
Bibliographic Control and Agriculture

SARAH E. THOMAS

ABSTRACT

BIBLIOGRAPHIC CONTROL OF agricultural publications is a complex and costly process. Many key agricultural documents are difficult to catalog because they contain insufficient information about the author, publisher, or other essential bibliographic elements that are needed to describe the item. There is also no single source in which to locate all citations to a particular topic in agriculture. AGRICOLA, AGRIS, and CAB ABSTRACTS are the three major databases that provide journal-article-level analysis of agricultural topics. The overlapping relationships among these databases and their unique features are described. Although the labor to obtain control over the voluminous and expanding numbers of agricultural titles is great, the user is hindered in accessing this material by differing formats, varying vocabularies, and duplicative coverage manifested in the three competing bibliographic databases. Recent efforts by the database producers and others to cooperate to reduce overlap and to standardize vocabulary promise to improve the user's lot immeasurably. New technologies have also brought significant gains in the area of bibliographic control, and developments in automated indexing could make the process of providing bibliographic access and control more productive in the future.

Agriculture continues to be vitally important to global economy and stability. In industrialized nations and in less-developed countries, agriculture occupies a prominent position. Access to information about agricultural successes and failures is key to the furthering of research and development that strengthens the agricultural foundation on which the world community depends.

Sarah E. Thomas, Technical Services Division, National Agricultural Library, Beltsville, MD 20705

LIBRARY TRENDS, Vol. 38, No. 3, Winter 1990, pp. 542-61

© 1990 The Board of Trustees, University of Illinois

Bibliographic control provides a meaningful organization for the vast universe of published literature relating to both applied agriculture and scientific research in the field. Defined by *Harrod's Librarian's Glossary* (1987), bibliographic control is "the creation, development, organization, management and exploitation of records prepared firstly to describe items held in libraries or on databases and secondly to facilitate user access to such items" (p. 71). The challenges facing librarians seeking to impose control over agricultural materials are similar in many respects to those encountered by technical services librarians in general. That is, vast quantities of materials must be described and organized in a fashion that promotes access to the bibliographic item.

Oruma (1984) identified basic problems contributing to existing complexities in agricultural information management. One problem centers on the dichotomy between subsidized government research and private agricultural production. The agricultural documentalist must collect, store, retrieve, and disseminate information for two categories of users in the same subject area. A second problem is the multidisciplinary nature that makes agricultural information almost "unmanageably vast." Agriculture today includes not only traditional agriculture but also biotechnology, environmental sciences, human food and nutrition, rural development and sociology, and many other related subjects. Consequently, the meaning of the term "agricultural literature" has become indefinite and imprecise, leaving information managers a most complex task as they seek to organize bibliographic collections (p. 91).

The process of bibliographic description and analysis is still largely labor intensive. Despite incursions made through automation into the rather tedious process of transcribing data, much of the effort in the creation of the bibliographic record remains manual. Catalogers and indexers of agricultural literature must grapple with the inconsistencies of authors and publishers of documents, often struggling to ascertain responsibility for authorship or to determine who published an item and where or when it was published. If one of the principles of bibliographic control is to bring together all the works by a single individual or entity, what is the cataloger to do with an author whose monograph on fungi is attributed to A. Smith? In a large file, this author's books will be comingled with those created by others of the same surname, making it difficult to trace all publications authored by a given scientist. A similar problem may be found when a government agency changes its name. To provide authoritative access to all material emanating from a single source, librarians must often be detectives and genealogists as they track the history of particular corporate bodies.

Another common problem occurs when a journal changes its name, necessitating links to its previous titles. The ability of the user to identify all relevant material as unified by a particular series or journal can be hampered by inadequate references or access points in the

catalog, database, or other bibliographic reference tool. Insufficient bibliographic information contained in the item being cited can thwart description and analysis of a publication and hinder access. Although this deficiency is not unique to agricultural publications, it is probably more common in the sciences than in the humanities, and similarly more likely to occur in publications issued by corporate bodies and government agencies, many of which are important sources of agricultural materials. In addition, Third World nations, with their strong agricultural base, are often valuable resources for agricultural documents, but bibliographic control of publications from most developing countries is notoriously poor because they frequently do not follow standard bibliographic conventions. Thus, bibliographic control of agricultural documents is exacerbated beyond the routine because a greater proportion of the publications in the field is produced by those for whom bibliographic concerns are secondary or even irrelevant. They are often printed on poor quality paper, disseminated informally, limited in quantity, and intended primarily to meet an immediate need. Their long-term value is not always perceived, and rarely do the authors or publishers consider the importance of ensuring that an archival copy be retained. For example, even in the United States, extension service pamphlets are often considered ephemeral, and bibliographic control of these has been lax.

"Grey" or "fugitive" literature comprises a substantial segment of agricultural publications and is difficult to obtain and to manage. Grey literature, or nonconventional literature, is material such as reports that is not distributed through commercial channels. Such literature often is produced in only a few copies and is not sold by its publishers. It is cited infrequently in national bibliographies and is consequently not widely publicized. What the authors and publishers of agricultural literature fail to realize is that the lack of full publication or other bibliographic information impedes the dissemination of the information contained within their publications. Catalogers and indexers must devote scarce resources to unraveling the bibliographic mysteries they encounter, slowing the processing of other materials. Worse yet, they may fail to decipher an important clue and omit a key access point.

