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

ThS Lê Naoc Diêp

hư chúng ta đã biết, chương trình CDS/ISIS là một giải pháp lựa chọn để tin học hóa hoat động thư viên ở Viêt 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 dung, đặc biệt là khả năng tìm kiếm nhanh, chính xác thông qua têp đảo với các toán tử tìm tin. Tuy nhiên, chúng ta có thể thấy một han chế của CDS/ISIS là khó triển khai tra cứu trong môi trường mạng LAN, INTERNET, điều này làm giảm việc phổ biến tài liêu của thư viên cũng như viêc tìm kiếm, Trường Đại học Nội vụ Hà Nội

tiếp cận thông tin, tài liệu của bạn đọc.

Trong quá trình giảng dạy và nghiên cứu, cùng với những thử nghiêm 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 liêu từ CDS/ISIS sang phần mềm thư viên số Greenstone để đưa lên Internet tra cứu. Dưới đây, xin được giới thiêu những bước cơ bản trong quy trình dữ liêu của CSD/ISIS sang phần mềm nguồn mở Greenstone:

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

Ví du, chúng ta có môt CSDL với bảng xác đinh trường FDT như trình bày trong Hình 1:



Hình 1: Bảng xác đinh trường của CSDL giả đinh

Tiếp theo chúng ta tạo lập một format in để lấy dữ liêu ra dang text, thực hiên theo cách thông thường của CDS/ISIS. Ví dụ ta có đoạn format in như sau: V1': ', V2+|, |, 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 chon theo cách sau:

- Trong bước 1 của "Text Import Wizard", chon "Delimited" cho dữ liêu đầu vào (Hình 2).



Hình 2: Bước 1 của Text Import Wizard

THÔNG TIN và TƯ LIỆU - 3/2014

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

From Access	From Web	From From Oth Text Sources Get External Data	er Existin Connecti	g Refre	Connections	ens 2↓ [ ∡↓	Sort & F	Clear Reapply Advanced	Text to Columns E
	A1	- (-	f.s.						
100	A	В	C	D	E	F	G	н	1
2 3 4 5 6 7 8 9 10 11 12		This screen le below. Delimiters Delimiters Semicolo Comma Space Qther:	Text	delimiters yo Treat consec t gualifier:	ur data contains. 1 cutive delimiters as	rou can see one	how your text is	s affected in the p	review
13 14 15 16 17 18 19 20 21 22		Si.o trx Mát sà v Th- viôn	nh Linux v în G`chin : sè: Chơ m ""	μ phÇn m lắc cña c ôc vμ t×r	nguản më cang nghỗ tha n kiOm: S <sub>s</sub> ch o Cancel	ng tin vµ chuy*n kh	a truỷn th≪n N⊈o Back	ng Sç Nh-A Sç Nh-A Sç Qua	nuyô ng t Finish

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 bång tính Excel.

From Access	From Web	From From Other Text Sources - Get External Data	Existing Connection	Refres All *	Connections Properties Edit Links Connections		ort Filter	Clear Reapply Advanced	Text to Columns [
	A1	- (-	f <sub>x</sub>						
-	А	В	С	D	E	F	G	н	1
2 3 4 5 6 7 8 9		This screen lets y Column data for <u>© G</u> eneral <u>Text</u> <u>Date:</u> DN <u>© Do not impo</u>	1Y volumn (skip)	column an 'Gene remai	d set the Data Forma ral' converts numeric ining values to text.	at. values to	numbers, date	values to dates,	and all
10 11 12 13 14 15		Data greview						General	
16 17 18 19 20		Gi,o tr×nh Mét sè vÊn Th- viôn s	Linux vµ p ©'chân lão è: ChØ môc III	ohÇn mm c cña c∢ vµ t×m	nguản mà kng nghồ th≪ng kiOm: S,ch chu	tin vµ 1yªn kh¶	truýn th«n	phim Ng Sç Nh- A Sç Qua	uyô n, ng T +
21					Cancel	J [ < ₿	ack	Next >	Finish

Hình 4: Bước 3 của giao diện nhập văn bản

chúng ta cần bổ sung thêm một dòng vào đầu bảng tính. Mỗi cột tương ứng với một trường

Sau khi nhập (import) văn bản vào Excel, 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.

