

UNIVERSITY OF ILLINOIS  
UNIVERSITY OFFICE OF  
ADMINISTRATIVE INFORMATION SYSTEMS AND SERVICES

CHICAGO

URBANA · CHAMPAIGN

54 Administration Building, 506 South Wright Street · Urbana, Illinois 61801  
(217) 333-3840

6/9/89

TO: ILCSO Operations Committee  
FROM: Cathy Salika *Cathy*  
RE: Final Specification for Moving LCS Data to FBR

Last January, I distributed the first version of the specification for moving bibliographic data from LCS into FBR. Several members of the Operations Committee made useful suggestions for improvements to that specifications. These suggestions have been reviewed by the members of the ad hoc subcommittee which drafted the specification. Members of the subcommittee are Cecile Jagodzinski (IS), Barbara Klintworth (SS), Arnold Wajenberg (UC), John Whisler (EA), Don Wood (SC) and myself.

The final version of this specification is attached. I have marked with a bar in the margin the changes that have been made since the last version that you saw. Thank you to those of you who contributed to these changes.

LCS Restructuring Plan Phase I  
Adding LCS Records to FBR

Phase I of the LCS Restructuring Plan involves copying all records from LCS into FBR, except those records that are already represented in FBR. This is the first step towards eliminating redundant data from the two systems. It will simplify the search strategies which will make both the user interface and the command mode more understandable. It will also increase the accessibility of some bibliographic records by allowing keyword searches by title and corporate author as well as boolean combinations.

This process should be envisaged as being done for one school at a time since the LCS database is designed this way. The first step will be to translate all of the LCS records for that school (or, possibly, only the records that aren't linked - more on this later) into the MARC format. The second step will be to identify which of these records already exist in the FBR database. The third step will be to add scopes, holdings and links and, for those records which do not match a bibliographic record in FBR, the bibliographic record. Between steps 1 and 2, there is an optional step of ACCR2 upgrade for the main entries in the short MARC records.

These three steps will be described in great detail below, but there are 4 questions that we should look at first: Which LCS schools will want to make use of this process? Do we want to do this process on all records in a given database. How will this work affect our participation in national databases? And, of course, what about analytics?

This process will be run for all LCS schools. Schools that have not attempted full retrospective conversion will find this to be an economical way of loading their data into FBR. Schools which have attempted full retrospective conversion will get other benefits. This is the only way to assure that all LCS records are represented in FBR, which will give us cleaner results in Phase 3 of the restructuring plan. It will also clear up any remaining missing links, which will be a great help for some libraries. There may be one significant disadvantage to using this process. The matching algorithm that will be used to compare the LCS records to FBR will be very conservative to prevent false matches. As a result, it will miss some valid matches. The records that do not match will be added to the FBR database and those that should have matched will create redundant records. Someone will have to clean out these redundant records, which will

probably be a manual task, at least in part. Nevertheless, the benefits of success in Phase 3 of the restructuring are so significant that they should offset the problems caused by unmatched records.

There are 2 reasons why we might want to avoid putting some LCS records through this process. The first reason is that the record is odd. We may not want to put on order records, non-standard cataloging or other oddities into FBR. The analysis of the problems associated with odd records is underway and will appear in a separate document. The other reason for not putting an LCS record through this process is that we can take a short cut to determining whether it is in FBR yet. Such a short cut would be a great savings because the process is very detailed. Fortunately, such a short cut exists. We will check the FBR link file for the campus/title# combination. If an LCS record is linked, we are almost certain that it appears in FBR and there is no need to send it through the process. For analytics, there is an additional short cut. When making links, FBR recognizes analytics by the presence of "V.", "PT." or "NO." in the call number. If a link exists to an LCS record either with or without any of these analytic suffixes, we will not send the record through the process.

The ideal that we would like to strive for is to have the complete collections of all of the LCS schools represented in FBR by full MARC records with OCLC control numbers. It is also highly desirable to have the complete collections represented in OCLC or one of the other national databases. This process can contribute to both of these goals. At the end of this process each record from the LCS database will be treated in one of the following ways:

- 1) If the record from LCS matches a record from FBR and the school's holdings are present in FBR, we will verify that the links and scopes are accurate. Because there is a record with holdings, we can assume that these records are already in OCLC and there is no need to do anything special with them.

- 2) If the record from LCS matches the record from FBR but the school's holdings are not present, we will add appropriate holdings, scopes and links to FBR. Presumably, this record came from OCLC but the school does not have its holdings reflected in OCLC. In this case, we could send a tape to OCLC that would be used to add the holdings. There are costs involved in this because OCLC charges for this service, but it would be good to contribute holdings to a national database if possible. Aside from the benefit that this would have for resource sharing, it would have a benefit for individual libraries in the future. They could ask OCLC for full MARC records that correspond to all of their holdings, thus getting a larger proportion of their

collections in MARC format.

