

# AUTHORITY CONTROL SERVICES

## Library Technologies, Inc.

---

Key to the successful implementation of every local system is a high quality database; conversely, nothing undermines the credibility of a local system more than a poorly prepared database. As an experienced library automation vendor with a strong commitment to quality work, Library Technologies, Inc. provides a comprehensive program of professional database preparation services. These include a variety of duplicate record resolution options, MARC record updates and clean-ups, local system item field builds, item field re-mapping for local system migrations, smart and dumb barcode label production, media transfers, and batch and continuing authority control.

Whether your database contains ten thousand or several million records, at LTI the work will receive the same high level of personalized service and attention to detail. A database consultant works with library and local system staff to define a specification that not only satisfies immediate local system requirements, but also those that will be important should the library need to move its records to a different system in future years. LTI is dedicated to delivering the highest quality, most cost-effective database services possible.

LTI offers name and subject authority control based on the Library of Congress name and subject authority databases. LC authority files include 7.4 million name authority and 349,000 subject authority records. These files are supplemented by 2.3 million LTI name and subject authority records for which no LC authority is available.

In addition to LC name and subject headings, LTI also provides authority control over NLM MeSH, LC Children's, Sears, and selected genre subject headings. The difference between machine, or limited review processing, and full manual review is that in the latter *every* heading that remains unauthorized following machine processing is reviewed by an editor. Following that review, LC authority records will be found for some percentage of the headings that could not be linked by machine alone. While both limited and full manual review authority control options are available, for full manual review jobs we ask that the library submit its entire database for analysis prior to LTI's accepting the project.

LTI customer records receive the most advanced and rigorous authority control processing of any database vendor. Every batch or backfile authorization project undergoes some level of editor scrutiny for headings that could not be linked via computer. LTI is so confident of its authority control services that it guarantees 95% or more of a library's controlled headings will be linked to either an LC or an LTI authority record during limited review processing. If less than 95% of the library's headings link to an authority record, at no charge to the client we perform whatever additional editor review is needed to raise the library's overall heading link rate to 95%. This applies to all US libraries adhering to nationally accepted cataloging standards and practices.

Since 1989 LTI has authorized over two hundred million bibliographic records for hundreds of libraries. Local system vendors that LTI has worked with include: Ameritech Library Services, Best-Seller Library Systems, CARL Systems, Carlyle, CLSI, Comstow, Cuadra Associates, DOBIS, DRA, DYNIX, Endeavor Information Systems Inc., Ex Libris (USA) Inc., Follett, Gaylord, Geac, General Automation, HORIZON, INLEX, Innovative Interfaces Inc., The Library Corporation, multiLIS, NSC, Polaris Library Systems, Sagebrush Technologies, SIRSI, SydneyPLUS, TechLib Plus, Utlas, and VTLS.

Library Technologies, Inc. has evolved to become the only authority control specialist in the library automation marketplace. We do not develop or maintain a local system, nor do we perform retrospective conversion, record enhancements, or other cataloging services. We are a small company and stick to our knitting—focusing all software development on creating the most advanced batch and continuing authority control linking procedures of any library database vendor. LTI continues to improve its authorization services in linking controlled headings to Library of Congress authority records and takes satisfaction in delivering the highest link rates in the authority business.

A privately held company with a strong service orientation, Library Technologies, Inc. prides itself on competitive pricing, fast turn-around time, and accommodating the special needs of client libraries. We recognize the customer's financial investment in its

database and the importance of those records to the library's mission. LTI stands behind its services and promptly responds to any variation between the agreed upon specifications and the post-processed records.

## Introduction

Bibliographic control in a library catalog is accomplished by the assignment of a unique form of a heading and the use of cross-references from unused and related headings. Selecting a single form brings together in one place all the works of an author, editions of a work, titles in a series, and subjects dealing principally or exclusively with a topic. Through the use of *see* and *see also* references, authority control creates a syndetic structure that guides the user to the materials sought. The *see* reference tells the user that the information being sought is to be found not under that heading, but rather under a different (i.e., the "authorized") heading. *See also* references show relationships between headings. The two concepts of heading uniqueness and cross-referencing are the pillars of authority control. Together they make possible efficient access to library resources.

Why has authority control become increasingly important to libraries? In the past, library administrators have viewed authority control as a desirable element in the cataloging process, but one so costly in professional staff time as to be justifiable only in research institutions. These days the ubiquity of Library Management Systems (LMS) in all types and sizes of libraries belie the notion that authority control is a luxury.

Widespread implementation of on-line catalogs reveals what card catalogs so effectively concealed. Namely, that most card catalogs were plagued with "divided file" headings, archaic subject terms and place names, non-existent or inadequate cross-references, and otherwise misleading headings. Changing cataloging rules, reliance on shared cataloging from bibliographic utilities, and, until recently, the catalog's physical form, have all contributed to the problem of maintaining current and consistent name and subject headings. Many of the problems found in card catalogs were carried over into online catalogs, particularly if the library's records were never authorized, or authorized poorly.

Fortunately, Online Public Access Catalogs (OPACs) not only reveal problems that remained hidden in card catalogs, but the machine-readable records from which

the LMS is created make possible a level of authority control superior to anything that could be achieved in non-computerized catalogs.

Authority control is essential for effective local system searching. It improves access dramatically by providing consistency in the form of headings used to identify authors, place names, uniform titles, series, and subjects. Equally important are logical linkages between names and concepts. Keywords, combined with Boolean operators, offer powerful supplemental search capabilities, but they are no substitute for authority control. The collocation functions of a library catalog go well beyond its "finding list" functions and can only be met by traditional authority control.

Although library users are the principal beneficiaries of enhanced access resulting from authority control, authorized headings and display of cross-references within a local system reduce the work of technical services librarians responsible for building and maintaining the catalog. There is a price to be paid for good authority control, but an even bigger price to be paid for not having good authority control. Participation in resource sharing networks, cooperative ventures such as NACO, dependence on Internet connectivity, and recent trends in "outsourcing" technical services operations are also strong arguments for libraries to ensure headings adhere to national bibliographic standards. An ancillary benefit of authority control is the imposition of consistency on variant forms of the same heading that fail to link to an authority record.

## Steps In Authority Control Processing

Authority control generally takes place following all other database work, e.g., duplicate record resolution, creation of item fields. Libraries migrating from one local system to another often take this opportunity to re-authorize their bibliographic records and obtain up-to-date authority records prior to loading records into the new local system.

From a processing perspective, batch authority control is achieved through a series of clean-up operations. First, headings are "normalized" for spacing and punctuation to increase the probability of a link between a catalog record heading and an authority record heading. Next, the normalized headings are compared against a comprehensive index of authorized and variant-form headings generated from authority records. When a match occurs, a link is created

between the authority record heading and the catalog record heading. If the catalog record heading matches a *see* reference (tag 4XX) in an authority record, the catalog record heading is replaced by the contents of the authorized heading in the authority record (tag 1XX).

Ideally, all headings in the library's database should match an LC authority record, but in practice some headings do not link. The vendor responsible for authority control must do everything possible to increase links between catalog record headings and authority records while avoiding false matches.

After all possible headings are matched to an authority record, they are inserted back into the bibliographic records. That is, the authority record heading replaces completely the bibliographic record heading. Along with the post-processed database of bibliographic records, the linked LC authority records are extracted and returned with the bibliographic records for loading into the local system's authority module. Authority records returned to the library are collection-specific. The local system builds its index tables from both the controlled headings in the bibliographic records and the *see* and *see also* cross-references present in the LC authority records.

The number of bibliographic records returned is the same as the number of source records to be authorized. Prior to loading the new, comprehensive replacement authority records, the library will want to delete **all** existing LC and local authority records. In other words one wants to "overlay" the newly authorized catalog records, while authority records need to be completely replaced. During the period while the database is being authorized staff may continue to catalog new titles; however, they should avoid making changes to existing catalog records. This restriction does not apply to item record data—e.g., location codes, barcode numbers, etc.

Depending on the local system, the library may want brief or "provisional" authority records created for headings either not matching, or not fully matching, an LC authority record. These are skeleton authority records containing the unverified heading and default data. Provisional authority records do not contain cross-references and are of limited use since most modern local systems are capable of generating such authority records from unlinked headings.

Following batch authority control, a library is confronted with two challenges in keeping its catalog headings up-to-date and consistent. The first is to ensure that headings from new catalog records being

added to the database are authorized. Options for controlling new headings include: 1) authorize them manually at the time of cataloging prior to loading the new bibliographic records into the database; 2) make no attempt to authorize new headings at the time of cataloging or record load, but periodically export and submit newly added bibliographic records to the authority control vendor for batch authority processing; or, 3) use a FTP-based service to transmit new records to the authority control vendor in batches on a regular (e.g., weekly, biweekly) or irregular basis.

A second challenge in keeping a library's headings in harmony with authorized headings is controlling for new and revised authority records issued after the library's records were batch processed. To help catalogers keep abreast of these additions and changes, authority control providers offer services alerting customers to the issuance of new and changed LC authority records affecting headings in the library's database.

Unless the library views authority control as a one-time event (i.e., it has no interest in maintaining control over headings after the batch authority control is finished), the availability of a comprehensive update service should be an important consideration when selecting an authority control vendor.

