Aggregated Interdisciplinary Databases and the Needs of Undergraduate Researchers

Barbara Fister, Julie Gilbert, Amy Fry

Draft of a paper published in portal: Libraries and the Academy 8.3 (July 2008): 273-292

Aggregated Interdisciplinary Databases and the Needs of Undergraduate Researchers

Abstract

After seeing growing frustration among inexperienced undergraduate researchers searching a popular aggregated interdisciplinary database, the authors questioned whether the leading interdisciplinary databases are serving undergraduates' needs. As a preliminary exploration of this question, the authors queried vendors, analyzed their marketing materials, surveyed librarians and students, and examined what titles were being downloaded at 14 liberal arts institutions. While librarians are satisfied with these databases and vendors intend to continue the trend of adding more content, actual usage patterns suggest that these databases are not serving the purpose one might expect. Librarians should learn more about user experiences to influence the development of these products.

Introduction

When librarians at Gustavus Adolphus College compared notes after teaching introductory library sessions for fall 2006 first-semester students, we realized we all had found students more frustrated than in the past with the results retrieved in the aggregated interdisciplinary database that we had considered a good starting point. Though students would retrieve many citations, they were having a harder time than in previous years finding sources they felt were relevant. The inclusion of highly-specialized science, technology and medical (STM) articles, in particular, seemed noticeably higher than in previous years. As a result, it was growing harder to persuade our students that library databases could provide better resources for their research than Google.

This anecdotal impression led us to pose a question: do the current market leaders in aggregated interdisciplinary databases provide a product that serves the needs of undergraduates? We recognize that these databases serve needs beyond an undergraduate audience, but our question is focused specifically on undergraduate users. To approach this question, we queried vendors, analyzed vendors' marketing materials, surveyed librarians about desirable traits for interdisciplinary databases for undergraduates, conducted a limited survey of students, and reviewed usage patterns of interdisciplinary

databases at 14 undergraduate institutions. By examining the issue from multiple perspectives, we hoped to arrive at a clearer understanding of these resources and their suitability for undergraduates' research needs – or, at least, to identify areas that bear closer scrutiny.

Considering the importance of these databases and their cost to academic libraries (or, more commonly, to the consortia that provide them to an entire state or region), there is remarkably little published on the topic. Tiana French[1] and Naama Tal[2] questioned the usefulness of scholarly journal articles found in these databases for high school and community college students, who they believed are more likely to find overviews and less-technical information in books rather than articles. Péter Jascó called for clearer and more accurate information about the contents of these aggregated databases. [3] Shawn Kelly Blessinger and Maureen Olle described differences in the title lists of the leading interdisciplinary databases in 2001 and 2002 and recommended that librarians develop a more systematic understanding of database contents.[4] Shawn Lombardo and Kristine Condic reported that students had a great deal of difficulty finding articles in print format; as a result, nearly half of their subjects relied on full text articles for their research, a percentage that may well have grown since 2000, when the study was conducted. [5] Carol Tenopir found that students and librarians both value quality content and convenience, but librarians believe students in particular value full text articles and a familiar interface.[6] Janice Lewis and John McDonald, who felt the needs of undergraduates were being overlooked by libraries engaged in periodical cancellation projects, drew up a core collection of 2,100 serial titles for undergraduate collections; one of the three criterion they used was inclusion of the titles in an interdisciplinary database because they "actively target the undergraduate student population as a primary market."[7] Whether undergraduates remain a focus for these databases is open to debate.

Two critics suggested these general databases were simply too big. According to Scott Dennis, "[d]atabase vendors have a regrettable tendency to emphasize quantity of content over *quality*."[8] Mick O'Leary suggested "we may be at a point of diminishing returns with these mega-databases; they stretch the envelope so far we may be better off contracting it instead of expanding it further."[9] He particularly criticized them for "adding increasingly specialized and obscure journals."[10] Strikingly, these criticisms

were published in 1999 and 2001. The number of titles indexed in the largest interdisciplinary database produced by Gale has nearly doubled since that time; the number of titles indexed by EBSCO's most exhaustive product has nearly tripled.[11]

The Nature and Growth of Aggregated Databases

Before online databases were available for end-user searching in libraries, there were two distinct tools for finding articles. Specialized indexes and abstracts such as *Psychological Abstracts* or *Music Index* aimed for depth of coverage within a particular discipline. General indexes such as Wilson's *Humanities Index* and *Social Sciences Index* were gateways for the non-specialist and covered a limited number of scholarly journals – those most commonly found in libraries. Subscribing libraries voted on which titles were to be included, receiving ballots periodically by mail. These indexes were limited in scope for practical reasons; interlibrary loan was prohibitively slow and costly, and including more journals would raise the production costs of the indexes. But beyond practicalities, they were intentionally selective. Because they were aimed at a general audience, they sought out titles that were significant but not highly specialized.

As libraries adopted online content, the practical reasons for limiting the number of journals indexed were no longer pressing. Expectations of how much information should be available in a convenient form rose steeply with the popularization of the World Wide Web. Interlibrary loan became faster and easier for the end-user. Finally, the addition of full-text access changed the identity of databases fundamentally from a finding tool to a kind of aggregated subscription, a library-supported Web of previously-published articles.

At the same time, deciding which journals would be included was no longer left up to librarians. Publishers' role in deciding what they would provide, at what cost and under which terms, has shaped the contents of aggregated databases. All of these factors have contributed to the growth of databases.