3) Many records from LCS will not find matches in FBR. These records will be added to FBR. They will be readily identifiable because they will be short and have no lower case letters. More importantly, we will make up a record ID for them which will be in the form "cc#####", where "cc" is the LCS campus code and "#####" is the LCS title number. Because of this distinctive number it will be possible to extract these records from FBR in the future. The records could be sent to OCLC or another vendor for conversion to full MARC and the short records from LCS could be replaced by better cataloging. This procedure could be done several times with aid from several different vendors until a small number of unconverted records remain which could be handled by local cataloging staff. Please note that software does not presently exist to extract records on the basis of RID or to replace the short records with the long ones although there is a lot of code available that would be usable for these purposes. Libraries that think they have done a complete retrospective conversion may wish to approach this problem manually. During the conversion process, AISS staff can produce a list of all of the short records added to FBR from their LCS database. The list may include all of the data in the records or just the RIDs. Any library that wants such a list should request it by filing a work request before their database is converted.

Ah, analytics. The problem, of course, is that LCS represents analytics as a serial while FBR represents them as individual volumes in a series. As a result, the LCS records will never match FBR. The approach would be to add the MSET records to FBR on the grounds that it is a good thing to increase access by serial title. As a result, some serial titles would be searched with the FBR title (F T) search, some would be searched with the series title search (F SET) and some would be searched both ways. (As long as I don't have to explain this to patrons...)

\*\*\*\*\*  
\* Step 1 \*  
\*\*\*\*\*

The first step in this process is to convert all the records from an LCS database into short MARC records. As was noted above, these records will be entirely in upper case. They will all have the same number of variable fields, which will make it more convenient to process them in later steps. Each field is described below giving the location from which the data will be taken in LCS and the strategy for getting tagging and indicators right. Since authors pose a special set of problems, they will be described last.

The first decision that has to be made is which MARC format to use. This does not have an effect on the matching or tagging that will be done, but it would be good to make the 008 fields as accurate as possible. The serials format will be used if the SER flag in LCS is set. The maps, films and music formats will be used if the call number prefix suggests that one of these is appropriate. (Remember that FBR collapses the scores and sound recordings formats into a single music format.) Otherwise the books format will be used. A table showing how prefixes will be used to determine formats may be found in Appendix B.

#### Leader

The table gives the value for the various fields in the leader:

Byte	Use	How we will set it
0-4	Rec.Length	We will use the actual record length
5	Rec Stat	"n" (New record)
6	Type	"a" (Printed material)
7	Bib lvl	"s" if the SER flag in LCS is set but the MSET flag is not set "e" if the prefix indicates a map "g" if the prefix indicates a film "j" if the prefix indicates a recording "m" otherwise
8-9	blanks	blanks
10	ind cnt	"2"
11	subf cnt	"2"
12-16	Base addr	"00145"
17	Enc lvl	"M" (Less than full cataloging)
18	Desc	blank (Record is in non-ISBD form)
19	link rec	blank
20	len	"4"
21	offset	"5"
22	tran type	X'01' (Produce)
23	undef.	"0"

#### 001 Control Number

As was mentioned above, control numbers will be in the form "cc#####", where "cc" is the LCS campus code and "#####" is the title number. This field will be padded with 3 blanks on the end so that it is always 12 characters long. This will be handy in later steps when we find a match because we can replace the data in this field with the OCLC number with a minimum of processing. There are no subfields or indicators to think about.

#### 008 Fixed length data elements

Because the bibliographic level byte of the leader may vary,

several of the MARC formats may be used. The layout of the 008 field is different in these formats. The tables in the Appendix A of this document show our use of this field. An attempt has been made to define accurate values for each byte, but it is not always possible to do so. This should not be important, however, since none of these fields affect the indexing of these records in FBR at present.

#### 010 LC Card Number

The data in this field will come from the LCCN field in LCS. The data will be put in a single \$a. Both indicators are blank. It should be noted that some LCS schools put other data in the LCCN field via SUPERWYLBUR. Whatever they put there will go into the 010.

#### 040 Cataloging Source

It would be wise to make use of one of the fields that describes the source of the record to flag that these records came from a less-than-pristine source. We will use the 040 for this purpose. The indicators will be blank. There will be one \$a with "LCS" in it. There could also be a \$c with the OCLC holding symbol at the 3-character level.

#### 049 Local Holdings

This field is essential to the creation of scopes, links, and holdings. We will leave the indicators blank. The only subfield will be \$a, which will have a 4-letter OCLC holding symbol. Any symbol that the school uses will do.

#### 099 Local Free-Text Call Number

We will put all call numbers in this field regardless of classification scheme. Since call numbers do not appear in FBR, the tagging is not important. Both indicators will be blank. The call number will go in \$a with no special spaces or punctuation.

#### 699 Added Class Number

There is one special issue surrounding call numbers. If a library were to extract its short records from FBR and send them to a vendor for conversion to full MARC, it would be desirable to be able to send the call number as part of the record. For this reason we will put the call number in the 699 field as well. This field will not display in any but the complete (\$,c) display. Both indicators will be blank. The call number will go into \$a with no special spaces or punctuation added.

