7-1-1993

When Is the Same Database Not the Same? Database Differences Among Systems.

Carol Tenopir
University of Tennessee - Knoxville

Katie Hover

When is the same database not the same?: database differences among systems

Authors: Carol Tenopir and Katie Hover
Date: July 1, 1993
From: Online(Vol. 17, Issue 4.)
Publisher: Information Today, Inc.
Document Type: Article
Length: 4,268 words

Abstract:
Users of online databases frequently have several options available for searching the same database. Selection of the appropriate access method can be simplified through the evaluation of seven criteria. Price, time periods covered, and frequency and timing of updates should be considered first. Variations in file structure among online systems should also be evaluated. Finally, the online systems themselves should be evaluated for their support systems and search features.

Full Text:
Database Differences Among Systems

With hundreds of online systems available for searching, it doesn't make sense for a searcher to use just one or two. Even if one favorite system has all the databases you every use, chances are other systems will provide a better search at one time or another. Searching the same database on different systems may be less expensive, or better quality, or more powerful, or more comprehensive. Experienced searchers know this; research libraries (academic and special) in Canada and the U.S. access on the average 7.4 online systems [1]. Add the many CD-ROM systems and locally-loaded databases, and the number of systems probably double or triple.

One assumption many of us make is that once a database is created it is the same no matter what system it is on. Searchers are taught to assume that ERIC is ERIC, MEDLINE is MEDLINE, and PsycINFO is PsycINFO, no matter where it is mounted. Although this is often true, for some databases it is way off the mark. The same databases, created by the same producer but loaded on different systems can and do differ. An article by Nahl-Jakobovits and Tenopir focused on the differences between online and CD-ROM versions of two datases [2]. This article will take a different approach and look at the many ways online versions of databases can vary on different systems.

BACKGROUND INFORMATION

Before exploring the ways the same database on different systems can differ, some background information may help put things into perspective. Even the average of over seven oline systems and perhaps a dozen CD-ROM systems used in research libraries is small compared to the possibilities. A total of nearly 750 online systems are available, according to the January 1992 edition of Directory of Online Datases [3]. On these systems there are more than 5,000 databases, produced by over 2,200 databse producers (Figure 1).

Numbers like these are difficult to comprehend. When searchers hear 750 systems, one immediate reaction is to give up and rely on just one or two favorities. An instinctive reaction might be: "I can't possibly ever learn to search them all so why bother?"

It isn't actually as bad as it seems. Though there are hundreds of systems available, in reality only a handful of systems are heavily ued by intermediary searchers. Just seven online systems are reported to dominate use in the U.S. and Canadian library/information Center market: DIALOG, Mead Data Central, BRS, NLM, Westlaw, STN, and ORBIT [4]. CD-ROM expenditures in the early 1990s were concentrated on Information Access Company's InfoTrac, SilverPlatter, Lotus Corporation, and H.W. Wilson Company [5].

Out of thousands of online databases, again a relatively small number are heavily used. For all types of libraries, excluding law libraries, the most used online databases are MEDLINE, NEXIS, ERIC, World Patents, Predicasts files CA File (Chemical Abstracts with abstracts on STN), CA Search (Chemical Abstracts without abstracts on several systems), and FINIS [6].

Since special libraries are such heavy online users, this list is dominated by special library favorites. Looking at academic or public

libraries alone the favorites change somewhat. MEDLINE, ERIC, and Magazine Index are perenial favorites in public libraries. Academic libraries favor MEDLINE, ERIC, PsycINFO, BIOSIS, Health Planning and Administration, Dissertation Abstracts, INSPEC, CAB, Social SciSearch, PDQ, CATLINE, ABI/INFORM, National Newspaper Index, and Magazine Index.

Many of these titles as well as others appear on lists of popular CD-ROMs in libraries. Among thosee often mentioned include InfoTrac (one version of which is Magazine Index), ERIC PyscLIT, Books in Print Plus, Academic Index, MLA Bibliography, ABI/INFORM DISCLOSURE, Dissertation Abstracts, and Social Sciences Index.

