TEEAL: The Essential Electronic Agricultural Library—Getting the Literature of Agriculture to the Developing Countries

Mary Anderson Ochs

What Is TEEAL?

In one sense, TEEAL (The Essential Electronic Agricultural Library) is a self-contained library of the core journals in the field of agriculture. It is available at cost ONLY to research institutions in the developing countries. It is a system being produced at Mann Library at Cornell University with the cooperation of many scientific publishers, database producers and the Rockefeller Foundation.

In the larger sense, however, TEEAL is the solution to a problem. In the last decade, there has been a huge investment in research and education in the developing countries. The goal has been to make the research and education programs in these countries self-sustaining. In spite of the monumental strides made in improving these programs, little has been accomplished in bringing these countries access to the world's scholarly publishing. Without access to the scientific literature, researchers in the developing countries cannot know that solutions to the problems they are struggling with have already been found by others. Many students from the developing countries come to the United States and Europe to complete their graduate work. Upon returning to their own countries, they no longer have access to the literature that they have relied upon for their work.

Providing access to the literature is a key to the success of research programs in the developing countries, but providing access to literature in all fields was too great an undertaking. So which fields were the most critical? The research area most fundamental to the improvement of conditions in many of the developing countries is food production and nutrition. Thus, the choice was made to focus on this literature (Olsen, 1994). By gathering the opinions of over 600 scholars in the field of agriculture, with 250 scientists focusing on the developing countries, the most important literature was identified.

The work to identify the core literature of agriculture began at Mann Library in 1988, and beginning in 1991 a multi-volume series, *The Literature of the Agricul-*

Mary Anderson Ochs is deputy director, TEEAL Project, Cornell University.

tural Sciences (Olsen, 1991–1996), was published by Cornell University Press. This extensive bibliography defines the literature most critical to researchers in all areas of agriculture. The ultimate goal in identifying the most important current literature was to make its full-text available to researchers in the developing world, where research libraries are few and far between.

After visiting universities and agricultural research stations around the world, officials at the Rockefeller Foundation agreed that one of greatest barriers to the improvement of agricultural research in the developing countries was the absence of literature and libraries to support the research. There was a tremendous need to make this core literature available, but it had to be in a form that the libraries in the developing countries could afford, support technically, and maintain under less than ideal environmental conditions. TEEAL, a CD-ROM system, was the answer.

In today's presentation, you will hear a bit more about the selection of material for inclusion in TEEAL, the creation of the system, marketing the system and thoughts on measuring the impact of the introduction of TEEAL in the developing countries. I must say at times I feel a bit like the producer of a low-budget movie. TEEAL may not be quite as flashy as other information retrieval systems, but I believe it will have a powerful impact by bringing agricultural information to those who have never had it before.

Who Are the Players?

TEEAL is the result of a successful multi-player partnership. Jan Olsen, the former director of Mann Library, and Wallace Olsen, a research associate at Mann Library, are the originators of the idea. Six years ago, and again three years ago, they performed a marketing study to determine whether there was a need for TEEAL. Responses in both cases were positive, and a decision was made to proceed with the project.

Even though the need was determined, the system could not be built without the cooperation of the major publishers of the agricultural journals and some significant start-up money. The publishers were approached with the idea, and most of them, including the major scientific publishers, such as Elsevier, Academic, Kluwer, and Springer, agreed to provide free copies of the journals selected for inclusion in the system. With a list of 80 journals approved for inclusion in the system, production could begin. The Rockefeller Foundation agreed to provide the funding.

Last April Mann Library signed a contract with a scanning vendor to perform the digitizing of the journals, and they sub-contracted with an information handling company to build the information retrieval system. At this point, I will demonstrate the system so that my description of its creation will be more meaningful.

The system is designed to support the three main uses scholars make of the literature: current awareness in their field; subject searching to support research proposals, student papers, and ongoing projects; and "known item" searching, such as finding a specific article by a specific author. I will demonstrate each of these functions.

Why CD-ROM?

Perhaps the first question some of you may be asking is, "why did they use CD-ROM technology?" While online access is now ubiquitous in this country, wide-spread access to the Internet in the developing countries is still many years off. Even where institutions do have access to online information resources, connections can be very poor and costs astronomical. With CD-ROM, once the system is purchased, it stands alone. At some point we will be looking at DVD technology, but costs for DVD were still too high when we made the decision to use CD-ROM.

Creating the Database

The core literature of agriculture spans a wider range than what one might expect. The journals selected for TEEAL include journals in crops, animal science, veterinary science, soils, environmental management, human nutrition and diet, food science, genetics, and other subjects. No one of the agricultural databases indexed all 130 journals included in TEEAL 1993–1996.

We started the production of the index with a set of records from CAB International, one of the major agricultural databases. We made an assessment of how complete their coverage was of each of the journals included in the system. If they covered less than 90% of the articles in a volume, we supplemented the indexing with records from another database.

In the initial TEEAL set (1993–96) there are 47,808 CABI records, 21,748 Medline records, 6,248 BIOSIS records, 2351 EconLit records, 396 Agricola records, 275 Agris records, and 602 records created at Mann Library. All of these database producers provided records to include in the system either free or at a reduced cost.

In order to combine these records into one database, we had to carefully reformat the non-CABI records to match the standard CABI record structure. For each database, we analyzed the record structure to determine how, for example, the BIOSIS fields would be transferred into the CABI fields. Each field had to be transferred into an existing field in the CABI-based record template or a new field had to be created. All data elements had to be reformatted. For example, authors with names spelled out had to be cut to include just initials, and language codes had to be converted to the full name of the language.

