

## Chia sẻ kinh nghiệm

# GIẢI PHÁP CHUYỂN ĐỔI DỮ LIỆU CDS/ISIS SANG PHẦN MỀM NGUỒN MỞ GREENSTONE

ThS Lê Ngọc Diệp

Trường Đại học Nội vụ Hà Nội

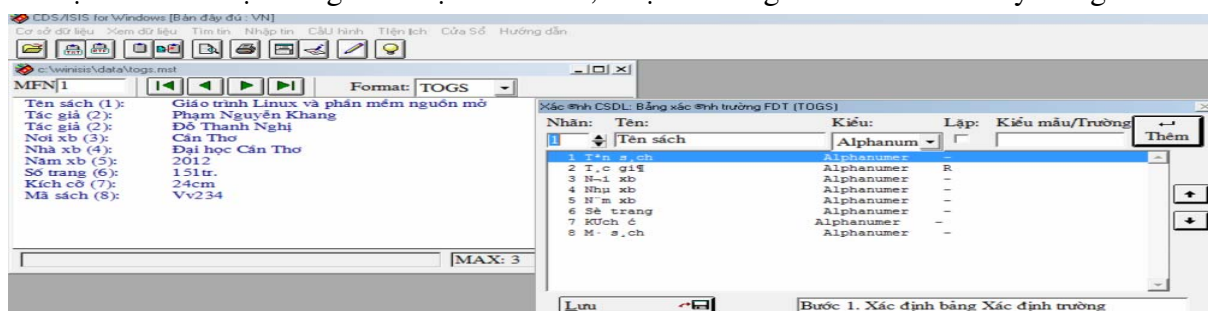
Như chúng ta đã biết, chương trình CDS/ISIS là một giải pháp lựa chọn để tin học hóa hoạt động thư viện ở Việt Nam trong nhiều năm qua, kể cả khi đã có nhiều phần mềm thư viện tích hợp trong nước và quốc tế như Libol, Ilib, VTLS... Phần mềm CDS/ISIS có nhiều ưu điểm như dễ cài đặt và sử dụng, đặc biệt là khả năng tìm kiếm nhanh, chính xác thông qua tệp đảo với các toán tử tìm tin. Tuy nhiên, chúng ta có thể thấy một hạn chế của CDS/ISIS là khó triển khai tra cứu trong môi trường mạng LAN, INTERNET, điều này làm giảm việc phổ biến tài liệu của thư viện cũng như việc tìm kiếm,

tiếp cận thông tin, tài liệu của bạn đọc.

Trong quá trình giảng dạy và nghiên cứu, cùng với những thử nghiệm thực tế, tác giả bài viết đã bước đầu thành công trong việc khắc phục hạn chế nói trên của CDS/ISIS thông qua chuyển đổi dữ liệu từ CDS/ISIS sang phần mềm thư viện số Greenstone để đưa lên Internet tra cứu. Dưới đây, xin được giới thiệu những bước cơ bản trong quy trình dữ liệu của CSD/ISIS sang phần mềm nguồn mở Greenstone:

### Bước 1: Xuất dữ liệu từ CDS/ISIS ra Excel

Ví dụ, chúng ta có một CSDL với bảng xác định trường FDT như trình bày trong Hình 1:

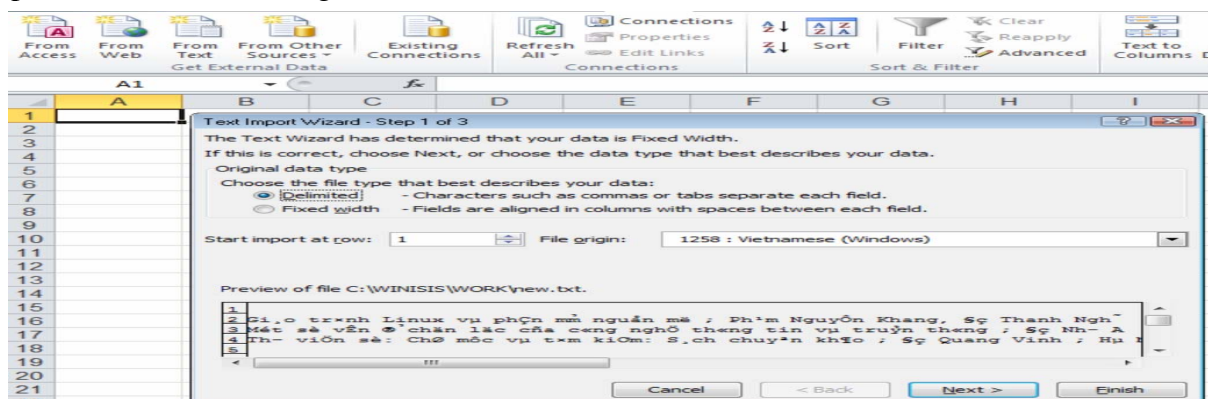


Hình 1: Bảng xác định trường của CSDL giả định

Tiếp theo chúng ta tạo lập một format in để lấy dữ liệu ra dạng text, thực hiện theo cách thông thường của CDS/ISIS. Ví dụ ta có đoạn format in như sau: V1' : ' , V2+l , l , V3' : ' , V4' : ' , V5' : ' , V6' : ' , V7' : ' , V8' : ' /. Dữ liệu của các trường trong CSDL/ISIS sẽ được phân cách với nhau bằng dấu hai chấm (:).

Sau khi có file dạng text được in ra bằng CDS/ISIS, theo format đã nói ở trên chúng ta mở file text bằng chương trình Excel bằng cách thực hiện các lựa chọn theo cách sau:

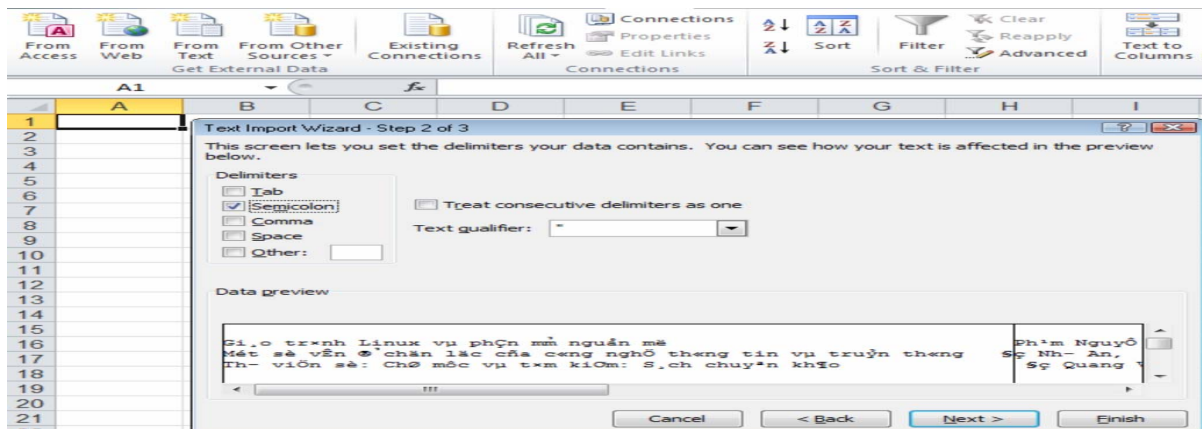
- Trong bước 1 của “Text Import Wizard”, chọn “Delimited” cho dữ liệu đầu vào (Hình 2).