For many years, the leading database vendors have competed on several measures. Vendors compare themselves to the competition based on how many of their titles are full text, whether there are embargoes on current titles and, if so, how long the

embargos last, how far back their backfiles go, how many of their titles are unique, and how many of their titles are peer reviewed. Telephone and e-mail interviews with vendor representatives confirmed that these measures remain important points of comparison.[12] Controversies over embargoes, exclusives, how "peer reviewed" is actually defined by aggregators, and other complexities erupt on library discussion lists from time to time.[13] Yet strikingly, the two leading aggregated interdisciplinary databases have consistently done one thing: they've gotten bigger, and seem to be constantly competing to offer the biggest interdisciplinary database.[14] (See Table 1.) Rather than focus on a limited number of core journals as the interdisciplinary indexes of the past did, the leading aggregated databases appear to be attempting to create gateways to as many publications as possible, building a metasearch product that includes highly specialized trade and scholarly publications. Representatives from both Gale and EBSCO indicated that they expect to increase the number of titles indexed and to increase the full text content in their databases. They also reported that both Gale's Academic OneFile and EBSCO's Academic Search Complete are focusing on enhancing subject coverage by including more content indexed in major subject databases and adding more STM publications.[15]

Table 1: Approximate number of titles indexed, October 2007[16]

Humanities Index in print (Wilson)	550 titles
Social Sciences Index in print (Wilson)	600 titles
General Science Index in print (Wilson)	280 titles
Reader's Guide to Periodicals in print (Wilson)	400 titles
OmniFile Mega Edition (Wilson)	4,000 titles
Academic Search Complete (EBSCO)	9,500 titles
Academic OneFile (Gale)	11,000 titles

At the same time, both companies have responded to the Google challenge by developing Web portals for their database content in order to reach users who do not begin a search through library Web sites. Gale's AccessMyLibrary debuted in June 2005;[17] EBSCO entered the field with EBSCOHost Connection in April 2006.[18] Contents of their databases can now be crawled by Web search engines, with links

connecting users who log on to their libraries – if those libraries have activated their accounts. However, as of this writing, these products have received scant attention from either librarians or end users and so far have had little visibility in search engine results.

Who Are Vendors' Customers?

Libraries are only one player in the economics of databases. A full text database put together by a vendor has to be attractive to both content providers (publishers) and to the potential audience (whose needs are represented by librarians and library consortia who make subscription decisions). Furthermore, vendors' parent companies produce products and provide services other than databases, and those markets may influence their business practices.

In the case of EBSCO, the vendor's relationship with publishers partially relies on the fact that the company is also the world's largest subscription agent, handling both print and electronic journal subscriptions for libraries. As Sam Brooks, Senior Vice President of Sales and Marketing, said in an interview in 2002, "we would not do anything to destroy our core business."[19] In 2004, Brooks warned libraries that canceling subscriptions to journals included in full text databases is a form of "misuse" that could have a devastating effect.

> "'Customer misuse' refers to customers who cancel a particular journal specifically because it is available in full text via the database. Databases offer many benefits, but this is not one of them. Vendors whose customers have 'misused' their databases have paid for it in the form of massive turmoil, i.e., full text journals becoming halted or removed completely from the database . . . Basically, full text databases can only continue to exist if they are a complement to journal subscriptions."[20]

In short, he argued that if libraries were to provide full text to their students, they should also continue individual journal subscriptions in order to subsidize content in aggregated databases, regardless of student preference. Recently, the company's Website began to reflect a changing perspective, arguing that e-content sales are a growing percentage of the market and so are crucial to publishers' interests.[21]

Cengage Learning, known as Thomson Gale until July 2007, owns Gale and its InfoTrac suite of databases. It also owns a number of K-12 and higher education textbook publishers that bundle with their textbooks limited "free" access to the same database content marketed to libraries. Students buying a college textbook from a Cengage Learning publisher get a password for four months' access to InfoTrac College Edition, a repackaging of a library product that the company promotes directly to students (or rather, to faculty who adopt the textbook), claiming the resource is notable because it's available 24/7 and free with a textbook purchase – though, unlike library subscriptions, access expires four months after activation.

The Website for this bundled product suggests it is better than what is provided by libraries. A professor is quoted saying he uses it for his research, "saving myself countless hours at the library or with other electronic searches which provide article abstracts only." According to another testimonial, "I find it difficult to take time to make a special trip to the library every time I need information. InfoTrac College Edition enables me to search easily from home or office, and at my convenience. In addition, as a doctoral student who is doing graduate level research, I am able to quickly put my hands on the most valuable and current resources."[22]

In effect, Cengage's textbook divisions use misleading statements about libraries to compete *with* libraries by offering the very same content libraries have licensed from the company on behalf of their students and faculty. The catalog copy for these publishers' textbooks claims that a purchase includes access to an "online university library," a staggeringly overblown description of a single interdisciplinary database that begs the question of why students shouldn't use the online resources provided by their own library. InfoTrac is not alone in this practice. EBSCO, too, has bundled their database content with textbooks through an arrangement with Pearson, the world's largest textbook publisher.[23]

It's worth noting that the practice of bundling content with textbooks has been criticized by the U.S. Government Accountability Office for raising textbook costs.[24] It must be aggravating for students to find that material bundled with an expensive textbook is already available in their library. Perhaps even more aggravating for libraries is having

vendors tell their students and faculty that their “free” product is superior to what the library offers.

What Do Librarians Think?