In addition to incomplete, inadequate, or confusing information relating to the bibliographic item, there are a number of other factors that complicate bibliographic control. The problems in providing access to agricultural materials are compounded as geographical and political boundaries are crossed. Since agricultural information is of worldwide significance, access to it needs to be as broad as possible. Yet this international demand for agricultural information raises important issues with regard to language of access, format, and even description of content. For example, if a controlled vocabulary is employed, what language is used? Because the primary producers of agricultural databases have been from English-speaking countries, the major the-

sauri have been published in the English language. Multilingual vocabularies exist but are not applied to all, or even a majority of, bibliographic records. Different countries have developed different conventions and formats for recording bibliographic data, thus hindering exchange and manipulation of data. In the United States, for instance, most research libraries adhere to the *Anglo-American Cataloguing Rules*, second edition, for cataloging monographs, serials, audiovisuals, maps, computer files, and other bibliographic formats. They use the MARC (machine-readable cataloging) format for communicating their cataloging records. In theory, if not entirely in practice, U.S. and British librarians share a common cataloging code and both embrace the MARC format. A number of fairly significant variations make sharing of their data slightly troublesome although far from impossible. Yet there are a number of other points on which practitioners in these two so closely related countries diverge, particularly in the area of classification and subject analysis. Cataloging records created in other countries are even less likely to follow the same convention.

In the United States, almost all research, most public, and many special and school libraries use a bibliographic utility such as OCLC (Online Computer Library Center, Inc.), RLIN (Research Libraries Information Network), or WLN (Western Library Network) for cataloging. These systems are supported by large mainframe computers and provide access to databases containing millions of bibliographic records. Located in Dublin, Ohio, OCLC started as a consortium of Ohio libraries in 1967 and has grown to serve over 6,000 libraries. The OnLine Union Catalog contains records for over 20 million items. RLIN, the operational arm of the Research Libraries Group, based in Mountain View, California, holds records for over 17 million bibliographic items in its database. These records were contributed by thirty-four of the nation's largest research libraries and other special and associate members. The Western Library Network, much smaller than OCLC or RLIN, is the bibliographic support system of choice in the Pacific Northwest. Its database comprises approximately 5.5 million records.

All three databases share several characteristics that are typical of bibliographic control of monographs, serials, maps, audiovisuals, and computer files in North America. First, and above all, they are databases created through shared cataloging endeavors. The Library of Congress (LC) MARC records serve as the foundation for these databases. The MARC Distribution Service of the Library of Congress provides almost 350,000 MARC records annually, and these records are used many times by participating libraries in the production of records or cards for their local catalogs. Cooperating libraries contribute original cataloging for items not represented in the database, and other institutions, in turn, take advantage of this member-contributed copy to reduce their need for original cataloging. The value of such cooperative enterprises is the increase in available bibliographic records and the consequent reduction

in the need for expensive, labor-intensive original cataloging (Bengston, 1984; Horney, 1984).

The success of cooperative cataloging has spawned other programs in bibliographic control. One major effort is CONSER, created as the CONversion of SERIALs and now the Cooperative ONLINE SERIALs program. Originally conceived as a project to create machine-readable records for serials through a retrospective conversion project, the program has grown to encompass much more. The database is a widely available source of authoritative bibliographic information about serials. At the end of 1988, it contained 437,623 records with 46,274 records added or newly authenticated during 1988. The National Agricultural Library is a CONSER participant specializing in providing bibliographic control of agricultural serial publications.

Another key bibliographic program is NACO (Name Authority COoperative Project). It was initiated in 1979 and is now known as the NATIONAL Coordinated Cataloging Operations. Participants contribute headings to a cooperative automated name authority file. Authority control is a vital aspect of bibliographic control particularly when libraries contribute records to large network databases. An authority file is a record of the authorized or established forms of headings or access points used in the catalog (Bengston, 1984). Without adherence to authorized headings, conflicts result and the user of a catalog or database may not locate all items relating to a particular subject or created by a specific individual or corporate body. NAL is a participant in NACO and contributes authoritative forms of names in the field of agriculture.

In agricultural literature, however, cataloging records represent a minority of items under bibliographic control. Indexing records comprise 85 percent or more of the large agricultural databases, underscoring the importance of journal literature in scientific disciplines. Database producers do not follow a single common standard; consequently, there are salient differences in the methods for describing bibliographic items. Journal title abbreviations vary, different elements are recorded, and the formats in which they are created differ markedly. Furthermore, while a tradition of sharing cataloging records has been established in the United States and is gaining strength internationally, no such model exists for indexing. This movement has led to increased standardization to facilitate the exchange of cataloging records. In the past decade there have been some attempts at increased cooperation in indexing, but the union catalog that exists for monographs has not materialized for journal articles. Most U.S. research libraries contribute records for monographs and other materials to OCLC or RLIN, but BIOSIS, Chemical Abstracts, CAB ABSTRACTS, and AGRICOLA (AGRICulture OnLine Access) records all employ varying formats, and their producers generally do not exchange or share records, a situation dictated by commercial considerations as well as technical issues. AGRICOLA records for U.S. imprints are shared with AGRIS (The

International System of the Agricultural Sciences and Technology), but they are subjected to several machine and human edits because of varying policies, procedures, and formats, and thus the transfer of records is not accomplished seamlessly as it is in the North American model for the creation and exchange of cataloging data.

Bibliographic control, as practiced in the creation of databases of citations to agricultural publications, has several other critical characteristics beyond the standards used for bibliographic description and analysis and the formats used to describe this content. One of the key factors in the management of the data is the extent of coverage. The enormous quantity of publications being issued means that essentially no single library or institution can afford to collect comprehensively in all aspects of agricultural literature and has provided the impetus for cooperative collection development. Similarly, no library possesses the resources to catalog or index all material being published in agriculture and relevant sciences. Arrearages, and how best to handle them, is a frequent topic of discussion among technical services librarians and ample testimony to the magnitude of the problem. Librarians have developed several strategies for coping with an almost unmanageable volume of publications. One method is to be highly selective about the titles one acquires, catalogs, or indexes. Another approach is to catalog or index more material but with a less than full record. Subject analysis may be omitted or curtailed. A third tactic is to work cooperatively with other libraries or documentation centers to create a shared database, distributing responsibility for bibliographic control. Many institutions employ a combination of these methods according to the priorities established for processing their acquisitions. They catalog or index core material fully but treat peripheral material less completely. Core journals may be indexed cover to cover, while other publications less central to the needs of the database's users will be analyzed selectively. Certain ephemeral materials may not be recorded at all. Domestic titles may be accorded a higher priority than foreign publications, or documents detailing research activities may be processed before popular reading material intended for the lay person. The priorities naturally reflect the mission of the institution.

