S1. PREMARC File/PREMARC Replacement Process

S1.1. Summary History

S1.1.1. Methodology Used to Create the File

In 1979, Carrollton Press and University Microfilms International issued a microform copy of the Library of Congress shelflist. In that same year, Carrollton Press announced that it would print a title index to this microform shelflist. The index would be generated from machine-readable records--from existing MARC records and from non-MARC records that would be converted to machine-readable form by Carrollton Press. The resulting body of machine-readable shelflist records was marketed by Carrollton Press under the name REMARC. Carrollton Press was acquired by UTLAS in 1986; UTLAS continued to market the records for several years until UTLAS itself was acquired by another company.

In 1980 the Library of Congress entered into an agreement with Carrollton Press to buy copies of the REMARC records. The non-MARC records, which were keyed in Scotland from hard copy "blowbacks" of the microform shelflist records, were machine-readable but contained little content designation. These unedited records were sent to the Library on tape for further processing.

The principal feature of this processing was a suite of programs, which together were called Format Recognition. Format Recognition took unedited cataloging data and produced a content-designated, MARC-formatted record. These programs were developed at the Library and were used as a part of LC's machine-readable cataloging (MARC) workflow until the advent of online input/update in 1980. Format Recognition automatically content designated records based on "clues," such as keywords and punctuation, which implied specific content designation. A simplified explanation of the way this processing worked follows: the record was scanned for various clues--the presence of the abbreviations "p." and "cm.", for instance--which indicated that the physical description had been located. Scanning backwards from that point, the programs used punctuation and keywords to identify the imprint, edition statement, and title, and assign the appropriate tags, indicators, and subfield codes. The record was also scanned forward from the physical description and tagging for fields beyond the physical description, including added entries, was supplied. The fixed field data were generated. The resulting records were then loaded into a separate file called the PREMARC file (originally called the Shelflist file) and were indexed through both SCORPIO and MUMS.

S1.1.2. PREMARC Input/Update System

Although the Format Recognition programs were very powerful, the results are not perfect because of keying errors, ambiguity in bibliographic data, and varying cataloging practices that were never incorporated into Format Recognition. During the 1970's, LC practice was to verify MARC records after Format Recognition. Since this verification step has not yet been undertaken for PREMARC, except on a very limited scale, the records have accumulated in the PREMARC file in their unverified state. An input/update system became available in April 1988 and many thousands of corrections have been made to individual PREMARC records by staff in the PREMARC Team, formerly in the MARC Editorial Division and currently in the Policy and Standards Division (PSD). In addition to various kinds of special clean-up efforts, records in the PREMARC file are corrected to reflect changes that affect access to material, usually changes in call numbers.

S1.2. Contents of the File

During 1978 and 1979, University Microfilms International microfilmed the Library's shelflist. Thus, the bulk of the records in the PREMARC file reflects the state of their counterpart shelflist cards as they were at that time. The "shelflist" actually had three components:

1) the principal card shelflist, housed in and maintained by the then Shelflist Section, Subject Cataloging Division;

2) the atlas and map shelflists, housed in and maintained by the Geography and Map Division;

3) the music shelflist, housed in and maintained by the then Music Section, Special Materials Cataloging Division.

In later stages of the project, records for the following materials were added:

1) non-MARC items cataloged 1978-1983, principally JACKPHY language materials, music and sound recordings;

2) non-MARC sound recordings cataloged before 1978;

3) filmstrips and motion pictures that are in the Library's collections and that were cataloged 1952-1968;

4) the unclassified law collection.

The PREMARC Team has added records representing items in the Priority 4 (P4) Collection

and records representing various kinds of collections of ephemeral materials. As of September 30, 1997, the PREMARC file contained approximately 4.76 million records.

S1.3. Missing from the File

Since the PREMARC file is based on the shelflist, categories not represented there are excluded, e.g., pre-1898 records, unadapted cooperative cataloging copy from cooperative cataloging programs in place prior to 1983.

Originally, dashed-on entries, which were keyed as separate records by Carrollton Press, were not added to the PREMARC file because they contain a control number that duplicates the one used for the main item. A technique has been developed to accommodate supplements and other similar material (essentially, a 500 general note preceded by "Supplemented by: ... " is added to the record for the main item). As they are discovered, they will be added.