In March 2007 the authors surveyed academic librarians about opinions of large interdisciplinary databases and their observations of how undergraduates use them. The survey, created with SurveyMonkey and announced on several online discussion lists (including COLLIB-L, ILI-L, Colldv-L, and SERIALST) received 565 responses.

A majority of respondents worked directly with students in reference and instruction; a majority also had collection development responsibilities. Respondents were asked about their satisfaction with one of what were at the time of the survey the three leading general, interdisciplinary databases: EBSCOhost’s Academic Search Premier, Gale’s Expanded Academic ASAP, or Gale’s newer product, Academic OneFile. (At the time the survey was conducted, EBSCO had not yet added Academic Search Complete to their product line.) Respondents were also asked about their perceptions of how well these databases meet undergraduate research needs. Though the majority (79.2%) of respondents addressed their responses to their experiences with Academic Search Premier,[25] frequency and select cross tabulations for each database found no significant difference between responses focused on Academic Search Premier alone and responses to all three databases combined. One out of five respondents reported their library was considering subscribing to another interdisciplinary database, either to augment or replace their current offerings, while over a third of librarians said they already subscribed to two or more of them.

Overall, librarians expressed a surprising level of satisfaction with interdisciplinary databases. It may be the case that vendors have listened to their librarian customers and have aligned the development of their products to customer feedback very successfully. Almost all respondents (91.3%) were satisfied or very satisfied with the database they use. Database search features met with satisfaction with over half of respondents, a third of them reporting they were “very satisfied.” A sizeable number of

librarians (40.9%) reported being more satisfied now than when they first used the database.

The feature librarians reported appreciating most in an interdisciplinary database was full text availability, though nearly half of respondents would like to see more full text content added. The second most appreciated aspect was the search features offered, with the interface itself coming in third. Among least favorite features, interface was the most chosen – though, perhaps as an indication of how satisfied librarians are with interdisciplinary databases, it was chosen by slightly under 18% of respondents; more than one third of respondents did not select any feature as their least favorite. Most respondents (over 84.1%) also felt that the number of journals indexed was about right. Eleven percent thought there were too few titles, and only 4% felt there were too many.

Presented with a list of enhancements proposed by the authors (see Table 2), librarians were generally not enthusiastic about them. Of the proposed features, librarians were most interested in the ability to limit a search to a particular discipline, with over half finding it either somewhat or very desirable. Librarians were less enthusiastic about other proposed enhancements. More than half of respondents either felt all but the first proposed limit were "not useful" or they were unsure that they would be useful.

Table 2: Desirability of Additional Features (in percentages)

FEATURES	Very Desirable	Somewhat Desirable	Not Useful	Not Sure
Limit by Discipline	23.1	39.8	13.3	23.8
Limit by Expertise	15.8	25.3	26.5	32.4
Limit to Core Journals	16.8	33.1	24.1	26.0
Limit to Locally-Determined Set of Journals	9.0	26.8	34.8	29.4
Results Ranked by Algorithm	11.4	30.4	28.2	29.9

These results suggest that librarians are happy with the databases they have, feel they are the right size, and do not require additional limits. However, open-ended comments raised concerns that focused on inadequate indexing and full text availability. The following comments illustrate frequently-mentioned issues.

- "Indexing is terrible."
- "Horrible subject heading consistency."
- "I wish all of them would offer better lists of subject headings as they did years ago. Not all students do well with key words."
- "Many titles are claimed as full text, but specific issues and articles are missing."
- "I hate the fact that many full text [journals] have a delay making it impossible to get current articles."
- "I find that their definition of 'scholarly journals' sometimes doesn't match librarians' or teachers' definitions."

What Do Librarians Think Undergraduates Want?

Our survey also explored librarians' observations about undergraduates' use of interdisciplinary databases. (See Table 3.)

Table 3: Features Librarians Consider Important to Undergraduates (in percentages)

Feature	Very Important	Somewhat Important	Not Very Important	Not At All Important	Not Sure
Large Number of Journals	60.7	32.1	5.7	0.2	1.2
Multidisciplinary	28.6	47.1	20.2	2.9	1.2
Full Text Articles	98.6	1.0	0.0	0.2	0.2
Current Articles	70.5	27.1	1.4	0.0	1.0
Smaller Number of Scholarly Journals	4.0	21.2	42.6	21.4	10.7
Familiarity	67.0	27.8	3.6	0.5	1.2
Simultaneous Search of Popular and Scholarly	24.0	40.7	26.7	5.0	3.6

Almost all respondents (98.6%) believed finding full text articles is important to students. Second in importance was having a familiar database to search and being able to find current articles. Being able to search a smaller set of scholarly journals than that found in more specialized databases was perceived as the least important of options.

We also asked respondents to rank the challenges students face when searching an interdisciplinary database (see Table 4.) The most commonly-chosen issues relate more to students' limited research skills rather than the characteristics of databases themselves. Almost all respondents (84.6%) agreed that undergraduates have difficulty choosing good search terms, with over half of respondents saying this is "very often" a challenge. The second most-cited problem was identifying relevant results, with over three-fourths of respondents finding this often or very often a problem. The majority of respondents (65.5%) believed students wrestle with having too many results for their searches. Over half of librarians, however, believed that retrieving articles that were too technical or not of high quality was only sometimes or rarely a problem for undergraduates. Consistent with their general satisfaction with interdisciplinary databases, librarians seemed to believe students' problems are attributable to the students' lack of sophistication as researchers, not the databases.