Database producers try to get the widest exposure possible for their products. The exclusive contracts of years ago (where a database was available only on a single online system) are uncommon today. Of the nineteen most popular online databases in libraries, only one is exclusive to a single online system (CATLINE on NLM). Five are on two systems, and the remaining thirteen databases are on three or more (Figure 2).

If we add CD-ROM or tape availability it gets even more complicated because almost all the most popular online databases are also on CD-ROM and many are on tape. This means that all of us are faced with the decision of what system to search every time we do an online search.

CATEGORIES OF DIFFERENCES

It is a myth that a database is the same no matter which system it is on. In reality the version of a database may differ slightly or drastically depending on the online or CD-ROM system that makes it available. There are at least seven key potential differences, which we have grouped into three broad (and rather arbitrarily named) categories (see box on page 22). The first three differences fall under the "easy or obvious" category; the next two are grouped as "database choice"; and the final two fall under "system choice."

EASY (OBVIOUS DIFFERENCES

Easy or obvious things are just as important, and maybe more so than the other categories. They are "easy" only because much of the information is easily available and may even be published in source directories such as the Gale Directory of Databases. With this article is a bibliography that lists some sources for finding this type of information.

Three main factors influence the "easy" part of database selection:

1. updating 2. dates covered 3. price (and no, price should not necessarily be the most important factor)

We will discuss each in turn. (Table 3 provides details of online system differences for several major databases.)

EASY (OBVIOUS) DIFFERENCES: Updating

Updating means two things: how often a database is updated on each system (updating frequency), and when, within that update period, does each system do the updates. Frequency is straightforward: PTS PROMT is updated daily on DIALOG, but weekly on Data-Star; INSPEC is updated weekly on ORBIT and CAN/OLE, but seminmonthly on Data-Star and BRS. The CD-ROM version of PsycINFO from SilverPlatter comes out quarterly, but online versions are available monthly. CINAHL from SilverPlatter is bimonthly; CINAHL from Cambridge is monthly. You can check this easily in directories or by calling the system or database producer.

With CD-ROM versions you can often save money by sacrificing currency, if you are willing to settle for quarterly or even annual updates. For EBSCO's Magazine Article Summaries, for example, you pay considerably less for less frequent updates. (But SilverPlatter's bimonthly version of CINAHL costs the same as Cambridge's monthly version. Logic often doesn't have anything to do with it.)

Online, the price usually doesn't seem to have a direct relationship with update frequency. DIALOG's daily updated PTS PROMT is just $2 more per hour than the Data-Star version that is updated weekly; BRS's weekly version of BIOSIS is $13 less than Data-Star's biweekly. Choose the most up-to-date version if you need it, regardless of cost.

Trickier is the issue of when a system actually loads the updates it gets from the database producer. MEDLINE may be updated weekly on DIALOG, BRS, and NLM, but (for example) DIALOG may load the tape on Tuesday, BRS on Wednesday, and NLM on Monday. That means the same tape with the same dates of records may be available days earlier on one system or another.

When updates happen is especially troublesome because this information is usually not published or advertised. You have to log and check the file banner, check the update field within the database, or call each online system and ask them to check for you. If timeliness is crucial to you, it is worth checking. It also may vary from week to week unless the database producer is also the online service. WILSONLINE can be expected to be more consistent with loading Wilson index updates, as can STN for Chemical Abstracts.

The shocking (although perhaps not surprising) story is that updates don't always happen as scheduled or planned. Updates are delayed for all sorts of reasons--database producers are late with a tape, the online system had a special higher priority project, there were sunspots that week...INVESTEXT on DIALOG is billed as being updated semiweekly, but when we checked it once last year it hadn't been updated for over three weeks. DIALOG said they hadn't yet gotten the tape from Investext. Investext has its own online system and always wants to get the updates up first there before sending tapes to competing systems.

