WORKING WITH THE MARC TOOLS

Terry Reese
Head of Digital Initiatives
The Ohio State University
reese.2179@osu.edu
ROADMAP

- What is the MARC Tools section
- MarcBreaker
- MarcMaker
- Charactersets
- MARCSplit/MARCJoin
- Batch Processing
- XML Functions
  - How does it work
  - Building/Editing new XML Functions
- Automatic Record Classification
- RDA Helper
MARC TOOLS

- What is the MARC Tools section
  - Access to the Making and Breaking functionality
  - Characterset processing
  - Access to the XML Sub-routines
MARC TOOLS
MARC TOOLS

- Built-in functions
  - MarcBreaker – Tool used to convert MARC records to the MarcEdit mnemonic format
  - MarcMaker – Tool used to convert MarcEdit mnemonic format to MARC
  - MARC=>MARC21XML – converts MARC to MARC21XML
    - Automatically converts data from MARC-8 to UTF8
  - MARC21XML=>MARC – converts MARC21XML to MARC
    - Doesn’t automatically convert data from UTF8 to MARC8 – will leave data in UTF8
MARC CHARACTER CONVERSIONS

- Supports moving between any known Windows Character set and MARC8.
- Can be run from the Breaker/Maker – or as its own standalone utility.
MARCSPLIT

- Utility used for splitting large MARC record sets into smaller files
Utility used for joining large sets of MARC data to a single file
BATCH RECORD PROCESSOR

- Allows MarcEdit to process “lots” of files.
- Files can be processed against an entire folder’s contents or by file type.
- Can utilize any built-in or derived XML Function transformation.
MERGE RECORDS TOOL

- Allows users to merge MARC data from two files
- Allows users to merge unique data, selected data and all data.
EXPORTING DATA

- Exports raw MARC data using TCP/IP connections
- Can export individual or multiple records (regardless of ILS system support for batch record import)
- If set in the preferences, will capture any data returned back from the server
LITTLE KNOWN FUNCTIONALITY

- Ability to process remote records from within the MARC Tools area
  - Can be either MARC or XML data
  - Can be zipped (if zipped, the file will be preprocessed)
LITTLE KNOWN FUNCTIONALITY

- XML Processing
  - MarcEdit’s MARC Tools supports multiple XML engines, settings.
    - MSXML (Microsoft .NET optimized XSLT engine)
    - Saxon.NET (.NET version of Saxon)

- Why do we care?
  - MSXML processor
    - Small and fast, Optimized for low memory footprint and speed and Supports only XSLT 1.0/Xquery 1.0
  - Saxon.NET
    - Supports XSLT 2.0, Xquery 2.0, Little slower, but more flexible and allows use of Saxon XSLT extension when writing XSLT documents.
Character conversion isn’t limited to known – pre-populated items. You can define your own character-sets for process.

- MarcEdit can utilize any known Windows Characterset when doing characterset conversion.

- Defined MS charactersets:
MARCEDIT AND BAD RECORDS

- Two MARC breaking algorithms
  - Strict MARC algorithm
  - Loose breaking algorithm

- Loose algorithm can heal MARC records (sometimes)
  - Structural errors
  - Missing field or record markers
MARC CONVERSIONS
MarcEdit model:

- So long as a schema has been mapped to MARCXML, any metadata combination could be utilized. This means that no more than two transformations will ever take place. Example: MODS → MARCXML → EAD
MarcEdit: Crosswalking Design

- MarcEdit Crosswalk model
  - Pro
    - Crosswalks need not be directly related to each other
    - Requires crosswalker to know specific knowledge of only one schema
  - Con
    - each known crosswalk must be mapped to MARCXML.
MARCEDIT CROSSWALKING MODEL

- EAD
- FGDC
- Dublin Core
- MARC
- MODS

Diagram shows the relationships between these standards.
What’s MarcEdit doing?

Facilitates the crosswalk by:

1. Performing character translations (MARC8-UTF8)
2. Facilitates interaction between binary and XML formats.
MarcEdit can leverage OCLC WorldCat to generate call numbers automatically for files

- Fields used:
  - 001
  - 010$a$z
  - 020$a$z
  - 022$a$z
  - 024$a$z
  - 1xx$a$
  - 776$w$z
OCLC Classify Service

Generate Classification

This service makes use of OCLC's Classify Web Service. Classification recommendations are generated by sending OCLC a control number from the record. OCLC's webservice will return back to MarcEdit either the Dewey Decimal or LC classification most often associated with this record.

About OCLC's Classify Service

Insertion Instructions:
- Add to Selected Record
- Add To All Records

Classifications:
- Dewey Decimal
- Library of Congress

Subject Headings:
- Always Add Facs: Headings
- Add FAST Headings When No Subject Headings Are Present

Options:
- Calculate cutters and dates
- Append Flag to Date
- Append Date Exceptions
  - Omit date for serials
  - Omit date for integrated resources
- Use subfield z for matching

Insert if the following fields are not present (example: 050;060;090;092)

Insert generated classifications into the following field:

090

Classify Close
RDA HELPER

- Add 336 -- Content Type
- Add 337 -- Media Type
- Add 338 -- Carrier Type
- Add 344 -- Sound Characteristics
- Add 345 -- Proj. Characteristics
- Add 346 -- Video Characteristics
- Add 347 -- Digital Characteristics
- Add 380 -- Form of Work
- Add 381 -- Other
- Evaluate 260/264
  - Always use ©, © symbols

- Qualifying Information (015/020/024/027)
- Modify 040 -- add $arda
- Delete GMD
- Generate GMD
- Expand Abbreviations

Source: 
Save: 

Select: All | None

OK | Close