X   🖌 🤊 • 🕫 - 🗋   =	3	Book1 - Microsoft Exce	el			
File Home Insert Page Layout Formulas Data	Review View					
Cut Ariai - 11 - A A =	≡ 📄 ≫* 📑 Wrap Text	General	•		i 🗧 🖷	Σ AutoSum *
Paste → Format Painter B I U + B + 🎂 + ▲ + ≡	👅 🗐 🗊 🗊 Merge & Center *	🥶 - % · 號	.00 Con Form	ditional Format Cell natting - as Table - Styles -	insert Delete Format	Clear * Filter * 1
Clipboard 12 Font 12	Alignment 78	Number	19	Styles	Cells	Editing
A1 - 🕼 🎵						
AB	C		D	E	F G	H I
1 TT Tên sách	Tác giả	N	lo'i xb	Nhà xb	Năm xb Số trang	Khổ cỡ Mã sách
2 1 Giáo trình Linux và phần mềm nguồn mỡ	Phạm Nguyễn Khang, Đỗ Th	anh Nghị Cầ	ần Thơ	Đại học Cần Thơ	2012 151tr.	24cm Vv234
3 2 Một số vấn đề chọn lọc của công nghệ thông tin và truy	ên thông Đỗ Như An, Bùi Đức Dương,	Đặng Hừu Đạo Hả	à Nội	Khoa học kỹ thuật	2008 482tr.	27cm Vv123
4 3 Thư viện số: Chí mục và tìm kiếm: Sách chuyên khảo	Đỗ Quang Vinh	Ha	à Nội	Đại học Quốc gia Hà Nội	2009 221tr.	21cm Vv564
5						

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

THÔNG TIN và TƯ LIỆU - 3/2014

<u>Bước 2</u>: 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 Charater set delimited format) có đuôi .CSV để hỗ trợ (như Hình 6).

	Book1 stsx - LibreOffice Calc									
Eile	e Edit View Insert Format Tools Data	a <u>W</u> indow <u>H</u> elp								
1	i · 🖻 🖃 🙆 📝 📓 📆 "	5 🐨 🖌 🖻 🗗 🔺 🗠	👔 👪 🟦 💣 🔽   🧇 🛍							
	Arial 💌 11 💌 🖉		- III 4 🔳 % 😘 🖑 I E 🖻 I 🗄	- 📰 - 🚈 -						
D16	5 💽 😹 Σ =									
	AB		C	D	E	F	G	н	1	J
1	TT Tên s	ách	Tác giả	Noi xb	Nhà xb	Năm xb	Số trang	Khố cỡ	Mã sách	J
2	1 Giáo trình Linux và phần mềm r	nguồn mở	Pham Nguyễn Khang, Đỗ Thanh Nghi	Cần Thơ	Đại học Cần Thơ	2012	151tr.	24cm	Vv234	
3	2 Một số vấn đề chọn lọc của cộn	ng nghệ thông tin và truyền t	Đỗ Như An, Bùi Đức Dương, Đăng Hữ	u Đ Hà Nội	Khoa học kỹ thuật	2008	482tr.	27cm	Vv123	1
4	3 Thư viên số: Chỉ mục và tìm kiế	ém: Sách chuyên khảo	Đỗ Quang Vinh	Hà Nội	Đại học Quốc gia Hà Nộ	2009	221tr.	21cm	Vv564	1
5		Export Text File		1						-
0		Field option	15						-	+
/		Charles options	No. Inc.		OK					÷
0		Character	rset Unicode (UTF-6)		Court 1					+
10		Eield delin	miter .		Cancel					t
11		Testelle			Help			1	-	t
12	2	ext delin	niter					-		t
13	8	Save o	Save cell content as shown					1		T
14	1	Save o	Save cell formulas instead of calculated values							Т
15										
16	5		all text cells							
17		Fixed o	column <u>W</u> idth							
18	8		1							

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

<u>Buốc 3</u>: 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 CSVplugin như sau (Hình 7):

- Kích chọn Input\_encoding là UTF8.

- Không chọn các mục Default\_encoding; extract\_langguge và default\_langguage.

Greenstone Librarian Interface 2.85 St	erver; v2 Mode: Librarian Collection: CSDL the mut	c (csdlthmc)		
Download Gather	🖾 Enrich 🔅 Design 🕌 Creat	e 🚯 Format	<b>()</b> нер	1
Document Plugins Search Indexes	Document Plugins	Configuring Arguments		even even even even even even even even
Partition Indexes	ZIPPlugin	Please co	nfigure the arg	uments for CSVPlugin.
Drowsing classifiers	GreenstoneXMLPlugin TextPlugin		CSVPlug	in 📫
	HTMLPlugin EmailPlugin	process_exp	(31)(\se)s	=
	PDFPlugin		ReadText	File
	WordPlugin	✓ input_encoding	utf8 - Either utf8 or u	nicode automatically detected. 🛛 💌
	PostScriptPlugin PowerPointPlugin	default_encoding	utf8 - Either utf8 or u	nicode — automatically detected.
	ExcelPlugin	extract_language		
	ISISPlugin	🗌 default_language		
	NulPlugin EmbeddedMetadataPlugin		AutoExtractM	etadata
	CSVPlugin -input_encoding utf8	🗌 first		
			AcronymExt	ractor 🛛
	Select plugin to add: BibTexPlugin	ОК	§	Cancel
	Add Plugin Co	onfigure Plugin	Remove Plugin	

Hình 7: Cấu hình CSVplugin

<u>Bước 4</u>: Thực hiện chuyển dữ liệu (từ các cột trong bảng file .CSV sang các yếu tố tương ứng của Dublin Core trong Greenstone) (gather) của Greenstone Librarian Inferface để chọn file dữ liệu CSV mà ta tạo ra từ Libreoffice để đưa vào bộ sưu tập (ví dụ: tệp Book1.csv) (Hình 8).

