Monthly TSzserv Run for Print, Audio, Video, and Software: Background and Procedures

I. Objective
The monthly automated Z-Server process is designed to locate full records containing more extensive bibliographic information to overlay matching minimal level records in Aleph. The resources described by these records have been assigned permanent call numbers and shelved for user access in the owning library. For this process, a full record is defined as one containing all of the following:

1. At least one LC subject heading (6## with second indicator 0)
2. A 300 field that does not contain the strings “p. cm.” or “p. ;$$c cm.”
3. Encoding level (LDR/17) is blank (^), 1, 2, 4, I, or L.
4. A 040 that does not contain $$b

This strategy is a primary component of the Technical Services effort to more efficiently process materials and make them available to users more quickly while also enhancing the bibliographic records with information readily available from peer institutions.

II. Background
The record processing workflow currently in place in Technical Services is designed around the use of some form of automated bibliographic record enhancement. Such a process was previously used for the former Monograph Cataloging Division’s Managed Backlog collection. The collection consisted of books either lacking full records or lacking bibliographic records altogether. Each book was assigned a temporary accession number-based and shelved in non-user space. The automated process used the Z-Server, a Z39.50 protocol-based method, to perform batch searches of other Z39.50-accessible catalogs in pursuit of full matching records to replace or overlay the existing records. This initial version of the Z-Server process extracted the first ISBN found in each Managed Backlog bibliographic record to search the Library of Congress catalog. If no record was found in the LC catalog, OCLC WorldCat was searched. When a record was successfully overlaid, the book was pulled from the Managed Backlog, local cataloging was completed, and the book was labeled for permanent shelving. Records lacking ISBNs were identified and manually processed since programming for title/author searches was deemed too complex at the time. Five was set as the maximum number of search attempts in a twenty-four month period. When that number was reached, the books were to be pulled from the Managed Backlog shelving for cataloging locally or through a third-party vendor. Marcadia, a service offered by Backstage Library Works (previously available from RLG), functions in a similar manner.

The new version of the Z-Server process, also known as TSzserv, is to be used for books, serials, music recordings, videos, and software. ISBNs (020 field), ISSNs (022 field), and
publishers’ numbers (028 field) will be the primary search keys used to find full records. The capability to use title searches (combined with author if 1XX is present) when these numbers are lacking is currently under development. Records will be searched monthly for eighteen months. After eighteen months, the materials will be pulled from the stacks and routed to the Original Cataloging Unit to receive complete bibliographic records. The Z-Server process is currently being used to overlay full records onto minimal level records for on order materials and received Korean language materials. Processes are also being developed to update multipart monograph records and overlay minimal level records for electronic resources.

III. Criteria for TSzserv Processing
Records for books, serials, audio CDs, DVDs, and VHS tapes are flagged for TSzserv/Z-Server processing in the Monograph Receiving Unit and the Classification Unit. Macro Express macros used in both units are written to flag a record by adding a 996 field containing TSzserv in subfield $c when a record lacks an LC subject or contains “p. cm.” in the 300 field. A staff member can also manually add the TSzserv flag if they notice an error in a primary access point field (e.g., author, title, subtitle, etc.). The macros do not automatically flag records for works of literature that contain complete LC P classification numbers and lack LC subject headings because subjects are not required for these records. This does include criticisms and histories of literature.

The TSzserv is not intended at this time to process records in non-Western European languages. However, unlike the Managed Backlog Z-Server process, this process does not check the language code in the records. The expectation is that only Western European languages materials are processed in Monograph Receiving and the Classification Unit. There is nothing in the macros or the TSzserv programming that checks languages. All records containing 996 subfield $cTSzserv will be processed regardless of the language code in the 008/35-37 positions.

Records flagged for TSzserv processing should be saved at CAT level 10 so that the records do not go through MARS processing. The macros used in Monograph Receiving and Classification automatically save bibliographic records at level 10 when the TSzserv flag is present in the bibliographic record.

IV. Process Sequence
1. Records matching the TSzserv criteria are automatically flagged with a 996 $$cTSzserv during processing in the Monograph Receiving and Classification Units. Selected records may be manually flagged.
2. Each month all records flagged with TSzserv are to be extracted for processing. There will no longer be a ‘ripening’ period. Records that were previously unsuccessfully
processed will contain TSzserv and a search counter. Records will be searched no more than eighteen times.

3. The first 020, 022, or 028 field will be extracted from each record. If the record lacks these fields, the title and author are to be extracted. These will serve as the search terms.

4. The search terms will be submitted to the Library of Congress catalog via the Z39.50 protocol. All records that match the search terms are to be retrieved and checked for fullness.

