



10-1-1990

Database Quality Revisited

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, "Database Quality Revisited" (1990). *School of Information Sciences -- Faculty Publications and Other Works*.

https://trace.tennessee.edu/utk_infosciepubs/322

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

Database Quality Revisited

I HAD ALMOST despaired. Every year for the past six or seven years I predicted that quality control would be the next "hot topic" of the year. Nearly every year there was some important event or publication that led me to believe that this was the year for quality. In 1983, Jeff Pemberton, then editor of *Database*, issued his call to searchers to name names and identify databases of poor quality; in 1984 and 1985, Anne Mintz, now librarian at *Forbes* magazine, published her germinal articles in *Online* and *LJ* on "Information Practice and Malpractice" (*LJ*, September 15, 1989, p. 38); and Barbara Quint began her popular "Super Searcher" forums and gripe sessions at the yearly Online meetings. Yet when I published a column on "Quality Control" (*LJ*, February 15, 1987, p. 124-125), the response was underwhelming.

Quality is important

Librarians know that quality control in information products has always been important, whether online or in print. The evaluation process for print reference tools is well established. With online there seems to be more of a tendency to accept answers or results, rather than to question, verify, and cross-check, and many end users still believe that anything out of a computer must be good. Had we just decided to let the buyer beware or accepted some database producer's arguments that quality problems are not their fault (or too expensive to fix)?

At last I feel vindicated in my initial judgment. Quality of databases was the main theme of the National Online Meeting held this May in New York and will play a prominent role in the Online/CD-ROM '90 conference in November in Washington,

D.C. It isn't just the outside observers who are speaking up, it is also the searchers themselves who rely on databases and online systems every day in their jobs.

The real impetus for the focus on quality is coming at the grass-roots level. The Southern California Online Users Group (SCOUG), one of the largest and most active of a diminishing species, held a retreat in August to formulate an action plan for rating (and thereby paving the way to improving) databases. Almost 50 online searchers, representatives from database producers, online systems reps, writers from many of the major online journals, and even the host of a weekly radio show called *Your Expanding InfoSphere* spent a weekend brainstorming to begin formulating criteria for evaluating databases.

A proposed rating scheme

SCOUG's task is to prepare a rating scheme for bibliographic, full-text, and directory databases from the *output* perspective; that is, to focus on those factors that cause the most problems to searchers and that hamper the ability to get the answers we need out of a database. The quality of the outputs from database searching is really dependent on a combination of three things: 1) the data themselves; 2) the database structure and organization; and 3) the system on which they are loaded. At the National Online Meeting, Martha Williams presented an expanded vision of this "information generation-database use chain." Problems with quality can come anywhere in a seven-link chain that includes, "1) author/originator, 2) primary publisher, 3) secondary publisher or database producer, 4) tertiary publisher or online vendor, 5) gateway to online services, 6) the searcher/analyst, and 7) the end-user or requestor." SCOUG's criteria focus more narrowly on the product created by the database producer and made accessible by the system.

Such a rating scheme may be used by anyone reviewing a database or it could be used to develop a separate rating publication that would be-

come the *Consumer Reports* of databases. Searchers would use the ratings to determine which databases on which online systems to search. Producers and vendors could use the ratings and accompanying evaluations to improve their products.

Rating scheme components

So far SCOUG has identified 11 main components that will help a professional searcher judge quality: 1) consistency; 2) coverage/scope; 3) timeliness; 4) value in terms of cost; 5) accuracy/error rate; 6) accessibility; 7) system performance/ease of use; 8) integration with other databases; 9) output; 10) documentation; and 11) customer support. (End users probably perceive value and ease of use to be most important, but in reality all of these factors affect the quality of the end user's search experience.)

Consistency describes the internal consistency of record structure and data elements within a database. For example, if abbreviations are used in some fields, are they always the same and are they logical and predictable? (Retreat attendees felt that if abbreviations are used, the online system should be able to automatically recognize spelled-out versions as well.) Are the same fields present in all records? Are bibliographic elements such as title consistently input as given in the original publication?

Coverage/scope encompasses 1) how well producers tell you what will be in a database by publishing a clear editorial policy, and 2) how well they follow that policy. If they claim cover-to-cover indexing of a group of journals, are they really indexing everything? Also, databases must be evaluated by how comprehensively they cover a given topic or span of time.

Timeliness refers to how frequently a database is updated on any system as well as what time lags there are before material is put into the database. A database may score high on one aspect of timeliness, but low on the other. Some bibliographic databases that are regularly updated monthly are a year or more behind in



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

their indexing. As with coverage, databases should also be ranked on how well they disclose their policies for priorities and how consistently they follow their policies.

Value in terms of cost is a more subjective criterion, but certain factors can be evaluated. Many of these factors are inextricably tied to the online system, demonstrating why each database must be evaluated separately for each online system it is on. For example, is there an understandable and fair pricing policy on the online system? Are there any upfront or subscription fees? Are there free output formats for this database that facilitate browsing or relevance judging? Can searchers select from a variety of cost options so they can control the costs? Do they get a lot of information for their money?

