Better Faster: Using MarcEdit to Process and Enhance CRDP Records - Transcript

Please stand by for real time captions.

Hello to everyone.

Hello, everyone. Welcome to the FDL P Academy webinar. Presented by Leslie and Colson. Our technical support is Don Lynn. I'm looking forward to the webinar today and hope you are , too. Before you start, we have a bit of housekeeping. If you have any questions you would like to ask the presenter, or if you have any technical issues, feel free to ask questions in the chat box which for your desktop Peter Cehlarik laptops are in the bottom corner of the screen. We will email a link to the recording for everyone who registered for the webinar. We will be hosting it

---

For those of you who are familiar with MarcEdit, I am not a power user, but I have found several tools and makes that MarcEdit can be used for to better enhance my CIDP records in a different way. I find that MarcEdit, think I know how to use it, is very useful that helps with enhancing and processing records in a quick way. So, let's talk a little bit about what I hope to be able to cover in this up in our. We are going to talk about how to validate MarcEdit, check fields for errors, consolidate in task manager, and URI, which are useful data, and remove records from sets. I'm going to talk a little bit through my PowerPoint about the context for each of these different tasks. And, I will be leaving the PowerPoint are going to demonstrate all of these in MarcEdit. Let me tell you a little bit about the context in which I work my environments. Our library management system is Alma, and I can do a lot of these things that we do in MarcEdit and Alma, but I do find it to be a little bit more clunky and cumbersome to use and that is why I do all of the processing in MarcEdit before I import them into Alma. We get four different studies from archive, we have the physical and online records split apart and then those are split into new and changing records. While I could import all of those in one file into Alma, I think Alma can Hondo -- and all of this. Because of the workflow know the things that we need and the different types of records, I have decided to have all of those split up into four different files so that I can process each one of those difficult -- differently in MarcEdit. And now, the adventure begins. So, MarcEdit is currently available for download in version 7.3. 7.5 is available in nature, if you are rave and daring and want to venture into that adventure. The environment that I will be demonstrating is 7.2.9 five. I have not upgraded yet to 7.3. So, let's talk about validation. Validation with the MARC will check for spaces of where they shouldn't be. Or lines in different MARC fields that should not be there. And then the checkmark using the rules files will check the MARC files against the MARC 21 settings. I'm going to go back just a second. No matter where I get the records from, I always validate them in MarcEdit. The things that I get from Marcive are usually pretty good when it comes to structure wise, but I also validate after I have been editing, because I have had structural errors through the blank line or not having enough spaces between the tag and the first subfield. So this helps me in case I have made many mistakes and to get the MARC structure correct before I tried to import into Alma so that I can fix those errors really quickly with trying to figure out why the file has the import. The next thing is checking the fields for errors. When we break a file, to get the file into the MARC editor in MarcEdit, you told how many records in the file. We will tell me how many fields a certain tag there are. In other words, how many 035 fields and I can look at this list and tell if there is any record that might be missing certain fields that I want to have or might have extra fields that I want to look at. Or maybe feels that I'm not familiar with and he wanted to look at the MARC 21 standard and see what the fields are and see if Ike want to keep them for the record or not. It might give me a sense for what the file is contained I'm looking at or cases I need to follow up on to fix mistakes or data that I don't want in their I want to use

the fund records missing field find the records that are missing. MarcEdit has a really wonderful feature. There many tasks that you can do in MarcEdit, but the wonderful thing is that I can read these consolidated tasks against that file, and, knock off several different tasks very quickly and not have to do them all manually. These are tasks that I'm going to use in these files every time, for instance, I want to remove all 012 fields and all 019 fields. So, these are going to be in a consolidated tasks I do not have to keep this in every time. And then, anything, any file that a run through MarcEdit, I ran the URI. They have a linked data environment behind-the-scenes. And, I think it's really great. I am glad that they've been very intense about moving into that link data environment. But there's no guarantee that we will be in Alma forever. We do migrate and it could be that will eventually migrate into a different LMS system or a fully linked data environment you want to make sure that our records are ready for that migration, so, I always add URI. Run all the records through adding URI to any of the headings that will take down. And then finally, sometimes we do remove records from sets. I will talk a little bit more about the particular use case when I demonstrate this. And, given your workflow and your environment, you may find that you need to remove specific records from a set before including into your LMS. I will walk you through how to do that very easily. If we have time, we have one more feature that I will demonstrate, so we will see. So, before I go onto the demonstration, are there any questions about what I'm going to be talking about?