Processing priorities determine the timeliness of the data available to the public or database user and are thus an important element in the management of the database. Different disciplines have differing perceptions of the necessity for timely data, with the medical and scientific professions placing a premium on very recent information, and humanistic areas more tolerant of a longer delay between appearance of a title and its citation. Much scientific information is presented at conferences and published in proceedings that are in high demand in the field of agriculture. Other scientific findings and research results appear in journal articles, which are more rapidly written, reviewed, and published than monographs. When time is of the essence for publication

and dissemination of results, timely citation of those results is equally important. Several studies document the length of time between publication of a document and appearance of a citation to that document in a major database, with the implication that a faster turnaround contributes to a higher quality, more valuable database.

Another policy decision affecting the utility of a database relates to the inclusion of abstracts. If a citation also includes an abstract, the researcher may be able to determine if the full text is required. Elimination of unnecessary requests for copies of journal articles or other library materials conserves the lending institution's resources and allows the user to spend his or her time more productively. On the other hand, incorporation of abstracts into the bibliographic record is an expensive process. Merely keying in the author's abstract, once the publisher's permission has been obtained for copyrighted materials, is costly. The creation of original analytical abstracts is obviously more costly. Thus the value of the abstract to the user must be weighed against the expense of inclusion, and the resulting equation must take into account the ultimate cost of using the database vis-à-vis its value to the user.

There are a number of sources that seek to organize publications relating to agriculture in formats that provide constructive access to the information they contain. Figuring prominently among these sources are three large agricultural databases that collectively index over 11,600 agricultural journals and provide access to hundreds of monographs, technical reports, and other materials annually. The three databases providing the most concentrated and comprehensive focus on agricultural literature are: AGRICOLA, generated by the staff of the National Agricultural Library of the U.S. Department of Agriculture (USDA); AGRIS, the database created through the cooperative efforts of those working through the Food and Agriculture Organization (FAO) of the United Nations; and CAB ABSTRACTS, the product of CAB International (CABI).

Several other databases offer access to literature of interest to agricultural researchers, but their focus is either more general or more narrow. For example, citations included in BIOSIS or CA Search often overlap with those in AGRICOLA, AGRIS, and CAB ABSTRACTS, but these databases also contain many other references that are irrelevant to agricultural study. Other bibliographies or databases may cover one topic in depth, such as a particular crop, but they do not attempt to index the whole of agricultural literature. Agriculture covers a broad spectrum of subjects and may include topics related to medicine and the biological sciences, science and technology, and energy and the environment. Among some of the more frequently consulted databases other than those named earlier are Aquatic Sciences and Fisheries Abstracts; CRIS (Current Research Information System); Food Science and Technology Abstracts; MEDLINE; GPO (Government Publications Office); NTIS (National Technical Information Services); PREDICASTS; and

the Dun and Bradstreet family of databases. Such databases provide access to literature that may be essential for the agricultural researcher or administrator but that may not fall into the scope of AGRICOLA, AGRIS, or CAB ABSTRACTS. In general, however, the three major agricultural databases are the best resources for most topics relating to agriculture. Each of these databases has a unique slant; consequently, they must be seen as complementary rather than duplicative.

The National Agricultural Library initiated what became known as AGRICOLA in 1970. CAIN (Cataloging And INDEXing system of the National Agricultural Library), as the file originally was called, included citations to journal articles, book chapters, conference proceedings, monographs, microfilms, patents, theses, and other items cataloged or indexed by NAL in machine-readable form. When NAL began issuing machine-readable tapes of its cataloging and indexing records in the MARC format in 1976 to conform to U.S. library standards calling for output using MARC, the database was renamed AGRICOLA. Throughout the years AGRICOLA has undergone numerous changes, reflecting shifting priorities within USDA, new directions in agriculture, and response to user needs. These changes—made to improve the database—have also resulted in inconsistencies in structure and content. Before 1987, for example, AGRICOLA included material in several subfiles such as AGECON (Agricultural Economics-USA) and FNC (Food and Nutrition). AGECON contained records with abstracts contributed by the American Agricultural Economics Documentation Center of the USDA Economic Research Service. When support provided through the American Agricultural Economics Association ceased in 1986, NAL assumed responsibility for indexing core agricultural economics literature, albeit without abstracts, and added these records to the main file rather than to the subfile. In general, the trend in AGRICOLA has been toward the discontinuation of separate subfiles in favor of a single integrated database with enhanced access that offers the capacity for creating subsets through retrieval techniques rather than the creation and maintenance of separate subfiles. Some other historical subfiles in AGRICOLA are: brucellosis, environmental impact statements, 4-H publications, adult extension, and parasitology. The FNC subfile, representing titles selected by the Food and Nutrition Information Center of the National Agricultural Library, is an ongoing subfile distinguished by its subject focus and by the fact that each record in it also contains an abstract or annotation.

All records in AGRICOLA are accessible through subject category codes which provide broad topical access. These codes, which have undergone three changes since their introduction, carry a numeric or (currently) alphanumeric notation. Since indexing and cataloging records in AGRICOLA employ different controlled vocabularies, the category code is the single unifying access point at the subject level.