A student discovered quite by accident two years ago that the NEXIS version of ABI/INFORM was missing seven weeks' worth of

records for seemingly random weeks across three years [7]. It seems that NEXIS occassionally forgot to load the ABI/INFORM tapes and no one bothered to check. It was news to even ABI/INFORM when we told them. (That relationship went downhill from there--ABI/INFORM is no longer on NEXIS).

It pays to check updates before each search if you want the most up-to-date version of a database. Soon patterns will emerge ass you regularly check a database with each system. One may be habitually late or may seem to have continual problems dealing with loading a database's updates. Then is the time to give up on that system for that database (letting he system and database producer know why).

EASY (OBVIOUS) DIFFERENCES: Coverage

Coverage has two parts, one trickier than the other. The easier part is how far the entire database goes back--that is, what backfiles are available. Some systems such as Dow Jones and CompuServe make a point of not keeping backfiles, but most research systems try to. Backfiles vary with the database, but a system may not always choose to mount the entire file. (Luckily, this information is readily available in directories and database documentation.)

CAB Abstracts on DIALOG begins in 1972, on CAN/OLE in 1983, on Data-Star in 1984. COMPENDX icludes 1970 forward on most systems, but only from 1976 on Data-Star. (Data-Star is a great bargain, but many of its fils don't go back as far as other versions. You may be paying less, but you are also getting less.

The second part of coverage is trickier--the coverage of individual journal titles indexed in bibliographic databases or included in multiple full-text databases. A database usually advertises title coverage by the earliest date in the database, not the average date or most frequent date. A database usually includes hundreds of sources, each of which may have its own coverage.

Trade & Industry ASAP includes over 200 journals and claims coverage from 1983 to the present. Here are ten titles covered in the database with their starting dates:

Arkansas Business, 1/89- Association Management, 01/92- BC Business, 1/89- Communications Daily, 1/91- Canadian Business Review, 1/89- Canadian Labour, 1/90- Food in Canada, 1/91- Journal of Accountancy, 01/90- Mass Transit, 01/91- PC Magazine, 1984-

Fewer than 100 titles in Trade & Industry ASAP go back to 1983, and when new titles are added databases usually do not cover that title retrospectively. This is standard practice for both full-text and bibliographic databases. Coverage dates heralded by the producer only mean that you won't find anything older than that.

Neither the online system nor database producers proclaim too loudly hat coverage of magazine titles in a database may differ depending on the online system. In 1987 Ruth Pagell wrote a revealing article for DATABASE on "how full is full" text? [8]. She discovered that the titles included in the various versions of Magazine ASAP and Trade & Industry ASAP were quite different.

This continues today. Forbes magazine is in the Data-Star and DIALOG versions of those files, but not on BRS. Most of the Canadian titles are not on NEXIS. What Pagell found really shocking was that not even all the magazine publishers knew that they were in one version but not another. (To be fair to Information Access Company, the different versions are differentiated by calling one Magazine ASAP, another Magazine ASAP II, etc.)

EASY (OBVIOUS) DIFFERENCES: Price

Online pricing gets more complex all the time. Many systems now calculate price differently (e.g., connect time or flat-free charge per database, or computer processing, or information retrieved, etc.). This makes comparison difficult, but offers more choices to the seacher [9]. Several years ago Martha Williams averaged all costs paid to online systems by users (no matter what pricing method the system used) to calculate an average cost per hour for major online systems. Although the exact numbers are out-of-date and pricing policies of some system have changed, Table 1 shows that searchers of multiple system can estimate what costs to expect if they compare previous searching costs on a variety of systems. Each library should calculate its own comparative costs based on current prices and usuage.

