước trên để tiến hành cập nhật vào DSpace



Hình 8: Màn hình thực thi lệnh cập nhật vào DSpace

Nghiệp vụ-Kỹ thuật

Khi đó quá trình cập nhật dữ liệu vào DSpace liệu sẽ được cập nhật vào phần mềm.
sẽ bắt đầu, chúng ta chờ cho quá trình hoàn tất, tài



```
C:\WINDOWS\system32\cmd.exe
Schema: dc Element: creator Qualifier: none Value: Trần Hữu Nghuyễn
Schema: dc Element: date Qualifier: none Value: 14-15/4/2006
Schema: dc Element: title Qualifier: none Value: Kinh nghiệm thực tế triển khai áp dụng ISO 9001:2000 và kiểm soát chất lượng ứng dụng CNTT trong hệ thống quản lý hồ sơ dự án lập Hồ sơ dự án
Processing contents file: c:\exported_DSpace\HASH011bc6092eb2ad570444e50f\contents
Bitstream: Truong DHDL Hai Phong.pdf
Processing handle file: handle
It appears there is no handle file -- generating one
4 HASH011bc6092eb2ad570444e50f
Started: 1410409642062
Ended: 1410409655109
Elapsed time: 13 secs (13047 msecs)
C:\dspace\bin>
```

Hình 9: Màn hình hoàn tất quá trình cập nhật vào DSpace

3. Kết luận

Với quy trình này, các đơn vị hiện đang sử dụng phần mềm Greenstone có thể dễ dàng chuyển dữ liệu đã có trước đây sang phần mềm mới mà không phải tốn quá nhiều thời gian cũng như chi phí cho việc xử lý lại từ đầu những dữ liệu đã có. Giải

pháp cũng cho phép các đơn vị mạnh dạn hơn trong việc chuyển đổi sang sử dụng phần mềm mã nguồn mở DSpace - một phần mềm được đánh giá cao trong việc quản lý tài liệu số và được đồng đảo cộng đồng thư viện trên toàn cầu sử dụng.

Tài liệu tham khảo

1. Phan Ngọc Đông (2012), DSpace - Giải pháp xây dựng thư viện số, Tạp chí Thư viện Việt Nam, 3(35), tr. 39-41.

Nam, 3(35), tr. 39-41.

2. Phan Ngọc Đông (2013), Tập bài giảng Cài

đặt và sử dụng phần mềm DSpace, Đà Lạt, Trường Đại học Đà Lạt.

3. <http://www.dspace.org>

4. <http://www.greenstone.org>