

An In-Depth Look at LTI's Authority Control

Library Technologies, Inc.

Overview

Each step in LTI's authority control process is designed to maximize authority record links and to eliminate incorrect links. Where the full heading in the bibliographic record cannot be validated or linked to an authority record, LTI attempts to link portions of the heading. In addition, every batch or backfile receives limited review by professional librarians for links that could not be machine-identified. LTI guarantees that 95% or more of the controlled headings will be linked to either an LC or LTI authorized access point, presuming that the library has tried to adhere to national cataloging standards. Typically, about 93% of the headings link to LC authority records and 4% to LTI authority records. If less than 95% of the library's controlled headings link to an authority record, LTI editors perform further review to raise the overall heading link rate to 95% at no cost to the customer.

LTI's batch authority control is achieved through a series of clean-up operations:

1. Headings are "normalized" for spacing and punctuation to increase the probability of a link between a bibliographic record heading and an authority record heading.
2. The normalized headings are then compared against a comprehensive index of authorized and variant-form headings generated from authority records.
3. When a match occurs, a link is created between the authority record heading and the bibliographic record heading.
4. If the bibliographic record heading matches a variant form (tag 4XX) in an authority record, the bibliographic record heading is replaced by the contents of the authorized access point in the authority record (tag 1XX).

In most library databases, we find that with minimal processing about 70% of the normalized headings will match exactly with a 1XX field, or a 4XX pointing to a 1XX in an LC name or subject authority record. In other words, provided the authority control vendor does not introduce problems that destroy links, with minimal effort any vendor should be able to link seven out of ten library headings to LC authorized access points. What distinguishes quality authority control is how the vendor handles the other 30%. LTI's high link ratio of library headings to LC authority records results from over 25 years of collecting and analyzing several hundred million bibliographic headings.

The LTI Method

LTI compares, validates, and updates library headings by computer programs and fix tables. They use a combination of fuzzy logic and "fix and loop" routines to manipulate and repeatedly test headings against authorized access points. Depending on the type of heading (personal name, preferred title, topical subject, etc.), unlinked headings go through a cycle of routines designed to make high precision full or partial links to authorized access points.

Often headings can be validated fully with an authorized access point only after one or more fixes are made to subordinate units.

LTI's fix procedures and tables are based on well-defined rules of what constitutes a valid and invalid heading, as well as on empirical analysis of errors that have appeared in library headings processed by LTI. By nature, authority files are dynamic, with records added, deleted or revised. LC weekly authority record files and current cataloging rules are scrutinized also for changing patterns of usage. LTI monitors RDA rule changes, LC/PCC policy statements, and the weekly distribution of records to keep up-to-date.

Headings that cannot be validated against an LC authority record are checked against 4.8 million LTI-created cross-references from incorrect headings to the authorized LC access point. For example, our files include over one hundred cross-references (chiefly variant forms and misspellings) for *Tchaikovsky*. Similarly, the subject subdivisions *United States* and *Description and travel* are each represented by hundreds of variations (e.g. *Untied States*, *Descripton & travel*) in subject subdivision fix tables. If an LC link cannot be made, the heading is matched against a supplemental file of 2.4 million LTI authority records.

Diacritics and special characters are retained in the match key. LTI's normalized headings also retain delimiters and subfield codes. This assists programs that check for links with and without all subfields in the heading. For example, if a personal name heading that contains \$c and \$d fails to link with an LC authority record, the library's heading is tested for linkage without the subfield \$c, and the \$d is also checked for permissible variations. For instance, the birth dates *\$b. 1952*, *\$dborn 1952* and *\$d1952-* are all considered equivalent.

LTI's experienced librarians have devoted years to analyzing why headings do not link, what processing is needed to achieve links, and what changes are needed to bring unlinked headings into conformity with current cataloging standards. We also recognize that every backfile authorization project benefits from some level of editor review. High frequency unlinked headings are always examined. Single-occurrence unlinked headings are reviewed if they fall into one of many categories of software-identified "problem" headings, which include a range of data and content designator errors. Examples are detection of a music subfield code (\$m, \$r, \$o) without the presence of a title subfield, a series that lacks a title, a birth date that falls after a death date, etc.

As indicated above, every batch or backfile authorization project undergoes limited editor review for headings that could not be linked via computer. In full manual review, every unauthorized heading that remains after machine processing is reviewed by an editor. Given LTI's high level of linking with limited editor review, full manual review is not a cost effective approach for most libraries. If during limited review processing 96% of the library's subject headings are validated against authorized headings, there is little to be gained, while incurring substantial additional expense. Prior to providing a cost quote for full manual review authority control, LTI requires that the library submit its entire database for a no-charge evaluation.

For all authority control processing, controlled headings are extracted from bibliographic records and first run through automated processing. Only after a library heading fails to link

to an authorized heading does it become a candidate for review by an editor. The critical point is that if a heading is mis-linked during machine processing, that heading will never come to the attention of an editor because it has successfully (albeit incorrectly) linked to an authorized access point. Editors do not check every linked heading in every record to verify that a proper and correct link has been made. Instead, they examine only those headings that failed to link to an authorized heading during machine processing. If editors reviewed every heading linked during machine processing, the authority control vendor's costs might easily exceed one dollar per record. Few libraries could afford authority control at that price.

Blocking Links

Thousands of bad or ambiguous LC cross-references need to be "blocked" prior to linking headings during machine processing. LTI blocks any heading containing five or fewer characters from linking during the initial machine link. Selectively, some of these headings are unblocked [e.g., Asia, Iran, Iraq, etc.] where there is no likelihood of an incorrect link being made to an authority record. If blocked by LTI, these headings appear in the unlinked headings list and, if the library believes the authority record is important to its catalog—for example when it contains a useful cross reference or explanatory note—the authority record can always be downloaded from LC.

