**Preprocessing Notes – v2.4, 10/18/2023**

**Preprocess 0 – Splitting MARC records to create a single Work with multiple Instances in BIBFRAME**

Unlike the other processing documents and the specifications for marc2bibframe, Preprocess 0 addresses the treatment of MARC records *before* they are converted to BIBFRAME Works and Instances. This process is a set of instructions to split a single MARC record into multiple MARC records based on data in the 007, 3XX and 856 fields.

After the marc2bibframe specifications are applied and the conversion processes are run, the final result will be a single BIBFRAME Work with multiple BIBFRAME Instances.

For example, consider a MARC record that describes a paper map and a digitized image of that map. The information about the digital version exists in fields 007 and 856. Preprocess 0 will convert the single MARC records into two – one for the paper map and one for the information related to the digitized image of the map.

If the MARC record has one 007 field, a record split is not needed and the MARC record is converted to BIBFRAME using the existing specifications and programs.

Initially, Preprocess 0 will only convert records that describe single-unit or collection resources (L/07=”a”, “c”, “d”, “m”). The program will be modified as needed to include additional types of bibliographic resources in future releases.

* 1. Common data elements in all records created by this process include:
* Leader: same in all records
* 001 field: unchanged for the first MARC record; existing number+increment for the new MARC record(s).

**Examples**:

001 21860952

001 21860952-02

001 21860952-03

* 005 field: same in all records
* 008 field: same in all records
* 040 field: same in all records

**0.2)** The newly-created MARC records contain a 758 field that links to the original MARC record. It contains the URI of the BIBFRAME “instanceOf” property in $4 and a URL link to the Work in $1.

**Examples:**

758 ## $4 <http://id.loc.gov/ontologies/bibframe/instanceOf> $1 <http://example.org/21860952>#Work

758 ## $4 <http://id.loc.gov/ontologies/bibframe/instanceOf> $1 <http://example.org/12577907#Work>

**0.3)** If the MARC record has no 007 field -- but an 856 field with a second indicator of “#,” “1,” or “8” exists -- the record is split, unless the 856 field contains “$3 Table of Contents …”. See also Process 3.

* The first MARC record contains all MARC fields except for the 856 field.
* The second MARC record contains the 856 field, a default 007 field, and the common data elements.

**Example:**

000 01447ccm a22003494a 4500

001 12983422-01

005 20190207063621.0

007 cr |||||||||||

008 021028s1869 ohusga n eng

040 ## $a DLC $c DLC $d DLC

856 41 $d mussuffrage $f mussuffrage-10183 $u <https://hdl.loc.gov/loc.music/mussuffrage.mussuffrage-10183> $q h

758 ## $4 [http://id.loc.gov/ontologies/bibframe/instanceOf $1](http://id.loc.gov/ontologies/bibframe/instanceOf%20$1) <http://example.org/12983422#Work>

**0.3.1)** URLs containing the following patterns should not be processed:

\*books.google.com\*

\*eblib.com\*

\*ebrary.com\*

\*ebscohost.com\*

\*exlibrisgroup.com\*

\*firstsearch.oclc.org\*

\*galegroup.com\*

\*mylibrary.com\*

\*overdrive.com\*

\*proquest.com\*

\*sagepub.com\*

\*serialssolutions.com\*

\*springerlink.com\*

\*springerprotocols.com\*

\*ezproxy\*

\*login\*

\*username\*

**0.4)** If the MARC record has two 007 fields where the second 007 field begins with “c”, and an 856 field exists, the record is split.

* The first MARC record contains all MARC fields except for the 007 field beginning with “c” and the 856 field.
* The second MARC record contains the 007 field beginning with “c” and the 856 field, along with the common data elements.

**Example:**

000 03248ckm a2200529 a 4500

001 12154212-03

005 20161117095820.0

007 cr|||

008 000824q18141820dcunnn | kneng

040 ## $a DLC $c DLC $e gihc

856 41 $3 color film copy transparency $d cph $f 3b51810 $u <http://hdl.loc.gov/loc.pnp/cph.3b51810>

758 ## $4 <http://id.loc.gov/ontologies/bibframe/instanceOf> $1 <http://example.org/12154212#Work>

**0.4)** If the MARC record has two 007 fields and a single 300 field that includes accompanying material in $e, the record is split.

* The first 007 field is paired with the 300 field data that appears before $e.
* The second 007 field is paired with the 300 field data that appears after $e.
* Subsequent MARC records contain each paired 007 and 300 field combination, along with the common data elements.

**Example:**

000 02930cjm a22003857a 4500

001 13091110-02

005 20060112141656.0

007 v| ||||||

008 030205s2001 io fmn chi

040 ## $a DLC $c DLC $d DLC

300 ## $a 1 videodisc (VCD)

758 ## $4 [http://id.loc.gov/ontologies/bibframe/instanceOf](http://id.loc.gov/ontologies/bibframe/instanceOf%20$1) $1 <http://example.org/13091110#Work>

**0.5)** If the MARC record has multiple 007 and 300 fields, the record is split.

* The 007 and 300 fields are paired – the first 007 field is assumed to match the first 300 field, the second 007 field is assumed to match the second 300 field, etc.
* The first MARC record contains the first 007 field, the first 300 field and all remaining non-007 and non-300 fields.
* Subsequent MARC records contain a paired 007 and 300 field combination along with the common data elements.

**Example:**

000 03366cgm a22006017a 4500

001 11510602-02

005 20120822113119.0

007 mr bf dnnaztnnau

040 ## $a DLC $c DLC $d DLC $e amim

300 ## $a 1 film reel of 1 (358 ft.) : $b si., b&w ; $c 16 mm. $3 non-viewing pos pic.

758 ## $4 [http://id.loc.gov/ontologies/bibframe/instanceOf](http://id.loc.gov/ontologies/bibframe/instanceOf%20$1) $1 <http://example.org/11510602#Work>

**0.6)** If the MARC record has multiple 007 fields, at least one 300 field, and the other 3XX fields contain information in $3, the record is split.

* The 007 and 300 fields are paired as above.
* The text strings in $3 are matched and the 3XX fields are placed in the applicable MARC record.

**Example:**

000 03651cjm a2200745 i 4500

001 21930318-02

005 20220503152518.0

007 vd cvbizq

008 210308p20171970enkrcn| i | eng d

040 ## $a OCP $b eng $e rda $c OCP $d OCLCO $d OCLCF $d B@L $d DLC

300 ## $3 videodisc $a 1 videodisc (approximately 84 min. )$b sound, color $c 4 3/4 in.)

336 ## $3 videodisc $a two-dimensional moving image $b tdi $2 rdacontent

337 ## $3 videodisc $a video $b v $2 rdamedia

338 ## $3 videodisc $a videodisc $b vd $2 rdacarrier

344 ## $3 videodisc $g stereo $g surround $2 rdacpc

344 ## $3 videodisc $g Dolby Digital Stereo $g Dolby Digital 5.1 surround $g DTS Surround Sound

346 ## $3 videodisc $b NTSC $2 rdabs

347 ## $3 videodisc $a video file $2 rdaft

347 ## $3 videodisc $b DVD video

347 ## $3 videodisc $e all regions $2 rdare

758 $4 [http://id.loc.gov/ontologies/bibframe/instanceOf](http://id.loc.gov/ontologies/bibframe/instanceOf%20$1) $1 <http://example.org/21930318#Work>