There are no questions in the chat at this time.

Great. I will share my screen. And you should be able to see MarcEdit. Customize this little bit in terms of the color of the screen, because the default white that goes for is too bright for my eyes, so moves to great. That's not really problematic but sometimes it is hard to see the cursor, so I apologize for that. I have already had some files that are open in MarcEdit. I just wanted to save the time for doing that. So, this is a file for online records. And, so I will just walk through the process. To validate, go to reports. And, then I go to MARC validator. You can see there already the rules file in place. Again, this is a rules file that comes with MarcEdit and is updated on a regular basis. You can update if there certain feels that your library uses that may not be standard MARC 21 fields, and you can add those to the file. So, the first am going to do is checkmark using the rules file. I can groups results by error, which I will demonstrate to you. But to begin with, don't do that because I just want to get a sense of whether there are any errors or not. But as a caveat, I want to say that the stylist amusing today, I am demonstrating what they look like. These come from archive and do not come with mistakes, so want to make it clear that these are not mistakes that came because they were given to us by Marcedit, there mistakes that I put in there. As you can see does processing it it does a pretty quickly. The first tells me is that there is some potential. I like to follow up on these by using find. I will search for just to see why MarcEdit thinks it was a duplicate. Often what happens is that it looks like a duplicate to because it is matching on basically the 245 field and wanted this finding is that I have multiple volumes. So, I have a four volume set, so four records, one for each. They could be a different title, or they can have a subfield. But to there are times when I have found an actual duplicate record. So I will just follow up on those very quickly to see why MarcEdit thinks there is an issue.

And if this is a really long list, I could, and I will do this, I guess, I can copy it to the clipboard very quickly and then open Notepad and I will leave this up for a second. I will paste it to the clipboard. This way can have this list available for me to have it next to my MARC editor file and work on fixing the errors just from the list. That is one way of doing it. The other ways that I can just click on this link that will take me right to record 18. So, does telling me that the error is in the 336 field and the sub field G here. Sometimes I find this to be the case that it will say go to record 18 and is not actually the right record, so, the other day when I was testing it, it worked fine. It actually took me ahead of the record that I

need to go to, as you can see. I just remove that and fix that mistake . There's no other way can do this. I'm going to close this and go back to my MARC validator. Going to group these results because what I have found is that sometimes there are errors that are common. There many of the same errors in the file and if that is the case, then it is something that I can batch change. Rather do that than manually go into each record and fix them, so I'm group them together to see if that's the case. As you see that through the records that have the 336 subfield G. I know that with those pictures I can do the search and find and see if that is the same mistake and if it is, I can fix it by batch. I have two other errors, one is the 260, invalid data, and another one invalid field format . It looks like they are with the same records, so I'm going to copy that title and fix that. After the show you the second part of validation. So I will close that. So that is one step in the validation process and that was checking the MARC rules file. What the next is to validate the structure. To see if there any spaces or dissing spaces or incorrect field coals and things like that. And it is actually telling me that there is an error. There is a field formatted incorrectly, we must have two spaces between the field and the field data, so nicely, it tells me specifically what the error is. And, it is in the same one that I copied the title on before. So, again, I can copy this to the clipboard and because of already copied the title, I'm going to close that with the search and edit. Sign and paste. As you can see, I cannot see this cursor very well because of the gray background so I apologize for that. This is the title that I want to look for. I will highlight that and jump to the record. And, the error was in the field to 60. As you can see, it is missing a space. There should be two spaces and then the indicator. So, I am going to save the file now. Go back to my report to validate and there are no errors. It was that?. Now not going to go through cleaning up all the other errors that were in the file, but hopefully you get a sense of how easy it is to find errors and fix them. So, that is the validation process. Are there any questions about the process?

No questions yet.