## LTI's Authority Control

Automated batch authority control processing has advanced significantly since the 1980s when first introduced. In the early days there were few vendors and their procedures evolved from controlling headings in the production of printed book and microform catalogs.

Because the demands of computerized catalogs far exceed those of card and book catalogs, LTI's authority control procedures have been designed to meet the more stringent authority requirements of the online environment. Given the enormous variation that exists among vendors in the thoroughness and quality of their authority control processing, LTI takes the position that customers have the right to know in advance what is being purchased. LTI takes the guesswork out of authority control by guaranteeing a minimum link rate of library headings to authorized headings for US libraries following national cataloging standards.

LTI's authority control processing includes filing indicator fixes; pre-processing fixes and MARC

updates; machine linkage of headings to authority records, including limited manual review of headings that remained unlinked following machine processing; optional full manual review of unlinked headings; final re-linking; and writing of bibliographic and linked authority records to separate files for transmittal to the library or its local system vendor.

Each step in the authority control process is designed to maximize authority record links and to eliminate incorrect links. Where the full heading in the catalog record cannot be validated or linked to an authority record, LTI attempts to link portions of the heading. Two types of batch authority control are available: limited manual review and full manual review. Hereafter, limited manual review is referred to as "limited review," and full manual review as "full review."

Limited review compares, validates, and updates library headings by computer programs and fix tables. Formerly referred to at LTI as "machine" or "machine-only" processing, all limited review batch authority jobs receive 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 "problem" headings. Problem headings are software-identified and 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, presence of digits in subfield *\$a* of a personal name, a birth date that falls after a death date, etc.

In limited review processing, LTI guarantees that 95% or more of the customer's controlled headings will be linked to either an LC or LTI authorized heading. This assumes, of course, 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 performs further review to raise the overall heading link rate to 95% at no cost to the customer.

In addition to a guaranteed link rate, distinguishing features of LTI's authority control service include: 1) a supplemental file of 2.3 million LTI authority record headings for which no LC authority record has been distributed; 2) extensive machine clean-up of headings prior to linking; 3) direct linkage of headings via 4 million LTI proprietary cross-references from incorrect headings to the LC authorized heading; 4) for full review jobs, systematic search in the nationally distributed authority files of every unlinked heading by professional librarian editors; 5) for current cataloging,

an FTP-based authority control service (called *Authority Express* or *AEX*) that turns around files of up to 10,000 records within half an hour after receipt at LTI; and 6) an update service (called *Authority Update Processing* or *AUP*) alerting the library of heading changes caused by new LC authority records and changes to previously issued records.

While there is no requirement to use either continuing service (*AEX* or *AUP*), for clients that do use them, there will never be a need to do a batch re-authorization of the entire database. Moreover, controlled headings always remain in synch with national cataloging practices. In effect, every *AUP* run represents a total reauthorization of all the library's bibliographic records.

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 headings.

What distinguishes superior authority control from mediocre processing is what the authority control vendor does with the 30% of the headings that have failed to link initially to an authorized heading. An important factor in LTI's high link ratio of library headings to LC authority record headings is the large number of headings linked via machine based on years of collecting and analyzing several hundred million bibliographic headings.

LTI uses a combination of fuzzy logic and "fix and loop" routines to manipulate and repeatedly test headings against authorized headings. Depending on the type of heading (personal name, uniform title, topical subject, etc.), unlinked headings go through a cycle of routines designed to make high precision full or partial links to authorized headings. Often headings can be validated fully with an authorized heading 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 over the past 20 years. For example, the subject subdivisions "United States" and "Description and travel" are each represented by hundreds of variations in subject subdivision fix tables. Headings that cannot be validated against an LC authority record are checked against 3.8 million LTI-created cross-references from incorrect headings to the authorized LC heading.

Similarly, our files include over one hundred cross-references (i.e., in this case variant forms and misspellings) for the composer "Tchaikovsky." If an LC link cannot be made, the heading is matched against LTI authority records. Matching against LTI's enhanced LC cross-reference file occurs after LC authority record links have been made but prior to LTI authority record linking.

Diacritics and special characters are retained in the match key. LTI's normalized headings also retain delimiters and subfield codes. This permits the construction of 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 example, the birth dates \$db.1952 and \$d1952- are considered equivalent.

Wherever possible we have replaced editor review of unlinked headings with computer-based correction and matching techniques. Writing programs to optimize linkages while avoiding mismatches is a balancing act that requires expertise both in cataloging and in manipulating headings in MARC records. We also recognize that every backfile authorization project benefits from some level of editor review. 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 modern cataloging standards. LC weekly authority record files are scrutinized for changing patterns of usage. In this way LTI is able to implement current LC cataloging decisions sometimes well in advance of notification in LC's *Weekly Lists* or *Cataloging Service Bulletin*.

### *Limited vs. Full Manual Review*

Some argue that full manual review of headings that failed to link, or only partially linked, to an authority record during automated processing is an essential component of any serious authority control service. Surely, there can be no question that manual review will always correct some headings not linkable via machine processing. On the down side, manual review of unlinked headings extends time and represents a major cost in delivering authority control services. This cost is passed along to the customer and is apparent in the wide price difference between machine processing and manual review processing.

Whether full manual review is critical depends to a large extent on the effectiveness of the vendor's machine processing. If full manual review corrects only a percentage point or two of the library's total headings, its considerably higher price may not be justified. For example, if during machine or limited review processing 96% of the library's subject headings are validated against authorized headings, the potential of increasing the link rate by several percentage points may not merit the additional expense. On the other hand, if the vendor's machine processing validates 80% of the library's subject headings and another 17% are validated during manual review, then manual review becomes a necessity with that particular authority control vendor.

In the absence of commonly accepted qualitative standards measuring the effectiveness of batch authority control services, libraries should consider two benchmarks prior to contracting for authority control. First, the vendor's anticipated validation rate of library name and subject headings to authority records; and, second, the ratio of authority records to bibliographic records that the library receives from the vendor after processing is completed. These ratios vary with library type, size, and nature of the collection, but the vendor should be able to provide reliable estimates based on similar jobs already performed. Vendors should also offer guidance as to the most cost-effective and suitable processing options for a specific library.

Unless there are special circumstances, such as the library requesting that locally tagged subjects be re-tagged as LC subject headings prior to authority control, no US database adhering to nationally accepted cataloging standards is shipped from LTI unless at least 95% of the controlled headings have linked to authorized headings.

Looking at the second benchmark, for databases up to 250,000 catalog records, LTI returns about one LC authority record per catalog record. As the number of catalog records increases, the ratio of LC authority records to bibliographic records decreases because headings are more likely to re-occur. Public library databases are likely to yield fewer authority records than academic library databases because of the likelihood of the collection containing multiple titles by popular authors. For academic databases above one million records, the ratio of authority records to catalog records is .65:1. In other words, a million bibliographic records might result in extraction of about 650,000 total LC name and subject authority records. Cost benefit is yet another way to evaluate machine versus manual review processing. [These figures are based on statistics from LTI processing and are not

applicable to other authority control vendors.] On average, for bibliographic records taken from OCLC, RLIN, or other distributors of LC MARC records, roughly one percent of the library's controlled headings will be changed during full manual review that would not have been changed during limited review processing. For example, assuming a database of 200,000 catalog records with each record containing 3.8 controlled headings—i.e., a total of 760,000 controlled headings—full manual review processing might change about 7,600 headings that would not be changed during limited review processing. To achieve this 1% link rate improvement, the cost of a library's authority control project will be more than double that of a limited review job. For some libraries, particularly those with unusual collections or whose records are of questionable quality, the additional fixes resulting from full manual review are justified. For most libraries, selecting limited review processing is an opportunity to save money without paying a huge penalty in quality. For other authority control vendors full manual review may have real utility but with LTI's processing it is nearly always a waste of library resources.

Libraries are sometimes persuaded to select manual review for bogus reasons. They are led to believe that manual review is somehow going to eliminate or reduce greatly the number of mis-linked headings. The "Madonna myth" illustrates this well. An authority vendor might suggest that in machine processing the rock star "Madonna" could be confused with "Mary, Blessed Virgin, Saint." There is a certain logic to this. However, if one considers how full manual review authority control is in fact performed, the example makes no sense, other than to discourage libraries from requesting a less costly machine processing option.

Regardless of whether a library selects machine or manual review processing, controlled headings are extracted from bibliographic records and first run through machine processing. It is only *after* a library heading has not linked to an authorized heading during machine processing that the heading becomes a candidate for manual 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 heading. 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.

Like the Madonna example, there are thousands of bad or ambiguous LC cross-references that 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.

Taken together, in the LC name and subject authority files there are over 11 million *IXX* headings and *4XX* (*see from*) references, of which LTI blocks 163,000. As one might anticipate, four out of five of the blocks involve corporate/conference heading cross-references (*4IX*). For example, the initials *AAS* appear as cross-references in 20 LC authority records, including those for the *American Astronautical Society*, *African Academy of Sciences*, *Arkansas Archeological Society*, *American Arachnology Society*, *Aquaculture Advisory Service*, *Association for Asian Studies*, etc. Neglecting to block ambiguous headings and cross references can lead to some amusing mislinks.