Four out of five of the blocks involve corporate/conference heading cross-references (41X). For example, the initials AAS appear as cross-references in 20 LC authority records. Neglecting to block ambiguous headings and cross-references can lead to some odd mis-links. For instance, we received a bibliographic record with the heading:

110 2 \$aBiblioteca Estadual Celso Kelly.\$c(Musician)

The pre-authorized heading probably read:

100 0 \$aBeck\$c(Musician)

Similar problems result when tables are used to expand parts of headings during a pre-processing procedure, without taking into account the entire heading. For example, we have seen many records in which the geographic subdivision *\$zMelbourne* was changed to *\$zMelbourne (Vic.)* when in fact many of the headings referred to the city in Florida.

Unfortunately, once a heading has been mis-linked, a subsequent vendor will find it almost impossible to identify and fix. Only a chance encounter or the presence of an invalid subfield code will allow it to be corrected.

Details of LTI's Processing

LTI's authority control processing includes correction of non-filing indicators, pre-processing and MARC updates, programmed linkage of headings to authority records, editor review of a subset of headings that remained unlinked, final re-linking, and writing of bibliographic and linked authority records to separate files for return to the library or its local system vendor.

Record Load

Regardless of database size, FTP is the standard method to receive and return bibliographic and authority record files. This can be done directly through the LTI website, or by using an FTP client. Records are accepted in either MARC-8 or Unicode (UTF-8) character encoding format, based on library preference. Data verification checks are made immediately after the transfer of records to LTI to ensure that records are properly formatted in the MARC-21 Communications Format. Bibliographic records are returned to the library in the same character set as received. Internally, prior to matching headings against LC authority records, we first convert controlled headings from Unicode to MARC-8. LTI returns LC authority records only in MARC-8.

Non-Filing Indicators

Setting of non-filing indicators in eight title fields is one of several pre-authority control processing operations. Non-filing indicators specify the number of initial characters to be ignored during computer filing.

For the title field (tag 245), the only title field to which the language code generally applies, articles associated with the fixed field language code are compared against the initial text in the title field. Based on this comparison, the non-filing indicator is set to 0 if no match is made, or to its proper matched value. The program takes into account leading diacritics and special characters that precede the first actual filing character.

If the language code in bytes 35-37 of the 008 is either blank or does not match a language code, the algorithm compares the title (245) field's initial text against a table of common articles in dozens of languages, and sets the non-filing indicator to its proper value.

Because the language of title fields other than 245 (e.g., X30, 240) is not necessarily the same as the language code, LTI's program compares non-245 field initial text against the table of common articles to set these non-filing indicators.

Automated non-filing indicator fix programs are sometimes unable to distinguish correctly between when a leading letter or word in a title is used as an article and when it is used as another part of speech that should not be ignored in filing. LTI's software uses, when appropriate, up to four words in the title to help determine if the initial word is actually used as an article. Examples of where the non-filing indicator is set properly to 0 based on an analysis of the second or subsequent word of the title are listed below:

A is for apple

A la orilla del viento

Das ist mir lieb

El Dorado, ciudad de oro

Lo que usted necesita saber sobre

Un de Baumugnes

While it is still possible for a non-filing indicator set correctly in the source record to be re-

set to an incorrect value, it is unlikely. LTI creates an ASCII text report showing the before and after settings along with the relevant title text. For libraries receiving corrections, the report provides reassurance that its non-filing indicators are being correctly set. Should a library insist on retaining its original non-filing indicators in the 245 title field, LTI can preserve them upon request. These libraries also receive the report, listing the changes that need to be made locally. Retention of incoming indicators is not an option in controlled title fields where the removal of initial articles is controlled by explicit LC authorized headings.

Resource Description and Access (RDA) instructs that initial articles be included in access points (RDA 6.2.1.7). However, LTI follows LC/PCC practice and applies the alternative instructions to omit initial articles in the formulation of headings. LC authority records do not contain non-filing indicators in 1XX fields, e.g., the geographic heading is *Dalles (Or.)*—not *The Dalles (Or.)* (n 82036146).

Preliminary Processing

Authority control at LTI begins with a generalized database clean-up program which increases the probability of bibliographic record heading matches against authority headings. To achieve consistency with the current MARC standards, bibliographic records are updated to reflect the latest *MARC 21 Format for Bibliographic Data* tagging and coding conventions.

Headings are normalized to correct a variety of typographical, punctuation, and spacing errors. Many changes are made at the subfield code level: additions (inserting \$f, \$l, \$s, and \$k in title fields, \$c and \$d in personal names, \$b in corporate names, and \$v in series), conversions (changing \$b to \$n in conference names, and correcting errors caused by the omission or improper assignment of \$c, \$d, and \$e in personal names). Certain non-controlled heading fields are also revised. Obsolete fields are deleted (e.g. 023, 039), or converted (e.g. 301/305 to 300) and selected obsolete elements or subfields revised (e.g. \$e removed from 041/052). An exhaustive table of authority control pre-processing fixes is found in the document *LTI MARC Update Changes* or on the web at: <http://www.librarytech.com/marcupd>.