#### 245 Title Statement

The title appears in the same field as the edition and the place of publication in LCS. We will put a \$a at the beginning of the LCS title field. The edition and place of publication are each preceded by a dollar sign. The text following the last dollar sign will be treated as the place

of publication. If there is another dollar sign, the text between it and the last dollar sign will be treated as the edition. If, there is a colon, semicolon or equal sign anywhere in the title, a \$b will be inserted. This coding scheme is not always bibliographically accurate, but it gives keyword access to the entire title (except for titles that have a dollar sign in them, but they are a minority). The first indicator will be set to 0 since there is no title added entry. The second indicator is a count of non-filing characters. We will use the following list of non-filing words to set this indicator: a, an, das, der, l', la, le, les, the.

#### 250 Edition

If there are 2 or more dollar signs in the LCS title field, the text between the last 2 of them will be used to create an edition statement. This text will be put in \$a. Both indicators will be blank.

#### 260 Imprint

As was mentioned above, the place of publication is found after the last dollar sign in the title field in LCS. If there is a place of publication, we will put it in the 260 \$a. UIUC used to abbreviate place names and it would be good to spell them out now. The following list of conversions represents UIUC practice. We will make the conversions for UIUC and probably for other schools as well, since they are likely to be useful.

BALT -> BALTIMORE  
BER -> BERLIN  
BOST -> BOSTON  
B AIRES -> BUENOS AIRES  
CAMB, ENG -> CAMBRIDGE, ENGLAND  
CAMB, MASS -> CAMBRIDGE, MASS.  
CHGO -> CHICAGO  
ENG CLIFFS -> ENGLEWOOD CLIFFS, NJ  
IND -> INDIANAPOLIS  
LEN -> LENINGRAD  
LOND -> LONDON  
MADR -> MADRID  
MEX -> MEXICO CITY  
MONTR -> MONTREAL  
MOS -> MOSCOW  
OX -> OXFORD, ENG.  
PHILA -> PHILADELPHIA  
SAN FRAN -> SAN FRANCISCO  
TOR -> TORONTO  
WASH DC -> WASHINGTON, DC

The publisher's name does not appear in LCS so there will never be a \$b. This is not a problem since FBR does not require that there be a \$b. If so, we will change that

table. The date of publication has its own field in LCS. It will be put into \$c. There are a few errors in the date fields in LCS and this would be as good a time as any to clean them up a bit before we put them into FBR. We will change the letter "L" to the digit "1" and the letter "O" to the digit "0". Next we will look at the whole field. If it has 4 digits or a mix of digits, question marks, and hyphens, it should be accepted. Otherwise, it may be one of the following: "ND ", "N.D.", or some other garbage. We will accept the first 2 possibilities, but if it really seems that we have garbage, we will omit the date completely. Since no publisher is present, the first indicator will be set to 1. The second indicator will be set to 0 for serials and blank for other formats. The distinction among formats was made when we set up the leader.

### 1xx Main Entry

There is an author field in LCS that usually contains a main entry. Some LCS records have no author, which is quite understandable. Some LCS records that came from OCLC have used an added entry in the absence of a main entry. There is no way of detecting that this was done, however, so we'll just have to accept whatever we find there.

The treatment of authors is the trickiest part of converting the LCS data into the MARC format. In the first place, we have to start by determining the 1xx tag. This is important because it affects the way that records are index in FBR. It also affects our ability to recognize matches in the FBR authority file. We want to maximize the matches in the authority file to minimize the number of new and probably spurious headings we create there. There's an odd fact about FBR that makes the matching easier. When the FBR software is trying to determine whether a heading matches a heading in the database, it considers the subfield delimiters, but it usually does not consider the subfield codes. As a result, if we incorrectly code a subfield, the FBR authority file will usually clean it up for us, but we still have to be as accurate as possible with the subfield delimiters.

For the same reasons, we hope to be more accurate about the subfield coding and indicators. There will be different procedures for setting these values depending on the tag that we assign to the heading, so we will start by determining the tags. For purposes of this discussion a "word" is a series of letters and hyphens. There may be blanks or punctuation marks before and after a "word". A string is also a series of letters and hyphens, but a string may be surrounded by letters, hyphens or other characters. For example, in the preceding sentence, "hyphen" is a string, but it is not a word because it is followed by a letter "s".

No matter how good the algorithm is, there are certain to be

a number of authors which are incorrectly tagged and therefore fail to match valid headings in the FBR authority file. When the loading is actually done in Step 3 of this grand scheme, a report will be printed listing all of the new headings so that manual review may be done to whatever level of detail the library desires. Also, to make the headings easy to recognize in the authority file, we will add "\$wLCS" to all of them. The \$w will not interfere with the matching of headings and it will not display except in the complete (\$,c) displays.

We will apply the following tests to each author in this order. Once we find a test that tells us what the tag should be, we will skip the other tests.