It is easy to identify databases where vendors have confused LC authority records with authority control within a given database. Actual examples of the mayhem caused by simplistic processing algorithms are listed below.

Catalog record heading as received from library:  
*610 20 \$aInternational Society for Augmentative and Alternative Communication,\$cthe patriarch,\$xJuvenile literature.*

In pre-authorized record, heading probably read:  
*600 00 \$aIsaac,\$cthe patriarch,\$xJuvenile literature.*

Catalog record heading as received from library:  
*110 20 \$aBiblioteca Estadual Celso Kelly.\$c(Musician)*

In pre-authorized record, heading probably read:  
*100 00 \$aBeck,\$c(Musician)*

Catalog record heading as received from library:  
*600 20 \$aIturralde Gomez, Antonieta,\$cPrincess of Wales.*

In pre-authorized record, heading probably read:  
600 00 \$aDiana,\$cPrincess of Wales.

Similar problems result when authority control vendors use tables to expand parts of headings during a pre-processing procedure, without taking into account the entire heading. For example, for years one vendor routinely changed the geographic subdivision \$zMelbourne to \$zMelbourne (Vic.) when in fact many of the headings referred to the city in Florida.

Consider the following examples from an academic library database previously authorized by an established authority control vendor. As LTI received the catalog record for Joanna Cole's *El autobus magico en tiempos de los dinosaurios* [*Magic school bus in the time of the dinosaurs*] (1995), it contained two unjustified personal name/title subject added entries (tagged 600 – 2<sup>nd</sup> indicator 0), one for:

\$aOliver,Rupert.\$tDinosaurios\$xLiteratura juvenil.  
\$2bidex

[and another for]

\$aWilson,Ron,\$d1941-\$tDinosaurios\$xLiteratura juvenil.\$2bidex

When examining the LC author/title authority records for *Oliver, Rupert.\$tDinosaurios* [n 96117774] and *Wilson, Ron,\$d1941-\$tDinosaurios* [n 97004704], one finds 430 cross-references from the title *Dinosaurios*. In this case the library's source record—i.e., as delivered to the first authority control vendor—contained the Spanish language subject *Dinosaurios\$xNovela\$2bidex* incorrectly tagged as *LCSH*. Because the vendor found multiple cross-references in the LC index files from *Dinosaurios* to both the Oliver and the Wilson authority records, it hedged its bets by adding both author/title subjects to the library's catalog record. Presumably, had there been 27 cross-references in LC authority records having a see-reference from *Dinosaurios*, the vendor would have inserted 27 new subject headings—none having any relationship to the work at hand. Identical sets of the Oliver and Wilson bogus author/title subjects appear in other records in this database, including Spanish language editions of Michael Crichton's *Jurassic Park* and Syd Hoff's *Danny and the dinosaur*. If nothing else, these records demonstrate the ease with which computers can trash a library database.

In another record LTI found a subject heading for a Russian author who had no relationship to the publication. The catalog record describes a state

government report issued by the Alaska Department of Fish and Game, Division of Wildlife Conservation, on the topic of wolves. The authority vendor added a spurious heading because there is a cross-reference from "Wolf" in the LC authority record having the control number *no 96032788*.

In still another catalog record, this one a handbook, one finds a puzzling uniform title (630) subject heading *Handobukku\$xWaterfowl management*. What happened here makes sense when examining LC authority record "nr 93048295," where there is a 430 reference from *Handbook* to *Handobukku*. In all the above cases the headings are structurally good—they just have zero relevance to the catalog records in which they occur. Most catalog users coming across them would not exert the effort to understand *why* they appear in the catalog records, thinking perhaps that the cataloger knew something that they did not.

Unfortunately, once a heading has been mis-linked, a subsequent vendor will find it almost impossible to identify and to fix such headings. They are encountered only by chance or the presence of invalid subfield codes. Manual review authority control is of no value if editors doing the review do not have easy access to the bibliographic record from which the unlinked heading was extracted. A large percentage of headings that remain unlinked following machine processing require that the editor examine both the source bibliographic record and the authority record in search of clues that will determine a link to be made between the two.

Finally, full manual review should not be confused with either "choice of entry" or re-cataloging. Similarly, descriptive cataloging is unaffected by authority control and, while some vendors do offer various record enhancement services that can result in the addition of subjects and added entries to the library's bibliographic record, authority control *per se* is focused on controlled headings already present in the library's source records.

Prior to offering a full manual review authority control cost quote, LTI requires that the library submit its entire database for a no-charge evaluation. If analysis shows that the library's cataloging is not consistent with U.S. national cataloging standards and practices, full manual review is not available as an option.

## Details of LTI's Processing

## Record Load

Regardless of database size, FTP is the standard method to receive and return bibliographic and authority record files. MARC format records can also be received and returned on a variety of magnetic tape formats, but FTP transmission accounts for nearly 100% of all data exchanges.

Data verification checks are made during and immediately after the transfer of records to LTI's computers to ensure that records are properly formatted in the USMARC communications format. Checks are also made to ensure that the record directories only contain numbers, that controlled heading fields do not contain non-MARC characters, etc.

## 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 diacritics and special characters associated with an article, but preceding the first actual filing character.

If the fixed field 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, composed from any languages, and sets the non-filing indicator to its proper value or to 0 as appropriate.

Because title fields other than 245 (e.g., X30, 240) do not necessarily correspond with the fixed field language code, LTI's program compares non-245 field initial text against the table of common articles and sets the non-filing indicator to its proper value or to 0 as appropriate.

As a rule, in controlled fields LC practice is simply to delete initial articles in the formulation of headings. LC authority records do not contain non-filing indicators in *IXX* fields, e.g., the geographic heading is "Dalles (Or.)"—not "The Dalles (Or.)" (n 82036146).

Automated non-filing indicator fix programs are sometimes unable to distinguish correctly between when a leading letter (e.g., *A*) or word (e.g., *Lo*) 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 to 0 based on an analysis of the second word of the title are listed below:

*A is for apple*  
*A B C of power brakes*  
*El Salvador 1932*  
*A la orilla del viento*  
*A la recherche du temps perdu*  
*Das ist mir lieb*  
*Un de Baumugnes, etc.*

Failure to take into account punctuation and other symbols found at the beginning of titles [e.g., ,/:/"/(\$/\*] can also cause problems. While it is still theoretically possible for a non-filing indicator set correctly in the source record to be re-set to an incorrect value, the chances of that happening are remote. LTI in-house studies show that for every non-filing indicator set incorrectly, roughly one hundred incorrect indicators will be set properly.

For those that insist on preserving their original non-filing indicators in the 245 title field, LTI can preserve them on request. This is not an option in controlled title fields where the removal of initial articles is controlled by explicit LC authorized headings.

LTI creates an ASCII text report showing every change made to a title non-filing indicator. This report is presented in multiple columns, showing the before and after settings along with the relevant title text. For libraries profiled to correct non-filing indicators, the report provides reassurance that its non-filing indicators are being correctly set. It is just as useful for libraries that have chosen not to have LTI set their non-filing indicators because it lists the changes that should have been made.

Optionally, the first indicator in the 245 field can be set based on the presence or absence of a *IXX* field. In other words, the first indicator is checked and if necessary changed to 0 when there is no *IXX* field present in the record, or to 1 if there is a *IXX* field. Most libraries select this option based on LC rule interpretation [*LCRI 21.30J*] that titles proper should always be traced.

## *AuthPrep and MARC Update Processing*

Authority control at LTI begins with a generalized database clean-up program, *AuthPrep*, whose purpose is to increase both the probability of catalog record heading matches against authority headings and to upgrade unauthorized headings to current forms.

*AuthPrep* normalizes headings to correct for a variety of typographical and punctuation errors. These include elimination of leading and trailing blank spaces, compression of multiple blank spaces to a single space in *1XX/4XX/6XX/7XX/8XX* fields, and deletion of blank spaces on either side of subfield codes.

*AuthPrep* makes many types of changes to headings, some of which correct the omission or improper assignment of content designators while others correct punctuation or bring headings into compliance with *AACR2*. Examples of changes at the subfield code level include additions (such as inserting, where appropriate, *\$f*, *\$l*, *\$s*, and *\$k* in title fields, *\$c* and *\$d* in personal names, *\$b* in corporate names, and *\$v* in series); conversions (such as changing *\$b* to *\$n* in conference names, correcting errors caused by the omission or improper assignment of *\$c*, *\$d*, and *\$e* in personal names); and deletions (optional removal of subfields *\$e*, and *\$4* from name headings). Unless instructed otherwise, obsolete OCLC-supplied *AACR2* verification (*\$w*) coding is always deleted.

Complex Bible, music, and other uniform title headings are parsed and updated to conform with current cataloging rules. Leading non-filing articles are removed from uniform titles and title portions of author/title headings and unnecessary parentheses and brackets are deleted from name headings. If not already present, brackets are added surrounding GMDs in controlled title fields. *AuthPrep* even updates certain non-controlled heading fields—e.g., the OCLC control number preface is changed from *ocl7* to *ocm0*, *301/ 305* fields are retagged as *300*.

