

Full-Text Aggregation: An Examination Metadata Accuracy And the Implications For Resource Sharing.

By Joel Cummings

Submission to Serials Review October 19, 2002.

Author Information Page

Joel Cummings
Electronic Resources Librarian
Owen Science and Engineering Library
Washington State University

Address: Owen 139
Washington State University
Pullman, Washington
99163

Telephone: 509-335-6493
Fax: 509-335-2534
e-mail: [jcummings@wsu.edu](mailto:jcumings@wsu.edu)

Full-Text Aggregation: An Examination Metadata Accuracy And The Implications For Resource Sharing.

By Joel Cummings

Abstract

The author conducted a study comparing of two lists of full-text content available in Academic Search Full-Text Elite. EBSCO provided the lists to the University College of the Fraser Valley. The study was conducted to compare the accuracy of the claims of full-text content, because the staff and library users at University College of the Fraser Valley depend on this database as part of the libraries' journal collection. Interlibrary loan staff routinely used a printed list of Academic Search Full Text Elite to check whether the journal was available at UCFV in electronic form; therefore, an accurate supplemental list or lists of the libraries electronic journals was essential for cost conscious interlibrary loan staff. The study found inaccuracies in the coverage of 57 percent of the journals sampled.

Full-Text Aggregation: An Examination of the Full-Text Content

Introduction

Electronic databases containing aggregations of full-text serials are nearly ubiquitous resources in libraries and a great benefit. They represent an ease of use and speed of access for library users that cannot be found with traditional print collections. For libraries with modest serials, these aggregation databases offer a large serials collection at a reasonable cost. Large libraries can also expand their serials collections through to the cost-effectiveness of electronic access to serials that are provided by such aggregators.

Even with aggregation problems exist for managing library collections and services. These problems are common to most bibliographic databases, such as (1) incomplete indexing of journal titles, (2) inconsistent indexing of individual serials, (3) incomplete or missing information about the contents of the index, and (4) inaccurate metadata itself. Full-text aggregators add a series of issues related to content that is no longer limited to metadata, such as articles that are incomplete (missing tables, graphs, pictures etc) or reduced quality of high-resolution images. Although these problems can occur with imperfect metadata in non-full-text databases, the consequences are different because the index is the collection and not restricted just to metadata. For example, incomplete articles that include only text and lack images and figures may require library users to resort to document delivery services to obtain the necessary information or the library user is quite likely to use what is found in the database and miss potentially relevant information.

Additionally, the contents of databases have a dynamic character. Journals may be frequently added or removed from these databases. This perpetual flux may create concerns for librarians. Some research has shown there is reason to caution against the canceling of print serial subscriptions because of the title's availability in an aggregator.^{1,2} Many libraries create databases, lists or title-level access to journal aggregators in the library catalog. They may also use OpenURL enabled technology such as SFX to maintain and create access to library materials regardless of format.

Transient content is another problem inherent in aggregation databases. Unlike the original publisher of the full-text content found in the aggregator's database, the aggregator does not necessarily have any direct relationship with the authors. . This introduces a level of uncertainty and impermanence into the agreements between publishers and aggregators and the aggregator will not necessarily obtain licenses for permanent archival access. Most notably, the recent United States Supreme Court decision that required Lexis-Nexis remove content from the New York Times that had been written by freelancers.³ Competition amongst aggregators and competition or fear of competition between aggregators and publishers lead to access barriers such as 'moving walls', where content is available to the aggregator only after a contractually agreed upon time period after the initial publication of an issue of a serial.

The UCFV is a small multi-campus college that has recently evolved from a community college and vocational training institution to expand its mission as an undergraduate degree granting institution. UCFV, has three libraries and five campuses. At the time of this study the total number of students enrolled in credit courses was 6,519 amounting to 4,146 FTE. UCFV also offers a large number of non-credit courses⁴. With the expansion of the academic program at UCFV the student population has also increased steadily. The number of faculty has also increased and many new faculty positions have required to have a Ph.D. with research as an expected part of their responsibilities at UCFV. Prior to the institution's shift from a junior college to a 'University College' the instructors were not expected or required to do research and frequently held only master's or bachelor's degrees. This change in mission also created growing demands on the library's serial collection and interlibrary loan services. A multidisciplinary database containing significant amounts of full-text scholarly journal literature was a great asset to a small undergraduate college library. UCFV depended heavily on interlibrary loans and document delivery to supplement its traditional and electronic journal collections.