#### Test 1

The names of some corporate bodies begin with the name of a city followed by a state, province or country. If one of the words below is the second word of an author field, we will tag it 110. Assume that all of the U.S. states and Canadian provinces are on the list, abbreviated as in AACR2 B.14. Also, we will include the abbreviations for all of the geographic names listed below as given in AACR, Appendix 3G.

AUSTRIA, AUSTRALIA  
BELGIUM, BRAZIL  
CANADA, CHILE, CHINA, CUBA  
ENGLAND  
FRANCE  
GREECE, GERMANY  
HONDURAS, HONG KONG  
ITALY, INDIA, IRELAND, IRAN, IRAQ, ISRAEL  
JAPAN  
KOREA  
MEXICO  
NORWAY, NETHERLANDS  
PAKISTAN, PANAMA, PHILIPPINES, POLAND, PORTUGAL, PUERTO RICO  
REPUBLIC  
SAUDI, SCOTLAND, SINGAPORE, SOVIET, SPAIN, SWEDEN,  
SWITZERLAND

#### Test 2

If the first word in the author is followed by a comma, it is probably a personal name. We will tag it 100.

#### Test 3

Uniform title main entries are often used for sacred, ancient or medieval works. A list of such works is given below. Treat the names of the books as a string. If one of these strings occurs at the beginning of an author, we will tag it 130.

ARANYAKAS	CHANSON	KORAN	TALMUD
AVESTA	HAGGADAH	MAHABHARATA	TOSEFTA
BEOWULF	JAINA AGAMA	MISHNAH	UPANISHADS

Test 4

FBR indexes corporate authors and conference authors in exactly the same way, so we do not have to distinguish between them. These authors usually have words in them that reflect a level of organization. The following list of such words comes from the LC CSB #20 and AACR2 Ch. 24 as well as other sources. We will treat these terms as strings. If any of these strings occur anywhere in the author, we will tag the author 110.

ASSOCIATION, ADMINISTRA, ACADEM, AKADEM, ASSEMBL, AMERICA,  
ADVISORY, AGENC, AUTHORITY, ASESORIA  
BUREAU, BOARD, BURO, BUERO, BIBLIOT  
COMMISSION, COMISION, COMISAIA, CHURCH, CONGRES, CONFERENC,  
CABINET, COMMITTEE, COMITE, COORDINACION, COLLEGE, CENTER,  
COLLOQUI, COUNCIL, COUNSEL, COMPANY  
DIPUTACION, DELEGATION, DELEGACION, DIRECTION, DIRECCION,  
DIRECTORIA, DEPT, DEPART  
EUROPEAN  
FISCALIA, FACULT, FACULD  
GROUP, GROUPE, GRUPE, GRUPO, GABINETE, GERENCIA, GESELLSCHAFT  
HOUSE  
INSPECTION, INSTITUT, INTERNAT, ISTITUT  
JEFATURA, JUNTA, JOINT  
LIBRARY  
MEETING, MINISTR  
NEGOCIADO, NACION, NATION, NAUK  
OFFICE, OFICINA, ORCHESTRA, ORGANIZA, ORGANISA  
PARTY, PARTI, PANEL, PARLIAM, PARALAMENT, PARLEMENT  
REGIMENT, REGT.  
SOCIE, SECRETAR, SENATE, SERVIC, SUPERINTEND, SCHOOL,  
SEMINAR, SYMPOSI  
TASK FORCE  
UNITED, UNIVERS, U.S., U. S., U.N., U. N., U.K., U. K.  
WORKING  
ZENTRALBLATT

Test 5

Some corporate bodies may be recognized by the presence of a geographic name. The following names will be treated as strings. Assume that all of the U.S. states and Canadian provinces are on the list. These will be tagged 110.

AUSTRIA, AUSTRALIA  
BELGIUM, BRAZIL, BRITAIN  
CANADA, CENTRAL, CHILE, CHINA, CUBA  
DUTCH  
ENGLAND, EAST, ENGLISH  
FRANCE  
GREECE, GERMAN

HONDURAS, HONG KONG  
ITALY, INDIA, IRELAND, IRAN, IRAQ, ISRAEL, ISLAND  
JAPAN  
KOREA  
MEXICO  
NORWAY, NORTH, NETHERLANDS  
PAKISTAN, PANAMA, PHILIPPINES, POLAND, PORTUGAL, PUERTO RICO  
RUSSIA, REPUBLIC  
SAUDI ARABIA, SCOTLAND, SINGAPORE, SOUTH, SOVIET, SWEDEN,  
SPAIN, SWITZERLAND  
TERRITORY, TER.  
WEST

Test 6

Names of personal authors often begin with a preposition once the names have been inverted. These words will be treated as strings. A name that begins with one of them will be tagged 100.

