



5-1-1992

File Reloads

Carol Tenopir
University of Tennessee - Knoxville

Follow this and additional works at: https://trace.tennessee.edu/utk_infosciepubs



Part of the [Library and Information Science Commons](#)

Recommended Citation

Tenopir, Carol, "File Reloads" (1992). *School of Information Sciences -- Faculty Publications and Other Works*.

https://trace.tennessee.edu/utk_infosciepubs/345

This Article is brought to you for free and open access by the School of Information Sciences at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in School of Information Sciences -- Faculty Publications and Other Works by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

ONLINE DATABASES

BY CAROL TENOPIR

File Reloads

HOW OFTEN HAVE you logged on to do an online search and been greeted by the message "file reloaded"? File reloads seem to happen every month to one or more files, but I suspect many searchers aren't too sure about what it really means. If your favorite file is reloaded, do you have to learn the ins and outs of searching it all over again? Are any of your old search aids still of use? If you search a file occasionally do you have to memorize and remember all of the information in the online reload news?

More than an update, but less than a completely new file, file reloads can mean some important changes that affect how you search a database. But not all reloads are created equal, so the impact on the searcher can vary. I decided to investigate what happened in the reloading of files from both the producer and vendor side to see just what a file reload means and doesn't mean for searching.

What happens in a reload?

In a file reload, the entire file, including all of the old records whether they are changed or not, are taken off of the online vendor's computer and completely replaced. New tapes of the entire file with all desired changes are supplied by the database producer. These new tapes are first processed by converting the information supplied by the producer into a form compatible with the vendor's system. The records are then run against the vendor's database creation software that creates all of the access points and searchable features for that database.

This lengthy process of conversion and machine indexing takes a lot of computer time, but users are usually not aware it is happening. A reloaded file does not replace the old one until it is complete and tested. Regular up-

dates may continue to take place to the old file. The drain on computer resources is minimized for large files by batch processing over days or weeks.

File update reload

Reloads are not all of the same complexity. The simplest type of reload is the file update reload. For many directory files (and some small bibliographic databases), every time the file is updated it is reloaded. This makes sense with directories because they replace most of their old information with new, more up-to-date information. Old records are not left to accumulate in the file as they are in a bibliographic database that is updated with new records rather than reloaded. In the case of directories, it is often simpler to take off the old file and reload a completely new one.

This simple type of reload may take the online vendor only a few days or weeks to do. To the searcher sometimes there are hardly any noticeable changes in field structure or necessary search techniques. Your search will result in newer data, but the way you get that data may remain exactly the same as before the reload. Documentation may not even change.

File maintenance reloads

Another similar and relatively simple type of reload is the file maintenance reload. Over time, errors may accumulate in a database if it is not regularly corrected. Newer files typically have quarterly or semi-annual correction runs when the database producer submits a correction tape.

In some old files on DIALOG (those loaded prior to the DIALOG2 software change in 1985), these kinds of corrections are not regularly made because the way the files were structured under the older version of the software makes it difficult. The entire file may be reloaded instead to correct a lifetime of accumulated errors. Subsequent corrections are usually easier to make once the file is reloaded with the new software conventions. If record error correction is all that is done in the file maintenance reload, the reload is almost unnoticeable to most users.

Database producer-initiated reloads

Most reloads of bibliographic or full-text databases are more complex than update or simple error-correction file maintenance reloads. Database producers may initiate a reload when they have made database enhancements that they want to be searchable throughout the file—retrospectively as well as in current records.

These enhancements may be changes in the controlled vocabulary indexing terms or in a classification system. For example, MEDLINE is reloaded every year on most online systems to reflect the yearly changes made by the National Library of Medicine (NLM) to Medical Subject Headings (MeSH). NLM incorporates MeSH changes throughout the file.

PsycINFO completely revised its classification system in 1991, then went back and automatically reclassified the entire file. Data-Star was the first system to reload PsycINFO after the revisions, so, for a while at least, it is the only online system to reflect this filewide change. Database producers try to reload their changed files at the same time on all the systems that have their database, but it may not always fit into the vendor's reload schedule.

Another common type of retrospective enhancement made by the database producer is the addition of a field or fields to improve the searchability of a file. When Trade & Industry ASAP was reloaded this year on DIALOG, a new industry subfiles field was added to every record in the database. The subfiles are broad subject categories that are automatically assigned according to what journal an article is in; for example, records from advertising journals get an ADV subfile, those from library journals get an LIB subfile. This new field allows a search for just those records that fall into a general area of interest.

