Analysis of MARC export functionalities in ArchivesSpace

Overview:

In anticipation of implementing ArchivesSpace ("ASpace") in NYU’s Special Collections, the functionality of ASpace’s MARC export was analyzed and compared against current functionality available through Archivists’ Toolkit ("AT"). This analysis was performed to discover new functionality in ASpace that corrects known issues from AT, identify gaps in functionality in ASpace, and assess the continued need for post-export record cleanup. Multiple MARC records were exported from the same resource records (e.g. data for fully-described collections) in both AT and ASpace version 1.1.2. In addition, new resource records were created in ASpace to analyze data fields and options not previously available in AT.

A note on how this analysis relates to Aeon is in Appendix 1.

A note on results of re-testing the MARC export in v. 1.4.2 is in Appendix 2.

Current Functionality in Archivists’ Toolkit:

Presently with AT, a Special Formats cataloger batch processes exported MARC records using Regular Expressions in MarcEdit and PyMARC Python scripts for analysis and cleanup of known MARC export errors. These errors include:

- Collection-level dates entered as single years or entered incorrectly do not properly populate the Date1 and Date2 bytes of the 008 (fixed length data) field (e.g. “18961896” rather than “1896\\” for a single year). In addition, for single year dates, the Date Status code in the 008 field remains coded for inclusive dates (“i”) rather than single year dates (“s”).

- The second indicator of the 245 (title) defaults to a blank, indicating there are no non-filing characters leading the title. However, the second indicator for collection titles beginning with non-filing articles should indicate the number of characters the system should skip to properly sort the collection title in an alphabetical list (e.g. second indicator “4” for a title that begins with “The”).

- Both the abstract and collection-level scope and content note are exported in the 520 (summary) field with blanks for the first indicator, whereas abstracts should have a first indicator of “3” and scope and content notes a first indicator of “2”.

- Biographical/historical notes (545 field) and scope and content notes (520 field) are both included in the exported MARC record, though often these notes will exceed character limits for the catalog record fields.
Syntax in subject headings (6xx fields) requires conversion from the form it is entered into AT (e.g., “United States |x History”) to proper MARC formatting (e.g., “United States$xHistory”). Subject headings in AT are entered with this formatting to accommodate for the single text entry box for subject headings.

Name headings (600, 610, 700, and 710 fields) do not export with proper punctuation between subfields unless the punctuation is expressly entered during AT data entry. In addition, name heading subfields do not export in MARC order.

Core functionality in AT MARC export does not include the creation of 853, 863, and 949 fields for item records and barcodes. These fields are necessary in records both for offsite transfer of archival collections as well as future implementation of Aeon as a circulation management system in Special Collections. The NYU Exporter plugin for AT adds necessary functionality for the creation of 853 and 863 fields, though 863 fields are created for every instance type (e.g., item, folder, etc.) rather than top-level containers (e.g., box or volume).

Lastly, use of MarcEdit is necessary to convert the AT-exported MARC .xml files into .mrk format. MarcEdit also allows the cataloger to combine multiple exported files into a single MARC file for batch processing and upload.

New Functionality in ArchivesSpace:

MARC export in ASpace corrects several deficiencies from AT export, including:

- Single year collection dates are now properly entered into the Date1 and Date2 fields in 008 (e.g., 2008\).\).
- Abstracts and scope and content notes now export with the correct first indicator coding of abstracts and scope notes in 520 field.
- Subject heading (6xx) subfield indicators are now properly created for each subfield, as ASpace now allows for granular data entry of subject headings.
- Name heading (600, 610, 700, and 710 fields) subfields now export in correct order.

Gaps in Functionality in ArchivesSpace:

While many of the export issues from AT were addressed in development of ASpace, the new system continues to export errors, including:

- For single year collection dates, the Date Status code does not change from “i” (inclusive) to “s” (single).
- Name headings (600, 610, 700, and 710 fields) do not export with proper punctuation between subfields unless the punctuation is expressly entered during
ASpace data entry.

- The current Agents module does not accommodate conference/meeting names as controlled name headings (111 fields). These headings, when used, will have to be entered as corporate names (110) and manually edited after export.

- The second indicator of the 245 (title) defaults to a blank, indicating there are no non-filing characters leading the title.

- The date subfield in name headings (1xx, 6xx, 7xx) currently does not export. Further, the date field for name headings (“Agents” in ASpace) does not appear as it is described in ASpace user documentation. This problem is resolved in version 1.3.0.

- The Preferred Citation (524) field is currently mapped to the Original Version Note (534) field. This is a known bug that has been reported to ASpace developers.

- Basic bibliographic unit and item records (853, 863 and 949 fields) are not created for the exported MARC record; therefore item records and barcodes are not included in the record.

It is pertinent to note that all fields export in the order they appear in the ASpace resource record, and ASpace exports every note field included in the record. As a result, subject headings and added name entries (6xx, 7xx fields) will appear before notes (5xx fields), and all collection-level notes will export in the order they were entered into the Resource record.

ASpace MARC records, as with AT, are exported in MARC .xml format and will require the use of MarcEdit for conversion into .mrk format.