AM, AUS'M, AL, ABU, ABD  
BEN  
DE, DEL, DU, D', DES, DA, DI, DELLA  
EL  
LE, LA, LI, LO, LAS, L'  
O  
TEN  
VAN, VON, VOM  
ZUM, ZUR

Test 7

Some compound surnames in Spanish contain the word "Y". If there is no comma after the first word and the second word is "Y", we will tag the name 100.

Test 8

Some compound surnames are made up of 2 words separated by a blank and followed by a comma. Authors that begin with this pattern will be tagged 100.

Test 9

Some personal authors can be identified as such by the presence of a title of nobility or religion. If one of the words listed here occurs anywhere in an author preceeded by a comma and a space, we will tag it 100:

ABBOT, ABBE, ABBESS, ASHIN, ARCHBISHOP, ABP.  
BEN, BISHOP, BP., BARON, BARONESS, BHIKSHN, BROTHER, BRO.,  
BART.  
CONTE, COUNT, COUNTESS, CHIEF, CARDINAL, CAPTAIN  
DAME, DEAN, DUKE, DUCHESS, DR., DOCTOR, DUCHESS, DIACONUS,  
DOKTOR  
EARL, EMPEROR, EMPRESS, ESQ.  
FREIHERR

GUSTI, GRAF  
HWASANG  
JR., JUNIOR  
KING  
LORD, LADY, LEDI SAYADAW, LANDGRAF  
METROPOLITAN, MOTHER, MR., MRS., MISS, MAHARSHI, MUNI,  
MARGRAVE, MARGRAF, MADAME, MME., MISTRESS  
NI  
PERE, POPE, PRINCE, PRINCESS, PATRIARCH, PHRAYA  
QUEEN  
RABBI, RADEN, REV., RITTER, REVEREND  
SIR, SAINT, SAINTE, ST., STE., SISTER, SWAMI, SR., SENIOR  
THERA  
VISCOUNT, VISCOUNTESS

#### Test 10

Some very prolific personal authors use only a single name. If one of the strings below occurs at the beginning of an author, we will tag it 100:

AESCHYLUS, AESOP, AESOPUS, ARISTOPHANES, ARISTOTLE,  
ARISTOTELES  
BOETHIUS  
CATULLUS  
DEMOSTHENES  
EPICTETUS, EUCLID, EUCLIDES, EURIPIDES  
GALEN, GALENUS  
HERODOTUS, HESIOD, HESIODUS, HIPPOCRATES, HOMER, HOMERUS,  
HORACE, H.D.  
JUVENAL  
LIVY, LUCAN  
MARTIAL, MENANDER, MOLIERE  
PINDAR, PINDARUS, PLATO, PLINY, PLOTINUS, PLUTARCH,  
PLUTARCHUS  
SALLUST, SAPPHO, SOPHOCLES, SUETONIUS, SOCRATES, STENDAHL  
TERRENCE, THUCYDIDES  
VIRGIL, VOLTAIRE  
XENOPHON

#### Test 11

Roman names of the classical period often include a number indicating the position of the individual in the family. If one of these words occurs anywhere in the field, the 100 tag will be used.

PRIMUS, SECUNDUS, TERTIUS, QUARTUS, QUINTUS, SIXTUS,  
SEPTIMUS, OCTAVIUS, NONUS, DECIMUS

#### Test 12

If all else fails, we tag the author as 110 because FBR does more extensive indexing done on 110 fields than on any other.

Coding within personal authors

1) The second indicator will always be set to "0". The first subfield will always be \$a.

2) Look at the string up to the first comma. This is probably the surname and we can use it to make a first attempt at setting the first indicator. Use these rules to set the indicator; use the first rule that applies: If the string contains the word "FAMILY", then code the first indicator as "3". If the string contains a blank or a hyphen, then code the first indicator as "2". Otherwise code the first indicator as "1". (We will change some of the "1"s to "0"s in a moment.)

3) If there is a second comma in the field, go to 4). If there is no second comma, look at the string between the first comma and the end of the field. If there are any digits in this string, they are probably dates and there is no surname. Insert a \$d at the beginning of this string; if the first indicator is set to "1", change it to "0"; the tagging is complete. If there are no digits in the string, we are probably looking at the forenames. Go on to 4).

4) Starting with the second comma, a new subfield should begin after every comma. The choice of subfield code is not important to FBR for purposes of indexing, but we would like to be as accurate as possible to increase the number of matches with the authority file. There are 4 rules for determining which subfield code to use:

- > If there are digits, use \$d.
- > If the first character is "(", use \$q.
- > If the first character is "X", "V", or "I", use \$b.
- > Otherwise use \$c.

If this sounds a bit too quick and dirty, consider this. If this record got into LCS via OCLC, subfields \$x and \$w are eliminated. If the record came from somewhere other than OCLC, it probably has no information in it that would not appear on a shelf list card. Most of the exotic subfields are unlikely to be present.

#### Coding within corporate/conference names

The choice of first indicator will not affect indexing or matching in the authority file, so we will arbitrarily choose to code it as "2" with one exception: If the corporate name begins with a place name (from tests 1 or 5) followed by a period, we will code the first indicator as "1". The second indicator will always be set to "0". The first subfield will always be \$a.