Complex Bible, music, and other controlled title headings are parsed and updated. Unneeded spaces are removed. Leading non-filing articles are removed from added titles and title portions of author/title headings, and unnecessary parentheses and brackets are deleted from name headings. If the records contain GMDs in controlled title fields, and they are not enclosed by brackets, the brackets are added. Cancelled subject subdivisions such as *Addresses, essays, and lectures* and *Collected works* are removed. The letters l and O are converted to 1 and 0 respectively in date subfields, a check is made to ensure that subfield code \$d precedes dates in personal names, and second indicators in 1xx fields are set to blank.

Processing options chosen by the library are implemented: removal of \$4 from name headings, deletion of selected heading types, and conversion of headings tagged as Children's or Sears headings to LCSH.

Other Pre-processing Routines

Changes to cataloging rules and the MARC format require special processing of series, conference names, and titles prior to authority record linkage.

Obsolete series fields (400/410/411/440) are retagged as 490 fields with a first indicator of 1, and the content copied into appropriate 800/810/811/830 field. Removal of initial articles, capitalization changes, and adjustment of filing indicators are frequently necessary as part of this processing. To illustrate, the series:

440 4\$aThe series in computer science

is retagged:

490 1\$aThe series in computer science

and an RDA series field is added to the record using the LC/PCC alternative to eliminate initial articles:

830 0\$aSeries in computer science

If the original 4XX series begins with the pronoun *His*, *Hers*, *Its*, or *Their*, the pronoun is replaced in the 8XX field with the full heading from the bibliographic record's 1XX field. Series tagged as 840 are retagged to 830.

In older conference name headings, the order and punctuation of data elements in \$b, \$c, \$d, and \$n are updated to current practice. In the 111/611/711/811 fields, the obsolete \$b is converted to \$n and the number, place, and date are placed in parentheses with proper subfield coding and punctuation. To illustrate, the conference heading:

111 20 \$aPermanent International Altaistic Conference, \$b12th, \$cBerlin, Germany, \$d1969

is converted to:

111 20 \$aPermanent International Altaistic Conference \$n(12th :\$d1969 :\$cBerlin, Germany)

Controlled title fields are checked for proper punctuation and subfield coding. Omitted subfield coding, including subfield \$l before languages and \$f before dates, is inserted. In records containing a GMD in \$h, the required corrections will be made.

Extraction of Controlled Headings

Following preliminary processing, headings in fields subject to authority control are extracted from bibliographic records. A unique, sequentially assigned number is appended to each field as a link for reinsertion of the authority controlled heading into the bibliographic record when processing has been completed.

The table below lists MARC record fields and subfields checked by LTI's authority service. With the exception of subfields \$u, \$w, \$4, \$5, \$6, and \$9, all subfields in bibliographic record headings are matched against all appropriate subfields in LC authority record headings. Volume designation in \$v in 8XX fields is validated and corrected wherever possible, e.g., when it has been miscoded as part of \$a or miscoded as \$n or \$p, or when other clear errors in formatting occur. In addition, subfield \$v data is corrected based on the 642 field of the linked authority record. For LC subject authority control, only subject fields with a second indicator of 0 or blank are authorized. LTI offers optional authority control of LC Children's subjects, NLM's MeSH subject headings, and some genre headings.

100	\$a q b c d e k t n p l f g	
110	\$a b e n d c k t p l f g j	
111	\$a q e g k t p l f j	
130	\$a t n p l f k s g d m o r h	
240	\$a n p l f k s g d m o r h	
*400	\$a q b c d k t n p l f g v	
*410	\$a b n d c k t p l f g v	
*411	\$a q e g k t p l f v	
*440	\$a n p v	
490	1 st indicator 0 (recommended but optional)	
600	\$a q b c d k t n p l f m o r s h g v x y z	
610	\$a b n d c k t p l f m o r s h g v x y z	
611	\$a q e g k t p l f s h v x y z	
630	\$a t n p l f k s g d m o r h v x y z	
650	\$a b v x y z	
651	\$a v x y z	
655	2 nd indicator 0 or 7 if \$2 = MeSH, LCSH, LCGFT, or GSAFD (optional)	
700	\$a q b c d e k t n p l f m o r s h g	
710	\$a b e n d c t p l f m o r s h g j	
711	\$a q e g k t p l f s h j	
730	\$a t n p l f k s g d m o r h	
800	\$a q b c d k t n p l f m o r s h g v	
810	\$a b n d c k t p l f m o r s h g v	
811	\$a q e g k t p l f s h v	
830	\$a t n p l f k s g d m o r h v	
840	\$a h v	*converted to corresponding 8XX

Table I. MARC fields and subfields validated by LTI's authority control service

Authority Record Matching

After preliminary processing is completed, compressed “match keys” from bibliographic record headings are compared first against an index of match keys extracted from 1XX/4XX fields in LC/PCC authority records, and, secondarily, LTI authority records.

To increase authority record links, certain tag variations in bibliographic record headings are ignored during match key comparisons. For example, if a corporate name (110/610/710/810

field) has been improperly tagged as a personal name (100/600/700/800 field), the link will not only be made with the proper form of the heading, but the incorrect tag will be corrected automatically.

Some headings do not link because of incorrect subfield coding, as when a chronological subdivision (\$y) is miscoded as topical (\$x) and vice versa. LTI scans for and corrects such problems.

A common reason headings fail to match an authority record is simply because no authority record exists for the heading. Name authority records date back only to 1977 and subject authority records are a reflection of LCSH rather than an exhaustive file of subjects and subject subdivisions appearing in MARC records. Very few name headings and only a fraction of the thousands of combinations of topical, form, chronological, and geographic subdivision headings receive their own LC subject authority records.