Table 4: Librarians' Perceptions of Challenges Facing Undergraduates (in percentages)

Challenges	Very Often	Often	Sometimes	Rarely	Not Sure
Too Many Results	31.2	34.3	29.8	4.3	0.5
Trouble Identifying Relevant Results	42.5	34.9	20.0	2.3	0.2
Articles Are Too Technical	12.9	19.3	51.2	15.0	1.7
Results Not of High Quality	5.0	14.3	55.8	22.3	2.6
Difficulty Choosing Good Search Terms	53.7	30.9	14.3	1.0	0.2
Topics Too Broad/Articles Too Focused	27.3	32.8	33.0	4.8	2.1

Finally, some open-ended responses reiterated librarians' perceptions that interdisciplinary databases are particularly well-suited to an undergraduate audience.

- "They are a good starting point for students starting out doing research."
- "They are excellent for the inexperienced searcher because pertinent results are easily retrieved when using fairly good key words and limits. Being able to retrieve usable results is especially encouraging to novice students."
- "The general interdisciplinary databases are as close to a Google-like experience as you can get with a single database, and Google is where the majority of our students start their searches."
- "Students tend to use whichever database they try first for everything, whether it is appropriate for that or not. Undergrads like the broad databases; they feel comfortable with them."

What Do Students Say?

Though conducting a similar survey of undergraduates fell outside the scope of this project, we were able to include some questions about database use and preference in a survey of Gustavus students enrolled in four sections of a freshman-level public speaking course in May 2007. Responses from 46 students tended to support librarians' characterizations of undergraduate approaches to database searching, though this small group of students reported more interest in finding high quality sources and less concern about currency than one would expect based on the survey of librarians. The majority of students reported they started their research for projects in their public speaking course with Google, but nearly 40% reported they started with a database. This finding may reflect explicit preferences stated by course instructors as well the impact of the librarian-led instruction session each section received.[26]

When asked in an open-ended question what they liked about their favorite database, the word "easy" was most often used, though many students mentioned being able to find high-quality credible sources and being able to find sources on almost any topic in one place. Slightly over half of students reported an aggregated interdisciplinary

database – Academic Search Premier – as the database they used the most. (See Table 5; it's worth noting that Gale's academic interdisciplinary databases are not available at our campus, so were not included.). LexisNexis was surprisingly popular, being cited as the first database choice of 28% of students. Curiously, students who reported they started their research with Google were more likely than others to choose LexisNexis as the database they used the most; students who started with a database were more likely to use Academic Search Premier the most. The factor that was most important to them in choosing a database was its familiarity, with convenience and authority rated as more important than currency (see table 6).

Table 5: Which database do you like the most? (choose one)

Academic Search Premier	25
Communication and Mass Media Complete	0
JSTOR	4
LexisNexis	13
Proquest Newsstand	1
Other (ERIC supplied by respondents)	2
Other (Wikipedia supplied by respondent)	1

Table 6: Which factors most influence you? (choose one)

Least amount of **time** to track down information	7
Most **convenient** – at place/time (e.g. dorm or lab)	8
Most **current** – I need the most up-to-date information possible	3
Most **authoritative** – gives the most reliable, complete information	8
Most **familiar** – "tried and true," has worked for me in the past	18
Most reliably **available** – no waits or hassles	0

When presented with a list of possible frustrations, getting too many results was the most commonly selected, though students were nearly as frustrated by having difficulty coming up with good search terms and having trouble obtaining the articles once identified. Twenty percent of students were frustrated by retrieving irrelevant articles. Only four students reported they were frustrated by having too few results. (See table 7.)

Table 7: What frustrates you? (check all that apply)

It's hard to get the actual articles	15
I get too many results to sift through	17
It's hard to come up with search terms that work	16
I don't get enough results	4
The articles are too technical	3
The articles are too narrowly-focused for my topic	2
The articles aren't relevant	9

Though the small number of responses and the fact that all of the students were enrolled in one course at a single institution limits the conclusions one might draw, these findings do suggest that, while students and librarians agree on some issues – for example, that choosing search terms is a challenge and that familiarity breeds database acceptance among students – they may disagree on the value of database size.

A large-scale study of undergraduate experiences in various courses on a number of campuses would be valuable. For example, it's interesting that many of the students in our limited survey were frustrated by getting too many results in a database search, given we rarely hear complaints that Google offers too many results; a larger survey could compare student experiences and levels of satisfaction with search engines that may return millions of hits (as a Google search often does) and large, interdisciplinary aggregated databases. Another interesting issue would be to compare undergraduate experience of Google Scholar with their use of interdisciplinary aggregated databases, given both products attempt to provide an ever-growing collection of academic resources, with many (but not all) linked to full text.

Because of the limits of our student survey, we were unable to confirm our hunch that interdisciplinary databases, as they add more content, are actually becoming less useful for undergraduates than previously as a place to start their research. However, an examination of searches and full-text downloads from an interdisciplinary aggregated database at 14 libraries serving undergraduates did provide some surprising results.

Which Full-Text Titles Do Students Actually Use?

In addition to impressions about interdisciplinary databases from vendor, librarian, and student perspectives, we were curious about how these databases are used. Vendors were able to give us an idea of how many publications were currently included and in general terms their plans for the future: more titles, more full text, more peer-reviewed journals, more subjects covered, more backfiles, more full text content for journals indexed in other databases, and more interconnectivity with Web search engines. But which titles are actually used by undergraduates?