Hình 2: Bước 1 của Text Import Wizard

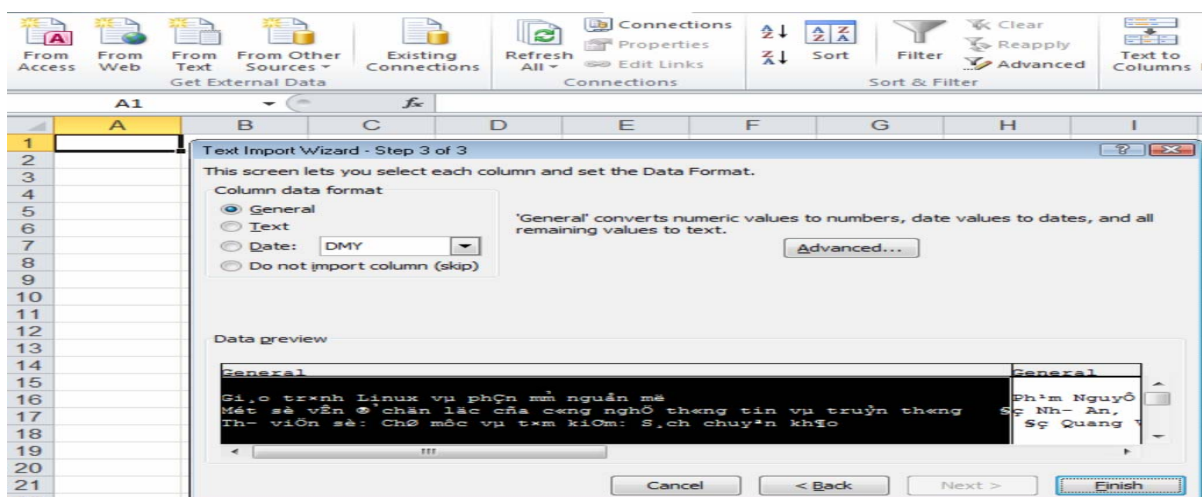
## Chia sẻ kinh nghiệm

- Ở bước 2, chọn “Semicolon” (hai chấm) là ký hiệu dấu phân cách để đưa văn bản vào từng cột (Hình 3).



Hình 3: Bước 2 của giao diện nhập văn bản

- Tại bước 3, chọn General (Hình 4) sau đó chọn Finish để đưa dữ liệu vào chương trình bảng tính Excel.



Hình 4: Bước 3 của giao diện nhập văn bản

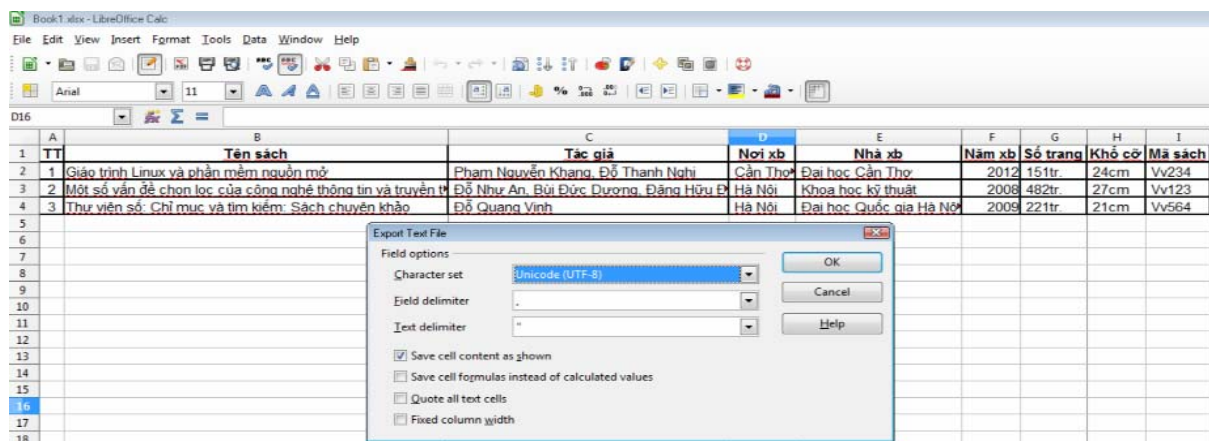
Sau khi nhập (import) văn bản vào Excel, chúng ta cần bổ sung thêm một dòng vào đầu bảng tính. Mỗi cột tương ứng với một trường của CSDL CDS/ISIS. Kết quả chuyển đổi dữ liệu từ CDS/ISIS được thể hiện trong phần mềm Excel được thể hiện trong Hình 5.