Other initial linkage failures are caused by typographical errors, headings constructed under earlier cataloging rules, variant treatment of names with prefixes, and direct versus indirect division of geographic names. Most of these headings are corrected and linked to an authorized heading. To achieve these links, bibliographic record headings are submitted to repeated manipulations and checks by computer.

Name Authority Matching

Many personal name headings (100/600/700/800) fail to link with an authority heading because of variations in the fullness of birth and death dates or variations in “titles and other words associated with the name” (subfield \$c) information. Tests are applied to personal name headings that disregard minor variations to maximize authority record links. For example, when the match key created for the personal name heading *Allingham, Helen Paterson, \$c"Mrs. William Allingham,"\$d1848-* fails to link with an authority record, the heading is checked without subfield \$c data. If the heading still does not link, another match is tried which accepts any death date in \$d, if the birth date matches. This second match key refinement allows a valid link with the LC authority heading *Allingham, Helen Paterson, \$d1848-1926*. When a link is found to an authority record, the form in the authority record replaces the form in the incoming heading.

Title Authority Matching

Title portions of name/title headings (X00/X10/X11 fields containing \$t or \$k) and preferred title fields (130/240/630/730/830) undergo special treatment to increase matches. In each case the title portion of the field is checked for “floating” elements which are corrected and validated separately. These include dates (\$f), languages (\$l), medium designator (\$h), and versions (\$s). Music headings also have the arrangement statement (\$o), key (\$r), and medium of performance (\$m) corrected and validated. LTI processing attempts to link the heading fully to an authority record. If the entire name/title heading cannot be matched, trailing subfields are omitted to find the most complete match possible. Often only the name portion of a name/title heading can be linked to an authority record.

Untraced Series Processing

In 2006, the Library of Congress discontinued controlled access to series in bibliographic records and stopped creating series authority records. Since then, the number of series authority records has grown less rapidly, although there remains a substantial backfile of existing LC series authority records, along with new national-level records created by NACO libraries. Over the years LTI has also created many series authority records in-house, used to match incoming headings in the absence of a national-level record. While untraced series (i.e., those tagged as 490 with first indicator 0) were formerly excluded from standard authority control processing, the changed LC policy makes it important that libraries have a mechanism to continue to treat series as authorized access points.

To ensure consistency in access to series, LTI's no-charge "untraced series" option includes 490 0 fields in authority control. If the first indicator in a 490 field is set to 0, LTI changes the indicator value to 1 and generates an appropriate 8XX (traced series) field. If the newly created 8XX field is linked to an authorized access point during authority control, both the 490 1 and authorized 8XX fields are retained.

If the LTI-created 8XX field remains unlinked following authority control, libraries can opt to either delete the LTI-created 8XX field and change the 490 first indicator back to 0, or, retain the 490 (formerly, "traced differently" but now defined as "series traced in 8XX field") series with a first indicator of 1 and the LTI-created (but unlinked) 8XX series heading. When the form in the 8XX field cannot be matched to an authority record, LTI recommends that the 490 1/8XX combination be retained. For libraries retaining the 490 1/830 series, we further advise that 490 1 fields not be indexed in "title browse" indexes. The same advice would apply to 490 0 fields but, under this option, none of these fields will remain in the database. Should an authority record later be distributed, it will be provided to the library when the heading is authorized when it is next encountered in a continuing authority control run, whether *Authority Express (AEX)* or *Authority Update Processing (AUP)*.

The 2008 MARBI decision to cease use of the 440 field for a traced title series clarified the line between transcription and access points. It also redefined the 490 first indicator value 1 to "Series traced in 8XX field." Unless otherwise instructed, LTI processing converts obsolete 440 fields into the currently defined 490 1/8XX paired fields.

Subject Authority Matching

All subject headings are first processed through a "subject fix" program that corrects common errors in subject headings and fixes incorrect and obsolete topical, chronological, and geographic subdivisions.

Using tables of commonly occurring names, jurisdictions, subjects, and subject subdivisions, abbreviations are expanded and subjects and subject subdivisions are updated. For example:

Gt. Brit. becomes *Great Britain*.

Women, British becomes *Women\$zGreat Britain*.

U.S.\$xRace question becomes *United States\$xRace relations*.

The West\$xBiog. becomes *West (U.S.)\$xBiography*.

Headings are then matched against the LC subject authority file. Because most subject headings with subdivisions do not have separate authority records, LTI maintains tables of “free floating” topical, form, chronological, and geographic subject subdivisions. Subdivisions in these tables are used to validate thousands of headings that would otherwise not link. Headings that have not fully linked are broken into their component parts and tested against subject authority records and tables of validated floats.

Unlinked personal, corporate, conference, or geographic name (600/610/611/651 fields) headings are checked against the LC name authority file (ignoring subfields \$v, \$x, \$y, and \$z) to validate the form of name wherever possible. Later, subfields \$v, \$x, \$y, and \$z are validated against float tables. Where libraries have used double dashes in place of appropriate subject subdivision coding, LTI's software recognizes the subdivision and inserts the proper subfield code.

Obsolete LC subject headings replaced by two or more valid headings (i.e., “split” headings) present special authority control problems. For example, the topical heading *Negroes* was replaced by the headings *Blacks* and *Afro-Americans*; then, in 2001, *Afro-Americans* was changed to *African Americans*. In some cases the correct division of split headings can be deduced from other parts of the heading. Thus, LTI automatically converts headings having the structure *Negroes* \$z[State] to *African Americans* \$z[State], e.g., the heading *Negroes* \$zMississippi is changed to *African Americans* \$zMississippi. The heading *Negroes* followed by a country other than United States is changed to *Blacks*.