Obsolete subject subdivisions such as *Addresses*, *essays*, and *lectures* and *Collected works* are eliminated. As part of this preliminary clean-up the letters *l* and *O* are converted to *l* and *o* respectively in date subfields, and a check is made to ensure that subfield code *\$d* precedes dates in personal names. *1XX* fields with a second indicator of '1' generate the appropriate *6XX* subject heading. All *1XX* second indicators are set to blank. Format integration changes are also made at this time.

To achieve consistency with the current MARC standards, catalog records are modernized to reflect the

latest *MARC 21 Format for Bibliographic Data* tagging and coding conventions. LTI's MARC Update service is an important clean-up step for records created prior to 1987. An exhaustive table of authority control pre-processing fixes is found in the document *LTI MARC Update Changes*.

Changes include deletion of obsolete fields and subfields; conversion of obsolete tags, indicators, and subfields to current usage; and conversion of outdated fixed-field element codes. Examples of MARC Update changes are deletion of *039* fields, conversion of subfields *\$d* and *\$e* in *245*, *246*, and *247* fields of the serials format to the currently defined codes *\$n* and *\$p*, conversion of the obsolete *705* and *715* fields to *700* and *710* respectively. Several MARC Update options are offered, including deletion of LC Children's, Sears, genre headings, or NLM MeSH subject headings. Unless instructed otherwise, LTI deletes the obsolete OCLC-generated *87X* fields.

## *Other Pre-processing Routines*

Changes to cataloging rules due to *AACR2* require special processing on series, conference names, and titles prior to authority record linkage.

Series statements (*400/410/411/440*) are retagged as *490* fields and assigned a first indicator of *1*. An *AACR2* series entry is then added in the appropriate *800/810/811/830* field. Removal of initial articles, capitalization changes, and adjustment of filing indicators is frequently necessary as part of this processing. To illustrate, the traced title series:

*440 4\$aThe series in computer science*

is retagged

*490 1\$aThe series in computer science*

and an *AACR2* series field is added to the record

*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 catalog record's *1XX* field. In addition, *840* series fields are tagged as *830* fields.

In conference name headings, the order and punctuation of data elements in subfields *\$b*, *\$c*, *\$d*, and *\$n* are changed to conform to *AACR2*. In the *111/611/711/811* fields, the obsolete subfield *\$b* is converted to subfield *\$n* and the number, place, and date are reformulated in parentheses with proper

subfield coding and punctuation. To illustrate, the conference heading:

*111 20\$aInternational Conference on Elizabethan Theatre, \$b1st, \$cUniversity of Waterloo, \$d1969.*

is converted to:

*111 2 \$aInternational Conference on Elizabethan Theatre \$n(1st :\$d1969 :\$cUniversity of Waterloo)*

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. Brackets are removed temporarily from media qualifiers in subfield *\$h*. Media designators are checked to ensure they do not prevent an authority record link, e.g., *\$hPhonorecord* and *\$hPhonodisc* are changed to *\$hSound recording*, *\$hMachine-readable data file* and *\$hComputer file* to *\$hElectronic resource*. In accordance with current LC and OCLC practice, following authority control: (1) brackets are added surrounding GMDs (*\$h*) in 245, 246, and 740 fields; and 2) GMDs (*\$h*) in these same fields are corrected to current AACR2 forms.

### *Extraction of Controlled Headings*

Following *AuthPrep* processing, headings eligible for authority control are extracted from catalog records and a unique, sequentially assigned number is added to the end of controlled heading fields. This number provides a link to permit reinsertion of the authority controlled heading into the catalog record later in the processing cycle.

Table I 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 catalog record headings are matched against all appropriate subfields in LC authority record headings. Subfield *\$v* in 4XX and 8XX fields are validated and corrected wherever possible, e.g., when volume designation information 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 referenced authority record. For LC subject authority control, only subject fields (6XX) with a second indicator of 0 (i.e., LC subject headings) are validated. A blank second indicator is treated as if it were 0. LTI offers optional authority control of LC Children's subjects, NLM's MeSH subject headings, Sears subject headings, and some genre headings. LTI ensures that the most

up-to-date authority records are used during processing by loading weekly authority files received via FTP from the Library of Congress.

### *Authority Record Matching*

After preliminary processing is completed, compressed "match keys" from catalog record headings are compared against an index of match keys extracted from 1XX/4XX fields in LC and LTI authority records. Links to LC authority records take precedence over links to LTI authority records.

To increase authority record links, certain tag variations in catalog 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 because a Library of Congress authority record does not exist 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. Almost no name headings and only a fraction of the thousands 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, catalog record headings are submitted to repeated manipulations and checks by computer.

### *Pseudonym Preservation*

In some situations libraries may want to override links to LC authority records. For example, the library may want titles written under the pseudonyms *Peter Curtis*

and *Juliet Astley* to appear in the catalog record under these names rather than *Norah Lofts*.

Public libraries having extensive collections of older fiction in particular may prefer title page cataloging because fiction is generally shelved by the author's name as it appears on the book's title page. LC's decision to adopt a revision to AACR2 Rule 22.2C after 1987 means that a separate authority record is issued for each pseudonym used by an author. Pseudonyms found in post-1987 authority records, as well as authority records corrected after that time, do not appear in LTI's preserved pseudonym list since they are already preserved without any intervention by LTI.

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
111	\$a q e g k t p l f
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	<i>1<sup>st</sup> ind. 0 - optional</i>
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
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
711	\$a q e g k t p l f s h
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

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

For libraries wishing to retain an author's name as it appears on the book's title page, LTI maintains a file of pseudonymous authors. At this point there are fewer than 80 names on this list and it is quite possible that separate authority records will never be distributed for most of these authors.

Inclusion on this list prevents the library's heading from being linked to an LC name authority record. Main

entry (*IXX*) headings in records are checked against the pseudonym file during the final linking process. Headings on LTI's preserved pseudonyms list are not converted to the "real" name present in pre-1988 LC authority records. LTI software recognizes most variations of the pseudonyms and corrects those variations to the standard form of the pseudonym.

### *Name Authority Matching*

Initially, 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, corporate, and conference 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. For many years, libraries that deviated from Library of Congress practice and added known death dates to headings for famous people had either to accept the replacement with an open date or to perform ongoing local maintenance to such headings. With the adoption of *LCRI 22.17*, authority records for several hundred well-known names had death dates added in a project that spanned most of 2006, with new additions weekly. Libraries will no longer have to re-edit authorized bibliographic records to add this information, nor be concerned with the loss of this data in the first place.

### *Title Authority Matching*

Title portions of name/title headings (*X00/X10/X11* fields containing *\$t* or *\$k*) and uniform 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.

### *Special Series Processing*

While untraced series (i.e., those tagged as 490 with a first indicator of '0') have generally been excluded from standard authority control processing, LC's decision effective June 1, 2006 to discontinue providing controlled access to series in bibliographic records and to cease creation of series authority records makes it important that libraries have a mechanism to continue to treat series as controlled headings. The number of series authority records has grown less rapidly since LC ceased to create them. However, the backfile of existing LC series authority records, as well as new national-level records being created by NACO libraries, continues to be available for authorization of series headings. 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.

Libraries are encouraged to take advantage of LTI's no-charge "untraced series" option that allows for the authorization of 490 0 fields. Using this option, 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 heading 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: (1) delete the LTI-created 8XX field and change the 490 first indicator back to '0'; or, (2) 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. Should an authority record later be distributed, it will be provided to the library when the heading is authorized during a continuing [AEX or AUP] authority control run. 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.

Given the 2008 MARBI decision to make the 440 field obsolete and to use instead a 490/8XX combination, LTI also suggests that libraries no longer restore unchanged 830 fields back to 440. The decision to no

longer use 440 as a traced title series clarifies the line between transcription and access points. It also redefines the 490 first indicator value '1' to "Series traced in 8XX field." Libraries using LTI's AUP service can easily convert all obsolete 440 fields to the new usage style.

### *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 modernized. For example:

*Gt. Brit.* becomes *Great Britain*.

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

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

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

Headings are then matched against the LC subject authority file. Because most subject subdivisions are not represented as separate LC authority records, LTI maintains tables of "free floating" topical, form, chronological, and geographic subject subdivisions.

Pattern headings 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.

### *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\$yCatalogs.

and the heading:

650 0\$aGermans in Waterloo Co., Ontario.  
is changed to  
650 0\$aGermans\$zOntario\$zWaterloo County.

Through limited review processing, LTI is presently linking on average between 96% and 99% of subject headings for academic and public library databases.

Obsolete LC 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*” were followed by a “*Biography*” subdivision, the new heading would be changed to “*Criminals.*” LTI’s handling of splits is not affected by whether the job is limited or full review.

### *Form Subdivisions*

In early 1999 the Library of Congress began to issue authority records with form subdivisions in \$v, heretofore coded \$x. LTI analyzed LC’s form subdivision usage patterns and created processing guidelines for the assignment of \$v. Briefly, when a heading contains a subdivision that is valid for use as a form subdivision and it is the last subdivision in the heading, it will be coded as \$v, except in certain special cases,

e.g., when there is a full link to an LC authority record containing the subdivision in \$x. Conversely, if the subdivision can be \$v, but is not the last subfield in the heading, it will not be changed to \$v. Exceptions to the latter rule are if it has been identified as a special case which will take a \$x following a \$v, e.g., \$v*Dictionaries*\$x[*language*] or when it is permissible to follow \$v with another \$v.