Book1 - Microsoft Excel								
File Home Insert Page Layout Formulas Data Review View								
Clipboard Font Alignment Number Conditional Formatting Styles Cells Editing								
A1	TT							
	A	B	C	D	E	F	G	H
1	TT	Tên sách	Tác giả	Nơi xb	Nhà xb	Năm xb	Số trang	Khổ cơ
2	1	Giáo trình Linux và phần mềm nguồn mở	Phạm Nguyễn Khang, Đỗ Thanh Nghị	Cần Thơ	Đại học Cần Thơ	2012	151tr.	24cm
3	2	Một số vấn đề chọn lọc của công nghệ thông tin và truyền thông	Đỗ Như An, Bùi Đức Dương, Đặng Hữu Đạo	Hà Nội	Khoa học kỹ thuật	2008	482tr.	27cm
4	3	Thư viện số: Chỉ mục và tìm kiếm: Sách chuyên khảo	Đỗ Quang Vinh	Hà Nội	Đại học Quốc gia Hà Nội	2009	221tr.	21cm
5								

Hình 5: Kết quả chuyển đổi dữ liệu từ ISIS vào bảng tính Excel

## Chia sẻ kinh nghiệm

**Bước 2:** Mở file Excel bằng phần mềm font chữ Unicode cho tiếng Việt. Khi lưu nguồn mở LibreOffice và lưu dưới dạng file chúng ta phải lựa chọn giá trị file phân cách bằng dấu phẩy (Comma Unicode (UTF-8) cho mục Character set delimited format) có đuôi .CSV để hỗ trợ (như Hình 6).



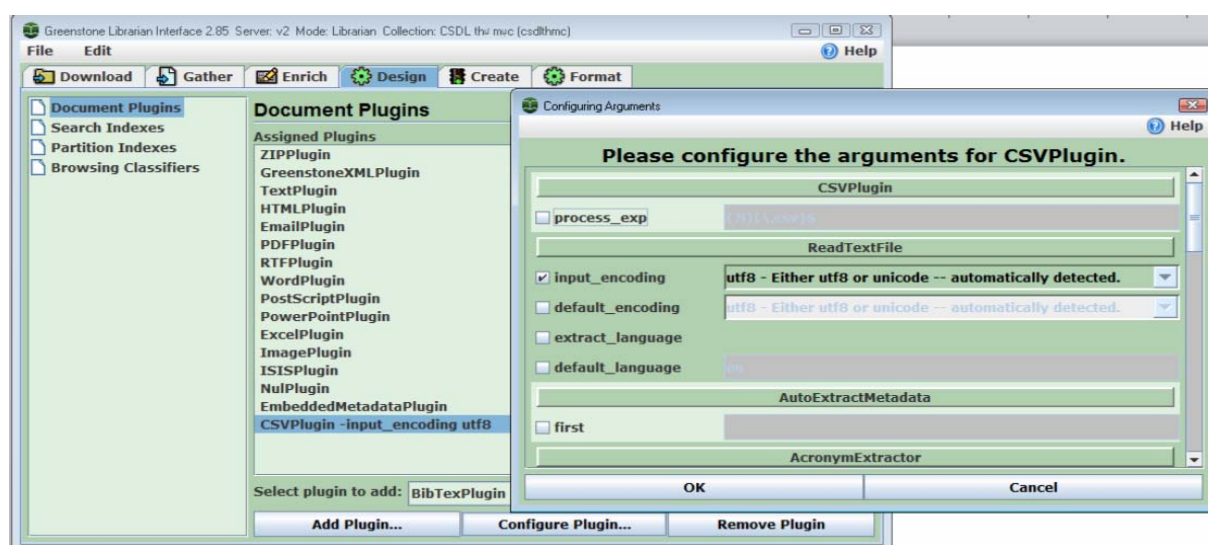
Hình 6: Chuyển dữ liệu từ Excel sang bảng tính của bộ phần mềm LibreOffice

**Bước 3:** Tạo bộ sưu tập bằng phần mềm nguồn mở Greenstone với dữ liệu có được từ Bước 2.

Trước khi đưa dữ liệu từ bảng tính LibreOffice, chúng ta phải tạo bộ sưu tập bằng phần mềm nguồn mở Greenstone theo cách thông thường. Một thao tác quan trọng cần phải thực hiện khi thiết kế CSDL trong Greenstone là bổ sung chương trình CSVPlugin (phần mềm Greenstone đã có sẵn

ISISPlugin nhưng chúng ta không sử dụng phần mềm này do không chuyển đổi trực tiếp từ CSDL CDS/ISIS vào Greenstone. Nếu dùng ISISPlugin thì không giải quyết được vấn đề về font chữ tiếng Việt hỗ trợ Unicode). Cấu hình code ngôn ngữ đầu vào của CSV-plugin như sau (Hình 7):

- Kích chọn Input\_encoding là UTF8.
- Không chọn các mục Default\_encoding; extract\_language và default\_language.