Periods are used in corporate conference names in 3 ways:

- > in acronyms ex. N. A. S. A.

> in abbreviations ex. Dept.  
> in hierarchical units ex. United States. Congress. Senate  
Acronyms are easy to identify because single letters are used. Abbreviations are relatively rare and we should clean them up anyway. The following conversions will be made.  
AMER. -> AMERICAN  
BRIT. -> BRITAIN  
DEPT. -> DEPARTMENT  
GT. -> GREAT  
U.S. -> UNITED STATES  
U.K. -> UNITED KINGDOM  
U.N. -> UNITED NATIONS

We will assume that all other periods are delimiters between hierarchical units, so we will put a subfield delimiter after them. We will also replace any double dashes with a subfield code. If the subfield ends with 4 digits, possibly followed by a hyphen, we will make it \$d. If there are no digits, we will make it \$b unless the previous subfield was a \$d, in which case we will use \$c. Got that? We are assuming that there is a \$c after every \$d.

If there was a \$d tagged in the heading, we need to take one more look at it. For pre-AACR2 conference headings, we have probably crammed several subfields into what we're calling \$d at this point. The pre-AACR2 headings will look something like this:

\$aInternational Conference on Art, \$n2nd, \$cUniversity of Wisconsin at Madison, \$d1982.

In this example, we would have the entire heading coded as \$d and we need to go back and sort this out. Here's how:

- 1) Start at the end of the "subfield" and back up over any digits or hyphens. Insert a \$d.
- 2) Back up to the previous comma. There should be a space after it. If the character after the space is a letter, insert a \$c before the letter. Otherwise, go back to where you were after 1).
- 3) Back up to the previous comma. There should be a space after it. If the character after the space is a digit, insert a \$n before the digit. If it's not a digit, insert a \$b.
- 4) If the first subfield is no longer \$a as a result of these changes, change it back to \$a.

#### Coding for uniform titles

Recall that we are dealing with 130 fields in section, not 240 fields. The first indicator is the count of non-filing characters. We will use the list of non-filing words that

was given with the specifications for the 245 field above. The second indicator will always be "0". The first subfield will always be \$a.

Periods are used between subfields in uniform titles just as they are for corporate and conference names. We will use the same rules for both kinds of heading when placing subfield delimiters, but there will be different rules for choosing the subfield codes for uniform titles. Use the first rule below that applies:

- > If the subfield completely matches one of these strings, code \$l. (That's the letter el.) ENGLISH., GERMAN., FRENCH., GREEK., LATIN., SPANISH., ITALIAN., HINDI., PORTUGUESE., RUSSIAN., SWEDISH., DANISH., DUTCH., HEBREW., NORWEGIAN., POLYGLOT.
- > If the subfield contains 2 of the languages listed above, separated by the word AND, code \$l. For example LATIN AND ENGLISH.
- > If the subfield begins with a "[", code \$h.
- > If the subfield is 4 digits, code \$f.
- > If the subfield contains a digit, code \$n.
- > If the subfield completely matches one of these strings, code \$k: SELECTIONS., PARAPHRASES.
- > If none of the above apply, code \$p.

That's it! We now have a reasonable MARC record!

\*\*\*\*\*  
\* Optional Step - AACR2 Upgrade \*  
\*\*\*\*\*

The FBR authority file is an important tool for both searching and database maintenance. It is desirable to keep it as clean as possible. The only data from the short MARC records created in Step 1 that are stored in the FBR authority file are the main entries. To minimize the number of incorrect headings that are loaded into FBR, we can send the short MARC records to an outside vendor for AACR2 upgrade before loading the records into FBR. This step is optional because it is contingent on the availability of funding.

If we decide to go ahead with this step, we can keep the cost down by sending to the vendor only those records whose main entries are not already in the FBR database. This can be done by using the standard heading matching algorithms in FBR.

\*\*\*\*\*

\* Step 2 \*  
\*\*\*\*\*

The second step of the process for loading LCS records into FBR is to attempt to find matches for the short MARC records in the FBR database. If a match is found, we will verify that the necessary scopes, holdings and links are present. If no match is found, we will load all parts of the short record. Essentially, the matching will be done on the basis of the LC card number with a reasonableness check to prevent false matches. There are several details that need to be discussed.

First of all, the LC card number has never been loaded into the FBR database. At one time, we thought of this number as being so unreliable that we did not even want to see it in FBR. We now realize that any control number, however homely, is useful for matching. We should modify FBR so that this number will go into the database in the future. This will be nice for the full MARC records, but it will be very important for the short records that originated in LCS. If libraries request copies of their short records so that they can send them to a vendor for conversion into long records, it will be important that the LC card number be available to the vendor for purposes of matching. Thus, we need to have the LC card number in the database for these records.

