



Importance and Use of Holding Links between Citation Databases and Online Catalogs

by Jerry V. Caswell, Fred H. Gulden, Kathy A. Parsons,
Dennis C. Wendell, and William H. Wiese

This article describes a statistical study of three databases, conducted at Iowa State University, on holding links between citation databases and the online catalog. It provides insights into the construction and quality control of citation databases and online catalogs, and identifies a new issue of continued maintenance of ISBNs and ISSNs for catalogs.

Academic libraries building integrated information systems that extend beyond the online catalog need to consider the importance and use of linking citations in citation databases to local library holdings. One of the advantages of mounting citation databases on a local library system is that it opens up the possibility of tying the contents of citation databases to the library's holdings. For the purpose of this article, citation databases refer to bibliographic databases, including those with abstracts, that index the literature in one or more subject areas. These are also commonly called abstracting/indexing tools.

Although citation databases have been available in machine-readable formats for two decades, until recently users have had to perform a separate check of their local holdings for titles that are of interest to them. By the end of the 1980s, vendors of automated library systems recognized the possibilities inherent in mounting citation databases on local library systems.¹ Several of them redesigned their systems to accommodate multiple databases and to link references in the databases to the holdings in an online catalog.² Linkages are usually made through matches on ISBNs or ISSNs. If a record in the citation database contains one of these numbers in a properly coded field, it can be matched against a corresponding number in the record in the online catalog. This match can then be used to generate a message about the existence of the linkage and display the library's holdings.³

As libraries move into the world of transparent intersystem linkages through Z39.50 protocols, the concept of linking citation databases to local holdings takes on a new dimension: citations in remote databases, whether they be at a university half way across the country or at a library service supplier, such as OCLC or RLIN, can be linked to local holdings, if supported by

the local system's software. This means that a library can present through a single interface a group of databases with links to local holdings, even though those databases may come from a variety of institutional and extra-institutional sources.⁴

"As libraries move into the world of transparent intersystem linkages through Z39.50 protocols, the concept of linking citation databases to local holdings takes on a new dimension: citations in remote databases, whether they be at a university half way across the country or at a library service supplier, such as OCLC or RLIN, can be linked to local holdings, if supported by the local system's software."

The obvious benefit of these linkages lies in facilitating the work of students and researchers at their local institutions. With holdings information available during the search process, students and researchers can modify their search process based on the results. Undergraduates writing term papers might decide to select another topic if the citations found in the database are not well represented in the library's holdings. Advanced researchers might expand a search if insufficient results are found, or they might plan a trip to a remote site which houses strong holdings in their area of interest.

For the library there are other outcomes of having these linkages available. One outcome is a strong tool by which to evaluate the citation databases using the local

*Jerry V. Caswell, Fred H. Gulden,
Kathy A. Parsons, Dennis C. Wendell, and
William H. Wiese are at Iowa State University
Library, Ames, Iowa 50011-2140.*

collection. It is doubtful that a library would want to mount a citation database which was not closely tied to its collections. Determining the proportion of linkages between the two could be an important tool in deciding what databases an institution presents to its users through the local system interface.

The same mechanism could also be used to evaluate the quality of the citation databases. Are the ISBNs and ISSNs coded properly? Are they current? Have they been correctly entered? If the answer to each of these questions is yes, it may indicate that the database was carefully prepared and consistently maintained. Problems in this area may indicate quality control problems with the database in general.

It was the issue of quality control in relation to ISBN/ISSN numbers that prompted the investigation of holdings links at the Iowa State University (ISU) Library. Between 1992 and 1993, the library mounted eight citation databases on its local NOTIS system, SCHOLAR. Seven Wilson databases⁵ were made available to assist the undergraduate population in their course work and research. AGRICOLA, a product of the National Agricultural Library (NAL), was mounted to assist students and researchers in the field of agriculture. Within a few months of its installation, public service librarians reported that many materials in the AGRICOLA database were not showing holdings links to the online catalog even though the materials were known to be in the library's collections.