Hình 7: Cấu hình CSVplugin

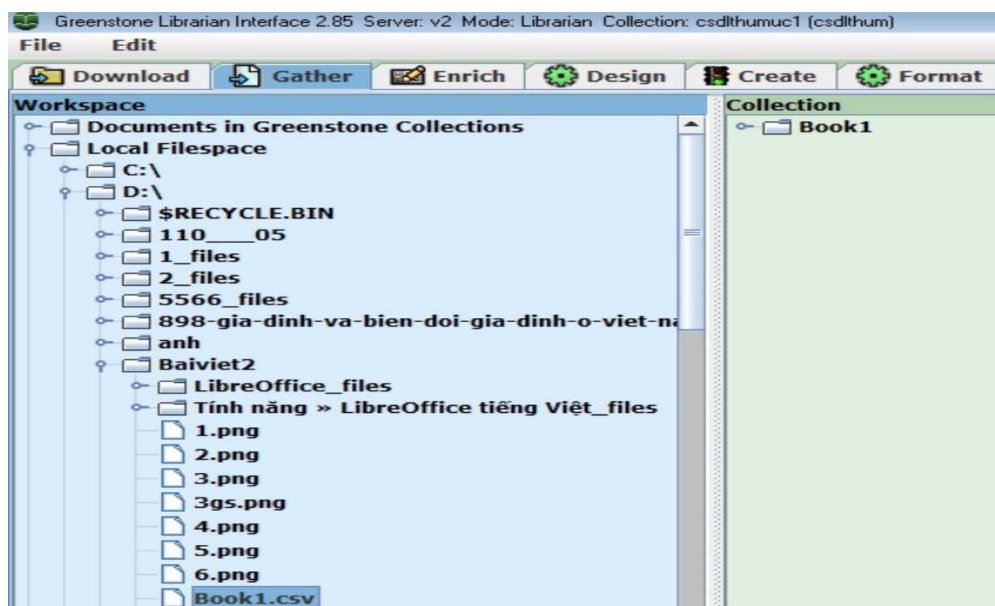


## Chia sẻ kinh nghiệm

**Bước 4:** Thực hiện chuyển dữ liệu (từ các cột trong bảng file .CSV sang các yếu tố tương ứng của Dublin Core trong Greenstone)

- Chúng ta sử dụng chức năng “thu thập”

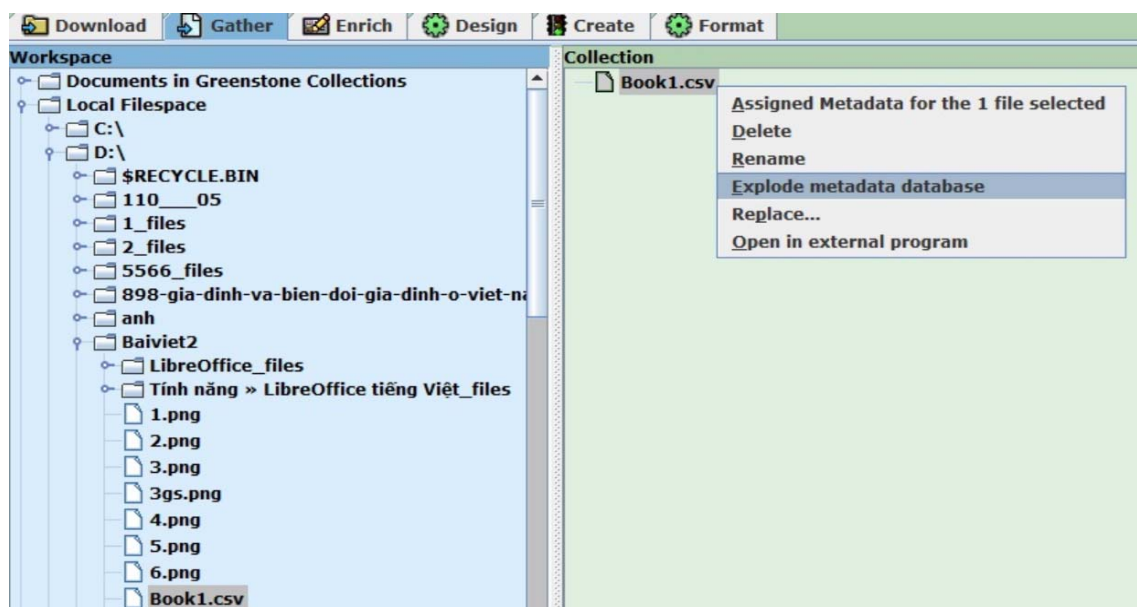
(gather) của Greenstone Librarian Interface để chọn file dữ liệu CSV mà ta tạo ra từ Libreoffice để đưa vào bộ sưu tập (ví dụ: tập Book1.csv) (Hình 8).