The University College of the Fraser Valley, British Columbia (UCFV) maintained a list of electronic journals at <http://www.ucfv.bc.ca/library/ej_home.htm>. Lists such as this one are

designed with the library users in mind but are used by the interlibrary loan staff to verify that loan requests are not locally available in electronic form. In the latter case, having an accurate list of the locally available electronic journals is essential. EBSCO provided the interlibrary loan staff at the University College of the Fraser Valley Library with a printed list of the database Academic Search Full-Text Elite. The staff checked for the local availability of journal and magazine articles. An online list of serials was also available but it did not allow for quick checking of individual titles; therefore, it was too time-consuming to be useful. Staff also used printed lists for the other full-text aggregation databases that the library had access to, such as ABI-INFORM and CBCA (Canadian Business and Current Affairs). The staff could then inform library users that the article was available in electronic form via the library. If the list contained a large number of inaccuracies two errors would result. First, interlibrary loan staff would be liable to assume articles that could not to be found in the library collection could be found in the library collection. Second, interlibrary loan staff would be liable to assume that articles that could to be found in the library collection could not to be found. In both cases the inaccuracies of this supplied information would create regrettable delays in service and in the latter case cost the library money. UCFV also participated in a user initiated document delivery program which compounds these problems.

When searching by subject the library user is not necessarily going to notice what is not being retrieved in a search, especially if the library user is looking at large result sets or the library user is very inexperienced. The library user or the interlibrary loan staff who are looking for known items may waste their time and experience frustration when searching for articles that are not in a database, even though the desired title and the issue are supposed to be included in full-text. Complete issues may be missing, full-text of a few indexed articles may be missing, and thereby create a situation analogous to the problems of irregular serial runs. This shortcoming of aggregation databases shares essentially the same consequences as incomplete indexing.

Unlike licenses negotiated through consortia or by individual libraries with publishers, the

aggregator may have great difficulty maintaining consistent access to titles within their database, potentially leaving the library without access to important local periodicals or titles that are in high demand. A library could develop claiming procedures for full-text aggregators, but since the library is not always subscribing to specified periodicals through their license agreement with full-text aggregators, such procedures are not likely to be as efficacious as with traditional methods of claiming. Some programmatic collaboration between publishers, aggregators and libraries may be able to efficiently resolve this question. If an issue of a serial were missing from a library collection, the library would be likely to claim missing issues or failing that the library would be likely to indicate in their catalog that the library is missing particular issues or volumes of a journal. For aggregation databases, the further problem of missing articles or sections of issues complicates the claiming endeavor.

Purpose of Study

This study is intended to compare the full-text coverage of journals found in Academic Search Full Text Elite versus two lists of the database coverage available to the librarians at the University College of the Fraser Valley (UCFV), library staff and UCFV's library users during March and April of 1999. The two lists are Academic Search's complete title list, that was available within EBSCOhost's gateway for registered users and a printed listed entitled "EBSCOhost Database Coverage List" that was being used by UCFV library staff and students. The URL to the former list required a login and therefore lacked direct access for all. Both lists were created by EBSCO and the printed list being used at the time was created in August 1998. This study is concerned with the accuracy of the lists of full-text coverage because the staff and library users at UCFV depend on this database as part of the libraries' journal collection.