Monographs and other cataloged material constitute 16 percent of

the 2.6 million records in the AGRICOLA database. Prior to July 1972, these records carried subject headings from the *National Agricultural Library Subject Heading List*. Since that date, Library of Congress subject headings (LCSH) have been employed. Before 1985, a controlled vocabulary for indexing records, which constitute 84 percent of the database, was not used except in the case of the FNC subfile. Subject access to the majority of records was available through the AGRICOLA category codes and through keyword (free-text) searching, which provided inconsistent and uncertain results depending on the adequacy of titles to represent article contents and of searchers to formulate effective search strategies. Recognizing this deficiency, NAL evaluated options for controlled vocabularies and in 1984 selected the *CAB Thesaurus*.

The *CAB Thesaurus* prevailed as the working tool of preference because of its extensive research-oriented vocabulary of over 48,000 terms and because CAB International, then Commonwealth Agricultural Bureaux, agreed to modify the second and subsequent editions of the thesaurus to reflect AGRICOLA indexing requirements. As a result of this agreement between CAB International and NAL, the second edition of the *CAB Thesaurus* was annotated to incorporate the so-called "Americanisms"—those terms or spellings used commonly in the United States. For example, the British "groundnuts" was supplemented by the American "peanuts," and "flavour" by "flavor." In addition, NAL indexers recommended new terminology for inclusion in the thesaurus to accommodate AGRICOLA's scope and coverage or to represent new concepts in agriculture.

Overall, the effect of the use of the controlled vocabulary has been the imposition of greater consistency on the indexing records being contributed to the AGRICOLA database. Compared to the practice of title enrichment, which had been followed before the introduction of the *CAB Thesaurus*, the new policy injected a greater rigor into NAL indexing practice, and, consequently, resulted in higher quality indexing records (Thomas, 1985). Indexers apply approximately eight to ten terms per record, providing greater access to the topic being analyzed. As a result of this shift in policy, online searchers and other users of AGRICOLA or the *Bibliography of Agriculture* have benefited. Hood (1988) calculated that use of a controlled vocabulary in searching AGRICOLA results in greater precision in retrieval and decreased costs. In performing an online search in response to a specific reference query, Hood retrieved 274 percent more relevant records when employing *CAB Thesaurus* descriptors than when searching titles and abstracts free text. Furthermore, the cost per relevant record decreased 82 percent in the controlled vocabulary search. Controlled vocabularies are especially useful in controlling synonyms and quasi-synonyms. Hood observed that "a controlled vocabulary facilitates economical searching by shifting costs from output (searching) stage to the input (indexing) stage of information retrieval systems" (Hood, 1988, p. 16). Since NAL imple-

mented use of the *CAB Thesaurus*, over 400,000 items have been indexed using CAB descriptors, creating a large selection of current records with access provided through a controlled vocabulary.

One immediate effect of the transition to a controlled vocabulary was a drop in the number of items indexed from about 120,000 annually to 80,000. Because there were inadequate resources to expand the indexing staff, NAL reduced the number of journals indexed. The pressure to eliminate titles from the list of journals indexed to counter the mounting backlog coincided with a new policy aimed at decreasing the extent of overlap with AGRIS. At the AGRIS Technical Consultation in June 1985 in Rome, Italy, and again at the seventh conference of the International Association of Agricultural Librarians and Documentalists (IAALD) in Ottawa, Canada, in 1985, agricultural information leaders called on the producers of the three most prominent agricultural databases—AGRICOLA, AGRIS, and CAB ABSTRACTS—to cooperate in the coverage of the world's agricultural literature by dividing responsibility for various aspects of agriculture. At the same time, the National Agricultural Library established a policy of eliminating from its coverage foreign language journals included in AGRIS and began encouraging countries participating in AGRIS to increase the extent of their coverage. As a result of this policy, the number of journal titles indexed in AGRICOLA decreased from over 5,000 in 1984 to slightly over 2,500 in 1989. The percentage of U.S. literature indexed in AGRICOLA rose from 29.88 percent to 55.19 percent during the same period as a direct result of NAL's implementation of its stated objective of being the premier source for U.S. publications in the field of agriculture, especially USDA and state agricultural experiment station and extension service documents. Coverage of U.S. literature is actually substantially higher than 55 percent when consideration is given to the number of international publications that have joint places of publication—such as London and New York—but are not counted as U.S. imprints in statistical analysis.

NAL and forty-seven land-grant university libraries have been participating in a program to increase the bibliographic control of state agricultural publications in AGRICOLA (Mathews, 1987, 1988; Thomas, 1988). In this program, each land-grant library assists NAL in identifying and verifying titles published in its state and, in some cases, arranges for the titles to be sent to NAL. If the library catalogs titles appearing in a particular series, it notifies NAL. Catalog copy from the land-grant library is used as the basis for NAL's cataloging record which is entered into AGRICOLA. NAL either catalogs or indexes items not covered by the land grants. As a result of the cooperative activity and increased priority given to processing this material, the AGRICOLA database logged the entry of over 5,800 state agricultural publications in 1988, a 20 percent increase over the number added to the database in 1984, the year the program was initiated. The number of state

agricultural publications indexed by NAL rose 85 percent over the same interval, from 2,837 in 1984 to 5,242 in 1988, as NAL increased its coverage of items not cataloged by land grants. In addition, catalogers are attempting to improve bibliographic control of state experiment station and extension service publications, most of which are issued in series, by communicating directly with the originators of the documents. Through a program of education about the requirements and advantages of bibliographic control, the catalogers are endeavoring to reduce the practice of publishers of providing inadequate or ambiguous bibliographic information that frustrates effective bibliographic control.

Retrospective conversion projects and other special cataloging efforts also have enhanced coverage of USDA and state agricultural publications in AGRICOLA. The University of Illinois at Urbana-Champaign Agriculture Library Project has received funding from the U.S. Department of Education to catalog selected USDA and state agricultural experiment station publications. Over the first four years of this project, the University of Illinois has provided access to over 40,000 titles in OCLC and AGRICOLA.