This prompted the Automated Systems Division to make a preliminary examination of the linking process to define the problems that prevented the links from occurring. Several conditions were identified, but it was not clear how pervasive they were in the context of the whole database. In order to provide substantive evidence, it was important that a statistically valid methodology be applied to the examination of the holdings links and resulting categorization of problems. To this end a study group with representatives from Public Services, Technical Services, and Automated Systems was formed. The group consulted with Professor William Q. Meeker of the ISU Statistics Department to determine what sampling methodologies, that have been applied to large populations, could be applied equally well to large databases. The answers, it turned out, were not complicated. Either random or systematic samples of 1,000 or more

records could be applied to an analysis of the databases. Larger samples could be applied if results warranted examination of ambiguous results.

The study group decided to apply systematic sampling because it was easy to generate the record numbers to be searched. Samples of 1,000 items per database allowed a manageable number of records to be searched and compared. The work was divided among several persons to ease the burden of searching. The results were tabulated and used to compute 95% confidence intervals (two standard deviations from the mean).

Over a period of six months, the study group sampled the holdings links for three databases, AGRICOLA (AGRI) and two Wilson databases, *Applied Science and Technology Index* (ASTI) and *Readers' Guide Abstracts* (RGAB). This enabled the group to study not only the character and proportion of linkage problems, but also to gain some insights into the differences in the structure and focus of the databases.

**"The process of correcting
ISSNs needs some prioritization
because it has to compete
with the large array of
responsibilities of catalogers."**

PRELIMINARY STUDY OF WILSON DATABASES

A preliminary study conducted while the Wilson databases were still in test mode indicated that three conditions prevented Wilson records from linking to online catalog records:

- A different ISSN was found in ISU catalog records;
- No ISSN was found in ISU catalog records; and
- No ISU catalog record was found for the title identified in the Wilson index.

The differences in ISSNs between Wilson records and ISU cataloging records appeared to derive from changes made to the ISSN over the course of time. These may have reflected changes in title, changes in publisher, or some other situation that required an updated ISSN number. ISU cataloging records had not been updated to reflect the changes because until now there had been no need.

The lack of an ISSN in ISU cataloging records was the result of several things. For one, ISU serials records had been maintained in a pre-MARC environment for many years as the means of producing a printed serials list. This environment did not require the detailed record structure of the MARC format; therefore, the serial records lacked many data elements including ISSNs. When the library implemented an integrated online library system in 1990, the pre-MARC records were converted to a MARC format and loaded into the new system. While some serial records have been cataloged via OCLC, not all of them have ISSNs, because the cataloging took place before an ISSN had been assigned and there was no need to update the record for that reason alone.

The absence of particular journal titles from the online catalog reflected the past collection development decisions of the ISU Library. The identification of titles which are present in the Wilson indexes, but not in the ISU collection, should help the Collections Division in the continuing review process.

It is important to observe that each holdings link failure in Wilson, observed in the sampling process, was due to only one problem. Insofar as the study group could determine, all problems were with the ISU online catalog, so correcting the problems is under the control of the library. Part of the complexity of the AGRI analysis hinged on the fact that more than one problem may prevent a link between a single citation record and the online catalog. Some of the problems that occur in this citation database are beyond the control of the library.

APPLIED SCIENCE AND TECHNOLOGY INDEX

The ASTI database contained 250,000 citations for a period from 1989 to early 1993. Use of the database during its first months of availability under SCHOLAR indicated that it was the third most heavily used of the seven Wilson databases mounted (after *Readers' Guide Abstracts* and *Social Sciences Index*) and supported nearly 10,000 searches a month during the academic year.

In the 1,000 record sample, 19% or 194 citations did not link to the online catalog. Of these 194 citations, 173 were represented by online catalog records while 21 of the citations were not in the ISU catalog. Fifty-one citations were represented by catalog records that contained different ISSNs; the remaining 122 contained no

Table 1
Holdings Links Sample Results

Category	AGRICOLA		ASTI		RGAB	
	Number	Percentage	Number	Percentage	Number	Percentage
Links OK	471	47.1	806	80.6	907	90.7
In OPAC	232	23.2	173	17.3	74	7.4
Not in OPAC	297	29.7	21	2.1	19	1.9
No IS?N in OPAC*	136	13.6	122	12.2	44	4.4
No IS?N in DB**	220	22.0	0	0	0	0
Diff ISBN/OPAC?	0	0	N/A	N/A	N/A	N/A
Diff ISSN/OPAC	20	2.0	51	5.1	30	3.0
Error in IS?N	5	0.5	0	0	0	0
Wrong Host Info	14	1.4	0	0	0	0
ISBN Qualifier	17	1.7	N/A	N/A	N/A	N/A
Links/IS?N Corr	163	16.3	173	17.3	74	7.4
Sample Size	1,000		1,000		1,000	

Notes: *IS?N is used for this category because AGRICOLA includes both ISBN and ISSN numbers but ASTI and RGAB contain only ISSNs.
**DB is used for citation databases.