Where there is insufficient information to make the split, LTI links all occurrences of the obsolete heading to the broader heading. For example, *Crime and criminals* will go to *Crime*. However, if *Crime and criminals* is followed by the subdivision *Biography*, the new heading is changed to *Criminals*.

Geographic Subdivisions

LTI maintains a file of 170,000 correct indirect geographic subdivision forms. In addition, software is used to convert direct geographic subject subdivisions into indirect form. Geographic subject subdivisions are identified and fixed even when \$z data has been improperly coded as \$x or \$y or appended to \$a. Where necessary, higher level jurisdictions are inserted. For example, the topical subject heading:

650 0\$aDrawing, Paris\$xCatalogs.

is changed to

650 0\$aDrawing\$zFrance\$zParis\$vCatalogs.

and the heading:

650 0\$aGermans in Waterloo Co., Ontario.

is changed to

650 0\$aGermans\$zOntario\$zWaterloo (Regional municipality)

Topical (\$x) or Form (\$v) Subdivisions

Certain subdivision headings may be coded as topical (\$x) or form (\$v). Authority control processing must attempt to determine the usage in each case. In general, when - such a subdivision is the last subdivision in the heading, it will be coded as \$v, except when there is a full link to an LC authority record with the subdivision in \$x. Conversely, if it is not the last subfield in the heading, it will be tagged as \$x, unless it has been identified as a special case which allows \$x to follow \$v (e.g., \$v*Dictionaries \$x[language]*) or when it is permissible to follow \$v with another \$v.

The Role of Added Cross-References (Variant Access Points)

An important component to LTI's ability to offer a guaranteed link rate is a supplemental file of over 5.1 million added cross-references. Name headings constitute 88% of these with the remaining 12% subject cross-references. Each reference links the form of the heading used in the bibliographic record to the correct form used in the LC authority record.

These references are not added to LC authority records, but rather cumulated in separate files that are processed after linking to LC authority records, but before linking to LTI authority records. For example, when an LTI editor links manually the library heading *Marquand, John P., \$d1893-* to the LC authority record heading *Marquand, John P. \$q(John Phillips), \$d1893-1960*, that cross-reference is saved and applied subsequently to other library databases. Since not all cross-references can be applied to every database, editor-created variant access points undergo a second review prior to being added to LTI's cross-reference file.

Because so many machine-readable records are derived from shared cataloging databases, these supplemental cross-references play an important part in LTI's authority control routines, serving two purposes. First, they correct the access point in the library database that prompted their creation. Second, when the same unlinked heading appears in another customer's database, that heading is automatically linked to the proper authority record.

LTI Authority Record Linking

LTI maintains a proprietary file of name and subject authority record headings currently numbering 2.5 million. These authority records, established in accord with RDA and LC cataloging practice, are created from validated but unlinked access points that have appeared in library databases. The purpose of LTI authority records is to provide consistency in the absence of a nationally distributed authority record, and reduce the number of unlinked library headings that need to be reviewed by LTI editors. By eliminating headings that have already been reviewed in a prior job, editors are able to focus on the unlinked headings that can most benefit from review. Libraries that plan to examine the unlinked headings report are also helped by exclusion of these valid headings from that report.

To qualify as an LTI authorized access point, the heading must meet four criteria:

- it has been searched thoroughly but not found in the LC authority files,

- it is coded properly (tags, indicators, and subfield codes),
- it conforms to current national cataloging standards, and
- it is sufficiently unique that it is unlikely to represent more than one entity.

If an LC authority record is later distributed for the heading, the LTI authority record is deleted. Library access points are matched against LTI authority record headings only after they have failed to link to an LC authority record. In a typical database, LTI authority records validate 3% to 4% of the headings and most of these represent no change to the source heading—i.e., the access point in the library's bibliographic record is identical to the LTI authority record heading.

Review of Unlinked Headings

After all possible machine matches have been made, LTI editors selectively search unlinked headings in the LC name and/or subject authority file. Typical targeted headings include those which occur with a high frequency and those with illegal data, whether characters or subfields. Editors have access to a fully indexed version of the library's database and can display bibliographic records to resolve ambiguous headings.

Incorrect tags and subfield codes, typographical errors, omitted or incorrect dates, and related problems are corrected manually and resubmitted for relinking. In some cases, the heading alone is ambiguous, matching 4XX fields on several different authority records, as with initialisms and acronyms. If all the bibliographic headings are, in fact, a single entity, editors temporarily force the correct link for that specific database.

MeSH, LC Children's, Sears, & Genre/Form Headings

LTI's approach to NLM MeSH, LC Children's, and Genre/Form heading authority control parallels procedures for LCSH headings. Separate files of authorized headings are maintained for each controlled vocabulary.

NLM MeSH Headings

Using the National Library of Medicine's MeSH authorities file, LTI offers a MeSH authority control service to medical and health sciences libraries. NLM headings are updated annually. Topical (650) and a few geographical (651) subject headings having a second indicator of 2 are extracted for MeSH processing. Major and minor descriptors, as well as subheadings, are included in the MeSH file. Conversion of MeSH to LCSH is not an option. Libraries not using MeSH headings may choose to delete them from bibliographic records. As of 2016, LTI offers the option for libraries to follow MeSH practice of using deconstructed subject headings in bibliographic records.

LC Children's Subject Headings

In 1996 LC began distribution of Children's authority records based on Library of Congress usage. These headings supplement LCSH, providing alternative terms to simplify the adult terminology. Authority records in this short list are primarily topical (650) with a handful of

simplified names (600, 610). In addition, bibliographic records may contain LCSH headings tagged as LC Children's headings but considered appropriate for children's use. Therefore, LC Children's headings are first matched against LC's file of Children's authority records, and then against standard LC subject and name authority files.