An unknown number of the duplicate records weeded out after Format Recognition processing are false duplicates. Most resulted from keying errors. The identification and disposition of these false duplicates are still unresolved issues.

S1.4. Characteristics of PREMARC Records

A PREMARC record usually does not reflect the complete catalog card as it appeared in the shelflist. The following data elements were candidates for inclusion in the record according to the original LC-Carrollton Press agreement:

001	LC control number: if not available, a new number was assigned with the prefix "unk" (unknown)
050	LC call number
051	LC copy statement
1XX	Main entry
240	Uniform title: not always input
245	Title: generally keyed only through the first "full thought"; statements of responsibility not input
250	Edition
260	Imprint: only the first place and name keyed when multiple places and publishers present; printers and/or distributors not included
300	Collation: only "p. cm." input in many records created in the early stage of the project
6XX	Subject headings
7XX	Added entries
8XX	Series added entries

041   Language codes
043   Geographic area code

The following data elements usually were excluded:

350   Price (field now obsolete)
4XX   Series statements
5XX   Notes
020   ISBN
022   ISSN
082   Dewey number
086   Superintendent of Documents number

Records in non-roman scripts were handled either by inputting only the romanized portions or by partially romanizing the records. A note (500) was added ("Romanized.") when the title was romanized from its original script. The imprint (260) often contains only the date.

Records input during the later stages of the project may contain more complete information; for example, most of the records for sound recordings contain all notes found in the original record.

Since all records are in the books format (the records were processed through Format Recognition which was designed for books) and since notes generally were excluded, records for items in some formats are camouflaged.

1) Maps. Scale information recorded in a note was excluded. Map records can be identified, however, by a call number in the span G1000-9999.

2) Serials. Much of the information that would indicate that a bibliographic record was created for a serial was excluded: beginning numeric or alphabetic designation, a note giving earlier titles, a note giving earlier issuing bodies. Thus, records appear at first glance to represent multipart items, since the imprint date is open or is a span of years. A note (500) was added ("Serial.") to some records at input; be aware, however, that this note was added erroneously to some records for multipart items. Other clues that might identify a serial record include the nature of the title itself, subject headings, and the presence of title added entries for other than partial or parallel titles. (But, earlier titles recorded on the verso of the ONC ("Official Name/Title Catalog" also referred to as "OCAT" ("Official Catalog")) main entry card as see references were not converted to added entries as part of the original input, since this catalog was not the source file for input.)

Another effect of including records for non-book items in the books format is that records for sound recordings and films have the header "Not in LC collection" when this is not the case in many instances.

S1.5. Search Strategy and Limitations

1) The PREMARC sign-on to the PREMARC input/update system is limited to those staff who need to work in the PREMARC file. If a cataloging team has such a need, consult PSD. While signed onto the PREMARC input/update system, one can retrieve non-PREMARC records by using the same file qualifiers used in MUMS (e.g., f=na, f=se). The exception is that a file qualifier is not needed for a control number search of other files while signed onto PREMARC.

   Possible search keys are PCRD, NCRD, CARD; PTK, PATK, PPNK; ISBN, ISSN; FIND whether signed onto the PREMARC input/update system or not. ONUM is not a search key that can be used in PREMARC.

2) While signed onto MUMS or another MUMS input/update system, it is only possible to retrieve PREMARC records by using the file qualifier "f=prem"; again, the exception is that a file qualifier is not needed for a control number search of PREMARC while signed onto MUMS or another MUMS file.

3) When searching for a title in PREMARC, be thorough. Since the non-filing indicator was not always set accurately, it may be necessary to include the initial article as the first element of a search key. The numbering designation may also need to be included in the PTK search key for series titles of less than 4 words.

   ```
   245        The book of the apple

       ptk    the,bot

   740        The debt of Texas.

       ptk    the,dot

   830        Pädagogisches Magazin, Heft 1341.

       ptk    pad,mh

   830        Forschungen zur Kolonialfrage, Bd. 3.

       ptk    for,zkb
   ```

   Sometimes diacritics were not assigned correctly, subfielding may be incorrect, or words were misspelled. Try different combinations.