### *Role of Added Cross-References*

An important factor to LTI’s ability to offer a guaranteed link rate is a supplemental file of 4 million added cross-references. Name headings constitute 89% of the cross-references; the remaining 11% are subject cross-references. Each reference refers from the form of the heading used in the catalog 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. Because not all cross-references can be applied to every database, editor-created cross-references undergo a second review prior to being added to LTI’s cross-reference file.

Added cross-references serve two purposes. First, they correct the heading 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, thereby reducing the number of headings that require manual review and allowing LTI to focus on the research and establishment of headings that are genuinely new. 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.

When the complete catalog record heading cannot be validated or linked to an authority record, an attempt is made to link portions of the heading. Through limited review processing, LTI is presently linking 90% to 96% of name, title, and series headings, either fully or partially to a nationally distributed authority record, for most academic and public library databases.

### *LTI Authority Record Linking*

LTI maintains its own file of name and subject authority record headings established in accord with AACR2 and LC cataloging practice. Presently, there are 2.3 million LTI authority records. These authority records are created from validated but unlinked headings that have appeared in library databases. To qualify as an LTI authority record heading, the heading must meet three criteria: 1) it has been searched thoroughly but not found in the LC authority files, 2) it is coded properly (tags, indicators, and subfield codes), and 3) it conforms to national cataloging standards. If an LC authority record is later distributed for the heading, the LTI authority record is deleted. Library headings are matched against LTI authority record headings only after they have failed to link to an LC authority record heading. 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 heading in the library's catalog record is identical to the LTI authority record heading.

A 2008 analysis of a one million record database showed that 3.2% of the controlled headings were validated against LTI authority records. Six out of seven of these headings were **not** changed by LTI—i.e., the headings appeared in the authorized catalog records exactly as they appeared in the library's source records. Where changes did occur, in 1,300 [0.0003%] out of 4.4 million headings, the variations were primarily attributed to content designation fixes and minor differences in the headings. This includes insertion of *\$d* (date) and *\$q* (fullness of name) in personal names, *\$b* in corporate names, and tagging errors in subjects. Again, the sole purpose of LTI authority records is not, in the absence of a nationally distributed authority record, to change a library's headings; but rather to 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 more attention on the unlinked headings that can most benefit from manual review. Libraries that plan to examine the unlinked headings report also benefit from exclusion of these valid headings from that report.

### *Review of Unlinked Headings*

After all possible machine matches have been made, LTI editors selectively search targeted unlinked headings in the LC name and/or subject authority file. Sometimes an editor needs to view the bibliographic record(s) from which headings were extracted to determine if they represent the same person or entity. Editors have access to a fully indexed version of the

library's database and can display catalog 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. Where the same heading is linked to obviously different bodies, as sometimes occurs with acronyms, the editors temporarily force the correct link for that specific database.

Unlike authority vendors that have a policy of treating the same name heading with and without dates as distinct headings, in an LTI authority controlled database one will almost never find variant forms of the same heading. In backfile processing LTI editors merge all variant forms of a heading whenever they are determined to be the same person or entity. Failure on the part of the authority control vendor to merge such headings defeats the purpose of authority control.

## **Authority Record Distribution**

Most libraries request files containing LC authority records that have linked to the library's database during processing. Authority records prepared by LTI are written in accord with the *MARC 21 Format for Authority Data*.

Separate authority records are extracted for each level of a multi-level heading. For example, three authority records would be extracted from the heading:

*English poetry\$yOld English, ca. 450-1100\$xHistory and criticism.*  
*English poetry*  
*English poetry\$yOld English, ca. 450-1100*  
*English poetry\$yOld English, ca. 450-1100\$x History and criticism.*

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. Above two million bibliographic records, the ratio stabilizes at about one authority record per two bibliographic records. Linked LC authority records are deduped so that the same record appears only once. LC name authority records are written to a different file than subject authority records. Name authority files contain only LC name authority records. Depending on

the library's local system, subject authority files contain both topical and geographic subjects, as well as name authority records for names and uniform titles that have been used as subjects. About 30% of a library's subject authority records are likely to be composed of LC name authority records that have been used as subjects (600/610/611/630 fields).

The default option is for a library to receive all LC authority records referenced in its bibliographic records. Upon request, libraries may instruct LTI to return only those authority records that contain a *see* (4XX field) or *see also* (5XX field) reference. About two-thirds of LC authority records contain a cross-reference.

In May 2007 LC announced its intention to begin creating and distributing "Subject Authority Records for Validation Purposes." The reasoning behind the project was to reduce the cost of subject cataloging by providing more subject string authority records for popular and frequently-assigned headings. Presumably, they would also make it easier for local systems to validate LCSH subjects. To date these authority records are limited to: (a) 651 subject heading fields for country names followed by free-floating subdivisions; and, (b) subdivisions found in *Free-floating Subdivisions: an Alphabetical Index* that appear after topical and geographic headings. They are barebones records containing no references and identifiable by the presence of a 667 field [nonpublic general note] that carries the text "Record generated for validation purposes." While this type of authority record may be useful in an ILS that requires an authority record for every controlled heading or for libraries that lack a comprehensive authority service such as **AEX** or **AUP**, some LTI clients have asked that they be excluded from their subject-related authority records. As their number grows and their value is questionable to some, LTI has responded by adding an option to not return subject authority records containing 667 with the validation purposes note.

Similarly, because a number of libraries have reported problems with the display of non-roman 4XX fields within their local systems, a profile option now exists for **AEX** and **AUP** users to elect not to receive updated authority records containing 4XX references in non-Latin scripts. LC began distribution of these records in mid-2008 and presently [May 2009] there are over one-half million of them.

### *Provisional Authority Records*

Upon request, LTI will create provisional authority records for unlinked and partially linked headings. After manual review and edit, these records can be upgraded by the library to full authority record status using the local system's authority module. Since most local systems having an authority module can automatically generate authority records for unlinked headings as the records are loaded into the system, few customers have a need for vendor-supplied provisional authority records. LTI's default is not to create provisional authority records.

### *"Deblinding" Authority Records*

LC authority records are structured so *see* references never point to a heading that is not found in the library's database. In other words, *see* references are already deblinded and do not require further processing on the part of the authority control vendor.

Deblinding can also refer to the removal of certain *see also* references. Deblinding at this level prevents 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*. LTI recommends that such references be retained 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.

The following summarizes a posting to the LTI Users 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."

### *LTI Authority Reports*

At the conclusion of an authority control job, libraries receive summary statistics showing authority record link rates for name, series, and subject headings. LTI's *Final Link Report* details the number and percentage of name, series, and subject headings that have linked fully, linked partially, or not linked to either an LC or LTI authority record. Included in the report is a table showing the frequency of each controlled heading field, the number and percentage of headings both validated and not validated against an LC authority record, and the number and percentage of headings changed following the final link procedure.

Every authority control job, be it batch or the ongoing **Authority Express (AEX)** service, generates an ASCII text file of unlinked headings. These files can be loaded into a word processor for review purposes. In reviewing the unlinked heading reports, one should keep in mind that in the vast majority of cases, the only thing "wrong" with these headings is that no LC or LTI authority record exists for the heading. Most clients do not systematically search unlinked headings in nationally distributed authority files.

While LTI does not authorize alternate graphic representation data in 880 fields, there are two reports available for batch and continuing authority control services that make it easier for libraries to update 880 field data. These reports are produced at no charge but they must be specifically requested by the library.

The first lists controlled headings with \$6 that were *changed* during authorization. Each change is shown when it happens but not repeated in subsequent processing—i.e., continuing authority control service reports. This ASCII text report may be reviewed by library staff to make needed edits to the corresponding 880 fields. Format of data shown on the changed 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. Note that the original data is displayed as raw ASCII.

A second report lists errors in 880 fields, such as bad parsing of \$6 data or the field linked to the 880 is not present in the catalog record. For **AUP** users receiving changed bibliographic records this report continues to list errors until they no longer exist in the copy of the database maintained at LTI—i.e., unless the "base file" is replaced, even if the library makes the needed corrections locally. Format of data 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 (again raw ASCII), and a brief explanation of the error found.

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

LTI's approach to NLM MeSH, LC Children's, Genre/Form headings and Sears 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 eligible 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.

### *LC Children's Subject Headings*

In 1996 LC started distribution of Children's authority records based on Library of Congress usage. Because LC Children's headings can contain adult headings considered appropriate for children's use, 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 the 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. These same processing options are available for Sears.

When Children's headings are to be converted to *LCSH*, during a pre-processing routine, second indicator codes of '1' in 6XX fields are changed

globally to '0' and the subject subdivisions *Juvenile literature*, *Juvenile fiction*, etc. are added to the converted heading. Following authority control, resulting duplicate headings are removed from records. For example, assume the library's source catalog record has the following two subject headings:

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

Following authority control two identical *LCSH* headings will exist—i.e., *\$aFishes\$vJuvenile literature*. One of these headings will be deleted because a final LTI check eliminates controlled headings that are truly identical after authority control.