Even assuming systems have the same pricing formula, such as DIALOG and BRS and ORBIT, the exact prices per hour or per print may differ widely. You must balance the connect time and print costs with the efficiency of the system (can you really get the same amount of information in the same amount of time?) and your skill in using the system.

Price also includes what discounts are availale. Remember Data-Star's 20% discount for North American usuage in Swiss non-prime time and NLM's non-prime discounts and high-volume use discounts and high-volume use discounts from the major connect hour systems.

Many libraries use after hours services such as DIALOG's Knowledge Index (now available through CompuServe) or BRS/After Dark to save money, because they offer access to many of the same databases that are on the full-priced daytime service, but at a much lower price. These services are intended for personal use or end-user search services. If a library doesn't have an end-userces. online search service, the differences in cost shown in Table 2 may be enough to make you start one!

Some of the best online bargains in North America are (in no particular order):

* Data-Star (20% discount)

* OCLC EPIC and FirstSearch for OCLC libraries

* BRS/After Dark (only for simple searches because the menus seem to take forever)

* CompuServe and othe "flat-free" services such as America Online

* Internet (for searching Uncover and other databases)

DATABASE CHOICE: Subfile Structure

Subfile structure is how a database is divided or subdivided on each system. It may be all in one big file such as MEDLINE in File 155 on DIALOG or File MESH or BRS, or it may be in many little files separated by date, such as the seven separate MEDLINE files on NLM (Table 3). The best option, offered by several systems, is offering a choice of one big file and smaller date-divided files. Databases may be divided by criteria other than date, such as the 36 subject subfiles of MEDLINE on NEXIS, or another NEXIS subdivision option of English only or human subjects only. EMBASE on CD-ROM forces you to choose from among a group of subject discs; while EMBASE online is all in one big file.

Subjectt subdivisions are aimed at end-users who may have very specific or narrow subject interests. Such subdivisions help reduce the number of documents retrieved by giving users just ones from the viewpoint or perspective of their research. Subject subdivisions are rarely useful for libraries that have a broader view of a topic and must serve a variety of users, because users run the risk of missing relevant information. The searcher looking in the "allergy" subset of MEDLINE on NEXIS, for example, may miss relevant items in the "dermatology" or pediatrics" sections. With subject subdivisions, the searcher doesn't know what he or she is not getting.

A better way to deal with subject subdivision is to have it as an optional field within the database records. MEDLINE users on DIALOG can restrict a search to just Index Medicus, or Index to Dental Literature, or International Nursing Index subfiles. CAB Abstracts online puts together in one file, 50 separate print abstracting journals. If a user is only interested in one (or a few) of these separate publications, they can be searched by using SUBFIELD CODE. The user who only wants rice Abstracts or Maize Abstracts need not search through the entire million-plus records from subfiles such ass Protozoological Abstracts or Index Veterinarius. In ERIC online a search can be restricted to just the Resources in Education (RIE_ or Current Index to Journals in Education (CIJE) subfile, but the default is to search them both.

Date segmentation is a different story from subject segmentation. Separate files by date is often to the searcher's advantage, especially in huge databases with millions of records such as MEDLINE. With date subdivisions you know what you are missing, so items aren't missed by mistake. If you only need current information or records back only a few years, response time will be much better on a subdivided file.

```
TABLE 1 Average Cost per Hour per System System 1982 1986 BRS $49 $69 DIALOG $93 $126 Mead $150 $141 NLM $28 $39 ORBIT $109 $130 STN $122 $203 Westlaw $30 $100 Average $103 $122 TABLE 2 Per Hour Connect Charges on Different System(*1) Database Knowledge Index(*2) DIAGLOG After Dark BRS AGRICOLA $24 $45 $16 $37 Books in Print $24 $60 $18 $71 ERIC $24 $36 $8 $31 International $24 $90 $39.50 $75 Pharmaceutical Abstracts SPORT $24 $114 $20 $93 MEDLINE $24 $36 $24 $33 (*) Not including CompuServe connect fees. Not including telecommunications charges and display charges.
```

[TABULAR DATA OMITTED]

Subfiling isn't all bad--response time is better and false drops are minimized. Searchers should pick the subfile arrangement that will be to their benefit. For lowest cost and best response time, pick the smallest subfile that best meets the question at hand--sometimes it may be on NEXIS, other times on Data-Star, NLM, DIALOG, or BRS.