Given that conflicts do occur between LCSH and LC Children's headings (e.g., LCSH *Swine* versus LC Children's *Pigs*), libraries may want to consult with their ILS vendor about the desirability of integrating the two controlled vocabularies in a single index.

LTI offers four options for processing LC Children's headings:

- 1) authorize them as LC Children's headings using LC's (*sj* control number) Children's authority records,
- 2) convert them to LCSH,
- 3) delete them from the library's bibliographic records, or,
- 4) ignore them completely during processing.

When Children's headings are converted to LCSH, second indicator codes of 1 in 6XX fields are changed globally to 0 and the appropriate subject subdivisions *Juvenile literature*, *Juvenile fiction*, etc. are added to the converted headings. Following authority control, resulting duplicate headings are removed from records. For example, a catalog record has these subject headings:

650 0 \$aFishes\$vJuvenile literature.
650 1 \$aFishes.

Following authority control two identical LCSH headings will exist: *\$aFishes\$vJuvenile literature*. One of these headings will be deleted in a final LTI check.

Children's headings having no LCSH equivalent—e.g., *French language materials*—are retained, though the second indicator is still set to 0.

Sears Subject Headings

While Sears subject headings may be authorized using a file of Sears authority records, the structure of Sears Subject Headings differs substantially from that of LCSH. Having both LC and Sears headings in a library's database can create confusion because of the inherent conflicts and inconsistencies. To avoid this, LTI recommends that libraries with a substantial number of Sears headings choose the processing option that converts Sears headings to LCSH.

Genre/Form Heading Processing

Genre headings in 655 fields are authorized using LTI's Genre Heading processing option. Headings may include LCGFT (Library of Congress genre form terms) or NLM MeSH. Libraries have the option to convert GSAFD (*Guidelines on Subject Access to Individual Works of Fiction, Drama, Etc.*) headings to LCGFT. LTI may authorize additional genre/form controlled vocabularies in the future when demand and authority record availability warrant it.

Implementation of genre/form headings has taken many years and several forms since first announced in 2007. LC's original approach was to identify LCSH headings that describe forms or genres for use in a field tagged: 655 _7 ... \$2l^{csh}. Later, to parallel subject heading use, the tagging was revised to prefer: 655 _0, with no \$2. In 2010, LC announced that it would, instead, create a new thesaurus for genres and forms, to be called *Library of Congress Genre/Form Terms for Library and Archival Materials (LCGFT)*. Headings taken from this list are assigned as: 655 _7 ... \$2l^{cgft}. The following year, LC replaced all the form/genre authority records that had been issued using the control number prefix *sh* with new records using the prefix *gf*.

In authority control processing, 655 headings with fields tagged in any of the above forms (655 7 ... \$2l^{csh}, 655 0 [no \$2], or 655 7 ... \$2l^{cgft}) are pulled from bibliographic records for processing. Headings are compared against LC's genre/form (*gf*) authority records and LTI-created authority records derived from official and semi-official sources. Content of the headings is updated to current usage, including correction of typographical errors in entry. If \$2 is absent, but the code is present following a space, headings will still be pulled, e.g. \$aAerial photograph l^{cgft} will be recognized and \$2 inserted as \$aAerial photograph\$2l^{cgft}. A variety of typos are also corrected: \$2l^{cgft}, \$2l^{gft}, etc. are corrected to \$2l^{cgft}. All LC-derived headings are converted to 655 _7 ... \$2l^{cgft}, reflecting current usage. Authority records for all form/genre headings that have a record are provided to libraries in a separate file.

In late 2015, LC completed a project to create authority records for general, music, and literature genre/form terms. Despite this, there remain some terms originally identified in LCSH but not yet represented by national authority records. LTI has supplemented the official headings with some 655 terms that are authorized as LCGFT although they lack authority records. Unless otherwise instructed, LTI tags these terms in the same manner (655 7... \$2l^{cgft}).

Some libraries may still be using the GSAFD thesaurus for genre/form terms. This thesaurus has not been updated since 2001, and the majority of terms are now covered by LCGFT headings. LTI strongly recommends conversion of GSAFD terms to LCGFT.

Prior to the implementation of genre/form headings, many libraries assigned the corresponding terms in LCSH as topical subject headings. For example, to aid access to videos, many libraries assigned the topical heading *Videorecordings* in a 650 field. LTI offers an additional option to convert such headings to 655 fields using the current LCGFT term. The latter conversion occurs only with stand-alone \$a genre/form headings. If an LCSH topical heading includes another subfield, such as \$v, \$x, \$y, or \$z, the heading is not considered a form heading.

MARC Record Distribution

Following an authority control project, libraries receive several files of MARC format records. The primary file has all of the library's bibliographic records, with corrections in place. Unless authority control is done during a system migration, records in this file are

used to overlay the existing, pre-processed bibliographic records in their entirety, so that all revisions, however minor, are made to the database. Other files provided contain LC authority records that have linked to headings in the library's database during processing, written in the MARC-8 character set and formatted according to the *MARC 21 Format for Authority Data*. All existing authority records should be deleted from the library's ILS prior to loading the new, comprehensive replacement authority records. In LC authority records, cross-references are always found on the "top level" authority record and not repeated on records with added subdivisions. Therefore, separate authority records are extracted, when available, for each level of a multi-level heading.

For example, three authority records would be extracted from the heading:

\$aEnglish poetry\$yOld English, ca. 450-1100\$xHistory and criticism.