To approach this question, we posted a message on the OBEREF e-mail list and then systematically contacted librarians at the 80 liberal arts colleges that are members of the Oberlin Group,[27] requesting statistics on interdisciplinary database use. Because these databases do not offer any way of limiting usage data to a specific class of users, we used this group of libraries because they tend to focus their services and collections on undergraduate users. Though these colleges emphasize fields of study that are traditionally considered "liberal arts," many offer a variety of pre-professional programs as well, such as majors in nursing, education, social work, athletic training, management, and accounting. Furthermore, though member institutions do not make claims that their students are, like Lake Wobegon's children, "all above average," it seemed a safe assumption that their students would not be unusually underprepared; in other words, we felt these libraries would not provide data that falsely under-represent undergraduate research sophistication.

The data we requested from these libraries were COUNTER Journal 1 and Database 1 reports for in Excel format for 2005 and 2006. The COUNTER standard (Counting Online Usage of Networked Electronic Resources) originated in 2003 and provides a set of international guidelines for normalizing the recording and reporting of usage statistics from online resources. Journal 1 reports include numbers of full text downloads for each title in a database, by month. Database 1 reports include numbers of searches (user-executed database queries) and sessions (discrete connections to database resources).[28] Because these numbers are calculated identically for each product, using COUNTER reports was the only way to be certain that usage across different institutions

was being recorded the same way. However, COUNTER reports have only been available for EBSCO products beginning in calendar year 2005, so we were limited to examining only two years of data.

We were able to gather and compile usage data from 14 institutions with enrollments ranging from approximately 1,500 to 4,700 FTE, four of which have some graduate programs.[29] Eight are located in the Midwest, 5 in the Northeast, and one in the Southwest. Out of these, only one sent usable statistics for a Gale product, so our findings by necessity focus on searches conducted in the interdisciplinary database most commonly found in these libraries, Academic Search Premier. The number of searches conducted at the various colleges in 2005 varied from under 9,000 to over 114,000 and the range of articles downloaded ran from approximately 13,000 to almost 92,000. We were surprised to see that, though the number of sessions increased at all but three libraries in 2006, the number of completed searches and downloads overall did not. In fact, the number of searches completed fell from 2005 to 2006 at half of the libraries (though the average for all 14 institutions rose slightly). More surprisingly, the average number of downloaded articles declined more than 10% overall. Eleven of the 14 colleges saw the number of downloaded full text articles fall, even though the number of indexed and full text titles included in the database rose (see Table 8.)

Table 8: Averages at 14 Colleges Subscribing to Academic Search Premier

	Searches	Sessions	Downloads
2005	60,392	25,693	53,182
2006	61,568	29,409	47,164

Several factors may contribute to this startling finding. It is possible that, though more database connections were made, students found the citations returned from their searches in Academic Search Premier more dissatisfying in 2006 than 2005 and thus chose to view and save fewer articles from their search results. On the other hand, the data may indicate that, due to increased user sophistication because of improved library instruction programs and database-side enhancements in search and retrieval technology,

students' searches were more successful and they were better able to readily identify relevant choices in their results, resulting in fewer downloads overall. (Were this the case, however, we might expect to see more downloads coming from more scholarly titles, rather than a continuing reliance on popular magazines.)

Academic Search Premier may also be leading users to useful citations that are not full text in that database but which students then obtain through another source (whether it be online, in print or through interlibrary loan), facilitated through open-URL linking. In such an instance, a decline in full-text downloads does not indicate that the database is less successful at leading students to desirable sources. We are also not sure what impact federated search systems may have on these numbers; when results from general databases are compiled with those from subject-specific resources, students may find more relevant results elsewhere. Unfortunately, COUNTER data does not allow us to see any specifics of database usage beyond total connections, searches and full text downloads from within the database itself, so further research is needed to pinpoint why use appears to be decreasing even though database content (both indexing and full text access) is increasing.

To examine which journals were receiving the most use, we combined the COUNTER Journal 1 reports from all 14 schools (taking care to match titles that, because of spelling or other discrepancies in the title list, did not collate correctly) and examined those that accounted for the majority of full text downloads – those that fell in the top 50% overall - to see what titles were downloaded most often, whether or not these titles are peer reviewed, and what subject areas encompass most full text downloads. These results also confounded our expectations.

In 2005, a mere 178 of the 4,700 full text titles included in Academic Search Premier accounted for over 50% of articles downloaded from that database. Fewer than 90 of these titles were scholarly (determined by looking at their classification in the database's title list); articles from these scholarly titles only accounted for 36% of the articles downloaded in the top 50%. Likewise, in 2006, 185 journals accounted for over half the database's full text downloads, 95 of which were classified as scholarly, with downloads from these journals accounting for 38% of full text downloads in this group. (It should be noted that the number of downloads from publications included among

these “peer-reviewed” titles is actually lower since a number of titles such as *Natural History* and *The Humanist* appear on the list as “peer-reviewed” when, in fact, they are not.) In 2006, over 1800 journals did not have a single full text download at any of these libraries. In other words, nearly 40% of full text titles did not have a single article downloaded at *all* of the 14 institutions combined.