The library did not systematically use 856 fields in the bibliographic record to provide to links to electronic journals. The interlibrary loan staff routinely used EBSCO's printed list of Academic Search Full Text Elite to check whether the journal was available at UCFV in electronic form; therefore, an

accurate supplemental list or lists of the libraries electronic journals was essential for cost conscious interlibrary loan staff

Anne Knowlan, the quondam UCFV Library's Head of Public Services, submitted the results of this study to the British Columbia Electronic Library Network (ELN). ELN was the organization responsible for negotiating affordable site licensing and consortium discounts for electronic information resources, such as Academic Search Full-Text Elite, for partner libraries across British Columbia. This study was designed to give feedback to ELN about the products that they licensed for member libraries and potentially to provide feedback to EBSCO about their product.

Methodology

The full-text journals to be examined were selected from the Academic Search 'Journal search' screen within the database. This is a separate list from the title list on EBSCOhost's website. The journals covered in Academic Search could be listed alphabetically and displayed twenty titles at a time, with the first title being the last title on the previous screen. Therefore, nineteen new titles are listed on a page. The first title on each screen that had some full-text content listed in EBSCOhost's complete title list was studied to determine whether both aforementioned lists were accurate. One title per screen was selected. Each journal's coverage was compared to the two lists' claims. The range of both the indexing & abstracting and the full-text coverage found in the database was compared to the lists.

Results:

The basic question of this study is: are the claims of full-text content from the two lists provided by EBSCO accurate or not.

- A total of twenty-eight journals were examined.
- Eleven of the journals contained differences of coverage from the UCFV library's printed list

and/or EBSCOhost's electronic list.

- Seven of the journals contained more coverage than the lists indicated.
- Four of the journals contained less coverage.
- Four of the coverage differences were inaccurate when compared with UCFV's printed list only.
- Six of the coverage differences were inaccurate in both UCFV's and EBSCOhost's lists.
- One of the coverage differences was inaccurate on EBSCOhost's list yet accurate on UCFV's printed list

Please see Table 1 for details about individual journals. Minor errors were noticed, such as missing issues or lack of full-text for an issue(s) or an article(s) or incorrect bibliographic information. Some errors were observed in a total of sixteen of the twenty-eight journals, or fifty-seven percent journals tested.

Conclusions

The coverage of Academic Search often included more full-text than was claimed. UCFV library may have incurred interlibrary loan costs for articles that were available within the UCFV library system because of the inaccuracies. Using the database itself as a guide was far more time consuming than using the printed lists; therefore, the staff opted to use the printed list. The UCFV library would have benefited from a more frequently list, as this was presumably a source some of inaccuracies between the lists and the database. An electronic list that is easy to use online or printable and is continually updated by EBSCO and is more accurate than the list available then would have presented a solution. Simple and efficacious lists are now available online⁵ in both Excel and HTML files. Currency is still a concern, since the list is updated quarterly and many changes can occur during a single quarter.

A second concern is the high rate of missing issues or missing full-text of an issue and other inconsistency that are noted in the results. A third concern is the problem of icons not appearing on the

screen in one search yet if the search is repeated these icons appear. Some public service librarians working at UCFV at the time of this study have noticed missing print/email/save button, refine search button, result list, and the full-text icons on occasion. Newer versions of EBSCO's search interface seem to have eliminated this problem -- based on anecdotal observations by the author.

This study was limited in its scope. No attempt was made to determine the completeness or consistency or accuracy of indexing or abstracting within the database. No patterns were discovered that indicated why individual articles that were supposed to be full-text were not full-text in the database. In general, no pattern was observed for missing individual issues of full-text. Although in some cases, the coverage ranges were off by a year or a few issues, either starting before the listed date or ending before the stated date.

To provide seamless access to electronic collections libraries need to develop methods of tracking and providing access to those electronic materials. Full-text aggregation databases represent a significant source of information for library collections and present libraries with significant challenges to provide reliable consistent access to information⁶. Providing library users with a convenient source of information about all serials available via a library's collection is a high priority for libraries. This study has demonstrated one such challenge, the inaccuracies that exist in the metadata of one aggregation database.

Table 1

The term accurate refers to date range of full-text content. Other errors, such as single missing issues are noted under comments.