One distinguishing characteristic of AGRICOLA is that it provides holdings information and, in almost all cases, call numbers, allowing the searcher to locate the source of the citation for examination of the full text. For many years, every item cited in AGRICOLA was held by NAL. In 1985, NAL began a cooperative cataloging program in which records are contributed by land-grant university libraries for agricultural titles not held by NAL. This program recognizes the breadth of agricultural collections across the United States and acknowledges that NAL depends—just as other U.S. research libraries do—on resource sharing rather than physical ownership for access to a comprehensive agricultural collection. The decision to permit organizations outside of NAL to enter records into AGRICOLA was a step toward increasing the scope of the database and providing access to more agricultural publications. Accordingly, some universities have begun including records for their doctoral dissertations and master's theses, a category of material not systematically collected by NAL; other institutions supply citations to titles in subject areas in which they maintain the premier collection (Thomas, 1989). Cornell University's Mann Library cataloged extensive Chinese language holdings on agriculture and made these records available for inclusion in AGRICOLA (Acosta, 1989). The AGRICOLA database has been significantly strengthened by these efforts.

AGRICOLA also is notable for its coverage of a variety of formats. In addition to bibliographic records for monographs, serials, and journal articles, AGRICOLA contains references to audiovisual material such as filmstrips, videotapes, slides, photographs, and audiotapes; to computer software and laser discs; to manuscript collections; to maps; to patents; and to realia such as three-dimensional representations

intended for instructional purposes. This extensive array of subjects, formats, and types of materials makes bibliographic control of agricultural material in AGRICOLA complex and challenging.

NAL contributes bibliographic records for documents published in the United States to AGRIS, one of the three large agricultural databases. AGRIS, which became operational in 1975, is a cooperative database produced under the auspices of the Food and Agriculture Organization of the United Nations. Over 127 national centers and 18 regional and international centers prepare records for AGRIS with each participating country assuming responsibility for providing bibliographic control over its national production of agricultural publications. AGRIS numbers over 1.5 million references and is growing at an annual rate of about 120,000 records. U.S. records have accounted for 42 percent to almost 50 percent of contributed records in recent years. The German Federal Republic, France, Japan, and the Netherlands are the next largest contributors, with 10 percent, 5 percent, 5 percent, and 4 percent, respectively, of the records coming from these national centers in 1987 (Lebowitz, personal communication, 1988). Although developed countries provide more than two-thirds of the citations, most of the centers preparing records represent developing nations. It is the potential of these less-developed countries to provide access to the agricultural information within their geopolitical boundaries that makes the concept of AGRIS so unique. AGRIS participating centers follow procedures developed in consultation with FAO's AGRIS Coordinating Centre located at FAO headquarters in Rome, Italy. Manual and automated input to the database is handled through the AGRIS Processing Unit at the International Atomic Energy Agency in Vienna, Austria.

It should be noted that the AGRIS database that is marketed in the United States (called AGRIS International) differs from the worldwide edition of AGRIS primarily in that it does not contain records with U.S. imprints which would duplicate records in the AGRICOLA database. Furthermore, all records contributed by NAL are not included in the AGRIS database because AGRIS considers aspects of certain subject areas which are core to AGRICOLA—such as human ecology—to be outside of its scope.

To facilitate access to data in many languages, references are indexed using AGROVOC, a multilingual agricultural thesaurus prepared by FAO in conjunction with the Commission of the European Communities. AGROVOC, first published in 1982, is a thesaurus of approximately 9,000 terms. English, French, German, Italian, and Spanish editions of AGROVOC already exist, and other versions, such as Portuguese, are being developed. The focus of the AGRIS database is on worldwide information in scientific, technical, and socioeconomic publications relating to food and agriculture. In general, the mission of AGRIS is to serve less-developed countries as well as developed nations. The agricultural literature of developing countries, often omitted from

databases produced in developed nations, is of particular relevance to agriculturists in other developing countries (Samaha, 1988).

CAB ABSTRACTS, the third major agricultural database, began in 1972 as an outgrowth of CAB printed publications. Produced by CAB International, which is based in Wallingford, Oxon, United Kingdom, CAB ABSTRACTS now contains some 2 million citations covering agriculture, forestry and forest products, and allied disciplines. CABI editors and associates scan over 9,100 serial titles annually, resulting in a yield of about 130,000 references being added to the database each year. Items are selected for indexing on the basis of their research value and their relationship to one of several of CABI's abstract journals. CABI consists of fourteen bureaus and institutes concentrating on a range of subjects from helminthology to forestry and crop science to recreation and leisure. When the Commonwealth Agricultural Bureaux underwent a transformation in the 1980s from being a Commonwealth organization to being an international group, countries outside the British Commonwealth were invited to join the renamed CAB International. An arrangement was made for the Chinese Academy of Agricultural Sciences to contribute indexing records for Chinese agricultural publications. More recently, Hungary has joined the organization, and further non-Commonwealth countries are considering membership. These associations allow CABI to maintain a database rich in foreign language materials. Approximately forty languages are represented in CAB ABSTRACTS. Perhaps the most prominent feature of CAB ABSTRACTS is its reliance on original abstracts created by subject experts on the CABI staff for about 85 percent of the material being cited (Porta, 1986). These lengthy abstracts offer a depth of subject analysis highly valued by users of the database. In addition to the abstracts, subject access is available through the application of *CAB Thesaurus* terms and more general subject codes.