Popular publications, including magazines, newspapers, and library trade journals, account for the majority of downloads. The title with the most downloads in both 2005 and 2006 was *The Economist*, downloaded 37,948 times total, and topping the list at most schools. (See Table 9.) Less unexpectedly, the two most prominent U.S. news weeklies, *Time* and *Newsweek*, were also on the list of the ten most downloaded titles; in 2006 *US News & World Report* joined them as it rose from 11th place to 10th. Also among the top ten are newspapers (*USA Today* was the second most downloaded title both years;

Table 9: Ten Most Frequently Downloaded Titles

2005 Titles	**Downloads**	**2006 Titles**	**Downloads**
Economist	21,563	Economist	16,385
USA Today	12,911	USA Today	12,395
Library Journal	10,741	School Library Journal	8,608
Time	9,125	Library Journal	8,013
School Library Journal	8,963	Foreign Affairs	7,532
Newsweek	8,649	Newsweek	7,392
Foreign Affairs	7,915	Time	7,004
Chronicle of Higher Education	7,197	Chronicle of Higher Education	6,434
Christian Science Monitor	6,728	Social Work	5,995
Social Work	6,236	U.S. News & World Report	5,190

Christian Science Monitor dropped from ninth to eleventh place in 2006). Perhaps most surprising of titles in the top ten were the *Chronicle of Higher Education*, and two library publications: *Library Journal* and *School Library Journal*. We surmise that, in general, popular publications with short articles published frequently are likely to return more

results than scholarly journals that publish fewer articles less frequently. We also surmise that the two library titles are on the list because of the sizable number of brief book reviews they publish on a wide variety of subjects. How well that content satisfies users' needs is a question worth asking.

Only two scholarly titles appeared in the top ten: *Foreign Affairs*, which ranked sixth in 2006 and fifth in 2005 and *Social Work,* which ranked tenth in 2005 and ninth in 2006.

Looking at all of the titles in the top 50% of titles downloaded both years, about half of those labeled "peer-reviewed" in Academic Search Premier are journals in the social sciences; the remainder are spread among science and medicine, the humanities, and interdisciplinary areas. Among popular magazines, the majority of titles are non-scholarly publications focused on specific topics (e.g. science, health, religion, education, or business), though news and opinion magazines and newspapers dominate the most downloaded non-scholarly titles.

While looking beyond the top half of downloaded titles would doubtlessly show increased granularity in terms of subject areas and number of titles used at succeeding levels, such an exercise is unlikely to lead to different conclusions about the nature of the content that students are using – overwhelmingly more popular than scholarly, and more geared towards social sciences than humanities, despite any potential curricular differences across the institutions surveyed. In fact, the patterns of use were remarkably consistent across the institutions, regardless of their different populations and courses of study.

Our examination of usage statistics raises many questions that deserve further study. What impact do library instruction programs and the availability of subject-specific databases and federated search tools have on aggregated database use at individual institutions? How do curricular differences and the nature of assignments influence database use? Would we find students are choosing more scholarly sources using aggregated databases than it appears if open-URL statistics (linkthroughs to full text and interlibrary loan requests) were examined? Are the titles that are not used at undergraduate institutions more frequently used at research institutions, or do advanced researchers rely more on specialized databases? Are there discernable patterns among the

nearly 40% of full text titles that weren't downloaded? Should the drive for increasing full text content in these interdisciplinary aggregated databases take a backseat to other improvements, such as improving subject access for the titles already included? We need more insights into information-seeking preferences and actual usage patterns in order to better understand and evaluate these large interdisciplinary databases.

Conclusions

Having examined the suitability of general interdisciplinary databases for undergraduate research from a number of angles – what vendors tell us through their marketing materials and in interviews, what librarians report, what a limited number of students have to say about databases, and by examining which titles are actually used – we can make some preliminary observations and suggest avenues for further exploration.

First, vendors are in a race to add indexing and full-text content to their product lines. Both EBSCO and Gale plan to continue building relationships with additional publishers to provide them greater visibility and impact by being represented in these databases. Relationships with participating publishers play a greater role in determining the size and contents of interdisciplinary databases than in the pre-digital age of general indexes. Both EBSCO and Gale assure librarians that they are committed to covering more subject areas and adding more scholarly titles to their databases; in turn, librarians want more full text content and are pleased with the size of these databases, which have grown enormously in the past few years. It would be wise for librarians to be more knowledgeable about the contents of these databases and their actual use as these products are developed.

Second, academic librarians by and large do not share the authors' impression that interdisciplinary databases are becoming less well-suited to undergraduate researchers' needs. Librarians are extremely satisfied with these products. They recognize that students have difficulties with a variety of search tasks, especially with selecting effective search terms and identifying relevant results, but they feel being able to search a large number of journals is important to students. The one area of dissatisfaction appears to be in the inadequacy of controlled vocabulary, which may relate to the reported difficulty

our small sample of students had in coming up with effective search terms. Evaluating the controlled vocabulary available in these databases and how it affects users' search success would be a worthy avenue of research.

Whether these databases, as they grow in size, remain particularly well suited for undergraduates as a place to find information easily is open to debate. Though students surveyed did express a desire to have a single easy-to-use database that could be searched for any subject, they also were frequently frustrated by retrieving too many results. When we examined how these databases are used at 14 primarily undergraduate institutions, we learned that only four percent of full-text titles included accounted for half of the articles downloaded. Ten times as many titles did not have a single article downloaded at all 14 institutions, leading us to question the value for undergraduates of continually adding more titles without first exploring why so many are not being used. Comparisons of students' attitudes toward and use of Google Scholar and their experiences with these all-purpose databases would particularly interesting. How do these interdisciplinary aggregated databases compare with free options? Do their methods of retrieval and full-text contents compare well or poorly with Google Scholar? Is the free competition improving as more scholars and publishers embrace open access initiatives and, if so, how should that affect our decision-making about interdisciplinary aggregated database subscriptions?