(Although file subdivision is categorized here as a "database" choice, often the system gives the producer little or no say in the matter. We can't always blame database producers for poor subdivision choices.)

DATABASE CHOICE: Field Subdivision

How a database is divided into fields can also often differ from system to system. If a system publishes AidPages or BlueSheets or whatever they call them, use them. Though the database producer provides the same informatiuon to each system, what is searchable, how it is limped, and how it is searchable vary.

MEDLINE on BRS, for example, has a separate field for Gene Symbol, while DIALOG lumps it in with Enzyme names and numbers in the Identifier field. Data-Star has created KEYWORD SUPERLABELS that automatically include all the subject-related fields except full text in IAC full-text databases such as Health Periodicals Database. This greatly helps the searcher avoid false drops. Readers' Guide to Periodical Literature on BRS has 24 fields; the WILSONLINE version has only 21.

Develop a preliminary strategy, then look at the documentation for each system to see which has structured the database in the best way for your question.

DATABASE CHOICE: Content or Inclusion

The second topic under database choice is "content or inclusion." This is one of the most frightening categories, because most searchers assume at least that the same records and the same content will be on the same database on different systems. We all assume it every day we seach--it is a fundamental assumption of searching.

You will be relieved to know that usually it is a safe assumption, but not always. Scarier yet, is that the database producer or online system doesn't always acknowledge the differences.

It is no secret (but people seem to ignore it) that certain types of information are left out of CD-ROM versions. PsycLIT (the CD-ROM version of PsycINFO) does not include the dissertations or technical reports that are online. Sociological Abstracts online includes a variety of document types, ondisc it only has journal articles and dissertations.

The Trade & Industry ASAP and Magazine ASAP journal coverage differences and ABI/INFORM problem on NEXIS mentioned earlier fall into this category as well. What you get from a search on the same database on a different system in these cases will be different. There is no way, using your search strategy or your skills, that you can retrieve information that is just not there!

Sometimes differences in content are not so bad; the Marcive CD-ROM version of the U.S. GPO file cleaned up the dirty source information, eliminating duplicate records. Other versions did not.

SYSTEM CHOICE: Support Features

The last category of differences is System Choice. This includes two things--what databases support features are available from each system and what system search features are available. Support features include general customer support such as help desks staffed by experienced searchers, the resources to call on subject or database specialists, and good database documentation. It also includes some fancy system features that affect searching--like loading online versions of database thesauri, or linking full-text files with bibliographic files.

At the 1992 Special Libraries Association meeting, an informal poll of searcheers voted Mead's customer support desk as "most helpful." Dow Jones News/Retrival's lack of a toll-free number caused it to be voted "least helpful." Data-Star folks are friendly and answer the phone promptly. Most searchers said they find DIALOG to be helpfulo if they can get through! Busy signals and long waits are the rule there during prime time.

If you are searching a database that has a good controlled vocabulary, like MEDLINE or ERIC or any of the Wilson indexes, use a version of the database that has the thesaurus online. DIALOG has thesauri online for at least 45 databases. If you search MEDLINE on BRS, DIALOG, or NLM you can EXPLORE or MAP and generally exploit the thesaurus online. On Mead and some other systems you are on your own.

More users want full-text retrieval, so choose systems that offer document delivery for a database. Better yet go with the systems that have full text online linked to the database. That means selecting BRS MEDLINE for its LINK feature to the Comprehensive Core Medical Library (CCML), database, and Magazine Index (tied to Magazine ASAP) and Trade & Industry Index (tied to Trade & Industry ASAP) on DIALOG. These combinations give you the precision advantages of bibliographic databases and an easy way to use full text for document delivery.