Since the LC card number is not in the database, we will have to scan the ILLINET tapes to get it. We will create a dataset that can be thought of as a table with three columns: the LC card number, the other fields that are part of the reasonableness check, and the OCLC number. This table will be sorted by the first two columns and duplicate entries will be removed.

There is one additional thing that will actually be done as part of Step 1 but is easier to describe here. Before each of the MARC records, we will append the LC card number and whatever part of the record that we choose to use for the reasonableness check.

It should now be easy to see how Step 2 will proceed. We will sort the file with the short MARC records by the data that is appended to the front of each record. We will then compare the file of short records to the table that has the LC card numbers and the OCLC numbers. If we find a match based on the LC card number plus the reasonableness check data, we will replace the dummied up number in the 001 field with the OCLC number from the table. Recall that we padded the 001 field with blanks so that it will always be 12 bytes long. Recall too that there is a constant number of MARC fields in all of the short records. Because of this, the OCLC number can be added to the short record without messing with the record directory. The variable part of the 001

field will occur at a fixed offset within the record, which is also convenient. Finally, whether a match is found or not, the data that was appended to the front of the short MARC records will be stripped off and the record will be written out to a dataset. When we are done with this step, the contents of that dataset will be the short MARC records, some of which will have OCLC numbers in the 001 field.

We need to think about what data to use for the reasonableness check. The purpose of this check is to assure that false matches are not made if we come across an error in an LC card number. We don't want to use the author field as part of the reasonableness check because its vocabulary is controlled. (For example, we could easily find Clemens records in LCS that will correspond to Twain records in FBR.) Neither do we want to use long sections of any field that we choose. The longer the sections, the greater the chance that minor variations in keying will unreasonably interfere with the reasonableness check, while short sections will do an adequate job. So much for the theory. We will use the following fields in the reasonableness check:

- 1) The first 5 letters of the first significant word of the title will be used. The LCS stopword list will be used to define insignificant words.
- 2) The date field from 260 \$c will be used. Only the first 4 characters will be used.
- 3) The place of publication from 260 \$a will be used. The first 5 characters of the field will be included in the reasonableness check.
- 4) The first 3 characters of the edition statement (250 \$a) will be used.

All four of these fields will be padded with blanks if necessary to make them fixed length.

\*\*\*\*\*  
\* Step 3 \*  
\*\*\*\*\*

Now that the matching is done, we can simply load these short records as if there were nothing unusual about them. There is sufficient information in the 049 and 099 to create holdings, links, and scopes. If a short records "happens" to have an OCLC number in it, the holdings, scopes, and links will be added to the existing record in the database. If the short record does not have an OCLC number, its control number (cc#####) will be unique and no record loaded will ever match it. This means that if 2 LCS schools do not have full cataloging for the same bibliographic item, there will be 2 short records in FBR. The down side of this is that, until retrospective conversion projects get under way, there will

be duplicate records in the database. The up side is that we are able to use the RID field to extract a school's records and do replacements in the future.

There is a report that we do not usually use that comes out as part of the loading process. It lists all of the new headings that are added to the database as part of the load. All of the tags and indicators are included. If a heading should appear on more than one record, it will be listed only once. We will print this report when we load the short records. If necessary, we will sort the headings before the report is printed to make it easier to use. Libraries may scan the report to identify headings that failed to match because they were incorrectly tagged. These errors may be fixed using the regular FBR maintenance procedures. Schools for which this process is run will either need to have staff trained in FBR database maintenance or make arrangements with staff at a maintenance site to have this work done. Since erroneous tagging in the authority file affects access for everyone, it would be in the interest of the entire ILLINET user community to do whatever cleanup is necessary.

## Appendix A

## Layouts for the 008 Fields

For the books format

Byte	Use	How we will set it
0-5	Entrd	The date on which we create the record
6	Dat tp	"s" (single known date)
7-10	Date1	Same as the date in 260 \$c
11-14	Date2	blanks
15-17	Ctry	"xx " (country not known)
18-21	Illus	blanks (no illustrations)
22	Int lvl	blank (unknown)
23	Repr	blank (not a reproduction)
24-27	Cont	blanks (nature of contents not specified)
28	Govt pub	blank (not a government publication)
29	Conf pub	"0" (not a conference publication)
30	Festschr	"0" (not a Festschrift)
31	Indx	"0" (no index present)
32	MEBE	"0" (main entry not in body)
33	Fic	"0" (nonfiction)
34	Biog	blank (not a biography)
35-37	Lang	"eng" if NENG flag in LCS is not set "und" if NENG flag in LCS is set
38	Mod rec	blank (not modified)
39	Source	"d" (non-LC source)