Okay, we will move on to checking the field for errors. So again, I go to reports and then field count. I know that my file has 493 records in it. That will give me a sense of whether have the right number of fields to respond with that. And, usually it is for these records, and look at all of the fields, but I'm particularly looking at the specific ones as well. I want to make sure that every file has, every record has a 086 the field, which is the docket number. The reason we need that in every record is because our students will put these records into different collections in Alma. Because we want to have all of our government documents findable in a collection in the discovery layer. We put the been different locations based on the number or the department that publishes them. It must have that 086 field. If there any records that were missing, it will be much quicker for me to find them here, to find out that they're missing that here. Go out and find out what the number should be a put into the record as well. And get a in MarcEdit before it is reported into Alma. We want to make sure that we have enough 300 fields and as you can see, therefore hundred 93 records, but, I am missing, there is a record that is actually missing one of the 300 fields. The other things that I look for again, so look for fields that I may not be familiar with and I have been catalog enough that I'm pretty familiar with what a lot of these are, and with the new RDA field, learning what those are, so for instance, I will see that there is a 513 and 518. Not really too sure on what those are, so, I would publicly go and look those up just to make sure that I know what they are and that I want them in the record. The other thing I want to note about the field, for instance, the 515, there is only one record that has this field and that is kind of a heads up to me to take a look at this field to see why there is only one record that has this field that I want to keep. With some of these records for online resources, there is a field, I think it is a 550 field that has licensing data and sometimes the licensing data could be specific to an institution or specific to an environment that does not work for us. So, I will remove that field if it is not pertinent to our environment. So I scrolled down again, the whole list, checking to see what is there. I'm also looking, since I am, a live look

at the online resources and the physical resources, for the set that is typical resources, want to see if I want to have any 536 fields. Sometimes might remove that where we might have the at four. This is a file for online records and we can see that I have more 856 fields than I do records. That tells me that there are many records with multiple 856 fields and that is interesting to me because I want to see if I want to keep all of those, why there are multiple fields, and then there is the 92 field which tends to be a local field so I will took a look at that as well. I usually take note of which field I want to be looking at if they are not my usual ones, and after this step, I will go and look at them and I will show that really quickly. I want to look and find, and it's a search for a tag, I to find it in the ISBN or the other number, but if you put the = before it, it will look for the tag. When I am in the online records, I want to make sure that all of the 300 fields start with one online resource. They can include the pagination or not, they can include the color illustration maps and all that information. I want to make sure that they do all include the online resources. That sorts them in alphabetical order, which is a really quick way for me to -- because, all of the same ones are grouped together so I can quickly scan to the bottom to see if there any that don't. In this particular file, there's just this one. I know that I am missing a field. A 300 field, some going to go to edit shortcut. Field edit. Then find records missing field. In this search, you do not have the equals sign. Because it says it is looking for specific field tag and it takes me right to the record that is missing the 300 field and I will just edit room?. All right. I have fix that. I save my file after every change that I make. Due to a bad experience. I would encourage you to do that as well. I do want to look also at the field because there were duplicates there. And I could just scroll down over here, they have the record number and I could just scroll down looking for duplicates and depending on how large my file is, that can take a lot of time. I do see that there are two here and the record number 37. I can see that one is the PDF version and one is the text version. I am fine with keeping both of those in the record. The other option is that they will sort in alphabetical order and I can see the ones that have this field, what kind of format or resource are they pointing to. It will explain why there are multiple 856 fields in this particular record. And then I will scroll down and again, I'm going to expand this a little bit so I can see this full URL, I'm looking for anomalies in these URLs. It is just changing from PDF to text. Which is why it caught my eye. And, then it changes from text to noting field three. And then I have this one within notes. Going to take a look at the note to see if I want to keep it. Microsoft excel required to do the appendix. This is something I'm going to want to check after it has been imported to make sure that this note is showing up. I'm not quite sure how our import file is configured for adding notes to the portfolio using subfield C and I want to make sure this is showing in that portfolio. So I will probably, if I was doing this in my real-time, I would go into this record, jot down the number so I can look at it later. Not just going to continue to scroll down, looking for anomalies. Usually, there aren't any. Hopefully there won't be any. But, there is. So, this is missing the GPO number. This actually did come from archive this way, so, this URL actually takes me to a page that says this URL is no longer assigned to a resource. So, again, I would jump to this record, I would find with the resource is, the continuous stream discharge, copy it and paste it in Google search for's and then at it, and the complete URL while I am in MarcEdit, before I import this into Alma. So, that , the context and how I check the fields for errors, I there any questions about the process?