5. Records found to be full will undergo two additional checks. First, the 040 field will be checked to ensure it does not contain subfield $b. The subfield $b typically indicates the record was cataloged by a non-English speaking cataloging agency. Because the subfield $b puts the accuracy of the LC subjects in question, the record will be rejected and considered not full. Second, the 245 field is checked for subfield $h. If the field contains subfield $h and the GMD matches one found in the minimal record’s 245 field, the record will be accepted. If the GMD does not match or there is no subfield $h in the minimal record’s 245 field, the record will be rejected and considered not full.

6. Search terms that failed to find in LC’s catalog at least one full record that could be used for overlay will then be submitted to the OCLC WorldCat catalog via Z39.50. The retrieval and evaluation process will repeat.

7. If two or more full records are selected from either LC or OCLC for the same search term, the records will be scored based on the number of fields present in the records. The record with highest score is selected for the overlay. Currently all fields have the same value but certain fields can be given more weight in determining the score. Possible weighting: LC subject (6XX X0) = 3 points, Added entry = 2 points, All others including Notes = 1. Should encoding level and number of holding institutions (available in 948 field via OCLC’s Z39.50; this function is active yet in Aleph) also be utilized to weigh the records?

8. If the two or more records have the same score, the overlay candidate records are to be loaded into Aleph’s MiU02 database with the minimal record’s MiU01 SysID in 035 field. The MiU02 SysIDs of the overlay candidate records will also be included in the 996 field of the minimal record.

9. The 245 $a and $b subfields of the best match full record will be compared to the 245 $a and $b subfields of the corresponding minimal level record. Because the 245 fields of minimal records often do not include subfield $b but do contain subtitles, subfield delimiters will be stripped from both 245 fields so the actual contents of the fields are being compared, not the coding. If the 245 $a subfields of the two records differ, the 245 of the minimal record will be moved to a 246 subfield $a in the full record. It will be preceded by subfield $iOrder title.
10. All subfields (subfield $3, subfield $u, and subfield $z) in 856 4^, 856 40, and 856 41 fields in the full records will be changed to subfield $x. This will be done to prevent URLs that are invalid for UM users from displaying in Mirlyn. Staff in the Electronic Cataloging Unit will evaluate the suppressed 856 fields and reactivate the links or delete the fields.

11. All 856 42 fields in the full records will be deleted.

12. The minimal level records will be overlaid by the selected full record. Any existing 006, 007, 520, 533, 590, 856, and 998 fields in the minimal record will be retained. Information in all other fields, including non-LC subjects, MeSH, and added entries, will not be retained unless it is also present in the overlaying full record.

13. All 440 or 4901 fields in the newly overlaid record are converted to 4900. When subfield $n and subfield $p are encountered, the contents of those subfields are moved to subfield $a.

14. All 800#0, 810#0, 811#0, or 830^0 fields are deleted.

15. The 996 field is deleted. A new 998 with overlay in subfield $c and the current date in YYYYMMDD format is added for statistics purposes. The record is saved at CAT level 20.

16. The holdings XPR subfield $a is set to OCLC and the subfield $b is updated to the current date.

17. All overlaid records will go through the MARS processing, including those that may have been prematurely saved at cat level 20.

18. A report of the number of records overlaid, with subtotals for the number found in LC and OCLC, will be provided by Library Systems.

V. Multiple Full Records
There will be occasions when multiple full records with the same score will be retrieved for minimal level records. This is due to the existence of duplicate records in WorldCat, the occasional practice by publishers of reusing their identifier numbers (ISBNs, ISSNs, and publisher numbers), and the relative imprecision of title/author searches. As described in IV.7 above, these overlay candidate records are to be saved to Aleph’s MiU02 database with the corresponding minimal level record’s MiU01 SysID in a 035 field. The SysIDs of the overlay candidates will also be added to the 996 field of the minimal level record in a subfield $f. Authorities & Database Management staff will review these records, select a record based on the information available in the minimal level record, and perform a manual overlay in Aleph.

VI. Failed Searches
When the search for a full record fails, the TSzserv flag will be retained in the 996 field. A counter integer with the initial value of 1 will be added to the field in subfield $n. This counter is to be increased by one each time the search fails. When the counter reaches
18, the indicator that the record has been searched for 18 months, the record will be reported for review. The language, publication date, format, and classification of these records that are not overlaid by full records will be analyzed to generate profiles of their characteristics. This data will be used to determine which materials should be pulled from the stacks to receive fuller cataloging as well as those for which fuller cataloging is unnecessary.

VII. Evaluation and Quality Control
All overlays that contain the order title in a 246 field are to be reviewed since this will indicate some variation in the 245 strings of the minimal level record and the full record that overlaid it. Records found to have clear disagreement between the 245 and 246 will be counted as incorrect overlays. That number will be used to determine the overlay accuracy of that particular batch. The minimum level of accuracy of any given monthly batch is to be 96 percent. Records found to be incorrect overlays will overlaid with the correct full record by Authorities & Database Management staff. Incorrect overlays for which no full record is available are to be referred to the Head of Quality Control & Database Management for further investigation.