The EMBASE reload on BRS added a journal title field and a publication type field. PsycINFO's reload on Data-Star added a Composite Age Group field, which allows searchers to select broad human age groups such as child, adolescent, adult, and elderly. Nursing & Allied Health (CINAHL)



Carol Tenopir is Associate Professor at the School of Library and Information Studies, University of Hawaii at Manoa, Honolulu

ONLINE DATABASES

on DIALOG added identifier and corporate source fields in its March 1992 reload. ISI's files (including SciSearch, Social SciSearch, and Current Contents) recently added Keywords+ and author abstracts fields.

Sometimes database producers make changes on just the new records coming into files without going back and changing old records. In these cases, simple updates rather than reloads are done, but the searcher has to remember both the old and the new ways to search. For example, when the government changed the Standard Industrial Classification (SIC) codes several years ago, some database producers chose to retain the older codes for old records and begin using the new codes only with new records. Other producers decided to convert all of the old codes into the new ones, necessitating a reload. Such retrospective changes and reloads make life easier for searchers who only have to remember one way to do something all the way back in a file.

Vendor-initiated reloads

Online vendors may also initiate file reloads. A common reason is to allow files to take advantage of system software enhancements that can only be implemented file-by-file as they are reloaded. An example of a system feature reload is the addition of the KWIC and HIGHLIGHT output options on DIALOG. Each file has to be reloaded before it can use these features because the features require additional information in the machine-produced indexes. When any full-text or most bibliographic files are reloaded on DIALOG for any reason, the new, more complex index structure needed for KWIC and HIGHLIGHT may be added.

The capability to do sorting, limits, or ranging are features added by the online vendor when a file is reloaded. If numbers were not originally loaded as numeric data, ranging won't work without a reload to change their data type. Limit features can be implemented several different ways, requiring a reload if there is to be an additional method or change in method.

Other examples of system features that require individual file reloads include the BRS Medspell (British/American spelling equivalencies) feature and its link feature that links bibliographic records in one file with their corresponding full-text records in another file.

Sometimes customized search features are added for a database, requiring both a reload and some programming on the part of the online vendor. When ISI's Current Contents was reloaded recently on BRS, the file was split into two files: one containing Table of Contents records, the second containing regular bibliographic records. The Link feature was customized to allow Table of Contents browsers to immediately request an interesting bibliographic record and vice versa. Chemdex on ORBIT was recently reloaded to provide chemical fragmentation searching.

Often changes in an online vendor's philosophy over the years necessitates reloads. DIALOG is trying to provide as much interfile consistency as possible to help searchers use multi-file search features like OneSearch.

If record error correction is all that is done in the reload, it is almost unnoticeable to most users

File reloads are changing all publication year fields to PY, for example, rather than having some YR, some PD, and some PY. Corporate source is being standardized as CO.

BRS and DIALOG now try to make records more understandable, with fewer codes. When BRS reloaded PsycINFO, it added explanations to various coded fields within the records. The Trade & Industry reload on DIALOG added a table of STC code meanings that display next to the SIC numbers in an EXPAND list.

Combined reloads

Because a reload is time-consuming and resource intensive, it is a time when database producers and online vendors ask themselves, "What is everything we want to do to this file?" One vendor representative told me, "A reload is an opportunity to do all those things we've wanted to do to the file. We can make all the changes that people want in the way the file looks and the way it is searched."

Sometimes a reloaded file looks quite different. There may be new

fields, updated vocabulary, new search features, and new display options. In these cases, online system documentation and sometimes database producer manuals are redone. System search aids like DIALOG Bluesheets or BRS AidPages are the first to be reissued, often at the same time as the changes are described in the system newsletter. The systems try to issue these important aids online and in print at the same time a reload is complete because they are necessary for effective searching of the reloaded file. More detailed documentation, such as database chapters, may take longer to come out, but revised versions should be purchased when available.

Database producers may or may not redo their manuals. If they are issued in loose-leaf form, revised pages may be all that is required. Details of the reload are usually given in newsletters, which typically are distributed free. Attending a database training session may be advisable for extensive changes in a file you use often.

Other times the changes are mostly operational and hardly affect the searcher. The only discussion may be in the system or producer newsletter and in the logon file banner.

Not every database is reloaded in its lifetime, and reload schedules are planned at least a year in advance by online vendors. Reloads take anywhere from two weeks to over a year from beginning to end, depending on how much is to be changed. Most (other than simple update or maintenance reloads) take an average of four to six months, more if reprogramming is required. During this time database producers and vendors are discussing changes, sending test data tapes, perhaps programming, testing, fixing, and retesting.

According to one person I spoke with, "any number of tragedies can occur in a big reload: new data tapes might not match the specifications, tapes can be misplaced, the system crashes. But most of the time it works."

Many file reloads take time, computer resources, and staff effort. They may mean a searcher needs to get new documentation and learn some new search techniques to best use that file. But reloads improve searching and offer consistency throughout a file and, often, across a system. One database producer summarized it nicely: "Reloads mean better searching—that's the whole idea."