How our instruction efforts affect student use of these databases is another intriguing question. It might be argued that the problem is not that a large portion of a database is unused, and that scholarly sources in particular are overlooked, but rather that our instruction efforts are failing to guide students toward a more diverse range of scholarly sources. Additionally, it may be that changes in faculty expectations for student research are having an effect, perhaps driving greater student use of specialized databases. Certainly, both the librarians' observations of student search behavior and the list of sources actually used underscore the need to help students gain a better grasp of how to frame their questions and how to select sources. But in any case, given the preponderance of non-scholarly articles downloaded, our findings challenge the received wisdom that students who use library databases are automatically more likely to use

scholarly and high-quality sources in their research in contrast to students who conduct their research on the Web.

Finally, in spite of their overall satisfaction with aggregated interdisciplinary databases, librarians do not know very much about them. As we attempted to gain an understanding of trends in these interdisciplinary databases, we found it surprisingly difficult to track relatively simple measures, such as how many titles were indexed and how many were included in full text over the years. Not only are the ways these numbers are arrived at disputed, they simply are not available in any readily-retrievable form. As one vendor representative told us "we do not maintain copies of older, out-of-date subject title lists."[30] To complicate the picture, both companies offer a variety of different products and both have launched ambitious expanded products within the past two years. Neither EBSCO nor Gale was able to tell us how many academic libraries subscribe to their products, instead relying on vague generalities. And naturally, the cost of these databases is not publicly known since prices are set through undisclosed negotiations, quite often through consortia. All of these factors make determining return on investment difficult – because libraries know neither the investment nor the return.

When asked about use of periodicals, both print and electronic, over 77% of librarians responding to our survey told us that the use of their print collections was declining. Sixty-six percent said use of interdisciplinary databases was increasing at their library, and 19% reported they weren't sure. Only one respondent reported use of general databases was declining. These impressions are inconsistent with one of the most startling discoveries we made when looking at how general databases are used at 14 undergraduate institutions: the number of searches completed declined at half of the schools and the number of titles downloaded was down at 11 out of 14 institutions, with an overall drop of more than 10%.

Vendors conduct research to inform their product development, but for reasons of competitive advantage it is rarely shared with the librarians who work directly with users.[31] Librarians, who as a rule value making information openly available, have not insisted on clear and consistent information from vendors, nor have they routinely examined students' use of these products with the tools available.

It is unfortunate that we were unable to obtain data from more schools covering more years to gain a better sense of what the apparently drop-off in use means, or to have data that factors in the use of other databases, but that is another symptom of librarians' lack of knowledge. Comparable usage information using the COUNTER standard has been available from one of these major database vendors only since 2005. To add to the problem, tracking and analyzing this data is not easy; most of the 80 Oberlin Group libraries were unable to provide comparable month-by-month statistics. Additionally, consortial contracts, often made at a state-wide level, offer advantages to libraries and their users, but may have the unintended effect of distancing librarians from cost/benefit decisions, allowing us to simply make assumptions rather than informed decisions. Given the disparity we have shown between librarians' impressions and how these databases are actually used, librarians should strive for greater accountability from vendors – and from ourselves.

Academic librarians are dedicated to helping students learn how to find, evaluate, and use information. It's ironic that we are so ill-informed about the contents of the aggregated interdisciplinary databases we provide, often at substantial taxpayer expense, and how our students use them. Perhaps it's time to apply the information literacy skills we teach as we make decisions about our resources so we can do a better job of defining our needs, evaluating our options, and making responsible information choices.

---

[1] Tiana French, "Scholarly Articles and Community and Junior College Students: An Uneven Match?" *Community and Junior College Libraries* 13, no. 1 (2004): 13-23.

[2] Naana Tal, "Which One to Choose? A Comparison Between Three Aggregators," *Knowledge Quest* 34, no. 3 (2006): 24-29.

[3] Péter Jascó, "A Proposal for Database 'Nutrition Ingredient' Labeling, *Database* 16, no. 1 (1993): 7-9.

[4] Kelly Blessinger and Maureen Olle,"Comparison of Three Primary Aggregator Databases," *The Serials Librarian* 45, no. 1 (2003): 53-58; "Content Analysis of the Leading General Academic Databases," *Library Collections, Acquisitions, & Technical Services* 28 (2004): 335-346.

[5] Shawn V. Lombardo and Kristinie S. Condic, "Convenience or Content: A Study of Undergraduate Peridical Use," *Reference Services Review* 29, no. 4 (2001): 327-337.

[6] Carol Tenopir, "Database Use in Academic Libraries," *Library Journal* 124 (1999): 36, 38.

[7] Janice Steed Lewis and John D. McDonald, "Defining an Undergraduate Core Journal Collection." *Serials Librarian* 43, no. 1 (2002): 45-59.

[8] Scott L. Dennis, "Aggravating or Aggregating? Providing Effective Access to Contents of Aggregator Dtabases: A Reference/Collection Development Librarian's Perspective," *Cataloging and Classification Quarterly*, 28, no. 4 (1999): 17, emphasis in the original.

[9] Mick O'Leary, "Big Databases Pose Big Questions," *Online* (May/June 2001): 82.

[10] Ibid, 83.