4) Several problems have been identified that stem from incorrect tagging or problems with Format Recognition; only a few are listed here. (Many records exhibiting some of these

problems have been corrected.) If a record that should be in the PREMARC file is not retrieved, as a last resort use the FIND command without field identifiers.

a) name with a word the same as a title of address tagged as uniform title

```
130      King, Mary Perry.
```

b) some personal names tagged as corporate bodies

```
110      Eastman, Francis Smith, d 1846 or 7
```

c) some uniform titles tagged as corporate bodies

```
710      Die Klage
```

d) some corporate bodies tagged as personal names

```
700      Suffolk institute of archaeology and natural history, Bury
         St. Edmunds
```

e) some call numbers in a 110 or a 245 field with main entries in next field

```
110      PJ1425Y6
240      Young, Thomas, 1773-1829
```

f) some subordinate bodies coded as titles and some titles in name/title added entries coded as subordinate bodies

g) some fields missing or tagged incorrectly with other fields compressed into one field

```
050      Z152.09M2
245      Madan, Falconer, 1851-
300      A chart of Oxford printing '1468'-1900: [Oxford] Printed at
         the Oxford university press, February, 1904.  50 p. illus.,
         vii facsim. (incl. front., fold. chart) 28 cm.
```

h) many subject headings have incorrect 6XX tags (650 terms as 610, etc.); it may be more prudent just to use "find s [term]."

5) **Reminder:** The PREMARC file does **not** show usage since statements of responsibility and series statements are not included.

6) **Reminder:** When searching by control number, use a leading zero whenever the year prefix is a single digit. (This requirement is not limited to the PREMARC file but applies more often in PREMARC.)

   e.g. pcrd 04-32573 (for a number printed as 4-32573)

S1.6. Uses of the File

At the Initial Bibliographic Control (IBC) stage, the PREMARC file is searched in cases in which previously the manual ONC would have been searched. **WARNING: Since the file is still incomplete, absence of a record is not *ipso facto* evidence that LC does not hold the item.**

Descriptive catalogers should not search the PREMARC file routinely. As noted above, usage is not available since statements of responsibility and series statements were not included. However, in cases of material published before about 1970, it is prudent to search the PREMARC file and, if the item is not found, one of the manual files, to preclude cataloging a duplicate. This is particularly the case for older material that LC is likely to hold.

Access points are not necessarily correct; if the ONC card was corrected rather than reprinted, the shelflist card was not corrected. Since consistency in the form of heading in PREMARC does not necessarily signify the correctness of that heading, the PREMARC file cannot be used to determine the form for a linking reference in an authority record.

However, the PREMARC file can be used as a resource for catalogers, e.g., to confirm that a record exists in the ONC, to determine the probable entry element of a heading for searching in the ONC, to supply information for resolving conflicts in personal name headings, to aid in identification of authors and their subject areas, to supply for subject purposes a call number and subject headings for a serial continuing under a new title, etc.

S1.7. Upgrade Strategy for the File as a Whole

S1.7.1. Manual Upgrade (PREMARC Staff)

Studies were undertaken in the late 1980's and a general PREMARC upgrade strategy applicable to the entire file was developed. That strategy depended upon the PREMARC staff updating each PREMARC record based on the original source card in the Official Catalog. The plan was never implemented owing to a lack of resources and changing priorities.

S1.7.2. Machine Upgrade (Replacement Records From An External Source)

In 1992 an alternate strategy for upgrading the PREMARC file began to be pursued. This strategy was based on sending candidate records from the PREMARC file to be matched against records in the OCLC database. Successfully matched records would be returned to LC to be loaded into the BOOKS file; this would be possible because the records in OCLC had, for the

most part, been made compatible with AACR2 either through efforts to update the database by computer or through contributions from OCLC members. After an initial experiment with OCLC, the Cataloging Directorate issued a formal "Request for Quote" (RFQ) in 1993 to select, through competitive means, a source for PREMARC replacement records. PSD prepared the RFQ and managed the selection process. OCLC was selected as the source in 1993, and detailed and coordinated planning was begun by staff from the Automation Planning and Liaison Office (APLO), PSD, and Information Technology Services (ITS). Approximately 1.475 million PREMARC replacement records were loaded into the PREMARC file in April 1998 replacing counterpart PREMARC records that previously resided there. This is called the "PREMARC Replacement Process" (PRP) and the replacement records themselves the "PREMARC Replacement subset." (Various factors, including some technical ones, influenced the decision to load the replacement records back into the PREMARC file instead of the BOOKS file as originally planned.) See DCM S1.9 for more information about the PRP.