ISSNs in ISU online catalog records (see Table 1).

The 173 unlinked citations represented 42 periodical titles; the 21 citations without corresponding catalog records represented 10 periodical titles. In order to discern whether certain titles accounted for a disproportionate number of failed links, we made a ranked list of journal titles by number of citations that did not link. While the majority of titles accounted for six or fewer failed links, three journal titles accounted for a large number: *Aviation Week & Space Technology* (21), *ENR* (18), and *New Scientist* (23). The good news resulting from this analysis is that it should be possible to correct ISSNs in online serial records that represent approximately 17% of the records in ASTI. This would result in a 98% holdings link rate for citations in this database.

The process of correcting ISSNs needs some prioritization because it has to compete with the large array of responsibilities of catalogers. The ranking of periodical titles by number of citations that did not link can provide guidance for catalogers about which titles to correct first. Each online catalog record, that is updated with a current ISSN, has the potential of enabling links for hundreds (or in some cases thousands) of records in the citation databases. Perhaps the most outstanding example is *Aviation Week & Space Technology*, which accounts for 17,833 citations in ASTI, 7,538 records in *Readers'*

Guide Abstracts, and 6,640 records in *Wilson Business Abstracts*. All of these records would be affected by the correction of the ISSN in one online serials record. There are few activities in the field of cataloging that could have such immediate and dramatic impact on the use of the system.

In sum, the problems that cause holdings link failures between ASTI and the online catalog appear eminently manageable. The correction or addition of up-to-date ISSNs to a small number of ISU catalog records will enable 98% of the ASTI records to link to the OPAC. The quality of the Wilson citation records appears to be extremely high. None lacked ISSNs, and the ISSNs present seemed to be complete and up-to-date.

READERS' GUIDE ABSTRACTS

The RGAB database contained 636,000 citations for a period from 1983 to early 1993. During its first months of availability it became the most heavily used of the seven Wilson databases and supported over 13,000 searches a month during the academic year.

In the 1,000 record sample, 9% or 93 citations did not link to the online catalog. Of these 93 citations, 74 were represented by online catalog records while 19 of the citations were not in the ISU catalog. Thirty citations were represented by catalog records that contained different ISSNs; the remaining nine contained no ISSNs in ISU online catalog records.

The 74 unlinked citations represented 21 periodical titles, while the 19 citations without corresponding catalog records represented nine titles. A ranked list made of journal titles by number of citations that did not link indicated that only one title accounted for a large number of failed links: *Aviation Week & Space Technology* (22). The majority of titles accounted for seven or fewer links.

As with ASTI, it appears that correcting ISSNs in online catalog records will result in a high holdings link rate for citations in RGAB. The difference in the number of non-linking titles between the two indexes is due in large part to the number of titles indexed. ASTI indexes more than 500 titles compared to RGAB's 236.

PRELIMINARY STUDY OF AGRICOLA

AGRICOLA contains citations for journal articles, patents, dissertations, monographs, audiovisual materials, computer files, and government publications in the fields of agriculture, forestry, biotechnology, food and nutrition, and other related areas. It includes materials acquired by the NAL and cooperating institutions, such as the Food and Nutrition Information Center. It is a primary source for publications of the U.S. Department of Agriculture, Agricultural Experiment Stations, and Cooperative Extension Service.

Because of AGRI's size and complexity, it was expected that there would be a

variety of problems associated with it. A preliminary study was conducted to define the holdings links problems. After a keyword search on the term "biotechnology" was performed, the first 300 of the resulting 8,147 citations were examined for linkages. In addition to the three types of problems discussed with the Wilson products, the study revealed the presence of five more types of problems:

- A different ISSN was found in ISU catalog records;
- A different ISBN was found in ISU catalog records;
- The ISBNs or ISSNs do not match because they contain an internal error;
- An error in the NOTIS matching routine prevents matches when an ISBN has a textual qualifier;
- No ISBN or ISSN was found in ISU catalog records;
- No ISBN or ISSN was found in the AGRI citation record;
- The Host Item field (773) in the AGRI record does not point to the entity cataloged in the ISU catalog; and
- No ISU catalog record was found for the title identified in the AGRI citation record.