[11] It's surprisingly difficult to trace database size over time. Companies refashion databases, develop products with different names, and do not retain outdated title lists online for comparative purposes. Growth figures from 2001 can be found in O'Leary and from title lists on company Web sites referring to EBSCO's Academic Search Complete (undated title list, http://ebscohost.com/thisTopic.php?marketID=1&topicID=633 , accessed October 6, 2007) and Gale's Academic OneFile (title list dated October 1, 2007, http://gale.cengage.com/title_lists/, accessed October 6, 2007). Based on these sources, in terms of titles indexed, EBSCO's most complete database nearly tripled in size (from 3,390 indexed titles to 9,524) and Gale's nearly doubled (from 6,111 to 10,962).

[12] E-mail interviews were conducted with Jason Bass of Thomson (at that time the parent company of InfoTrac) and Tim Heiges of EBSCO in March 2007.

[13] See the archives of Liblicense-L, for example, found at http://www.library.yale.edu/~llicense/ListArchives/.

[14] The notable exception to this rule is Gale's Expanded Academic ASAP, which appears to have shrunk in both the number of titles indexed (from 4,216 to 3,390) and full text titles (from 2,516 to 2,089) between 2005 and 2007. (The 2005 figures are found in an unpublished report of the Infotrac Alternatives Task Force of the Tri-College [Bryn Mawr, Haverford, and Swarthmore] Library Committee dated May 2005; the 2007 figures come from Gale's Website.) There is no explanation provided at Gale's Website for this change in coverage and, in fact, appears from a search of discussion list archives, to have gone unnoticed.

[15] Personal communications with Jason Bass of Thomson (at that time the parent company of InfoTrac) and Tim Heiges of EBSCO in March 2007.

[16] Figures were found on company Websites cited above and from http://www.hwwilson.com/sales/printindexes.htm and http://www.hwwilson.com/databases/omnifile.htm (accessed October 6, 2007).

[17] "AccessMyLibrary.com Puts Library-Only Content on the Web," *Search Engine Watch* June 16, 2005, http://blog.searchenginewatch.com/blog/050616-191102 (accessed October 6, 2007).

[18] "Update on EBSCOHost Connection," *EBSCO Support News* (September 2006) http://support.epnet.com/support_news/detail.php?id=289&page=4 (accessed October 7, 2007).

[19] Emily McElroy, "An Interview with Sam Brooks," *Serials Review* 28, no. 2 (2002): 144.

[20] Michael G. Enyart, "An Interview with Sam Brooks, Senior Vice-President of Sales & Marketing, EBSCO Publishing," *Journal of Business & Finance Librarianship* 10, no. 1 (2004): 36.

[21] "Tools that Drive Content Delivery," http://www2.ebsco.com/en-us/ForPublishers/contenttools/Pages/index.aspx (accessed October 6, 2007).

[22] *InfoTrac College Edition: User Comments,* http://infotrac.thomsonlearning.com/ (accessed October 6, 2007).

[23] *Pearson Education Launches Research Navigator ™ with The New York Times and EBSCO* (2003) http://www.pearsoned.com/pr_2003/010803.htm (accessed October 6, 2007).

[24] United States Government Accountability Office. *College Textbooks: Enhanced Offerings Appear to Drive Recent Price Increases.* GAO-05-806. Washington, DC: GAO (2005), http://www.gao.gov/new.items/d05806.pdf (accessed October 6, 2007).

[25] Fifty-one percent of respondents worked at libraries that did not subscribe to either of the Gale products. Of the 36% of respondents whose libraries subscribed to both EBSCO and Gale general databases, the majority chose to focus on the EBSCO product in their responses.

[26] That said, other studies are challenging the notion that students almost always begin their research with Google. A study conducted in early 2007 at Saint Mary's College of California found that a majority of subjects (upper division undergraduates) started their research by consulting course materials and library resources, with fewer students starting with a general search engine. See Alison J. Head, "Beyond Google: How do Students Conduct Academic Research?" *First Monday*12, no.8 (2007). http://www.firstmonday.org/issues/issue12_8/head/index.html (accessed February 16, 2008).

[27] The Oberlin Group was founded in 1986 when directors of a number of liberal arts college library directors were invited to meet at Oberlin College. It remains an unincorporated organization of liberal arts college libraries that "exists for discussion, the sharing of ideas, collegiality, the sharing of statistics, and other cooperative activities that these directors are empowered by their institutions to

undertake" according to Ray English and Will Bridegam in "A Brief History of the Oberlin Group," (2007), http://www.oberlingroup.org/brief-history-oberlin-group (accessed October 22, 2007).

[28] "About Project Counter," *COUNTER*, http://www.projectcounter.org/about.html (accessed October 6, 2007).

[29] The 14 colleges reporting data were Berea, Colgate, College of Saint Benedict/Saint John's University, DePauw, Grinnell, Gustavus Adolphus, Hamilton, Kenyon, Macalester, Sarah Lawrence, Simmons, Smith, Saint Olaf, Trinity University

[30] E-mail interview with Jason Bass of Thomson (at that time the parent company of InfoTrac) , March 2007.

[31] An exception was a presentation made by John Law, Joanna Markel, and Serena Rosenhan of Proquest who shared their research on student use of Web and library resources at the 2007 Electronic Resources & Libraries conference ; see Amy Fry, "Several Comments on Different Sessions," *Library Hi Tech News* 24, no. 4 (2007): 9-13.