

# **Crosswalk or Cross-stagger? : Using OCLC's “Digital Import” Feature from Connexion to CONTENTdm**

May 2008

Kate Kluttz & Eve Neville

cutternumber@yahoo.com

State Library of North Carolina

*Best Practices Exchange 2008*

*Helena, Montana*

# What is Connexion Digital Import?

- New(ish) feature in Connexion (OCLC's cataloging interface)
- It lets you do this:
  - start with a MARC bibliographic record in WorldCat
  - upload a file to your CONTENTdm collection
  - create a link in the WorldCat record to that file
  - crosswalk the MARC record to Qualified Dublin Core metadata in CONTENTdm.
- See OCLC's announcement here:  
<http://www.oclc.org/news/announcements/announcement247.htm>

# What is required?

- Connexion version 2.0 or higher
- Full level authorization status or higher in OCLC
- a hosted version of CONTENTdm!
- an OCLC authorization that includes CONTENTdm authorization
- an existing WorldCat record to attach digital content to

# Why did we use it?

- State Library of N.C. includes the state documents Clearinghouse:
  - Digital docs needed to be discovered, shared, preserved, etc.
  - Depository Library survey indicated our clients want us to continue full MARC cataloging of documents-- let's re-use that data!
  - Began project with already-cataloged paper docs that have existing electronic versions
- And, we were using CONTENTdm Hosted Services anyway

# How does it work?

The screenshot shows the OCLC Connexion interface. The title bar reads "OCLC Connexion - [Online WorldCat: A comparison of Russian and Italian honey bees]". The menu bar includes File, Cataloging, Authorities, Edit, Action, Batch, View, Tools, Window, and Help. The main window displays a MARC record for OCLC 61442711. A callout box points to the "Attach Digital Content" option in the Edit menu, with the text: "In Connexion, start out with a MARC record. Choose 'Attach Digital Content'".

**In Connexion, start out with a MARC record. Choose "Attach Digital Content"**

Field	Subfield	Value
040		NCS #c NCS #d NCS
029	1	CNTNT #b /p249901c
037		02690 #b NCTE
043		n-us-nc
086		A75 30:655 #2 ncdocs
090		#b
049		NCSS
100	1	Tarpy, David R.
245	1 2	A comparison of Russian and Italian honey bees / #c [prepared by David R. Tarpy and Jeffrey
260		[Raleigh, N.C.] : #b North Carolina Cooperative Extension Service, #c [2005]



# How does it work? (cont.)

The screenshot shows the OCLC Connexion interface. The title bar reads "OCLC Connexion - [Online WorldCat: A comparison of Russian and Italian". The menu bar includes "File", "Cataloging", "Authorities", "Edit", "Action", "Batch", "View", "Tools", "Window", and "Help". The "Action" menu is open, showing options such as "Add to Authority File", "Report Error...", "Submit for Review...", "Recall Submitted Record", "Lock Master Record", "Release Record Lock", "Replace Record", "Replace and Update Holdings", "Replace Update and Add Institution Record", "Lock Linked Authority Record", "Save Record to Online File", "Save Record to Local File", "Delete Record", "Copy Record...", "Move Record...", "Export", and "Set Status...". The "Replace Record" option is highlighted. A callout box points to this option with the text: "Then Replace the record. The digital file will upload to CONTENTdm." The background shows a table with cataloging data.

Field	Value
090	#b
049	NCSS
100	1 Tarp, David R.
245	1 2 A comparison of Russian a
260	[Raleigh, N.C.] : #b North
300	1 folded sheet ([4] p.) ; #c
490	1 AG ; #v 655
500	Caption title.

# How does it work? (cont.)

 **CONTENTdm Administration**

admin home server

:: approve : index : add : edit : find & repl

**Current collection:** State Publications Clearinghouse Temporary Collection

## Item administration

### Approve

Approve, edit or delete items in the pending queue.

### Index

Index the collection after adding, approving, editing and deleting items.

### Add

Add an item to the pending queue. Items are reviewed and approved before being added to the collection.

### Edit

Edit and delete items in a collection.

**Now, index the  
CONTENTdm  
collection.**

# How does it work? (cont.)

## CONTENTdm Administration

admin home

server

collections

items

:: approve : index : add : edit : find & replace : lock administration :

**Current collection:** State Publications Clearinghouse Temporary Collection

### Edit item

Edit metadata then click **save changes**. To remove this item from the collection, you must rebuild the text index.

**Magic!!! After indexing, the digital resource and the metadata record are available in CONTENTdm.**

Title	Comparison of Russian and Italian honey bees	
Other Title		
Creator	Tarpy, David R.	shared controlled vocabulary
Contributor	Lee, Jeffrey.; North Carolina Cooperative Extension Ser	shared controlled vocabulary
Date	2005	
Release Date		
Subjects	Honeybee--Breeding.	shared controlled vocabulary
Place		shared controlled vocabulary
Time Period		
Description	Caption title.; Includes bibliographical references (p. [3]).	

# Magic!!!

completely  
gratuitous  
picture of  
Stonehenge



# OK, maybe it's not all magic.

completely  
fake  
Stonehenge  
in Texas



- MARC and QDC are not quite the same...