When a different ISBN was found in ISU catalog records, it was usually because the library has a different edition of the monograph identified in the AGRI record. This is a problem inherent in linking monographic works to the periodical literature and is not easily solved nor is it a frequently occurring problem.

When ISBNs or ISSNs did not match because one of them contained an internal error, it was usually due to a data-entry error. This sort of error is easy to remedy, once the correct number has been verified.

When an ISBN had a textual qualifier such as "acid-free paper" or "soft-cover," an error in the NOTIS matching routine prevented a match from being made. NOTIS Systems Inc. has reviewed this problem and has modified the database loader to strip out the textual information. NOTIS has advised its customers that they must modify their online catalog records in the same fashion to make the matching routine work properly, a process that will involve a considerable amount of custom programming.

Large numbers of AGRI records lacked ISBNs or ISSNs because they represented non-commercial publications (i.e., gov-

ernment documents, dissertations, and maps) to which such numbers are not normally assigned and which are often not cataloged. Their presence in AGRI is a mixed blessing, which will be discussed in more detail in the next section.

The Host Item field problem in AGRI was the result of a policy decision made by the Indexing Section of the NAL.⁶ When a monograph belonged to a series, the Indexing Section had chosen to reference the series rather than the monograph in the 773 host item field; the monographic catalog information (without ISBN) was referenced only in a 500 note field. Trying to correct this situation from ISU's end would be problematic because adding the series ISSN to the ISU monographic cataloging record would link AGRI records for different monographs in the series to the ISU cataloging record.

"If the purpose of mounting a database on a local library system is to provide access to the world-wide agricultural research literature and to link that literature, with a high degree of accuracy, to library holdings, then AGRI does not fill that role very well. Individual libraries will need to assess their particular requirements in relation to mounting a low linking database on a local library system."

The eight problems identified in AGRI represent a more complex set of issues than those found during the study of the Wilson records. Unlike the Wilson indexes, many of the AGRI records lacked ISBNs or ISSNs or contained errors. This meant that neither the citation database nor the online catalog could be taken as an authority for the information lacking in the other. It also meant that more than one error may cause a holdings link failure. Correcting an error in one database may not necessarily enable a link to be made.

AGRICOLA

The 629,000 records in the AGRI database covered the period from 1987 to early

1993. The database has been in use on the SCHOLAR system since January, 1992, and at the time of the study it supported more than 11,000 searches a month.

In the 1,000 record sample, 53% or 529 citations did not link to the online catalog. Of these 529 citations, 232 were represented by online catalog records while 297 of the citations were not in the catalog. Only 20 citations were represented by catalog records that contained different ISSNs; the remaining 136 contained no ISBNs or ISSNs in the record. In addition, 220 AGRI records contained no ISBN or ISSN. Eighty-four of these had corresponding ISU online catalog records; the remaining number did not, thereby adding a level of complexity to the analysis.

No examples of different ISBNs turned up in the systematic sample, although this problem had been noted during the preliminary study of the database. The sample did turn up five citations that exemplified internal errors in ISBNs or ISSNs, 14 with wrong Host Item information, and 17 where textual qualifiers in ISBNs caused a failure to link. These are judged not to be significant problems, although they may affect some users more than others. In the preliminary study, which was specific to the subject of biotechnology, the wrong Host Item problem was found in more than 10% of the 8,147 records.

Because of the large number of linking failures, an estimate was made as to how many citations would link, if errors were corrected, missing ISBNs and ISSNs were added, and a software solution was found for the ISBN qualifier problem. In addition to correcting obvious errors, it was assumed that, if an ISBN or ISSN appeared in either the AGRI or an ISU catalog record, one was available. On this basis it was determined that 163 of the 232 non-linking AGRI records in the sample could be made to link to the online catalog. This would bring the total possible links in the sample to the 63% level. That still left a significant portion of the database without the possibility of linkages, either because the materials were not represented in the online catalog or because there were not ISBNs or ISSNs available to create linkages.

At this point, it is important to remember that the AGRI database is not a standard indexing tool, but a record of the holdings of the NAL with indexing information generated from its periodical collection. Because NAL has a highly specialized mission to serve the U.S. Department of Agriculture as well as to

collect English language materials relating to agriculture, food, and nutrition, the database is better understood as an expression of that mission. In practical terms, this means that it will always contain a large number of materials which academic libraries do not collect or catalog.