S1.8. Maintenance

S1.8.1. Basic Principle

Catalogers have no official responsibility for maintenance to the PREMARC file in connection with their regular duties. However, the following corrections may be initiated (but do not do extra searching to initiate them):

1) Change of form of heading to effect consistency between forms used as main and added entries within the same PREMARC subset (replacement; "other") but **not** across subsets (*cf.* DCM S1.8.3). (Headings are **not** updated solely to bring them in line with AACR2.)

2) Change to access point affecting retrieval (e.g., typographical/input error, incorrect tag).

3) Change affecting location of material (e.g., correction of error in call number).

4) Change of control number from "unk" number to real LCCN.

5) Cancellation of a PREMARC record, i.e., a PREMARC record duplicates cataloging for an item also represented by another record under a different LCCN in a file in the current catalog. (A cancellation is not required for an item undergoing "NEW INPUT" (DCM B5.12). A record is newly input into a file in the current catalog under the same LCCN as that of the PREMARC record; the PREMARC record is deleted either by the cataloging team or by the PREMARC team as part of that procedure.)

6) Notification that a bibliographic record should be added to PREMARC. (Double check this situation by searching title as well as LCCN, since the record could be in PREMARC

under an "unk" number.)

Other corrections would not normally be made. However, if another correction is deemed important, send an explanation of the problem via the team leader to: Team Leader, PREMARC Team, PSD.

S1.8.2. Procedures For Initiating/Making Changes

As an exception to the guideline in DCM C13.1.1 that unverified records cannot be changed, unverified PREMARC records are candidates for change. (Records were originally loaded into the file in an unverified state.)

S1.8.2.1. Changes to individual/multiple (less than 6) records

For categories 1)-5) in DCM S1.8.1 above:

1) Prepare a change request per the guidelines in DCM C13.2.2.d. Use the label "PREMARC CHANGE" in the upper right corner of the printout.

2) Forward change requests to: Team Leader, PREMARC Team, PSD.

For cancellations, include the LCCN of the record being retained. For records to be added to the file (category 6 above), send a photocopy of the ONC main entry recto. For additions to the file and corrections involving a call number or an "unk" card number, PSD staff will validate the call number and LCCN, usually in the shelflist, before the correction is made.

S1.8.2.2. Changes to multiple records (more than 5) for the same correction

For changes to more than 5 records for the same correction:

1) use a multiple record display if the changes can be conveyed clearly;

2) if the correction cannot easily be represented by a multiple record display, fill out a "Heading change" form (6-66) with the correction to be made, annotate the form "PREMARC," and send it to: Team Leader, PREMARC Team, PSD.

S1.8.3. Effect Of The PRP On Heading Structure

Prior to the PRP, the heading structure in the PREMARC file reflected, for the most part, pre-AACR2 rules. There are many exceptions to this in the form of inconsistencies with particular headings (more than one form for the same entity or concept) or the fact that some of the records added in the later stages of building the file contained records that exemplified

AACR2 forms (e.g., records for music). The heading structure is now further complicated by the PRP in that the subset of PRP records (identified by a 985 field containing "PREMARC REPLACEMENT" in subfield $e; the presence in the record of field 005) may reflect an AACR2 or AACR2 compatible heading structure. Thus the same entity or concept may be represented by both pre-AACR2 and AACR2 forms. Since the PRP subset is slated eventually to be incorporated into the current catalog sometime after implementation of the ILS, it does not seem prudent to expend the human resources required to reconcile cases of different forms of heading for the same entity or concept across the two subsets (replacement; "other"). Therefore, it is not necessary to notify the PREMARC Team of these discrepancies.

S1.8.4. Effect Of The Decision To Distribute The PRP Subset

S1.8.4.1. General

Originally CDS planned to make the PRP subset available for purchase an initial, one-time special distribution of the 1.475 million records in the subset. Various complications now preclude such a one-time distribution. This simplifies corrections to individual records in the PREMARC file, particularly those in the PRP subset. Corrections are made to all records in the PREMARC file in the usual way. The only special precaution is that if a record is in a verified state at the time it is corrected, it should be reverified. (If the record is being upgraded to reflect the original source card, it should be annotated accordingly and verified.)