All the earlier mentioned databases share some common elements. AGRICOLA, AGRIS, and CAB ABSTRACTS are all large files providing access to agricultural information and, as such, overlap to some extent. Numerous studies examining the degree of overlap have been conducted over the past decade (Brooks, 1980; Datta, 1988; Frank, 1987; Heise & Hood, 1987; Krabbe, 1979; Longo & Machado, 1981). One of the most recent, conducted under the direction of Norbert Deselaers of the Federal Ministry of Food, Agriculture and Forestry of the Federal Republic of Germany, indicated that, in 1983, there was substantial duplication among the databases. Deselaers (1986) reported a 32.4 percent overlap among the three. Other studies have found from 15 percent to 40 percent overlap. Using data from the World List of Agricultural Serials database at NAL, a comparison of the journal titles indexed in AGRICOLA, AGRIS, and CAB ABSTRACTS reveals that of the 9,190 titles scanned by CABI, 1,178 of them, almost 13 percent, are also indexed in AGRICOLA. AGRIS indexes 1,386, or approximately 15

percent, of titles scanned by CABI. Of those, 728, or 27.7 percent, overlap with titles covered in AGRICOLA. The higher proportion is not surprising since there should be almost a 100 percent overlap of the U.S. titles in AGRIS and AGRICOLA; U.S. indexing constitutes a substantial proportion of AGRIS. With regard to foreign titles indexed in both AGRICOLA and AGRIS, there is virtually no duplication: in 1986, a mere 9 percent of non-U.S. titles indexed in AGRICOLA were duplicated in AGRIS, and in the intervening years, further reductions have been made in this area of coverage. Of the entire list of 11,619 titles, only 450 are indexed in all three databases. Of the 2,510 titles indexed in AGRICOLA, 999 are not indexed by either AGRIS or CAB ABSTRACTS; 1,097 of the 2,816 titles indexed in AGRIS are unique; and 7,076 titles scanned by CABI out of its total list of 9,190 titles are not covered in either AGRICOLA or AGRIS.

Analysis of journal title overlap yields incomplete information. The National Agricultural Library and CABI have approximately the same level of production of bibliographic records, yet CABI scans more than four times as many journals. CABI selects fewer articles from particular journal titles whereas NAL indexes many journals cover to cover. U.S. imprints selected from AGRICOLA for contribution to AGRIS make up over 40 percent of AGRIS records in any given year. Overlap of records not linked to journals is even more difficult to estimate. Deselaers (1986) argued convincingly that "every third documentation unit in the agricultural sector was analyzed unnecessarily" (p. 24), and he estimated that \$2.5 million could be saved through an integrated agricultural information system (p. 25). NAL's decision to reduce overlap with AGRIS would change the basis of Deselaers's calculations in 1989, but there is no doubt that substantial savings could be realized through increased cooperation. Not all information scientists, however, regard database duplication as unhealthy or wasteful. Heise and Hood (1987) note that some researchers have demonstrated positive aspects of overlap for database searching. Onorato and Bianchi (1981), for example, identified the following benefits: records for the same items retrieved on different files may contain different information, and every duplicate, therefore, adds unique data; if several databases choose to index the same article, this may be taken as evidence of the value of the article; a retrieval of duplicates may confirm the validity of the search strategy used; the presence of duplicates helps to provide knowledge of areas in which files overlap and may thus support decisions as to which files to search for a particular topic; and comparisons of duplicates advance knowledge of indexing policies of different databases and may thus lead to improved search strategies in these databases. Buckland, Hindle, and Walker (1975) also cited the utility of overlap as a planning parameter. It appears likely that some overlap between and among agricultural databases will continue for the near future despite increased cooperation among the database producers. As each database

caters to a slightly different, but also overlapping, audience, the database producers will continue to respond to the needs of their users and to the mission of the larger organizations of which they are a part. Nevertheless, a unified database reflecting all constituencies would be a great boon for agriculturists and database searchers and producers. Considerable economies would be realized through enhanced collaboration in the production of a single comprehensive agricultural database. AGRICOLA, AGRIS, and CAB ABSTRACTS, while all covering some of the same ground, have many points of dissimilarity. Table 1 attempts to identify some prime characteristics of the three databases.

Much has been said about the duplication of effort in establishing bibliographic control over international agricultural publications. Deselaers (1986) contends that one-third of the records in the three major agricultural databases overlap, and that the cost of creating and maintaining these redundant files amounts to millions of dollars. Comprehensive searching necessitates the involvement of three or more files and substantial duplication of citations is encountered; the added cost associated with this activity is clearly undesirable. Ernest Mann, president of the International Association of Agricultural Librarians and Documentalists (Mann, 1986), urged a rethinking of definition of scope for the three major databases in a paper delivered in Ottawa in 1985. He proposed that AGRICOLA concentrate primarily on U.S. literature; AGRIS concentrate on nonconventional and some conventional literature from participating centers; and that CAB cover the "bulk of the world's conventional agricultural literature" (p. 6). Mann envisioned a single, integrated information resource using standardized indexing procedures and a shared indexing language based on a unified thesaurus (p. 8). Realization of this vision remains elusive, but the three major producers of bibliographic records have developed some areas of cooperation that are mutually beneficial and, more importantly, of service to the users of agricultural information.

One key area of cooperation lies in thesaurus building. Unfortunately, when there was a growing recognition of the need for an agricultural vocabulary to offer greater control and access to the increasingly large files of bibliographic records being created, the paths of the database producers diverged. The Commonwealth Agricultural Bureaux developed the *CAB Thesaurus*, a tool tailored to the audience of the CAB abstract journals, rich in terminology and leaning heavily toward the scientific and research side of agriculture. At the same time, AGROVOC was developed for use in indexing for AGRIS but with a different slant. It employed only about one-sixth the terms selected by CAB, choosing a simpler approach that relied on broader terms. AGROVOC too was targeted to a specific audience, one that consisted of a majority of non-native English speakers residing primarily in developing countries. The National Agricultural Library briefly considered generating its own unique agricultural vocabulary, customized to suit