For the films format

Byte	Use	How we will set it
0-5	Entrd	The date on which we create the record
6	Dat tp	"s" (single known date)
7-10	Date1	Same as the date in 260 \$c
11-14	Date2	blanks
15-17	Ctry	"xx " (country not known)
18-20	Leng	"---" (length not known)
21	Blank	blank
22	Int lvl	blank (unknown)
23-27	Acc Mat	blanks (no accompanying material)
28	Govt pub	blank (not a government publication)
29-31	Blanks	blanks
32	MEBE	"0" (main entry not in body)
33	Typ Mat	"u" (type of material unknown)
34	Tech	"u" (technique unknown)
35-37	Lang	"eng" if NENG flag in LCS is not set "und" if NENG flag in LCS is set
38	Mod rec	blank (not modified)
39	Source	"d" (non-LC source)

For the maps format

Byte	Use	How we will set it
0-5	Entrd	The date on which we create the record
6	Dat tp	"s" (single known date)
7-10	Date1	Same as the date in 260 \$c
11-14	Date2	blanks
15-17	Ctry	"xx " (country not known)
18-21	Relief	blanks (no relief shown on map)
22-24	Base	blanks (base not specified)
25	RecG	"a" (single map)
26-27	Blanks	blanks
28	Govt pub	blank (not a government publication)
29-30	Blanks	blanks
31	Indx	"0" (no index present)
32	Blank	blank
33-34	Form	blanks (no special format)
35-37	Lang	"eng" if NENG flag in LCS is not set "und" if NENG flag in LCS is set
38	Mod rec	blank (not modified)
39	Source	"d" (non-LC source)

For the scores & sound recordings format

Byte	Use	How we will set it
0-5	Entrd	The date on which we create the record
6	Dat tp	"s" (single known date)
7-10	Date1	Same as the date in 260 \$c
11-14	Date2	blanks
15-17	Ctry	"xx " (country not known)
18-19	Comp	blanks (form of composition unknown)
20	Format	blank (format not supplied)
21	Prts	"u" (unknown if there are parts)
22	Int lvl	blank (unknown)
23	Repr	blank (not a reproduction)
24-29	Acc Mat	blanks (no accompanying material)
30-31	LTxt	"n" (Literary text not applicable)
32	MEBE	"0" (main entry not in body)
33-34	Blanks	blanks
35-37	Lang	"eng" if NENG flag in LCS is not set "N/A" if NENG flag in LCS is set
38	Mod rec	blank (not modified)
39	Source	"d" (non-LC source)

For the serials format

Byte	Use	How we will set it
0-5	Entrd	The date on which we create the record
6	Pub stat	"u" (don't know if still published)
7-10	Dates	blanks (don't know when series started)
11-14	Dates	blanks (don't know when series ended)
15-17	Ctry	"xx " (country of origin unknown)
18	Freq	"u" (frequency unknown)
19	Regulr	"u" (regularity unknown)
20	ISDS	blank (no ISDS code assigned)
21	Ser tp	"p" if SER and PER flags are set in LCS " " if only the SER flag is set in LCS
22	Phys med	blank (no particular format)
23	Repr	blank (not a reproduction)
24-27	Cont	blanks (no specified contents)
28	Govt pub	blank (not a government publication)
29	Conf	"0" (not a conference publication)
30	Titl pag	"u" (title page availability unknown)
31	Indx	"u" (no index published)
32	Cum ind	"u" (cummulative index unknown)
33	Alphabt	blank (alphabet unknown)
34	S/L ent	"0" (successive entry)
35-37	Lang	"eng" if NENG flag in LCS is not set "und" if NENG flag in LCS is set
38	Mod rec	blank (not modified)
39	Source	"d" (non-LC source)

Appendix B  
LCS Prefixes Used to Determine Formats

<u>Films</u>	<u>Maps</u>	<u>Music</u>
AV-CART	ATLAS	AV-CASA
AV-FMST	ATLASES	AV-CASS
AV-TRAN	MAP	AV-DISC
AV-VID	MAPCASE	CAS
AV-VIDD		CASSETTE
CARTRDGE		CD
FILML0		CDISC
FILMLOOP		CDISCM
FILMM		DIS
FILMSM		DISC
FILMST		DISCM
FILMSTRIP		DISK
FILMSTRIP-CASSETTE		DSK
FILMSTRIP-TAPE		LP
FILMSTRIPM		MAGTAPE
FILMSTRIPRECORD		MUSIC
FILMSTRP		PHO
FL		PHONO
FLM-VDO		PHONOCASS
FLSTRP		PHONODISC
FMS		PHONOTAPE
FS		REC
MOVIE		REC-SP
MSLIDE		RECORD
SLIDE		RECORDS
SLIDE-CASSETTE		SCORE
SLIDE-RECORD-TAPE		SCORES
SLIDE-TAPE		SND-DSC
SLIDES		TAP
TRANSP		TAPE
TRANSPAR		TAPE-REC
TRANSPARENCY		TAPEM
TRANSPR		
VDC		
VICASSETTE		
VID		
VIDEO		
VIDEOC		
VIDEOCASSETTE		
VIDEODISC		
VIDEOTAPE		
VIDISC		
VIDRE		
VIDREC		
VIDTPE		
VITAPE		
16MM-FILM		
8MM-FILM		