I think everyone is following along closely. There no questions yet.

Great.

All right. Now, I am going to consolidate tasks. So, to do that, I go to tools. And, I will look in manage tasks. This is where individuals, interval tasks as well as consolidated tasks are listed and I'm going to have the name, and as you can see, I have three consolidated tasks. 14 new online, one for change online, and one for change print. We don't process the new print, because we wait until we actually

have the item in hand and are cataloguing it in Alma to see if we need to do any updates to the record there. So we do not process those in MarcEdit. So for this particular consolidated task, it is change online. I actually have three tasks in here. I am going to add, I'm going to show you how to add to this consolidated task list. I'm going to close this first and I'm going to create a new task, which I'm going to call -- might want to use this task in records other than just my doxologies online. My want to have it available to use all of my tasks or any the record records that I process in MarcEdit. I'm going to create this task outside of my task list and then add this task to that task list. Not going to put a description, because I think my task list name is descriptive enough. So, in this box, I'm going to add attacks. And to do that, I will have an action. This means add a task and keeping over what all the other icons down here for, and this brings up all of the options that you have for creating tasks. I'm going to add a new field task. And, 33S. Went to copy and paste this duty here. And then I'm going to add the field. Noted that you need to add the indicators. In that text box. I'm also going to add the 337 field. And this 338 field. These are specific to online data. I think I will change the name. I'm going to save that. And, that should be right here. I could edit the task. Yes, I can. There we go. So these are specific to online resource. Now going to go to my GovDocs change online. And I can choose from the task list. That is really hard to say. Will it let me?. Rough one. Add . And I'm going to and from the task list. Okay, and it brings them in here. I'm going to say save. I could run this on my file. I'm going to close this. Go to tools. Assigned task and this is a currently available task. That is how quickly it ran. 492 records in that time. And it tells me what it has done and how many fields were affected. I always get rid of the subfield H because we don't need that anymore. But we still get them in the records. And, we have the 336, 37, and 338. Here I made a mistake, so, I'm going to close this and normally, if I would, actually, I'm going to do that's. Because I have saved, this is why I recommend saving after every task you've done because sometimes you make mistakes and you want to undo them. You can go to edit and go to special undo, which will on do it. But if you've gone too far and you can't do that, you can closeout and then reopen the file and you will not lose a lot of the changes they have already made.'s let me show you what mistake I made when I created that task, I'm going to edit this one. I want to have it as and in the field, only if not present, because some of the records did have a field and some of them did not.'s I only want to add them if they are not present. So it is very easy to add and fix things if they are not working the way you thought they should work without having to totally redo them. Here, it only added 71 fields. That was the number of records that were missing those fields. All right. Any questions about consolidating tasks?

Not a question per se, but there was a comment, keep in mind that we can automatically delete unwanted fields of your choice.

That is good to know. I know that there are ways that we can customize our profile with Marchive. Because this file is over almost 500 records, but could use this file, I'm going to switch and all of those different boxes that popped up while I was editing, the kind of stay there behind the scenes. And I'm going to open one that is a smaller file with only 12 records in here, because adding URI can be a very lengthy process, so I want to work with a smaller file to do that. Said to do that, you can go to tools. Linked data, and build linked records. And, already, you can't see it because it is great, but has my source file and by save file. It has preselected adding the main entry and added entry subject and if I hover over this icon, it tells me what data it has actually detected. The different authorized types of, I'm going to use the term heading. So, that vocabulary is attached. So I usually just run it like this. I get all of the links. But you can also limit it and I am going to, for this purpose, I'm going to limit it's because those of the main vocabularies that we use. This will not to do anything with the three XX fields but I'm essentially going to have to run these twice if I want those fields to billing. So, then we just click on process and what I do is set this to process and then go on and do something else, because if I have the save the file, it will take quite a while. Hopefully with these 12, it will not take very long, but what is