TABLE 1
CHARACTERISTICS OF AGRICOLA, AGRIS AND CAB ABSTRACTS

	AGRICOLA	AGRIS	CAB ABSTRACTS
Size of database, 1989 (no. of citations)	2.6 million	1.5 million	2 million
Average annual growth (no. of records)	110,000	120,000	130,000
Established Abstracts	1970 15% (since 1988)	1975 10%	1970 85%
Vocabulary	<i>CAB Thesaurus</i> Library of Congress Subject Headings	AGROVOC (English, French Spanish)	<i>CAB Thesaurus</i>
No. of languages covered	56	38	48
Citations in English	71%	60%	68%
No. of titles indexed (1989)	2,510	2,816	9,190
Focus	U.S. literature, especially USDA and state agricultural publications	International with special emphasis on developing countries	worldwide research
Producer	USDA, National Agricultural Library	Food and Agriculture Organization of the UN, AGRIS Coordinating Center	CAB International
Availability	Online: BRS DIALOG DIMDI CD-ROM: OCLC SilverPlatter Print: Bibliography of Agriculture	Online: DIALOG (Non-U.S. titles only) IAEA IRS DIMDI CD-ROM: SilverPlatter Print: AGRINDEX	Online: BRS CAN-OLE DIALOG DIMDI CD-ROM: SilverPlatter Print: 27 main journals and 20 specialized journals published by CABI Monthly
Updated	Monthly	Monthly	Monthly

its user community but discarded the idea as excessively costly. After reviewing AGROVOC and the *CAB Thesaurus* during 1984, NAL entered into an understanding with CAB to use the *CAB Thesaurus* and to participate in the development of a revised version of the thesaurus that would meet most of NAL's requirements. Although a common agricultural vocabulary did not result from the appreciable investment of resources into thesaurus construction, at least searchers could benefit from having two vocabularies to provide access to the citations being added to the databases, and there were only two, not three, variations.

Since the appearance of AGROVOC and *CAB Thesaurus* in 1983, agricultural librarians and documentalists have made progress in bringing these two important resources into closer alignment. The National Agricultural Library has worked with AGRIS to develop tables for converting *CAB Thesaurus* terms to AGROVOC terms, and NAL routinely attends semiannual meetings on AGROVOC revisions in Rome. Proposed changes are carefully checked against CAB terminology, and every effort is made to reduce variation between the two vocabularies. AGRIS leadership "realized that the original restriction of 8,000 descriptors was unworkable, and ...[AGRIS] has been adding new descriptors, particularly for commodities and taxonomic terms..." (Lebowitz, personal communication, July 19, 1989).

Despite these cooperative endeavors, the hard reality remains that two disparate vocabularies are in use in the three large agricultural databases, and countless hours of effort go into their maintenance as well as the creation and maintenance of dozens of other specialized agricultural thesauri. If adopted, a comprehensive agricultural thesaurus could offer opportunities to streamline indexing activities and would clearly be a unifying force in the bibliographic organization of databases. Discussion of a detailed global agricultural vocabulary began again in earnest in December 1988. In May 1989, representatives of CABI, the Consultative Group on International Agricultural Research (CGIAR), and NAL examined the problems associated with the development of a single, integrated thesaurus. AGRIS thesaurus experts and management were invited to join a more formal meeting to consider such issues as content, structure, and governance. The group plans additional discussion on the topic with a wide segment of interested parties (André, 1989).

Another cooperative undertaking in the realm of bibliographic control of agricultural publications has been the World List of Agricultural Serials Project, a joint project of the National Agricultural Library and CAB International. In a memorandum of understanding of March 1988, NAL and CABI agreed to work together to produce a comprehensive listing of agricultural literature published serially. FAO's AGRIS Processing Unit has supplied a list of titles indexed in AGRIS, and the Commission of the European Communities (CEC) has provided partial financial support. NAL has created a machine-readable file of over

50,000 records; CABI will publish the finished product. The primary objective of the *World List of Agricultural Serials* is to produce a comprehensive, regularly updated list of agricultural serials, giving essential bibliographic information. The first product of the database will be an international union list of approximately 11,600 agricultural serial titles indexed in AGRICOLA, AGRIS, and CAB ABSTRACTS. Access to this information will enable database producers to review their indexing policies to determine if any overlapping titles should be dropped, and the full database will serve as a source of additional titles to be considered for indexing. Both the union list and the comprehensive list will function as useful reference tools for librarians and others seeking information on agriculture published in serial form (Thomas, 1987).

All of the earlier discussed projects would have been virtually unthinkable before the application of computer technology to the problem of bibliographic control. Modern information technologies have created opportunities for access to, and exchange of, information that were not possible when bibliographic records were created manually. There are several new technologies that offer the promise of breaking new ground in the effort to organize and manage collections. Scanning, optical character recognition, and expert systems are all being used to capture information and to process it in a more effective way. Retrieval software and CD-ROM products are also shaping the future of bibliographic control. Several evaluation studies are underway to test the application of these new technologies. At the National Agricultural Library, the Indexing Branch has been experimenting with the use of scanners in the transcription of data used in the bibliographic description of journal articles. Using a hand-held scanner attached to a personal computer, a library technician scanned the abstracts to journal articles into a file. The abstract was subsequently merged successfully with the remainder of the bibliographic record which resided in a separate file on a minicomputer. Results demonstrated that scanning is faster than keying despite the need to correct numerous errors in the scanned text. A second phase of the experiment will test the scanning of abstracts with a more powerful flatbed scanner (Edwards, personal communication, 1989). Eventually NAL would like to see the development of a scanner and software combination that could automatically identify authors, titles, imprint, and other such bibliographic elements for coding in the bibliographic record.