**Recommendations:**

ArchivesSpace offers several improvements upon MARC export functionality in Archivists’ Toolkit. The errors present in ASpace MARC records should not impede our ability to implement the system, but do require delineation of content standards and standard field usage—as well as possible plugin development—in order to meet existing functionality. Content standard decisions—addressing which fields should appear in exported MARC records as well as how data should be properly entered into fields in ASpace—made by Archival Collections Management can address almost all of the current gaps in ASpace export functionality. MarcEdit will continue to be a useful tool for conversion of files into .mrk format and batch upload, as well as correcting the small number of records that will continue to export with errors; single year collection date coding and non-filing characters in titles are two issues that will persist.
Very few MARC export related issues remain in the ArchivesSpace developers’ issue tracker. Only the incorrect mapping of the Preferred Citation field affected records created for this testing and analysis. We should report the date field discrepancy and export issue for Agents/name headings to ASpace developers.

After consulting with members of KADD, we have determined that the ideal workflow for moving MARC records between ASpace and the ILS will remain exporting .xml MARC records that are batch processed in MarcEdit before upload to Aleph. Other solutions considered were connecting the systems via the ArchivesSpace and/or ALEPH / PRIMO APIs. Since Aleph includes a number of legacy MARC records with either outdated descriptive information or dead links to finding aids, we also discussed with KADD members the potential to clear all existing MARC records from the catalog and batch export and upload MARC records exported from ASpace.

To match existing functionality and continue existing practice of uploading exported MARC records into Aleph, we will need to customize the ASpace MARC export to include additional fields for item records and barcodes. Before customization, ACM will coordinate with the offsite storage supervisor to determine what specific information Clancy requires in 863 and/or 949 fields to successfully transfer material offsite.
Appendix 1 - Relationship to Aeon

The Phase I deployment of Aeon will shortly follow the implementation of ArchivesSpace—the target deployment date for ArchivesSpace is Spring 2016, target deployment date for Aeon is late Summer or Fall 2016. Aeon will need to draw, via OpenURLs, on MARC-based data in BobCat to auto-populate information about the item(s) a user is requesting and minimize manual re-entry for the end user or special collections staff.

This OpenURL connection drawing on MARC info will be critical for allowing users to request, via a ‘Request’ button in Bobcat, special collections materials that are under bibliographic control (e.g. books, some AV) and do not have finding aids. These materials do not have archival metadata records and MARC data for these is created via other cataloging workflows. The quality of ArchivesSpace MARC export will therefore have no bearing on the Aeon - ILS connection for this material.

For archival collections we have several options:

- We may choose to provide ‘Request’ buttons in Bobcat at the collection level for archival collections that also have records in ArchivesSpace /Special Collections Search. Princeton has set up their requesting system in this way. In this case, basic data needed to submit such a request - collection title, collection call no., location / repository would be available in archival MARC records to support that. Users would then supply additional data to a request via an Aeon webform.

- We may wish to make available ‘Request’ buttons in Bobcat for only individual containers / items within an archival collection. Most of the information needed to print a call slip would be available from the MARC record and could pre-populate into an Aeon request form, specifically the collection title, collection call number, repository, container label, barcode, and indicator re. whether the collection is off-site. Two pieces of information that will be important for paging will not be available via the MARC record -- namely, container-level access restrictions and shelving location for on-site materials. Complex scripting or a manual lookup in ArchivesSpace for each request that comes through Bobcat may still be required.

- We may choose to simply re-route users from the collection-level MARC record to the finding aid to submit their requests (e.g. via a note such as ‘To request for use in the Fales Reading Room select a box in the finding aid’ that links to the finding aid). Yale has implemented their connection in this way.

Early in the Aeon implementation and before the production ArchivesSpace deployment (winter /spring), further consideration will need to be given to each of these options. At this point, however we do not see any problems for Aeon caused by the ArchivesSpace MARC export.
Appendix 2 - Results of re-testing in v1.4.2

In December 2015, the MARC export functionality of ArchivesSpace was re-analyzed using a more recent release of the software, version 1.4.2. This round of testing confirmed that some improvements in the MARC export have been implemented, such as the date subfield in name headings (1xx, 6xx, 7xx) exporting properly into subfield d, and the presence of a free text date field in the agents module. All other gaps in functionality identified in this report are still present, though they do not present a barrier to implementing ASpace.

These gaps include:

- For single year collection dates, the Date Status code does not change from "i" (inclusive) to “s” (single).
- Name headings (600, 610, 700, and 710 fields) do not export with proper punctuation within and between subfields unless the punctuation is expressly entered during ASpace data entry.
- The current Agents module does not accommodate conference/meeting names as controlled name headings (111 fields). These headings, when used, will have to be entered as corporate names (110) and manually edited after export.
- The second indicator of the 245 (title) defaults to a blank, indicating there are no non-filing characters leading the title.
- The Preferred Citation (524) field is currently mapped to the Original Version Note (534) field. This is a known bug that has been reported on the ASpace JIRA (AR-1189).
- Basic bibliographic unit and item records (853, 863 and 949 fields) are not created for the exported MARC record; therefore item records and barcodes are not included in the record. We are doing local development work to customize the MARC exporter to accommodate this.
- All fields export in the order they appear in the ASpace resource record, and ASpace exports every note field included in the record. As a result, subject headings and added name entries (6xx, 7xx fields) will appear before notes (5xx fields), and all collection-level notes will export in the order they were entered into the Resource record.

Developers have corrected the Preferred Citation field mapping error, and the fix should be included in a future release of ASpace.