\$aEnglish poetry

\$aEnglish poetry\$yOld English, ca. 450-1100

\$aEnglish poetry\$yOld English, ca. 450-1100\$xHistory and criticism.

Authority record files always include name/series (LCN) and LC subjects (LCS). Additional files of records may be provided based on a library's profile options: NLM MeSH authorities (LCM), LC Children's authorities (LCJ), and LCGFT (LCG). Some local systems require authority records for names used as subjects to be placed in the LCS file, which is easily accomplished using LTI options. In this case, about 30% of the subject authority records are likely to be for names and titles used as subjects (600/610/611/630 fields).

For databases up to 250,000 records, the library can expect to receive about one LC authority record (name or subject) per bibliographic record. The ratio of linked LC authority records to bibliographic records is inversely proportional to database size. For example, in a database of one-half million records the ratio is close to .65—i.e., $.65 \times 500,000 = 325,000$ authority records. At above two million bibliographic records, the ratio stabilizes at about one authority record per two bibliographic records.

Special Options for Authority Records

In 2007, a project was undertaken at LC to generate and distribute "Subject Authority Records for Validation Purposes." The project was to provide subject string authority records for popular and frequently-assigned headings, to assist local systems in validating LCSH subjects. These records are barebones, containing only the heading (taken from the LC bibliographic file) and a 667 field [nonpublic general note] that carries the text "Record generated for validation purposes." To date these authority records are limited to: (a) 651 subject heading fields for country names followed by free-floating sub-divisions; and, (b) subdivisions found in *Free-floating Subdivisions: an Alphabetical Index* that appear after topical and geographic headings. This type of authority record may be useful in an ILS that requires an authority record for every controlled heading or for a library that lacks a comprehensive authority service such as AEX or AUP, which parses headings and validates floats. However their value to an individual library may be questionable as no library has exactly the same collection as LC, with only the subject heading strings used in the LC

records. Because the original estimate of the number of these authority records ranged from several hundred thousand to a million, some LTI clients expressed concern about the number and utility of these validation records. In response, LTI added the option to exclude from library “LCS” files those subject authority records containing a 667 with the validation purposes note. This project appears to have stalled or been abandoned, as the number of these records plateaued some time ago at about 79,000.

LC authority records began to be enhanced by the addition of non-Latin cross-references in 4xx fields in 2008. Currently, about one-half million authority records contain non-Latin fields, with non-Latin fields now included routinely in new and updated authorities. Because not all local systems are able to handle non-Latin characters correctly, LTI offers clients the option to have the non-Latin 4XX fields removed from authority records provided to them by LTI.

“Deblinding” Authority Records

“Deblinding” usually refers to the removal of certain *see also* references (5XX) from authority records, preventing the occurrence of *see also* references when the *see also from* headings are not present in the catalog. For example, deblinding would prevent the *see also* reference *Baked products see also Bread* when there are no records with the heading *Baked products* but there are records with the heading *Bread*. Since the *see also* information may be helpful to patrons even when the *see also from* heading is not currently used in a catalog record, LTI retains all 5xx fields. Note also that, because of the dynamic nature of both bibliographic and authority files, the deletion of such a reference may prove problematic when the heading *Baked products* enters the bibliographic database at a future date.

The following summarizes a posting to the LTI-Users electronic mailing list and makes an excellent point.

Blind references in an online catalog do not result in wild-goose chases in the same way they might have in card catalogs. Take, for example, the reference *Ho, Ho see He, He*. In a card catalog the user would have to flip through cards to find out what, if anything, was filed under *He, He*. Some references would involve opening new drawers, and depending on the size of the catalog, covering a bit of ground. However, many online catalogs tell one right away what to expect, e.g., *Ho, Ho see He, He* (0 records). Even if the number of records does not display up-front, the correct form should only be a click away. Informing users what the authoritative heading is, and that there is nothing there, actually prevents wild-goose chases. This is especially true of subject headings, with their sometimes counter-intuitive forms.

Reports

After each LTI authority control job, libraries receive several standard reports. The “Final Link” report details the number and percentage of name, series, and subject access points that have linked fully, linked partially, or not linked to either an LC or LTI authority record. The report shows: frequency of each controlled heading field tag, number and percentage of

access points validated or not validated against an LC authority record, and number and percentage of access points changed in processing.

Two other reports are generated each time LTI authorizes a file of bibliographic records: the “Unlinked Headings” report listing headings in which \$a could not be linked to an LC or LTI authority record, along with the frequency with which the heading was found in the file, and the “Non-Filing Indicator Change” report showing title fields in which a non-filing indicator was revised.

For the great majority of headings listed as “unlinked,” there is nothing “wrong” with these headings. There simply is no LC or LTI authority record established for the heading. Generally, systematic searching of unlinked headings in nationally distributed authority files is not a cost-effective choice, given the required time. Selected follow-up on high frequency unlinked headings or obvious anomalies may be of some benefit, but many libraries do nothing with this report.

Several optional reports may also be requested. “Partially Validated Headings” shows controlled headings where \$a linked to an authority record, but one or more subsequent subfields could neither be linked to an authority record nor validated using float tables. Most of these entries are name/title combinations where the name was validated.

880 fields

Bibliographic records with 880 fields containing alternate graphic representation data (non-Latin characters), present a challenge for authority control. While LTI does not authorize these headings, several optional processes can help with their management, using two reports, which incur no additional charge but must be specifically requested.