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

Authority control can also be exerted over Sears subject headings (i.e., 6XX fields having a second indicator of '7' [MARC 21] or '8' [OCLC]). Sears headings are distributed by H.W. Wilson and are updated as new editions appear. For libraries that subscribe to Sears subject headings authority control: Note that the 19th edition (2007) of the Sears authority records allows the use of "Juvenile" subdivisions—e.g., "\$xJuvenile literature," "\$xJuvenile fiction," etc. These subdivisions are now validated as legitimate floats in Sears headings instead of being altered or deleted.

### Genre/Form Subject Headings

Genre headings in 655 fields are authorized using LTI's Genre Heading processing option. At this time the service is limited to Library of Congress and NLM MeSH subject headings, as well as those found in the publication, *Guidelines on Subject Access to Individual Works of Fiction, Drama, Etc.* Headings which include \$2lcsch, \$2mesh, or \$2gsafd are pulled from bibliographic records for processing and compared against 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 the subfield coding is absent, headings will still be pulled if the code is present following a space, e.g. *\$aFiction lcsch* will be pulled and \$2 inserted to create, *\$aFiction\$2lcsch*. Typos in the codes are also corrected: *\$2lsch* is corrected to *\$2lcsch*.

The complete file of *GSAFD* authority records is available for free download on the Internet. LTI may authorize additional genre/form controlled vocabularies in the future when demand and authority record availability warrant it.

If genre headings are authorized, they are returned in either the old or new style tagging. The old style uses a second indicator of 7, with a \$2 subfield identifying the controlled vocabulary. *GSAFD* headings are always coded in this fashion. The new style alternative returns genre headings in 655 fields with the second indicator identifying whether the heading is *LCSH* or MeSH.

Libraries have the option of converting *GSAFD* headings to *LCSH*. They also have the option of converting genre/form headings that may have been coded as topical subjects (650) to 655. The latter conversion occurs only with stand-alone *\$a* genre/form headings. If a *LCSH* topical heading includes another subfield—such as \$v, \$x, \$y, or \$z—the heading is not considered a form heading. Thus far relatively few LC genre/form authority records [tagged 155] have been distributed and these are limited to motion pictures, television programs and videos. When returned to libraries these authority records are placed in a separate file; they are not included with subject-related authority records.

## Scheduling

Processing time for limited review authority control is three weeks for databases containing 300,000 or fewer bibliographic records, four weeks for databases up to 800,000 records, and five weeks for databases of one million records. Scheduling for databases larger than one million records will be provided on request.

When offered, processing time for full review authority control for a database of 100,000 records is six weeks. As a planning guideline, allow one additional week for each 75,000 records, e.g., estimated processing time for a database of 250,000 records is eight weeks. Full review projects should be scheduled at least two months in advance of when the library's records are to be received at LTI.

Library databases are rarely available for processing at the time initially planned. Therefore, for scheduling purposes, the processing time for a job "starts" after two events occur: 1) receipt of a readable version of all the customer's machine-readable records in MARC format; and, 2) receipt of a completed *Authority Control Work Specification Profile (WSP)*. The *WSP*

should be completed and submitted online. Authority control options designated by the customer in the *WSP*, along with any attachments describing special instructions, becomes the official document describing what processing is to be performed. The *WSP* takes precedence over telephone and email communications.

### *While Your Records Are Being Authorized*

During the period in which the library's bibliographic records are being authorized, staff should understand that no changes are to be made to catalog records. This is not the time to enhance records by adding local notes, secondary added entries, subjects, etc. When the post-processed bibliographic records are returned after authorization, any changes made to records will be lost as part of the "overlay" process. Also, this is not the time to conduct a "weeding" project—that should be done in advance of when the records are exported.

There is no need for libraries to export item records as item fields (e.g., 949 fields) when the only database service needed is authority control. This means that, while bibliographic record data is "frozen" during the time in which the records are being authorized, the library may continue to make changes to item records. For example, library staff can make changes to call numbers, location codes, barcode numbers, copy and volume information, or any other data contained in the item records. Based on the bibliographic record's local system control number, item records are going to be re-linked to catalog records when the authorized bibliographic records are loaded back into the database following authorization.

It follows that a library can continue to create new item records, as well as delete existing item records, while its bibliographic records are being authorized. Similarly, a library can continue to add new bibliographic records, along with associated item records, while the database is processed through authority control. The library will want to keep track of the date its bibliographic records were exported so that catalog records added after that date can be extracted for whatever continuing authority control service the library may wish to use.

There are certain situations when it is necessary for a library to export item records (i.e., formatted as item fields) along with its bibliographic records. The most common reason is that the library is migrating to another local system and it will be necessary to re-build the item fields as item records within the new local system. Occasionally a library will want to dedupe and authorize its bibliographic records at the same time.

In this case item fields must reside in the exported bibliographic records or else 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 the database vendor.

### *MARC-8 vs. UNICODE*

LTI recommends that libraries use their ILS vendor's utility program to convert exported records to MARC-8 before sending files for authorization. However, we do offer an option to work directly with Unicode [UTF-8] encoded catalog records. At this time we return LC authority records only in MARC-8 and will continue that practice until LC begins to distribute authority records in Unicode. Internally, prior to matching headings against LC authority records, we first convert controlled headings from Unicode to MARC-8. Libraries planning to submit Unicode records should so advise LTI so that profile options are properly set.

### *Serving Two Masters*

By definition authority control demands a single "authority." If the library's local system has its own "authority module" that flips headings based on 4XX fields in authority records, one is inviting Chaos to rule supreme over headings. On a regular basis LTI users report strange things when they "turn on" their ILS authority control module. A primitive authority control module, even when it was offered as "free" or at a reduced price, is no bargain and will result in corrupted headings.

As with errors introduced by second-rate authority control vendors, detecting corrupted headings triggered by the library's local system is serendipitous. For example the library may come across an inordinate number of headings tagged as subjects but that are in fact uniform titles or series. While a series may on rare occasions be used as a subject, it is most uncommon, and it is never permissible along with the presence of form (\$v), topical (\$x), period (\$y), or geographic (\$z) subdivisions. Examples of mangled headings in topical subject [650 fields] caused by the library's authority module are listed below.

*\$aAmerican university studies.\$nSeries II,\$pRomance languages\$xGrammar, Comparative\$xRomanian.*

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

When one examines the authority record for "*American university studies. Series II. Romance languages*," one finds a 430 x-reference for "*Romance languages*." Pretty clearly at one time the *LCSH* heading was "*Romance languages*," but a simplistic automated routine "flipped" that *LCSH* heading to the series title "*\$aAmerican university studies.\$nSeries II,\$pRomance 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 in all databases where the library uses an authority control module made available by its ILS vendor. Mangled headings are not limited to any particular ILS. Regardless of local system, LTI strongly advises that ILS automatic changes to catalog record headings based on authority records be **disabled**. The library has paid for and received the best authority control available to libraries today. Superimposing a redundant level of ILS "authority control" serves only to undermine the advanced algorithms and tables that were used to authorize the library's database.

The culprit is that local system authority control modules rely on the presence of the "old" heading as a 4XX field in an authority record. Not only does LC frequently not carry over the "old" heading, even if it did there may be other issues in linking the "old" heading because of variations in capitalization, spacing, or punctuation occurring, in either catalog record headings or in LC authority records. We see and report dozens of such errors in LC records each week.

As illustrated in the cited examples, there are many instances in which a link should **not** be made. While LTI processing results in the highest link rate of any authority control vendor, avoiding bad links demands even more vigilance. Presently, we block over 163,000 headings in authority records (mostly 4XX fields) from linking to any catalog record heading because of the likelihood of introducing an error in a library's file. Other heading changes are limited to only certain types of linkages—i.e., there must be additional data present before the link is allowed. Software that allows wholesale revisions of 650 subject headings based on a 430 in a series authority record is not something a library wants to be in control of its headings.

LTI's substantially more refined and tested routines not only prevent bad changes from occurring, they also make thousands of needed revisions that are not possible with a simple routine that flips headings if a 4XX field with matching text is found in any authority

record. Internal studies have shown that reliance on the LC record alone misses one-half to two-thirds of the needed changes. In short, when judiciously used, your library's ILS authority control module is perhaps better than none at all, but is certainly no replacement for a comprehensive batch and continuing authorization service.

## Continuing Authority Control Services

Bibliographic control in library catalogs is neither easily achieved nor easily maintained. LTI's batch authority control processing exerts control of headings at the time the database receives authority control. However, changes to authority records, the distribution of new authority records, and the addition of headings from new catalog records all affect the integrity of the catalog. The burden of monitoring changes to authorized headings and controlling new headings from current catalog records can be automated, either fully or partially, using continuing authority control services.

Specifically, libraries are confronted with two different problems. The first problem is caused by the addition of new catalog records to the database. Headings in these bibliographic records need to be authorized and, if necessary, new authority records added to the local system. LTI's **Authority Express (AEX)** service provides ongoing authority control for new catalog records.

The second problem, retrospective in nature, is caused by the weekly distribution of new and revised LC name and subject authority records. Libraries need to be informed of updates so that headings can be changed appropriately, and the corresponding new or revised authority records added to their databases. LTI's **Authority Update Processing (AUP)** notifies customers of the issuance of new, changed, and deleted authority records distributed following batch authority control. Unlike **AEX**, where new records are transmitted to LTI for authorization, the **AUP** service involves LTI reporting to the library about records previously processed.