# Post-crosswalk editing—lots!

At first you feel like this guy getting trampled, but after a while it's not so bad



# OCLC controls the crosswalk

- You don't pick which MARC field crosswalks into which QDC field—OCLC does!
- This means that you may have:
  - Data crosswalking into a field in which you don't want it
  - Data you don't want at all that crosswalks anyway
  - Data you want that doesn't crosswalk anywhere

# the Crosswalk Map!

## Connexion Digital Import Metadata Crosswalk Map MARC to Qualified Dublin Core

Sorted by MARC fields  
(Last updated: 30 October 2007)

MARC tag	Indicators # = blank; ? = any	Subfields	Dublin Core Element
008	null	Language	<a href="#">dc:language</a>
020	##	\$a	<a href="#">dc:identifier</a>
022	?#	\$a	<a href="#">dc:identifier</a>
024	??	\$a	<a href="#">dc:identifier</a>
030	##	\$a	<a href="#">dc:identifier</a>
033	0?	\$a	<a href="#">dc:coverage temporal</a>
033	1?	\$a	<a href="#">dc:coverage temporal</a>
033	2?	\$a	<a href="#">dc:coverage temporal</a>
043	##	\$c	<a href="#">dc:coverage spatial</a>
046	??	\$j	<a href="#">dc:date modified</a>
050	??	<a href="#">\$a\$b</a>	<a href="#">dc:subject</a>
060	??	<a href="#">\$a\$b</a>	<a href="#">dc:subject</a>
080	##	<a href="#">\$a\$b\$x</a>	<a href="#">dc:subject</a>
082	0?	<a href="#">\$a\$b</a>	<a href="#">dc:subject</a>
090	##	<a href="#">\$a\$b</a>	<a href="#">dc:subject</a>
092	?#	<a href="#">\$a\$b</a>	<a href="#">dc:subject</a>
096	??	<a href="#">\$a\$b</a>	<a href="#">dc:subject</a>
099	#?	\$a	<a href="#">dc:subject</a>
100	?#	<a href="#">\$a\$b\$c\$d\$g\$j\$g</a>	<a href="#">dc:creator</a>
110	?#	<a href="#">\$a\$b\$c\$d\$g\$n</a>	<a href="#">dc:creator</a>
111	?#	<a href="#">\$a\$c\$d\$e\$g\$n\$g</a>	<a href="#">dc:creator</a>

# Going into the “wrong” field

- Language
  - Lang and 546 from the MARC record both crosswalk into DC.LANGUAGE
  - We have Language set up as a field that uses controlled vocabulary
  - The free-text note from 546 has to be moved from Language to Description

# Unwanted data (awww...)

- Call number:
  - 099/092/096 from MARC goes into DC.SUBJECT
  - This includes \$a (subject portion of call number) AND \$b (Cutternumber for shelflisting!)
  - Again, it's a problem if you're using a controlled vocabulary for the Subject field.
  - We have to delete this data manually from the metadata record.
- Why not strip it out before crosswalking?  
Because the record is replaced in WorldCat!

# Data that doesn't crosswalk

- Some fields or subfields just aren't included
  - 041 (language codes)
  - 780/785 (title replaces/replaced by fields) only certain indicator/subfield combos are crosswalked
  - 260 \$a (place of publication)
  - 245 \$c (statement of responsibility)

# Fields that don't exist in MARC

- We have to input lots of data directly in CONTENTdm, because the data simply doesn't exist in a MARC record
  - “Collection,” “Digital Format,” “Rights,” “Metadata Creator,” etc.
  - You can't apply a CONTENTdm template to a record that has already been crosswalked

# Controlled vocabulary issues

- Our CVs include using LCSH and LC name authorities in various fields
- We loaded these terms into CONTENTdm after pulling the data from our Voyager system
- If the WorldCat record has had authority headings added or changed since the time we put it into Voyager, those terms aren't in our CV
- In Admin module: new controlled vocabulary terms can't be added to the CV directly from the record (must be laboriously added through different steps)

# What the--?!

- Much of our data was crosswalking into the wrong QDC fields (e.g., physical description went into “Digital Format” field)
- This had to do with the field names and “nicknames” not matching; OCLC had to fix it. They say this shouldn’t happen to anyone else now.
- Test carefully!

# I can't replace the \*\$@&! (# record!

- Our OCLC authorization is Enhance level
- Some of the MARC records we need have already been upgraded to Elvl:[blank]
- Digital Import requires that the record be “replaced,” but we can't replace these records!
- OCLC is working on this. Suggestions of creating new record or using Bib Change Request form not really what we want to hear.

# Other workflow issues

- We have to crosswalk records into a “temporary” collection that is not viewable by the public
- Then clean up the metadata while it’s in the temp collection (so we don’t embarrass ourselves with it)
- Then the entire temp collection has to be merged with the “main” collection (this makes the “reference url” change! Goody!)

# ...and more whining about workflow

- Sometimes you successfully attach digital content/replace the record in WorldCat, but then it doesn't appear when you index the collection the first (or second, or third...) time
- Unspecified time lags may occur

# Happy thoughts

- Crosswalking does allow special characters/diacritics to come through from WorldCat (they can't be input when writing a record "from scratch" in the Acq Station) (though they don't display properly in the public view)
- It is awfully nice to be able to re-use data from the MARC records

# Potential improvements?

- Ability to use templates or something to apply “constant data” (maybe 9xx fields added in Connexion?)
- Streamline adding of controlled vocabulary terms directly from metadata record while working in Admin module
- Ability to attach digital content to all records (including CONSER/EIvl:[blank] records)
- Let CONTENTdm talk to the WorldCat authority file for controlled vocabulary
- Ability to suppress records from public view, instead of having to put them in a temporary collection and then merge.

# Sometimes weird but pretty useful



- Definitely a good feature for certain situations

really weird roadside attraction in N.C. called Shangri-La