S1.8.4.2. Moving PREMARC records to a file in the current catalog

Records are moved from the PREMARC file to a file in the current catalog under the NEW INPUT procedure (DCM B5.12). When this is done, the heading structure is reassessed and changes are made, as needed, to ensure that forms of headings reflect the current authority structure. Therefore, neither "PREMARC REPLACEMENT" nor PREMARC records are any longer identified as such.

When copying PREMARC or PREMARC REPLACEMENT records to a file in the current catalog, follow these steps **EXACTLY**:

1) COPY to the target file under a pseudo LCCN, e.g., co76-1; leave the record UNVERIFIED;
2) delete the PREMARC record using "ndel", e.g., ndel 76-1;
3) change the LCCN of the COPIED record, e.g., co76-1 to 76-1.

When moving a PREMARC record to a file in the current catalog, it is adjusted to reflect at least the data from the original source card and the current authority structure. Therefore, it is no longer necessary to retain any PREMARC or PREMARC REPLACEMENT identification.

The term "NEW INPUT" in the 955 field is sufficient to identify the reason for input. In copying such a record to a file in the current catalog, note the following:

1) the 035 field containing the OCLC control number, if present, is copied; delete it at the prompt stage to preclude the record from being identified as a record from the resource stream;

2) adjust the 040 field of any PREMARC REPLACEMENT record to reflect DLC in subfields $a, $c, and $d;

3) the 042 field containing "premarc" is not copied; do not add it back;

4) the 985 field containing "PREMARC REPLACEMENT" in subfield $e is not copied; do not add it back.

If, however, some other 985 field is present in the PREMARC record and it is judged of value to retain it, then it will need to be rekeyed at the prompt stage of the COPY feature.

NOTE ALSO that for multipart items, the imprint (260 field), LC holdings (300 field), and contents note (500 field), if present, need to be verified in the Official Card Catalog and the shelflist for the most up-to-date iteration of LC's cataloging and holdings (in PREMARC REPLACEMENT records, these fields were replaced and may not reflect LC's copy and holdings).

S1.9. Summary Description of the PREMARC Replacement Process

In 1997-1998 LC sent to OCLC copies of PREMARC records for books. The records selected were the most likely to have replacement counterparts in OCLC and were the least likely to cause complications upon replacement at LC (e.g., records representing microform replacement at LC were not included). OCLC ran the candidate replacement records through a series of programs that processed them against the OCLC database. To ensure that a replacement record was for the same title represented by the PREMARC record, a successful match could be declared if OCLC matching algorithms were successful **and** the LCCN in the 010 field of the matched record was the same as the LCCN in the 001 field of the replacement candidate. Of the 3,930,450 replacement candidates sent to OCLC, 1,475,092 were determined to match counterpart records in OCLC for a match rate of 37.5%.

Copies of the replacement records were returned to LC, and in April 1998 the replacement process was executed. During the replacement process, various checks were made to ensure that replacement was appropriate. For example, the LCCN of the replacement record was checked against the files that constitute the current catalog to ensure uniqueness of LCCNs. If the LCCN of a replacement record did not have a counterpart PREMARC record, replacement did not occur. Other "no load" conditions were:

PREMARC record contained 500 note with "Serial."
PREMARC record was in a "Verified" state
PREMARC record contained evidence in field 985 of being in one of various projects
    such as American Memory
PREMARC record contained multiple 050 fields
PREMARC record lacked an 050 field

The replacement process really consisted of a "merge/replacement" process; the following fields in existing PREMARC records were retained in the replacement record:

050	LC call number
051	LC copy, issue, offprint statement
5XX	Fields with subfield $5
510	Citation/references note
590	Local note
591	Local "with" note
710	Fields with subfield $5
952	Cataloger's note
955	Tracking field
985	Record history field

NOTE: Fields 260, 300, and 505 (if present) were replaced; therefore, for multipart items, they MAY NOT accurately reflect the LC copy.

The following fields were generated in the replacement record:

985	Contains "PREMARC REPLACEMENT" in subfield $e (a separate 985 added if one was already present in the record)
042	Contains "premarc" is subfield $a
035	Contains the OCLC control number of the replacement record
005	Set to date of load (note that other PREMARC records **not** in the replacement subset do not contain 005)

The following settings were made in the replacement record:

The record was verified
Record status was set to "new"
Date entered on file: the date in the replacement record, i.e., the date input into OCLC
Date of last verification: date of load
Date of last transaction: date of load
Date of previous transaction: the date input into OCLC