

Introduction to MarcEdit

? Overview

Teaching:**Exercises:****Questions**

- What is MarcEdit? What can it do?

Objectives

- Understand the basic history of MarcEdit
- Explain what the MarcEdit software does
- Understand some of the common tasks MarcEdit can be used for

What is MarcEdit?

MarcEdit is a comprehensive, MARC agnostic, library metadata processing software suite designed specifically to make editing library metadata easier and more accessible. Too often, library metadata projects get shifted away from metadata professionals because working with large sets of MARC data has traditionally required a large amount of custom programming or significant development expertise. MarcEdit was designed to bridge that gap. The tool provides a rich set of data processing functionality that enables users, regardless of technical expertise, the ability to complete complicated metadata projects – regardless of size or edits required. And as a metadata tool that is MARC agnostic, the application can be used by any user working with any flavor of MARC – whether that be MARC21, UNIMARC, FINMARC, RUSMARC, etc. – or any kind of data represented in the MARC format.

Over the years, the application has evolved as the library metadata landscape has evolved. MarcEdit can work with a wide range of metadata formats, including:

- RDA Processing
- JSON
- XML (with built-in translations for a variety of library formats)
- Delimited formats
- Excel
- SQL
- SPARQL
- Linked Data Reconciliation
- Z39.50/SRU
- Etc.

Since 1999, the tool has continued to evolve to meet the various needs of the global library community with the continued, underlying goal of making metadata creation and management available to everyone.

History of MarcEdit

The public version of MarcEdit was initially conceived in the summer of 1999 and finished largely in response to a major database clean-up project that had been planned for the Oregon State University's electronic catalog. The problem was simple; many Marcive records were entering the database with invalid call numbers. The problem had been discovered over the summer, and while the initial response was to correct the Marcive profile, the larger issue was that no one really knew how many incorrect records had made it into the library catalog. After a careful analysis of the data, it was determined that nearly 50,000 records would need to be corrected in the database.

Building of a set of programming components developed while an undergraduate at the University of Oregon, Terry Reese utilized these initial programming components to complete a workable MARC engine that could be accessed through the Windows Scripting Host (WSH) to provide a drop in programming object for use in a variety of tools. Using these tools, the Oregon State University Libraries was able to complete the data clean-up project in a matter of minutes, and after some encouragement by Kyle Banerjee, the first version of MarcEdit was released to the public.

Is MarcEdit Open Source?

A common misconception around MarcEdit is that it is an open source project. While a large amount of the project code has been (and continues to be) released at: <https://github.com/reeset> -- the project as a whole has not been open sourced. For information on why, please see: <https://marcedit.reeset.net/is-marcedit-open-source>.

What kind of software is MarcEdit then? MarcEdit would be more accurately classified as freeware, as in, free to use and free to distribute. There are a lot of open source tools in the library community – these are great tools that provide libraries the ability to build really interesting services. However, these tools tend to be geared towards developers and while open, often require significant expertise and local developers to support, install, and update the software. While libraries and library software has come a long way, there still exists a significant gap in software developed that is both reliable and easy to install and update. This is the niche MarcEdit has designed to fill. MarcEdit has been created to be a tool that anyone can use, by simplify that installation and maintenance steps with a large and active user community. Since its inception in 1999, the program has been open to anyone to use or integrate with existing tools and software.