Journal Name	Accurate versus the printed list	Accurate versus the online list	Comments
ABA Journal	No	Yes	Full-text started January 1993. Printed list it was listed as starting January 1994.
Adaptive Behavior	No	No	No full-text was available after Winter97/Spring98 issue. Full-text coverage was listed as until present.
Advanced Materials & Processes	Yes	Yes	
Agricultural Research	No	No	Full-text coverage started January 1993, but September 1993 issue was missing. Full-text was listed as starting January 1994.
Alternatives: Social Transformation & Humane Governance	Yes	Yes	
American Enterprise	No	No	Included two issues (July/Aug 1993 & Nov/Dec 1993). Full-text coverage was listed as from January 1994 to present.
American Journal of Agricultural Economics	Yes	No	Printed list stated full-text coverage is from January 1990. Online list stated coverage is August 1990. Full-text coverage started February 1990.
American Journal of Psychotherapy	Yes	Yes	
American Midland Naturalist	Yes	Yes	
American Sociologist	Yes	Yes	
Annals of the American Academy of Political & Social Science	No	Yes	No full-text coverage indicated on the printed list. Spottiness in full-text coverage was observed. One article missing in July 1990, November 1993 issue is not full-text and only part of May 1992 issue is full-text.
Anthropology Today	Yes	Yes	
Architecture	Yes	Yes	

Art Bulletin	Yes	Yes	Spotty full-text. One article in December 1997 was not full-text and a portion of the September 1998 issue was not full-text.
Asian Philosophy	Yes	Yes	
Austrialian Feminist Studies	Yes	Yes	
Australian Science Teachers Journal	No	Yes	The printed list does not indicate full-text listings.
Behavioral Medicine	Yes	Yes	The Winter 1998 issue was missing. The Fall 1995 issue was listed as Fall 1996.
Biosciene	No	No	Full-text was listed as starting January 1984. Indexing and abstracting started October 1984.
Boston College Environmental Affairs Law Review	Yes	Yes	Missing full-text for 14(2), 11(1) to January 1990. Missing full-text and indexing of Fall 1992.
Business & Society	No	No	Before the April 1994 issue only spring issues were in the database and not all articles were full-text. Before V31(1) some articles did not have complete bibliographic citations -- no date, volume and these articles were not full-text.
CATO Journal	Yes	Yes	Jul/Aug 1998 issue seems to have the wrong volume and issue number.
Cambridge Journal of Education	No	Yes	Full-text was missing for issue 28 (1). Printed list did not indicate any full-text coverage.
Canadian Ethnic Studies	Yes	Yes	Full-text was missing for 23(2)
Canadian Review of American Studies	Yes	Yes	Full-text was missing for 26(2), one article from both 23(2) and 23(1) were not full-text and all of issue 21(2) was not full-text.
Chicago Review	Yes	Yes	
Chinese Science Bulletin	No	No	Full-text coverage was greater than listed, but the full-text coverage was spotty. There is full-text for issues 41(20, 22-24) and 42(3,10,15,21-22,24).
Clearing House	Yes	Yes	

Notes

- ¹ Anna Grzeszkiewicz and A. Craig Hawbaker. Investigating A Full-Text Journal Database: A Case Of Detection. Database 19(6): 59-62, December 1996.
- ² Nancy Sprague and Mary Beth Chambers. Full-Text Databases and The Journal Cancellation Process: A Case Study. Serials Review 26(3): 19-31, 2000.
- ³ Cary Kenney. Newspaper publisher in the post-Tasini era. Searcher 10(3) 50-59, 2002.
- ⁴ "Enrollment" <http://www.ucfv.bc.ca/physics/enroll.htm> (November 22, 2002)
- ⁵ "Bibliographic & Full Text Databases : Academic Search Elite" <http://www.epnet.com/academic/asearchelite.asp> (October 2nd 2002)
- ⁶ William A. Britten et al. Access to Periodical Holdings Information: creating links between databases and the library catalog. Library Collections, Aquisitions, & Technical Services 24(1) 7-20, 2000.