One particularly interesting challenge was translating BIOSIS and Medline codes into CABI codes. We created translation tables and converted codes by machine, and then did a manual check of the results. One slight error in the translation table created some rather amusing subject coding, with animal behavior articles receiving human codes. There may still be some of these lurking in the system, so we hope our users have a sense of humor.

Inconsistencies in the databases also created problems. In writing the programs to convert the data, our student programmers based their programming on the expected appearance of the data. Any time the data did not conform to this standard, the program needed to be adjusted to accommodate this. Seemingly simple problems took much programming time to find and troubleshoot.

Because the index for the system was created using records from many different databases, and all records for each journal from the 1993–1996 date range were added to the database, we had to write a duplicate checking program. This program had to match all the records against each other and delete the duplicate records. The CABI records were considered the primary records, so if there was a CABI record, it was retained. Records from the other databases supplemented the CABI records. The duplicate checking program matched the records on issn, year, volume, issue number, pages and the first 10 characters of the article title.

The index in the final product, as in the demo, will appear to be one large index, as seamless as possible, considering the multiple sources of the data. Our records were sent to the information handling vendor responsible for the construction of the TEEAL information retrieval system. There they were massaged yet again, and inserted into the Folio software that runs the TEEAL system.

The Folio software is not a traditional DIALOG or BRS-style bibliographic database. It was designed to be a full-text search and retrieval system, so it really looks at our index as a huge full-text document. It is possible to scroll through the entire database record by record without performing any kind of search. The search system simply creates pointers to the hits in this huge text file. The Folio software was selected because traditional bibliographic database software could not accommodate the browsing functions of the system. Using Folio was a relatively inexpensive way to build a system with as many of the desired features as possible. A custom designed interface would have been superior, but would have cost a great deal more and taken much longer to create. We began the actual production of TEEAL only about a year ago, so it has come a long way very quickly.

Scanning

Any of you who have undertaken a large-scale scanning project know the pitfalls of such a project. The sheer volume of the TEEAL material, over 700,000 pages, presents a challenge, but the complexity of the material, with its complicated pagination and multiple numbering schemes, adds to that complexity. Linking each article to its corresponding index record has not been an easy task.

Behind the system, lives a set of index files, one for each journal issue, that provides a link between the name of each image file and its corresponding page number. It is the "lifeline" that runs between the Folio database software and the separate image viewer. Figure 1 shows a small portion of an index file.

The majority of the images are bitonal scans. They are TIFF files, scanned at 300 dpi. Many of the scientific illustrations required grayscale scanning. The grayscale and color images are JPEG files, scanned at 200 dpi. These resolutions are much lower than typical preservation scanning, which is often done at 600 dpi. However, with the volume of material to be included in a CD-ROM system, it was important not to create files that were too large. The images at these resolutions are of good quality, although certainly not perfect. We are expecting the 1993–1996 set to fill approximately 150 compact disks.

Marketing

Unlike most other academic library projects, the undertaking of TEEAL required a business and marketing

Figure 1:   Index file matching page numbers with file names.

Cover 1,13562001.tif

Cover 2,13562002.tif

i,13562003.tif

ii,13562004.tif

iii,13562005.tif

iv,13562006.tif

v,13562007.tif

vi,13562008.tif

1,13562009.tif

2,13562010.tif

3,13562011.tif

plan. We had to decide how much to charge for the system and how many sets we needed to sell in order to cover our costs at that price. The start-up money was provided by the Rockefeller Foundation, with the intent that the system would ultimately pay for itself in its second year. The price for the initial 1993–1996 set is $10,000. The cost of the actual subscriptions to these journals for four years would be $365,000. It is a bargain by US standards, but for institutions in the developing countries, $10,000 is a great deal of money.

Many of the countries eligible to buy TEEAL are in Africa. In Africa, the marketing of the product presents particular challenges. Marketing efforts must be directed at decision-makers, not users, who do not have control over how money is spent. Many of the institutions eligible to buy TEEAL will need to find development aid to use to buy the system. In African countries, $10,000 may be the entire budget for the library at some institutions.

The distances involved and the difficulties experienced in communication with the African countries have presented additional challenges. As a result, we have put in place a marketing coordinator working out of Zimbabwe. More and more, administrators are becoming convinced that putting money into an information product is worthwhile. It will be the role of our African marketing coordinator to work with administrators to convince them that TEEAL is a good investment for their institutions.

Conclusion

Throughout this project, librarians have taken on many non-traditional roles. The TEEAL system is the brain child of librarians who were committed to the goal of bringing the literature of agriculture to the developing countries. They found the money to build the system, they persuaded publishers to join them in their efforts, they hired staff that could build a system, and they marketed the product to its intended audience. What remains to be seen is the impact the system has in the developing countries. Before it is in place, we hope to be able to take a benchmark reading of how scholars in some of the institutions buying TEEAL currently get access to the research literature. Over the course of several years, we will try to measure the impact of providing quick and easy access to the literature through TEEAL. We are certainly counting on making a difference in the research and teaching in these institutions.

Bibliography

Olsen, Wallace C. "The Current Core Literature of the Agricultural Sciences." *Journal of Agricultural & Food Information* 2.3 (1994): 99–109.

Olsen, Wallace C. *The Literature of the Agricultural Sciences.* 7 vols. Ithaca: Cornell University Pr., 1991–1996. 7 vols.