SYSTEM CHOICE: System Search Features

The last criterion is system search features. System search features are things that are the same across all databases on a system, but that affect the database choice for a specific search. Most major commercial online systems are fundamentally the same--they took different but they do pretty much the same things. You are bound to be more efficient or more comfortable with some systems than with others (and don't underestimate that in your decision making.)

But some features should influence your choice on a database-by-database basis or search-by-search basis. Multiple searching with duplicate detection is one of the most important features for comprehensive searches requiring many files that may overlap. For any search where one database isn't enough, consider first the systems that offer multifile features such as DIALOG's OneSearch and Data-Star's StarSearch.

If you will be searching a full-text database for a search that requires you to actually search the complete text (looking for a company name or obsucre information likely to be buried in a text) proximity operators that allow you to specify within a certain number of words are helpful. The BRS and Data-Star "same sentence" operator is not as precise asd Mead and DIALOG's "within n words" feature. On the other hand, the ability to search within the same grammatical paragraph is useful for eliminating false drops in mixed type full-text databases such as MAGAZINE ASAP. BRS, Westlaw, and Mead also offer a "within the same sentence" operator.

If your free-text search must contain the words "most," "many," "more," "some," or "several" don't search on BRS--they are all stopwords on that system. The phrase "Does he or she have that" is unsearchable in Mead.

CONCLUSIONS

Making system choices comes down to two things:

1. There are differences in the same databases on different systems.

2. It is best to make the decision on a search-by-search basis. No searcher needs to be limited to just one or two systems in today's competitive environment.

Some differences can be fairly easily tracked through printed documentation, logon banners, or system newsletters. For other

differences you may have to rely on published reviews, shared information from colleagues, or simple tests for coverage and updates.

When in doubt, call the system or database customer support desks. Ask them to help formulate your search if they subfiles or fields are confusing. Ask them to get online and check on the latest update. Ask them to give you credit on your invoice if their goof fouled you up. For CD-ROM databases, always ask for a free trial subscription. Let systems know what you like and don't like about the way they load a popular file. If you decide on another version, tell them why you will be searching it on their competition in the future.

REFERENCES

[1] Tesnopir, Carol and Ralf Neufang. "Electronic Reference Options: How They Stack Up in Research Libraries." ONLINE 16, No. 2 (March 1992); pp. 22-28.

[2] Nahl-Jakobovits, Diane and Carol Tenopir. "Databases Online and on CDD-ROM0: How Do They Differ, Let Us Count the Ways." DATABASE 16, No. 1 (February 1992): pp. 42-50.

[3] Directory of Online Databases. Detroit: Cuadra/Gale, January 1992.

[4] Williams, Martha E. Information Market Indicators. Monticello, ILL: IMI, quarterly.

[5] Tenopir, Carol. "The Most Popular Databases." Library Journal 116 (April 1, 1991): pp. 96-98.

[6] Ibid.

[7] Tenopir, Carol. "The Same Databases on Different Systems." Library Journal 116 (May 1, 1991: pp. 59-60.

[8] Pagel, Ruth. "Searching Full-Text Periodicals: How Full is Full? DATABASE 10 (October 1987): pp. 33-36.

[9] O'Lery, Mick. "Flat-Rate Online: A New Online Era Begins." ONLINE 17, No. 1 (January 1993); pp. 34-38.

Source Citation (MLA 8th Edition)
Tenopir, Carol, and Katie Hover. "When is the same database not the same?: database differences among systems." *Online*, July 1993, p. 20+. *Gale Academic Onefile*, https://link.gale.com/apps/doc/A14089257/AONE?u=tel_a_utl&sid=AONE&xid=785e49ae. Accessed 9 Nov. 2019.
Gale Document Number: GALE|A14089257