While there is no requirement to use either the **AEX** or **AUP** continuing service, for libraries that do use them, there is never a need to do a batch re-authorization of the entire database and the library's access points remain in synch with national cataloging standards.

***Authority Express***

**Authority Express (AEX)** was developed in response to requests from libraries seeking an easy way to authorize new bibliographic records. Because the service is totally automated, turn-around time is rapid and scheduling is not an issue. **AEX** records receive the same processing as limited review, with the exception that there is no provision for editor review. Files up to 10,000 records are typically finished within one-half hour after they have been received at LTI. Larger files take proportionately longer. Any library having the capability to FTP a file on the Internet can use **AEX**. Although most users have had batch processing done by LTI, use is open to all libraries.

**AEX** users are provided a web interface on LTI's website—[www.LibraryTech.com](http://www.LibraryTech.com). Popular browser software packages are supported to send and retrieve bibliographic and authority records.

Libraries are assigned a secure user-name and password to access the **AEX** server. Newly cataloged records are then sent to LTI via FTP for processing. File naming conventions permit the ready identification of pre- and post-processed bibliographic record, authority record, and report files. Files returned to the library include: a) a file of output MARC catalog records, now fully authorized; b) an ASCII text report showing the results of the processing—i.e., the Final Link Report, c) an ASCII text file containing unlinked headings; and d) file(s) of MARC format new LC name and subject authority records. LC authority records from **AEX** runs are deduped automatically against authority records previously provided to the library, whether through previous batch processing or earlier **AEX** runs. Only the new authority records are returned to the library.

Another **AEX** feature is that new catalog records are added to LTI's files of all the library's previously processed bibliographic records so that files are kept current and complete for **Authority Update Processing**. While libraries using **AEX** will probably also want to use **AUP**, there is no requirement to do so.

**AEX** has recently been enhanced to allow libraries to make batch additions/deletions to LTI's copy of the library's authority files. Users submit text files of LC control numbers [prefaced with "n" or "s"] or of NLM MeSH control numbers [prefaced with "D"] so that LTI's tracking files can be kept in synch with those of the library. Separate files must be sent for additions and deletions.

### ***Authority Update Processing (AUP)***

As noted above, if a library wants to keep its headings consistent with LC, it needs to be informed of new and revised LC authority records that affect previously authorized headings.

Changes may be made to a heading due to:

- Changes/corrections in LC/LTI authority records.
- Matches to new LC/LTI authority records.
- Changes in LC cataloging practice incorporated into LTI software.
- Changes in "floating" headings used in LC subject headings.
- Improvements in LTI's software for automatic heading matching and correction.

Each week LC distributes about 7,600 name authority records and 400 subject authority records. Some subset of them affect headings in almost every library's database. The authority record may be "new" to a given library's database—i.e., either the authority record did not exist at the time the batch authority control was initially run or for some reason was not extracted—or it may represent a correction or deletion. About 30% of the LC weekly name authority records are corrected or deleted headings and about 50% of the weekly subject authority records are revised or deleted headings.

Libraries using **AUP** have the option of being notified of new and revised LC authority records monthly, quarterly, semi-annually, or annually. Following the run, via FTP, subscribers receive: a) a file of bibliographic records in which one or more headings have changed as a result of the processing; b) an ASCII text report of headings affected by new and changed LC and LTI authority records; c) an ASCII text report of any new, revised, or deleted LC authority records that linked to the library's headings; d) a file of deleted LC authority records in MARC format; and, e) MARC format files of new and revised nationally distributed authority records.

The great advantage of receiving the changed record file is that library staff do not have the burden of implementing the changes. Instead, libraries "overlay" the existing bibliographic record with a replacement record containing the corrected heading(s). For those libraries that prefer making the heading changes locally, there is no requirement to load the MARC formatted changed bibliographic records. Files of new and revised LC name and subject authority records, custom to each library, are available for pick-up from LTI's FTP server.

To use **AUP** each of the library's bibliographic records must have a unique control number so that replacement

records containing updated headings have a single match point for overlay. The control number for overlaying need not be from the *001* field. Instead, it could be a bibliographic record number in a *9XX* field, a local system control number (*035* field), or some other unique control number present in the record.

If a record has been enhanced by the addition of local notes, secondary added entries, or subject headings following the return of the record from batch authority control, those notes or headings will not appear in the replacement overlay record, unless the edited record has also been sent to LTI using the **Authority Express** service or a new "base" file has been sent. There is also no reporting feature for a library to notify LTI of catalog records it has deleted. For collections that are actively weeded, libraries have the option of submitting new "base" files of their bibliographic records. Although there is a fee for submitting a new "base file" of bibliographic records, some libraries do so with each **AUP** processing cycle.

Not all authority control update services are the same. Of the several reasons listed above why a heading may change, only changes or corrections in LC authority records are reported by services provided by local system vendors and other authority control providers. Other notification services are based on the principle that if an LC authority record from a weekly update file has the same control number as a record in the library's authority files and the *005* transaction date is later, that LC authority record is pulled for the update report. Moreover, LTI's is the only update service that returns bibliographic records. Other vendor services have the more limited scope of notifying the library of changes to authority records.

The important difference between **AUP** and other update notification services is that using LTI's service re-authorizes the library's entire database each processing run. Every heading in the library's database is checked against all current LC and LTI authority records, LC and LTI cross references, and dozens of name and subject heading "fix" tables. **AUP's** more comprehensive reporting increases consistency between library headings and authorized headings.

Update services that track only a static file of LC control numbers from authority records that existed at the time the job was processed are unable to notify libraries of newly distributed LC authority records—i.e., those not available at the time the initial batch processing was completed. While reducing significantly the work required of the notification service provider, this approach disregards new LC authority records—70% of the weekly LC name

authority records and 50% of the subject authority records.

### ***AUP Reports***

Two ASCII text file reports are made available via FTP. The first identifies headings that have changed since the last processing run; the second lists changes in authority records since the last processing run.

In the bibliographic heading report file, headings are first arranged by type (i.e., name/title, series, subject) and then subarranged by categories, e.g., major changes, changes in geographic subdivision, form subdivisions (i.e., changed from \$*x* to \$*v*), indicator changes, etc. The headings change report begins with a summary showing how many headings were in the file processed, how many changed, how many did not change and gives counts of a number of specific types of processing, e.g. subjects where the only change was \$*x* to \$*v*.

A second **AUP** report lists any new, revised, or deleted LC authority record that linked to a heading since the last processing of the library's headings. Unlike the changed heading report described above, this report is devoted exclusively to Library of Congress authority records.

Reports from the first processing run of a library's file may be substantially larger than subsequent reports since they include all headings that changed between the initial processing and the current processing run. Similarly, reports from an annual run will be lengthier than monthly reports on the same database. Because LTI updates its file of the library's controlled headings after each processing period, the library is only alerted once of a heading change. Within the local system, the library can elect to make or not make the reported changes.

Upon request, **AUP** reports can be generated for alternate graphic representation data found in *880* fields. See earlier discussion under "LTI Authority Reports" for information about *880* field reports.

### ***AUP For Databases Not Authorized by LTI***

For a library to use the **AUP** services described above, LTI must have performed authority control on the database. For libraries that had their authority control performed by a vendor other than LTI, and that want to avoid the expense of re-authorizing their records, LTI offers a special **AUP** variation. In most respects this service is similar to **AUP**, except that updates are

reported based on the library's base file of LC authority records, as opposed to bibliographic records.

As input to this service the library makes available to LTI via FTP its existing files of nationally distributed authority records. On a scheduled basis (monthly, quarterly, semi-annually, or annually), the authority record control number and transaction date (*001* and *005* fields respectively) are checked against LTI's up-to-date copies of nationally distributed authority records. Nationally distributed authority records that match on control number but have different transaction dates are pulled and made available for FTP retrieval by the library. Note that all authority records in the library's base files must have a valid *005* field.

In practice, most libraries using this special **AUP** variation have also submitted large batch files of bibliographic records for authority control and use **AEX** for authorizing new bibliographic records.

## RFP's and Tests

Of the several hundred authority control projects that LTI performs each year, relatively few involve responding to a *Request for Proposal (RFP)*, *Request for Quotation (RFQ)*, *Request for Information (RFI)*, etc. Creation of an *RFP* by the library, as well as the vendor's response to it, are both time-consuming and expensive to all parties. Too often libraries tend to "borrow" heavily from other library *RFPs* or a suggested set of specifications prepared by an authority control vendor. In some cases LTI is asked to meet specifications that have not been current in years or, even worse, to meet a competitor's limitations that have been cast as "desirable" specifications. Vendors may respond with their own stock responses lifted from their service program literature.

In some ways purchasing authority control services is similar to selecting a local system. Vendors are not able to re-write their software for each new customer based on how that library thinks authority control should or should not be done.