- Chúng ta sử dụng chức năng "thu thập"



Hình 8: Cửa số "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

THÔNG TIN và TƯ LIỆU - 3/2014

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

🔄 Download 🔄 Gather 🐼 Enrich 🔅 Design 🕌 Cru	eate 💮 Format			
Workspace Coll	lection			
<ul> <li>□ Documents in Greenstone Collections</li> <li>□ Local Filespace</li> <li>□ C:\</li> </ul>	Book1.csv			
← 🗂 \$RECYCLE.BIN				1) Help
<ul> <li>☐ 11005</li> <li>☐ 1_files</li> <li>☐ 2_files</li> <li>☐ 5566_files</li> <li>☐ 898_gia-dinh-va-bien-doi-gia-dinh-o-viet-ni</li> </ul>	Exploding this metadata fi irreversible process and t	le will produce individual r he original metadata file w	ecords with editable metadata. This is ill be removed from the collection.	an
← 🗇 anh	plugin	CSVPlugin - CSVPlugin		
<ul> <li>Gaiviet2</li> <li>GibreOffice_files</li> <li>Gina Tinh năng » LibreOffice tiếng Việt_files</li> </ul>	input_encoding	auto - Use text categoriz	ation algorithm to automatically iden	
- 🗋 1.png	metadata_set	Exploded Metadata Set (	exp)	*
- 2.png - 3.png	document_field	Tên sách		
395.pmg				•
5.png 6.png	Explo	de	Cancel	
DUUKI.CSV				

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

+ Sau khi kích chuột chọn **Explode**, một hộp thoại mới xuất hiện, cho chúng ta lựa chọn việc kết nối giữa các cột (trường) của file CSV với các Yếu tố (trường) của Dublin Core (trong ví dụ, kết nối giữa Tên sách với dc.Title) (Hình 11). Ta chọn nút **Merge** để kết nối.

+ Thao tác này được lặp lại với các trường còn lại.

🔂 Download 🚯 Gather 🔯 Enrich 🚯 Design 🗱 C	reate 🚯 Format			
Workspace       Co	Ilection Book1.csv Werging Action Required. Source metadata element:Tênsá Target metadata set: Dublii Target metadata element: dc.Tit	t: Tênsách Dublin Core Metadata Element Set, Version 1.1: Reference D V		
<ul> <li>5.png</li> <li>6.png</li> <li>Book1.csv</li> <li>Book1.xlsx</li> <li>buildingGS.png</li> <li>g1.png</li> <li>g2.png</li> <li>GIÁI PHÁP TRA CỨU CƠ SỞ DỮ LIỆU CD</li> <li>isistoexcelandflash.doc</li> </ul>	Add	Merge	Ignore	

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

Sau khi thực hiện kết nối tất cả các trường, chúng ta đã có dữ liệu cho các Yếu tố của Dublin Core. Phần tiếp theo là thực hiện các bước tạo lập một bộ sưu tập theo cách thông

thường của Greenstone để có được bộ sưu tập.

Bộ sưu tập này trong phần mềm Greenstone có thể được đưa lên và tra cứu trên Internet một cách dễ dàng.

<b>(</b> )	http://127.0.0.1/gsdl	?a=q&r=1&hs=1&e=p-010-00-off-csdlthn	nc00-100direct-1040-11	11-vi-5020-about00-3-1-00-040	0-0-11-10-0utfZz-8- 🏫 - C
<u>v</u> ! -	Q, danh bạ điện thoại	SEARCH 🔹 🔶 🖂 🕶	🎲 - 🗋 - 🚹 - 😹 🄀	- ன -	
	of the work work work work work work work work	CSDL THƯ MỤC BIBLIOGRAPHIC DATABASE			
			tim kiếm	nhan để	tác giả
	Tìm trong tác giả	✓ cho tất cả các ✓ của	các từ Đỗ Quang Vinh	Bắt đ	âu tìm kiếm
	_			các kết quả	
$\langle \rangle$	Số từ xuất hiện: Quan	g: 1, Vinh: 1, Đỗ: 3			
	1 tài liệu khóp vớ	i câu truy vân.			
2	Thư viện số: (00000003.m	Chỉ mục và tìm kiếm : Sách chuy ll)	ên khảo		

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

 http://www.youtube.com/watch?v=hksa7w NO3-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)