happening is that the MarcEdit is coming out and pinging the different surfaces . And, it's looking for all of the terms to see if there are matching authority records there. Then it is bringing the URI and associated with that particular heading. And then we are almost done. Okay. And then as you can see, we have a URI in this field. It is subfield H and it is for the Arctic region strategic aspects. It will only add a URI if the entire string is, if there's a matching authority record with the entire string, so as you can see in the field love that, the security international subfield, the Arctic region. Most regents do not have authorized authority records, and so, there's probably not an authority record for the security regions, even though this is a valid subject heading string, that is why there is no URI there, because both of them together do not match an authority record. And their ". There are all the URI that I can get to that I can import these into Alma. Are there any questions by adding URI?

Not at this time.

Okay. Now I'm going to close out of this. So the next thing I'm going to demonstrate is how to remove records from sets. And in MarcEdit drawer, which is where these are located. MARC editor is one part of the MarcEdit tool and it is where you do your editing. The records in this format are in a form that is readable by Please stand by for real time captions. -- MarcEdit editor. You have to break them into this pneumonic form that MarcEdit can re-. You have to have a binary file so I'm going to use one when I go to that tool to remove. So I go to tool and then select MARC records and I want to delete selected records. Now, you can either delete or extract. If you delete, it just eliminates the records. If you extract, it will create a separate file with the records that you extract. I'm going to delete . Now you have to go all the way down to the bottom. It is a little counterintuitive, but you have to go to the bottom to find the records, and it is going to the default. I want to go here. So, I have my records there. And, the display field, it is a default so the field that will display right appear for each record is the 245 subfield A. Actually want it does display subfield A. So, then, once I got into the right field, I want to import and it brings them all in here. I forget how many are in this set. 592. And, then the field is displayed. If I had forgotten to change that, it is no problem to just type it back in here, reimport and it will give me what I want. So, no big deal. And, I'm going to sort this by the numbers. To see if it has any records and it that start with this prefix. Or with this pseudo-number. Because, I want to remove any of the records that have this. So in our context, this is the situation for why we do this. Because of our profile with FDLP with what resources we get, we have set it up to get public lots which have numbers. They come in two forms. One of them is a sheet of paper, some of them are hundreds of pages long. So, they come individually and then we get them as a bound version of all of those one leaf and 500+ page things all come down together. In our environment, we have catalogued this particular title as a serial. So, while we want each of these individual things so that we can have the most current ones available, we keep them as a bound volume in our serial set. So, we don't want to catalog each of these individual ones as monographs but we do get monograph records from archives. The could be a way that we adjust our profile with Marchive to adjust these, but I have not explored that yet. I want to keep and have bound some of the items, I want to keep the records, but not have them in the files that I want to import. I guess that's why have not investigated getting rid of all of these records, but I want to get them out of this file, so I could click this box to the left for all of these. But, there are many. So, in a sweeping able to find them is to search the site field and it defaults by putting this hashtag . I don't quite know what that means. I'm just going to remove it and type in what I wanted to search for. And found an selected 44. So very quickly, I was able to select all of them. And, I can either generate every port, which I'm not going to do, or I could delete. And is going to delete those. It is going to create a new set. Or, a new file for this. That's coming this way, when I exit, the original file that I had used, this remove records now that I was using, all of those have been removed and now I can import my file into Alma. Are there any questions on removing the sort of things?

Yes. Mary said how did you get the delete selected button to not be gray? I missed the step before this one.

I will just walk through that process again. I will go to tools, select, delete, and import. I will display them. And it is not gray. So, it is great until you actually load the record in this particular tool. I hope that answers the question.

We have another question for you. Is there any way to remove records based on a list created in a different program?

I think the only way is to, well, the only way that I'm aware of is that if you deleted by a specific field, so, I don't think you can import your other data into MarcEdit and delete. So, I guess I would answer that, as far as I understand, there is no way to do that.

And thanks. Mary says thank you as well. Any other questions?

I'm not seeing any at this time.