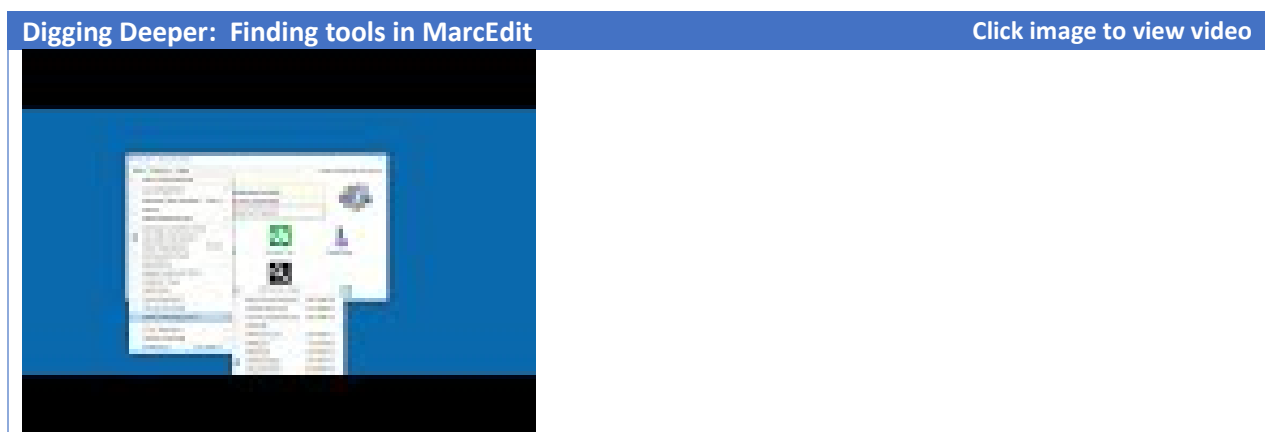
And what of the future? MarcEdit has a documented succession plan, which is reevaluated every couple of years to ensure that if something happened, the tool would find new maintainers or be made immediately open source.

What can MarcEdit do?

MarcEdit has often been described as a kind of metadata swiss army knife because the application is more akin to a suite of tools, rather than a single application. This is because libraries and catalogers receive and development library metadata in many different forms and get data from a wide variety of sources. MarcEdit can be used to standardized processing across this wide variety of formats and data sources, enabling metadata professionals to spend more time working with their data, rather than working with or building tools. MarcEdit can help users:

- Automate processing of vendor MARC records for e-records
- Facilitate large data clean up projects, like updating subject or name authorities

- Converting AACR2 data to Hybrid-RDA records (true RDA conversion isn't just about field changes, but requires a reassessment of the content)
- Convert data from Dublin Core, MODS, EAD, MARCXML, or other XML or JSON formats to MARC
- Convert data from MARC to Dublin Core, MODS, EAD, MARCXML, or other XML or JSON formats
- Reconcile MARC data to utilize Linked data concepts
- Validate records
- Join or Split large data sets
- Manage Character encodings and Unicode Normalizations
- Resolve inconsistencies in a data set through clustering
- Integration with OCLC and ILS systems
- Merging data between data sets or data files
- Retrieve data from remote sources through OAI, Z39.50, SRU, Atom Pub, etc.



The MarcEdit suite of tools supports a wide range of functionality to enable novice users to be immediately productive, while provide advanced power tools and options to support users interested in created highly automated or integrated workflows.

Some common scenarios where users might use MarcEdit would be:

- Receiving data from a vendor, the Library needs to convert the data to RDA, remove invalid URLs, and validate and update subject and main entry access points.
- A vendor provides a zip file with 1000 MARCXML records in 1000 individual files and a user needs to convert these records to MARC in order to view, validate, and edit the data for ingest into their catalog.
- One's library catalog requires UTF8 data to be normalized using the canonical normalization, but data coming from OCLC or other vendors includes mixed normalizations impacting display or indexing.
- A problem has been found in the catalog affecting hundreds of thousands of records. The issue could be corrected with a simple regular expression, but the catalog doesn't support this work.
- A vendor has provided a set of JSON files using a vendor specified format and your organization doesn't have the necessary developer resources to convert the data – but still need to translate the data to MARC [or other format].

MarcEdit includes hundreds of tools and utilities designed to make processing data easier and faster for user – including a number of wizards designed to bridge technical gaps for users need to process an unknown data type into or out of MARC.



Key Points

- MarcEdit is MARC agnostic meaning that it will work on any MARC data
- MarcEdit can work with most well-known library metadata formats
- MarcEdit can work with very large files or automate processing of large batches of files
- MarcEdit can help to clean up local data
- MarcEdit can integrate with many ILS systems as well as OCLC
- MarcEdit is really more of a suite of tools, than single piece of software, with hundreds of functions available to the user.