The recommended approach to selecting an authority control vendor is to read carefully each vendor's documentation and then clarify issues and questions via telephone or email. For specific questions, email has the great advantage of allowing the vendor to make a thoughtful response, and gives both parties a written record of the vendor's responses. Of course speaking with the library's local system vendor, as well as other

libraries that have used the authority control vendor's services, always adds valuable insights.

Should a library submit a test database to several vendors and compare the results? Asking vendors to run a test on several thousand sample bibliographic records prior to selecting a vendor can be a useful tool in deciding which vendor to use. Assuming the sample records are of sufficient number and drawn randomly, such a process permits one to make meaningful comparisons of how different vendors performed, as opposed to their marketing talents.

To be able to draw useful conclusions it is important that test records be selected in a random fashion—e.g., every *Nth* record from the database. If the library's systems people are unable to do this easily the library can FTP its entire database and request that the vendor select every *Nth* record. Once test records are extracted staff may want to include special interest records to see how vendors handle problematic headings. Keep in mind that, up to a point, the larger the database, the larger the sample should be. To ask an authority vendor to authorize a small test file of one thousand bibliographic records for a library having a database of one million records is largely a waste of time. Another aspect of sample size is that the library must have the resources to analyze and evaluate the work returned from the various authority control providers.

A complicating issue with giving vendors a small group of records to authorize is that the vendors are not going to treat the test as just another job. Since each vendor would like to win the job, test records are going to be processed with more scrutiny than the records from the entire database will receive. In fact, given a couple weeks to complete the test, a talented cataloger without any resources other than access to the LC authority records, could do a very credible job on a small database. One needs to remember that 70% or more of the headings will probably link to authorized heading without the vendor doing anything. If a library believes a test is necessary prior to awarding a contract, make it a requirement that the vendor complete the processing in as short a time as possible, e.g., certainly no longer than one week after the library's records have been received by vendors. That reduces, but not eliminates, the opportunity of the vendor to devote extraordinary attention to the test database records.

Pre-authority control tests can be useful in other ways. They can determine if: a) the library is able to extract its bibliographic records in MARC format; b) the local system ID or control number appears in every record (this will be the necessary match or overlay point when

the bibliographic records are re-loaded after authority control); c) the library can send and receive files of bibliographic and authority records via FTP; and, d) the necessary load tables are in place that will allow the library to re-load the post-processed bibliographic and authority records back into the library's local system.

This last issue is extremely important since it may come as a surprise to the library that its local system vendor requires purchase of a new "load table" that, for library databases under 50,000 records, can cost more than the authority control processing. Libraries considering an authority control project should begin the process by *first* contacting their local system vendor to find out if there are any special requirements or additional charges applied by the local system vendor.

Caution should be exercised prior to adopting authorization specifications that have been lifted from an authority vendor's model *RFP*. Incorporating vendor-created boilerplate into the library's specifications may seem innocuous but it distracts attention from evaluating specs that are genuinely important, versus those that are fluff and self-serving.

An inane, but always mystifying, specification that turns up regularly in many library *RFPs* is that the vendor should: "Supply only authority records from the LCSH, LCNA and MeSH authority files. "Provisional" records are not acceptable. If vendor cannot limit output of authority records to authoritative sources, it must be so stated." One can only guess what the intent of this requirement is. LTI has been offering authority control services for two decades and, during that period, for LC name and subject authority control, we are not aware of any authority control vendor that has **ever** provided other than LC name and subject authority records. Upon request, as an option, some authority control providers, including LTI, are asked to generate so-called "provisional" authority records—but these by definition do not conflict with LC authority records. They are in fact identical to what the library's local system would create—i.e., a skeleton authority record containing only the *IXX* for which there is no LC authority record distributed.

Another common suggested vendor requirement is to provide a report listing headings that match two or more authorized headings, as well as indicating the cost and format of the report.

Such a report relates to one's perception of how an authority control provider goes about its work of linking accurately as many library headings to authority records as possible, without making bad links. From LTI's perspective, by definition "two or more

authorized headings" for the same unique heading cannot exist. Following authority control, a heading is either linked to an authority record or it is not linked. In fact these so-called "multiple matches" reports are a construct of the vendor's inability to deal with the necessity to "block" ambiguous and identical cross-references prior to linking headings.

Admittedly, there are some instances where vendor-blocked cross-references prevent a link. In these cases LTI editors either review manually the high frequency unlinked headings, or such headings remain unchanged and appear in the unlinked headings report. On rare occasions the Library of Congress distributes multiple authority records for the same heading. These can be the same heading or a variant heading appearing in the *IXX* of the authority records. When this occurs, we notify LC in one of our weekly reports and LC resolves the conflict, often in the next weekly distribution of LC authority records. In summary, it does not seem reasonable that a specific vendor's limitations should become a "mandatory requirement" for another vendor that resolves the occurrence of multiple identical cross-references in authority records in a more rational and effective manner. Deluging a library with yet another set of reports simply removes the burden from the authority control provider and places it on library staff who must examine the reports and make their own decisions.

Computers excel at churning out reports but they are seldom a blessing. Unless a report presents summary data, its very existence implies that someone on the library's staff is going to have to review the information, either as a requirement to fix ambiguous or unlinked headings, or because the library believes it necessary to quality check the vendor's processing. Regardless, library staff are being asked to do work which the institution contracted with the authority vendor to do.

Perhaps more troubling is accepting a vendor's "Highly Desirable Requirements" on the assumption that the identified requirements define the playing field. Not to diminish the efficacy of consultative marketing, but a company cannot sell what it does not have. For example, LTI alone among authority vendors offers a comprehensive continuing authority control service that updates not only the new and revised authority records but one that also returns the library's updated catalog records for overlay, thereby removing this thankless and time-consuming task from the shoulders of library staff. But for inexplicable reasons this essential component of any authority control update service goes unmentioned, as does the more common sense need for some level of editor review in backfile authorizations,

under one-hour turnaround for authorizing new cataloging records, etc. There is a price to be paid for confining the library's requirements to what is available on one provider's menu.

## Re-authorizing the Library's Database

For libraries lacking the staff or resources to maintain headings following batch authority control, exporting and re-authorizing the entire database every few years may be the only way to keep controlled headings in synch with LC.

When a database that has previously been authorized is re-submitted for authority control, there is no need to export item record data or the library's existing authority records. All the authority control vendor requires is the bibliographic records. After the re-authorization the library should be prepared to delete all its existing nationally distributed and local authority records. LC, LC Children's, NLM MeSH, and Sears authority records are completely replaced as part of the re-authorization. Similarly, ILS software generates new provisional authority records for headings for which no nationally distributed authority record is available.

Following re-authorization, the library's catalog records are re-loaded into the local system database by "overlying" on the ILS's record ID control number. New authority records and bibliographic records are loaded and new indexes are built.

Some libraries may not want to delete their old authority records either because they have added cross-references or notes or removed cross-references that they consider not useful for their catalog, or because they are advised by their ILS vendor to "overlay" the new authority records. At least one local system vendor encourages the retention of existing authority records by imposing a high fee to delete the library's old LC and/or local authority records. Retaining old authority records rather than starting fresh with authority records directly referenced in those bibliographic records, will result in several kinds of conflicts as described below. The list is not intended to be exhaustive. Such conflicts may be minor or major, frequent or infrequent—but they are inevitable and the

library pays a price for keeping outdated and local authority records.

1. *Duplicate LC authority records*—Each week LTI reports to LC duplicate authority records. These are cases where the *IXX* heading is generally the same in both records, however the LCCNs of the two authority records are different. If the library's file contains the first of two duplicates and the file from LTI contains the second, two authority records will exist in the library's file. Though the authorized heading are frequently identical, there may be minor differences between the two, e.g. *\$d1900-* compared to *\$db. 1900*. There may also be variations in cross references.

2. *Deleted LC authority records*—The library may retain in its old authority file records deleted by LC. These may be in direct conflict with the current LC file.

3. *Unlinked LC authority records*—Unless the library has scrupulously deleted authority records from the file whenever the last occurrence of that heading was deleted from the bibliographic record file, the current authority file probably contains authority records for which no heading exists in the bibliographic file. The advantage of starting with a totally new authority file is the knowledge that all authority records apply to the bibliographic file.

4. *LC vs. locally created authority records*—If the library created a local authority record and LC later distributed an authority record for that heading, the forms of headings in the *IXX* may not agree. More serious problems arise when the heading in the library's local record matches a *4XX* in the LC record.

5. *LC changes pattern heading structure*—For example, several years ago LC replaced hundreds of headings of the type *\$aTrade unions\$x[Industry]*. The replacement headings were mostly of the type *\$aArtists\$xLabor unions*, but sometimes varied: *\$aTrade unions\$x College employees* became *\$aUniversities and colleges\$xEmployees\$xLabor unions*. In these situations, the new authority records often contain the old form as a cross-reference, resulting in a conflict.

One of the benefits of a newly authorized database of bibliographic records is the corresponding files of comprehensive and up-to-date LC authority records. Why compromise that by mixing old authority records with new ones?

*“Authority Control for the 21st Century”*

2300 Computer Ave. Suite D-19 Willow Grove, PA 19090 - 1736  
(215) 830-9320 Fax: (215) 830-9422  
(800) 795-9504 email: LTI@LibraryTech.Com  
**www.librarytech.com**  
**www.authoritycontrol.com**