Several other organizations are pursuing the development of automated tools that assist in the analysis of material to be indexed. Torben Friis, of Datacentralen in Denmark under the sponsorship of the CEC, has created a program that reviews the contents of abstracts, and based on the vocabulary used, prepares recommended indexing terminology using the *CAB Thesaurus* as a source (Friis, personal communication, October 1989). Vleduts-Stokolov (1987) described an indexing expert

system developed at BIOSIS which is intended to generate BIOSIS concept codes and biosystematic codes through the processing of natural-language titles. The system will offer online assistance to indexers. Automated indexing would allow more material to be indexed, reducing or eliminating the labor-intensive and costly human indexing process, or, more likely in the foreseeable future, allowing rudimentary access to materials that might otherwise be entirely without access. With the advent of text digitizing and the conversion of full-text materials into ASCII, some have predicted that the need for cataloging and indexing as it is known today would be obviated with word-by-word access replacing the requirement for a bibliographic citation and controlled methods of subject access. Initial explorations of this approach indicate that human intellectual intervention to organize the information remains essential if the searcher is not to be overwhelmed by voluminous hits and led astray by false drops. As a consequence, agricultural vocabularies, authority control, and other elements of more traditional bibliographic control remain valid, even in the context of revolutionary information technologies. Yet there is no doubt that bibliographic control will continue to undergo transformations as librarians and information scientists work together to exploit technology to provide enhanced access to agricultural publications.

REFERENCES

- Acosta, I. P. (1989). Chinese and Cornell retrospective conversion projects at NAL. *Research Libraries in OCLC: A quarterly*, 30(Summer), 13.
- AGRIS Coordinating Centre, Food and Agriculture Organization of the United Nations. (1984). AGRIS: Introduction. International information system for the agricultural sciences and technology. Rome, Italy: Food and Agriculture Organization of the United Nations.
- André, P. (1989). Universal agricultural thesaurus discussed. *Agricultural Libraries Information Notes*, 15(June/July), 12.
- Bengston, B. G. (1894). Bibliographic control. In I. P. Godden (Ed.), *Library technical services operations and management* (pp. 133-94). Orlando, FL: Academic Press.
- Brooks, K. (1980). A comparison of the coverage of agricultural and forestry literature on Agricola, Biosis, CAB and SciSearch. *DataBase*, 3(March), 38-49.
- Buckland, M. K.; Hindle, A.; & Walker, G. P. M. (1975). Methodological problems in assessing the overlap between bibliographical files and library holding. *Information Processing and Management*, 11(3/4), 89-105.
- Datta, V. K. (1988). Coverage of literature on mycotoxins by computer databases and comparison of this coverage with the TDRI in-house index facility. *IAALD Quarterly Bulletin*, 33(2), 61-75.
- Deselaers, N. (1986). The necessity for closer cooperation among secondary agricultural information services: An analysis of AGRICOLA, AGRIS and CAB. *IAALD Quarterly Bulletin*, 30(1), 19-26.
- Frank, R. C. (1987). Agricultural information systems and services. In M. E. Williams (Ed.), *Annual Review of Information Science and Technology (ARIST)* (Vol. 22, pp. 193-334). Amsterdam: Elsevier.
- Heise, D. A., & Hood, M. W. (1987). *Bibliographic databases: Coverage, overlap and time-liness, a literature synthesis*. (Unpublished).
- Hood, M. W. (1988). Using the CAB Thesaurus to search AGRICOLA. *Agricultural Libraries Information Notes*, 14(11/12), 15-16.
- Horney, K. L. (1984). Automation: The context and the potential. In I. P. Godden (Ed.), *Library technical services: operations and management* (pp. 43-88). Orlando, FL: Academic Press.

- Krabbe, H. (1979). The AGRICOLA and CAB databases: A Danish comparison (3rd International Online Information Meeting, London, 4-6 December) (pp. 245-49). Oxford, United Kingdom: Learned Information.
- Longo, R. M. J., & Machado, U. D. (1981). Characterization of databases in the agricultural sciences. *Journal of the American Society for Information Science*, 32(2), 83-91.
- Mann, E. J. (1986). Past, present, and future developments in the transfer and dissemination of agricultural information: The case for a single coordinated, world agricultural information system. *IAALD Quarterly Bulletin*, 30(1), 5-9.
- Mathews, E. (1987). Bibliographic access to state agricultural experiment station publications. *IAALD Quarterly Bulletin*, 32(4), 193-199.
- Mathews, E. (1988). NAL/Land-Grant university state agricultural publications program report and recommendations. *Agricultural Libraries Information Notes*, 14, 1-4.
- Onorato, E. S., & Bianchi, G. (1981). Automatic identification of duplicates after multi-database online searching. *Online Review*, 5(6), 445-451.
- Oviss Oruma, B. (1984). The problem of information management in agriculture. *IAALD Quarterly Bulletin*, 29(4), 91-94.
- Porta, M. (1986). Selected agricultural databases and computerized systems. *Computers and Electronics in Agriculture*, 1, 289-297.
- Prytherch, R. (Comp.). (1987). *Harrods librarians glossary and reference book* (6th ed.). Brookfield, VT: Gower.
- Samaha, E. K. (1988). *Sharing of the world's agricultural information*. Rome: Food and Agriculture Organization, UN.
- Thomas, S. E. (1985). Use of the CAB Thesaurus at the National Agricultural Library. *IAALD Quarterly Bulletin*, 30(3), 61-65.
- Thomas, S. E. (1987). NAL's cooperation with CAB International. *Agricultural Libraries Information Notes*, 13(12), 5.
- Thomas, S. E. (1988). A coordinated program for state agricultural publications. *College and Research Libraries News*, 49(9), 425-430.
- Thomas, S. E. (1989). MIT agricultural dissertations in AGRICOLA. *Agricultural Libraries Information Notes*, 15(5), 6.
- Vleduts-Stokolov, N. (1987). Concept recognition in an automatic text-processing system for the life sciences. *Journal of the American Society of Information Science*, 38(4), 269-287.