In order to determine whether the materials in the sample of 297 AGRI records that are not represented in the ISU online catalog are normally not cataloged, the study group spent some time categorizing them and established 20 categories. Each AGRI record was assigned to the first applicable category. Of these categories, the following contained more than 20 records: Conference/symposium (48), trade publications (42), research journals (38), government documents, U.S. (31), extension service publications (24), and non-English language publications (22). The large number of research and trade publications would lead one to expect a higher incidence of representation in the ISU catalog.

Government publications were strongly represented in AGRI. Because they do not normally carry ISBNs or ISSN, unless published by commercial sources, government publications in AGRI will not link even if records exist in the online catalog. As a result, users would have to conduct a separate search in the online catalog.

In summary, AGRI does not lend itself as effectively to holdings links as do the Wilson indexes. This is not attributable to any single problem, but to a group of issues that bear on the unique composition of the database. It is not realistic to expect as high a percentage of linkages from a database representing the holdings a research institution as from a database which indexes the periodical literature. If the purpose of mounting a database on a local library system is to provide access to the world-wide agricultural research literature and to link that literature, with a high

degree of accuracy, to library holdings, then AGRI does not fill that role very well. Individual libraries will need to assess their particular requirements in relation to mounting a low linking database on a local library system.

MISCELLANEOUS CONSIDERATIONS

Even for records that have holdings links, there are some problems with the linkage process that affect its utility. ISBNs and ISSN allow linkages to be created only at the title level, not the holdings level. This means that some records in a citation database which display the holdings link message may not be present in the library's collections either because the library did not receive that particular issue or the library has canceled the title. This situation requires some alertness on the part of the user to carefully peruse the holdings data of the library.

"As the possibility of linking remote databases to local holdings through Z39.50 protocols becomes a reality, holdings linkages will assume an even more important place in the functionality expected of a local library system."

SUMMARY

The sampling of three citation databases for holdings links revealed a number of important facts about the databases and the potential uses of links. First, it demonstrated variability in the quality and composition of citation databases. Second, it provided insights into the consistency, accuracy, currency, and scope of the databases. Third, it highlighted the need for continuous maintenance of ISBNs and ISSN in the library's online catalog. This

process can be expedited by appropriate sampling and ranking of non-linking titles.

As the possibility of linking remote databases to local holdings through Z39.50 protocols becomes a reality, holdings linkages will assume an even more important place in the functionality expected of a local library system.

REFERENCES

1. For a general discussion of issues relating to locally mounted databases, see William Gray Potter, "Expanding the Online Catalog," *Information Technology and Libraries* 8 (June 1989): 99-104; Emily Gallup Fayen, "Locally Loaded Databases: Some Factors to Consider," *Library Acquisitions: Practice and Theory* 14 (1990): 347-353. An alternative was suggested in David Cohen, "A National Networked Solution to Improving Access to Journal Articles," *Journal of Academic Librarianship* 15 (May 1989): 79-82.
2. Vicki Bloom Bakowski and Lisa E. Moeckel, "The Impact of Local Tape Databases on the Library: The M(I)DAS Touch," *Online* 14 (July 1990): 38-42; Ramona J. Steffey, "The NOTIS Multiple Database Access System: A Look Behind the Scenes," *Online* 14 (September 1990): 46-49.
3. This and other methods of linking citation database records to the OPAC are discussed in Susan Barnes and Janet McCue, "Linking Library Records to Bibliographic Databases: An Analysis of Common Data Elements in BIOSIS, Agricola, and the OPAC," *Cataloging and Classification Quarterly* 13 (1991): 165-170.
4. NOTIS' InfoShare product supports this. See NOTIS Systems, Inc., *InfoShare and MDAS: Expanding Your Access to Databases* (NOTIS Systems, Inc., n.d.).
5. The seven Wilson indexes were *Applied Science and Technology Index*, *Biological and Agricultural Index*, *General Science Index*, *Humanities Index*, *Readers' Guide Abstracts*, *Social Sciences Index*, and *Wilson Business Abstracts*.
6. Shirley J. Edwards, Memo on "New Indexing Policy" to Indexing Branch Staff, National Agricultural Library, March 19, 1987.