All right. So, we have 10 minutes. I'm going to show you very quickly one other tool. I need to make my, I'm going to make my online file. You need to do that as a file and then save it. Let me go back and find that. For those of you who have never used MarcEdit, this is how you do this. Now you verify URL. I can run that against my online file. And I'm going to create an output file. And, we probably won't -- it's not just one thread, this is time-consuming as well. It is going to check in for all of the URL in the record to see if there are any invalid URL, this has been very helpful, because, as you saw, there was one that had a link that was missing the GPO number. I think that might not have shown up in the verify URL process, because it did actually go somewhere. And, this file is probably too big to wait around for it to be done see could see the final products, but, their varying stages of verifying that it goes through and they will give you a list of how most of them are to say that they are all valid. There are usually couple at the top where either the resource was unfounded all, or the link went to a generic page. And to check those. Sometimes those the only place that I can go to. Has to search that page for a specific record. Sometimes he just points of the wrong page because, you know, maybe the government agency has changed the pages and now it's pointing at the wrong page or a different resource. So anyway, this is a very quick way, well, not quick to do if you have a mixed file, but it's a nice way to be able to check all those URL when you have a online record, just make sure that you are loading L URL that are going to load for your users. Now I'm going to stop sharing. We are not going to wait for that. And so, are there any final questions?

A couple, actually. Is there way to create a MARC file from a binary file?

MARC files come as binary. So, oh, yeah, they are already binary. When you break them, they become what is called a mnemonic file which can be read in MARC article. Then you move it back into a binary file. So if I'm not understanding your question, let me know.

Another one for you. How did you learn how to do these various MarcEdit tasks? Did you earn from a mentor, did you take a class?

All of the above. MarcEdit was really intimidating to me to learn the tool. I've seen enough demonstrations of what I could do you know that I wanted to learn how to use it, so I just gave myself a chance to do it and actually, I made a proposal to a conference to present a workshop on this which forced me to learn how to use it. So looked at the tutorials. I watch the videos. I read about some online help that has improved over the years, and then I just dived in there and figured it out. I don't really have a mentor for it, because actually became a mentor to everyone at the conference that I go to. I am the MarcEdit person. Figured out how to do that and help them, so, there is enough information out on the web that there is a lot of help out there. And plus, there is a listserv that I am subscribed to and whatever the question, I posted to that and their experts on that listserv that will help you, but Terry is one of the first people to answer the question, so, I highly recommend that if you're interested in using MarcEdit, just go to the website and download the tool and look at the website and all of the resources that are available to you. Go to YouTube and search for MarcEdit, and sign up for the listserv . These are all very helpful places to learn how to use them.

Is this MarcEdit L?

And I have another,, compile into MRC takes file from MRK to MRC.

Yeah, it is, it is called the setback after the file name. We have the binary file and the mnemonic files that can be read by MARC editors. File extension?

That is it.

Thank you.

Thanks. And, it is in response to what is the listserv.

It is available by subscription information is available right of the website for the MarcEdit website. Terry will have that information there.

Of the question is is there way to delete all fields except for one? I just want to keep the 035 .

Yes, so, let me share. Actually, I don't know. I was again going from -- yes, there is. There's always a case we have to figure out, what am I trying to do here?

It is still running. Let me think yes. If you go to the limited text transmitter, the source file, I do this all the time.. If I don't buy source file. Sorry, all files. That I will just run this line. The output file. Then I save. It will bring it as a text file, actually. So the limited tank. Then next. Yes, you just -- he has change this since I use this. No, I am sure it was the text translator.

Would work for you if we share the contact information for the person who asked the question?

Yes.

Because we just have one minute before the top of the hour. Thank you, I really appreciate that. And a few more comments, if the file has a different file extension, for example, connection expert has a different file, you can choose this to go into MarcEdit? And you can export files to separate files. I think for the demo, you want to export tabs. Then we have thank you very much for your presentation.

You are very welcome.

On behalf of GPO, thank you very much. We really appreciate it. There more comments coming in here at the top of the hour. So there is a shared URL for everyone. It is a satisfaction survey. If you could complete that, we would really appreciate it.

Thank you. I appreciate it. Will get back to the person you have the question about that shield.

Thanks so much, everybody. Look forward to seeing your satisfaction survey responses. Thank you. Have a good day. [Event Concluded]