Hình 8: Cửa sổ “Thu thập” (Gather)

Tiếp đến thực hiện kết nối (mở bung hoặc Explode) giữa các cột (trường) của file CSV tương ứng với các Yếu tố (trường) của Dublin Core theo các bước sau:

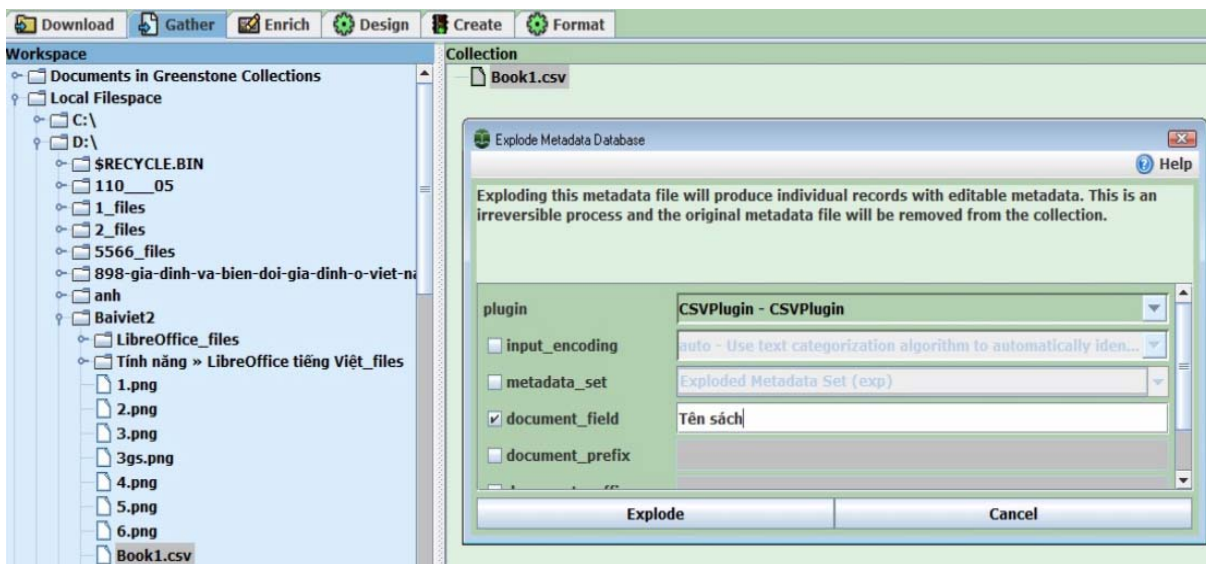
+ Kích chuột phải lên file CSV trong cửa sổ Collection (ví dụ ở đây file có tên là Book1), một hộp thoại xuất hiện, ta chọn mục **Explode metadata database** (Hình 9).



