

10 Ways to Improve Data Quality

With a coordinated effort, your library can make significant progress in cleaning up its online catalog **by Jeffrey Beall**

Have you ever found a book listed in your online catalog but couldn't find it on the shelf, even though it wasn't checked out and you looked for it over a period of weeks? Or have you had the experience of finding a book in the stacks that contains the information you were seeking, and you wonder why it didn't turn up in the searches you'd done? Both of these scenarios can occur when data is not properly maintained in library catalogs or when steps haven't been taken to ensure the books in the stacks match what's listed in the catalog. In many libraries, these glitches are often overlooked because they are no one's responsibility.

Dirty data in online library catalogs—including data with typographical errors, misspellings, and incorrect, excess, or missing information—can severely hamper patrons' access to library materials. Access is restricted when a catalog searcher does not retrieve sought-after items that would otherwise turn up in search results, or when a user retrieves a large number of false or duplicate hits. For example, if a patron is looking for a book about nutrition and doesn't know the author or title, and the subject heading on the bibliographic record for the book is misspelled as "nutritution," the record will not be included among the search results and the user may never find the item.

Bad and poorly maintained data in library catalogs causes lost time, money, and information. Time is lost when materials can't be located due to missing or incorrect data in the catalog. These materials may then have to be ordered through interlibrary loan, which costs money. And information is lost when it's not accessible in the catalog, or when listings for books and other materials are not retrieved in searches, but they would have been retrieved if the data had been entered and maintained correctly.

Most of what I describe here represents standard quality-control practice in library technical services. However,



JEFFREY BEALL is catalog librarian at the University of Colorado at Denver and Health Sciences Center's Auraria Library.

Bad and poorly maintained data in library catalogs causes lost time, money, and information.

because libraries often emphasize the fancy technological aspects of integrated library systems, we sometimes neglect to pay sufficient attention to the intellectual content and clerical accuracy of those systems—that is, the bibliographic data and its maintenance.

Bad data in an online catalog is the modern equivalent of a missing card in a card catalog, and although most of what I recommend is tailored to fully automated catalogs, the principles apply to all libraries. Whether catalog data is stored on catalog cards or on a computer server, it should be fully standard, up-to-date, complete, and accurate.

Low-quality data in library online public access catalogs is a topic not often discussed among librarians, but many of us have seen glaring typos and other errors when using our catalogs. Many databases are riddled with mistakes. Even the Library of Congress's catalog has errors. Other professions strive to reduce and eliminate mistakes, and librarianship should be no exception.

Fortunately, with a coordinated effort, a library can make significant progress in cleaning up its online catalog. To achieve this progress, I recommend that libraries take the 10 steps listed here. The first eight steps represent sound technical services practice and include tasks that are part of intake procedures, that correct past errors, that arise from changed and updated cataloging standards, and that involve general data cleanup, in that order. The final two steps are tasks not specific to technical services, but goals that all library staff can work together on.

1 Perform authority work on all controlled headings. This means checking to make sure every heading—authors, subjects, series—agrees with the established form on the authority record. Authority control can be manual or automated. It should be done for the headings on all new items entering the catalog.

2 Follow national cataloging standards in all aspects for all library materials. Doing this helps ensure colloca-

tion of materials in the online catalog and consistency in retrieval. Collocation means that similar items are located next to each other in online catalog displays—just like similar items are placed next to each other in the stacks—which facilitates research and access to information. And consistency means library users can be sure that their searches will always work the same for all materials the library holds.

3 Search for and correct typographical errors that have crept into your catalog. Fortunately, librarian Terry Ballard has compiled a list of the most common typos in “Typographical Errors in Library Databases,” available at faculty.quinnipiac.edu/libraries/tballard/typoscomplete.html. Library staff can search this list of misspellings in their local catalogs and correct the ones that turn up. Of course, some misspellings are correct in some contexts, and a misspelling in one language may be a correct spelling in another, but with minimal training, a library staffer using this list can help significantly reduce the number of errors in a library online catalog.

4 Make sure all your autobiographies also list the author as a subject. In the past, cataloging rules did not require that a separate subject heading for the author be included because the author heading functioned both as an author heading and as a subject heading. But coding and practices in how library catalogs index data have changed, and the authors of autobiographies must now be included as the subject.

5 Keep up-to-date with subject headings. For example, the subject heading string “Botany—Anatomy” was recently changed to “Plant anatomy.” The Library of Congress lists these subject-heading changes soon after they are implemented at www.loc.gov/catdir/cpsol/.

Also, libraries need to retrospectively add subject headings to their catalogs as they are created. LC recently created subject headings for individual earthquakes, such as “Alaska Earthquake, Alaska, 1964.” Previously, works on this topic were only given the subject heading “Earthquakes—Alaska.” So it is necessary to add the new subject heading to your library’s catalog for the records that correspond to books on this topic.

6 Find records without subject headings and add them as appropriate. There are two ways to do this: One is simply to locate the records by chance, as catalogers or others notice the omissions, and then fix or report them. Alternatively, many systems offer a feature to generate reports of records that meet certain criteria, such as records without subject headings.

7 Fix errors in initial articles. Catalogers code bibliographic records with a number that tells the system how many letters to ignore at the beginning of a title when it begins with an article such as *The*, *A*, *An*, or *Los*. But this coding system is error-prone: If the coding in an individual record is set at zero, the title may file on the word *The*. This coding should be checked on a regular basis.

8 Make sure the location for every item in your catalog is correct. For example, if the bibliographic record says a particular book is in reserves, be sure that book is actually in reserves. Providing incorrect locations for materials is practically the same as not having them.

9 Organize a librarywide shelf-reading/inventory project. Shelf reading means comparing what’s on your shelf to what’s in the catalog, a task that new wireless technology has made much more practical and efficient. Delete records for lost books, and add records for books not in your catalog. Replace call number labels that have fallen off or are unreadable. Also, make sure all library materials are in correct call number order.

In the past, libraries used their shelflists to do inventories. The shelflist was a separate card catalog, filed in call number order, with one card for each title. Today, online catalogs have virtual shelflists. By entering a call number search, you retrieve the library’s books and other materials in call number order. If your library has wireless capabilities, workers can put a laptop on a cart and bring it right to the shelf, or the information can be printed out and brought to the stacks. Library staff then compare this information to what is on the library shelves, minus any books that are reported lost or checked out. They next make a list of all missing items, change their status in the online catalog, and search for them at regular intervals. After a period of time, if the items still haven’t been found, their records should be removed from the catalog, and the library’s collection specialists should be informed so they can decide whether to replace the lost items.

10 Involve all library staff in database maintenance. Have staffers report errors to a central person, preferably someone in the cataloging department. Many library staff make extensive use of the library catalog and are in a position to observe errors. Take advantage of their catalog use and ask them to report the errors they find.

Library catalogs should be tools for research; we must not allow them to become barriers to retrieval. Libraries are obligated to make services as error-free as possible. A coordinated effort to eliminate errors in online catalogs demonstrates a strong commitment to quality service. The reward of this work will be better access to library materials for our patrons. ❖

Copyright of American Libraries is the property of American Library Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.