Accuracy/error rate is what most of us think of when we think of database quality. SCOUG members feel it is only one aspect of the broad picture of quality. This is an especially tricky area to measure. How can a reviewer accurately estimate the number of harmful typographical errors, for example, in a database with millions of characters? Even more difficult is judging accuracy levels in a directory database. Some of this judgment must be subjective impressions based on experienced searching. More objective measures can focus on the quality control measures used by the producer and its willingness to make corrections. Does an online system give credit to searchers who locate errors or who have unsatisfactory search results because of them? Does the system allow a searcher to view the inverted index through a command like DIALOG's Expand or Orbit's Neighbor so errors can be identified? The first online system that implements a free "Fixit" command to allow searchers to report errors online (and gives them rewards for doing so) earns a ten on this scale.

Accessibility can refer to factors of physical accessibility such as equipment required or special software required. It also refers to the more important area of information accessibility based on the fields that are present within a database, the availability of controlled vocabulary indexing, and the amount of information in a record. On the system side, search features such as proximity operations, truncation, automatic singulars and plurals, automatic term

equivalencies for common terms, and online thesauri all impact the quality of accessibility. For full-text databases, some ranking features were seen as especially important so large search results could be ranked according to likelihood of relevance.

Performance includes system-wide factors such as response time, baud rates supported, and hours of availability. Since ease of use is subjective, it mostly focuses on the variety of levels of interaction available.

The ideal system supports simple menus on up through fast commands for the experienced user. (Ideal in the future are systems that learn about users' knowledge and preferences each time they search.)

Integration refers to the ability of this database to integrate with other databases on the system. Consistency of field tags and field values certainly play a part here, as do integrated thesauri or linked vocabulary. The BRS LINK command that allows full-text

Avoid traffic jams in advance.



ADVANCE™, Geac's automated library system avoids pileups by keeping circulation information current. ADVANCE keeps lending, renewals, holds, recalls, returns and fine payments on the right road.

Geac

The Power of Working Together

United States (703) 836-0225

Canada (416) 475-0525

ONLINE DATABASES

and bibliographic records to be linked across databases is an example of good integration. The presence of sister files that are linked to each other, e.g., Information Access Company's Magazine Index and Magazine ASAP, rate highly. DIALOG's duplicate detection is another system feature that will rate high on integration.

Variety and flexibility of *output* options are essential to a successful search experience. Users must be able to define output formats, including no-charge formats that allow relevance judging or that encourage browsing. Records should be aesthetically pleasing (no all caps anymore!) and clearly presented. Output options such as sorting, stacking prints, choosing output media, and tagged records for downloading should be supported. In addition, full-text files should support Key-Word-in-Context and page and paragraph browsing. If graphics or tables are omitted, there at least should be captions describing the missing material.

Timely, accurate, and readable *documentation* should be both online and in print (and even distributed on disk for loading on a local micro). Internal documentation within records in a database helps explain editorial policies or link to related records. Journal lists, editorial policies, thesauri, and indexing procedures should all be documented and available to users.

A related issue is *customer service*. Toll-free numbers, accessible all of the hours the system is up, and with a knowledgeable person to answer questions is ideal. E-mail and online interactive help provide higher levels of service. System vendors and/or database producers should provide training classes as well as video- or microcomputer-based training. Low-cost practice files and learning incentives such as a free half-hour are valuable services. Database producers and online vendors must be responsive to customers' needs and willing to fix what needs fixing.

Using the rating scheme

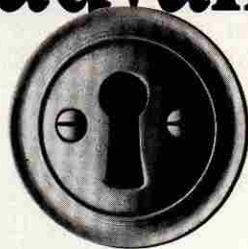
Details under each of these 11 factors will be given a numerical score (e.g., one through ten) to see how close to ideal the database and system comes. Much like an *Info-World* software rating, each database can then be given a composite score. One database may score low in a category such as timeliness, but make up points in another category such as

comprehensiveness of scope. A frequently updated database may not have the best indexing so the respective point totals will reflect that trade-off. Each database will be evaluated separately for each system it is on, so it may have different composite scores depending on the power and friendliness of each system.

SCOUG will present the first draft of its rating scheme and possibly some prototype evaluations at the SuperSearcher's "Quality Watch"

Forum at the Online/CD-ROM '90 meeting on November 7. All interested searchers are encouraged to provide feedback and to volunteer to review databases using the finished criteria. Contact Barbara Quint, 932 11th St., Suite 9, Santa Monica, CA 90403 or via DIALMAIL with comments, suggestions, or to volunteer to review databases. Martha E. Williams called quality the most important database issue of the 1990s. It's about time.

Find the key to cataloguing in advance.



ADVANCE's cataloguing feature holds the key to improved productivity. ADVANCE™ supports national and international cataloguing standards. ADVANCE opens other doors by letting you create additional formats to meet local needs.

Geac

The Power of Working Together

United States (703) 836-0225

Canada (416) 475-0525

Copyright of Library Journal is the property of Library Journals, LLC 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.