Hình 9: Hộp thoại chọn lựa phương thức mở bung dữ liệu

## Chia sẻ kinh nghiệm

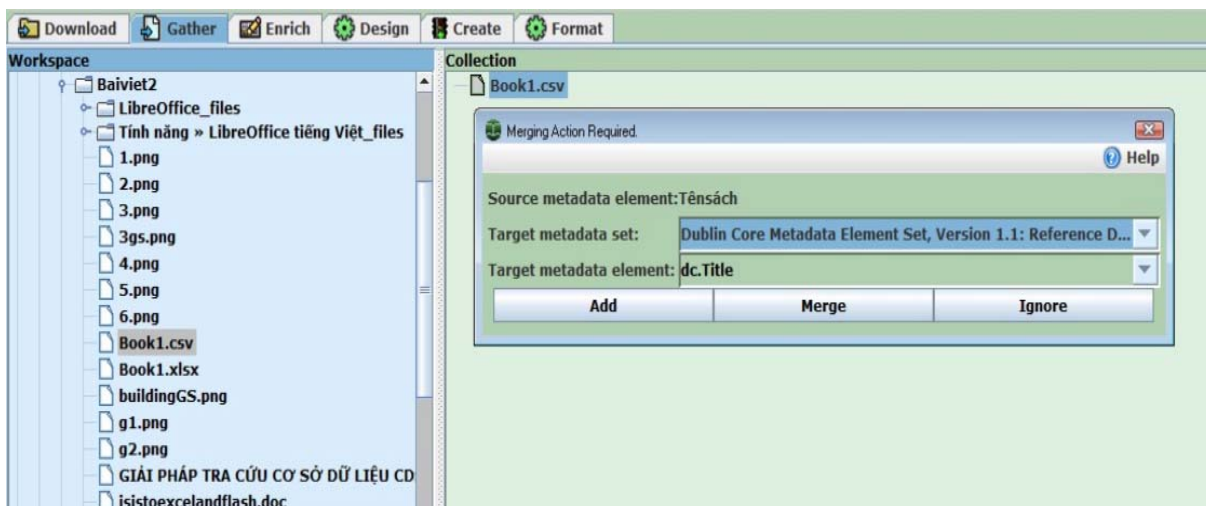
+ Sau khi chọn Explode Metadata một cột (trường) của file CSV (ví dụ nhập tên database, một hộp thoại mới xuất hiện, ta chọn cột Tên sách) (Hình 10) và kích chuột chọn chọn mục **document\_field** và nhập tên của Explode.



Hình 10: Thí dụ liên kết dữ liệu từ một cột của bảng tính

+ Sau khi kích chuột chọn **Explode**, một ví dụ, kết nối giữa Tên sách với dc.Title) (Hình 11).  
hộp thoại mới xuất hiện, cho chúng ta lựa chọn Ta chọn nút **Merge** để kết nối.

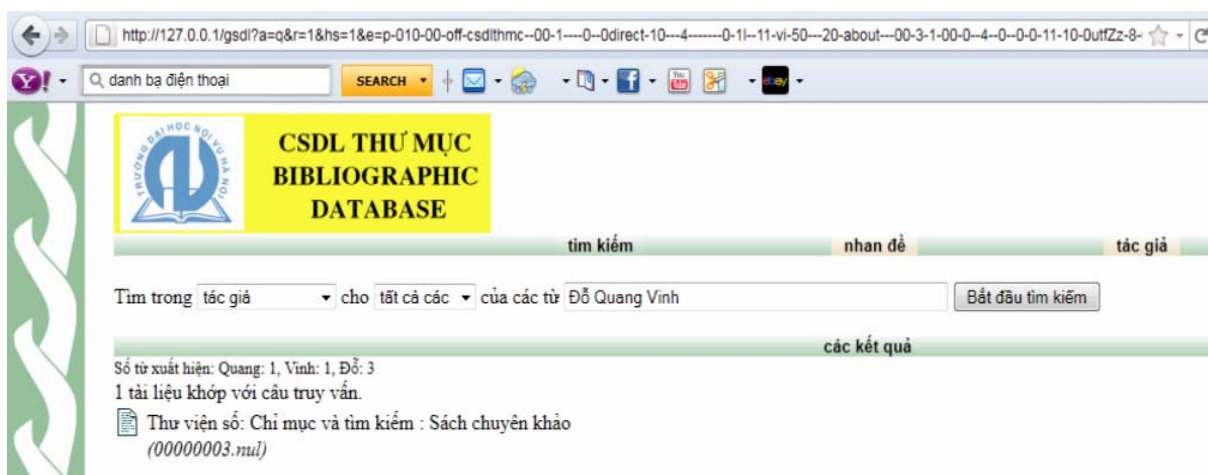
việc kết nối giữa các cột (trường) của file CSV + Thao tác này được lặp lại với các trường  
với các Yếu tố (trường) của Dublin Core (trong còn lại.