The “880-upd” report lists controlled headings that were changed during authorization but which show a linking field tagged as 880/\$6. Each change is shown only when it happens. To assist library staff in making needed edits to the linked fields, the report includes: the bibliographic record control number, the tag and content of the controlled heading field before updating, tag and content of the revised field after updating, and the content of the linked 880 field that may require local revision.

The “880-err” report lists errors in links between 880s and their corresponding fields, such as bad parsing of \$6 data or missing fields. The entry on the report will be eliminated when the record is corrected and re-sent to LTI, either in a new base bibliographic file or in an Authority Express (AEX) file. Each entry on the report includes: the bibliographic record control number, the tag and content of the controlled heading field with \$6, the tag and content of the linked 880 field, and a brief explanation of the error found. Authority Update Processing (AUP) users will continue to receive reports of each error until it no longer exists in the copy of the database maintained at LTI.

When RDA was adopted as the national cataloging code, the format of certain types of date subfields (\$d) in personal names was revised. To assist libraries with personal names in 880 fields, processing was enhanced to offer the option to revise the \$d to match the form in the linked, authorized personal name field. When this option is chosen, a third report, called

“880-sfd” is produced, listing each occurrence of this type of change. As with the previous reports on 880 data, the report lists: the bibliographic record control number, the tag and content of the controlled heading field, and a brief explanation of the change made.

All reports are provided as ASCII text, which can be loaded into a text editor or word processor for review purposes.

While Records Are Being Authorized

During the period while the database is being authorized, bibliographic records for new titles may be added to the database, however, existing bibliographic records should not be revised or deleted. When the post-processed bibliographic records are returned after authorization, any changes made to existing records will be lost as part of the overlay process. A log may be kept to track records that must be edited or deleted, so that these steps may be taken after the authorized database has been reloaded. This restriction does not apply to item data—e.g., copies, locations, barcodes, etc. which may be revised, added, or removed.

A library needs to export item-level data (formatted as item fields) along with its bibliographic records in only two situations. Most commonly, this is required when a library is migrating to another local system and this data must be used to build item records in the new local system. Occasionally a library will want to remove duplicate bibliographic records prior to an authority control project. In this case, item data must reside in the exported bibliographic records or item records will be "orphaned" when the bibliographic records are reloaded. Deduping a “live” database poses special problems and should only be done following close coordination between the library’s local system vendor and LTI.

ILS “Authority Control” Modules

By definition, authority control demands a single “authority.” LTI’s services provide the highest quality authority control available to libraries today. Using a local system’s authority control module will undermine the advanced algorithms and tables that were used to authorize the library’s database.

Such software modules rely on matching and updating bibliographic headings based on the presence of the “old” heading as a cross-reference (4XX) in an authority record – a limitation which can both miss needed revisions and introduce problems. LC frequently does not carry over the "old" heading, and even when it does there may be other issues in linking because of variations in capitalization, spacing, or punctuation occurring, in either catalog record headings or in LC authority records. Dozens of such errors in authority records are reported by LTI to LC each week

More problematic than changes that are missed are revisions that should not be made at all. Currently, we block over 201,000 headings in authority records (mostly 4XX fields) from linking to any heading because of the likelihood of introducing an error. Other heading changes are limited to only certain types of linkages, perhaps additional data must be present, or there must be an absolutely exact match, before the link is allowed.

LTI users report strange things when they “turn on” their local system’s authority control

module. For example, the library may find a disproportionate number of headings tagged as subjects but that are in fact titles or series. While a series may on rare occasions be used as a subject, it is uncommon, and rarely permissible along with the presence of form (\$v), topical (\$x), period (\$y), or geographic (\$z) subdivisions. If these rules are ignored,

Romance languages \$x Modality

can be converted to:

\$aAmerican university studies.\$nSeries II,\$pRomance languages\$xModality.

because the authority record for *American university studies. Series II. Romance languages*, has a 430 variant title of *Romance languages*.

LTI's processing protects against this type of incorrect tag level change and, if not linked correctly by our code or editors, at worst, the above headings appear in the unlinked headings list.

Such problems appear to be present to some degree in all databases where the library uses an authority control module made available by its ILS vendor, unless substantial staff time is invested in reviewing every such change. Regardless of local system, LTI strongly advises that ILS modules making automatic changes based on authority records be disabled.

Database Maintenance

Authority control processing represents an investment in a library's database. Even during the project, new bib records are being added to the database and the Library of Congress has been adding new authority records and changing existing authority records. Maintaining the library's authority-controlled database requires attention to each of these two processes so that the investment made may be protected.

LTI offers two services for continuing authority control, allowing libraries to maintain their files in an efficient and cost-effective manner. **Authority Express (AEX)** is used to authorize new bibliographic records added to a library's database. **Authority Update Processing (AUP)** generates new and revised authority records, reports on changes to bibs and authorities and – alone among authority control vendors – provides a file containing any of the library's bibliographic records in which a heading has changed. Details on these services are found in the LTI document, *Continuing Authority Control Services*.

The two LTI continuing authority control services work together to keep a library's bibliographic and authority databases up to date. They do different things, both of which are essential if a library wants to keep a complete and current authority file and have correct headings in their bibliographic records. When new cataloging is authorized consistently through *Authority Express* and regular runs of *Authority Update Processing* are received, a library need never again send out the entire database for batch processing.

Library Technologies, Inc.
“A Commitment to Quality”
2300 Computer Ave. Suite D-19
Willow Grove, PA 19090-1736
(215) 830-9320 (800) 795-9504 Fax: (215) 830-9422
LTI@LibraryTech.com
www.authoritycontrol.com

Copyright, Library Technologies, Inc., March 2016