Hình 11: Giao diện được kết nối giữa “Tên sách” với “dc.Title”

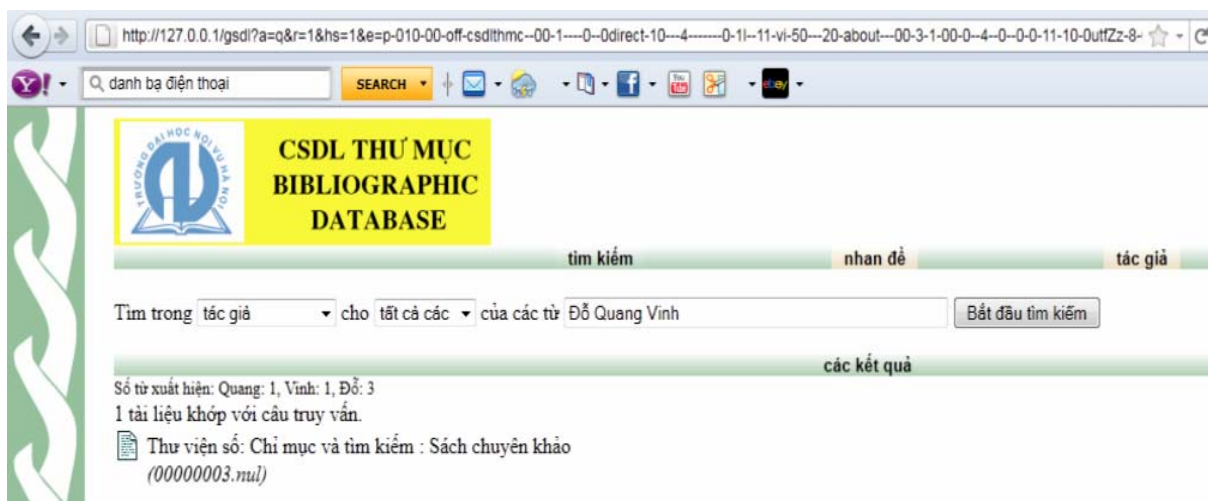
Sau khi thực hiện kết nối tất cả các trường, thường của Greenstone để có được bộ sưu tập.  
chúng ta đã có dữ liệu cho các Yếu tố của Bộ sưu tập này trong phần mềm  
Dublin Core. Phần tiếp theo là thực hiện các Greenstone có thể được đưa lên và tra cứu trên  
bước tạo lập một bộ sưu tập theo cách thông Internet một cách dễ dàng.

## Chia sẻ kinh nghiệm



**Hình 12:** Màn hình kết quả tra cứu từ sưu tập dữ liệu từ CSDL ISIS sau khi chuyển sang Greenstone

Kết quả xem chi tiết một biểu ghi (một tài liệu) trong sưu tập của Greenstone được trình bày như trong Hình 13.



**Hình 13:** Kết quả hiển thị một biểu ghi của ISIS sau khi chuyển sang Greenstone

Trên đây là một trong những giải pháp hữu ích dành cho những thư viện đang có dữ liệu từ CDS/ISIS, đã làm quen với phần mềm nguồn mở Greenstone và phần mềm nguồn mở Libreoffice trong việc giải quyết bài toán đưa CSDL thư mục từ CDS/ISIS lên tra cứu

trên mạng Internet. Giải pháp này cho phép cán bộ thư viện chủ động sử dụng các phần mềm nguồn mở nói trên mà không vi phạm về bản quyền, góp phần giải quyết được phần nào về vấn đề tài chính trong điều kiện của các thư viện Việt Nam hiện nay.

### Tài liệu tham khảo

1. <http://www.youtube.com/watch?v=hksa7wNO3-s> (Hướng dẫn chuyển dữ liệu CDS/ISIS sang Excel)
2. <http://frazernash-usa.com/digital/Greenstone-Imports.htm> (Hướng dẫn tạo bộ sưu tập bằng Greenstone với dữ liệu từ Excel)
3. <http://vi.libreoffice.org/tinh-nang/> (Giới thiệu về phần mềm nguồn mở Libreoffice)