

AgEcon Search: A case study on the differences between operating a subject repository and an institutional repository

Julie Kelly and Louise Letnes
University Libraries and Department of Applied Economics
jkelly@umn.edu, lletnes@umn.edu

Abstract

AgEcon Search is a subject repository containing the full text of working papers, conference papers and small press journals in agricultural and other areas of applied economics. In existence since 1995, it contains material from 215 organizations. Comparisons are made between the operations of a subject repository and those of an institutional repository, with each having easier and more challenging aspects. The field of economics has characteristics that contribute to the success of a subject repository, such as a pre-print culture and an interest in intellectual property and the economics of publishing.

Introduction and Background

At most colleges and universities, the energy, discussions and developments concerning repositories are mostly about institutional repositories. Hundreds of subject repositories exist, but they mostly have not caught the attention of administrators or librarians in higher education. Perhaps this is because many are narrow in scope, such as the Digital Bridges, <http://bridges.lib.lehigh.edu/> or OceanDocs, <http://iodeweb1.vliz.be/odin/>, although many have a loyal following.

While the most visible work in the open access movement focuses on large publishers, funders with mandates, and faculty resolutions, many subject repositories have been making positive contributions by capturing grey literature, hosting pre-prints and post-prints, and providing a free or very inexpensive electronic platform for small press journals, especially those from small professional societies.

AgEcon Search will be used as an example of subject repository that has been active on several of these fronts.

AgEcon Search, <http://ageconsearch.umn.edu> is a free subject repository with full-text of working papers, conference papers and small press journals in agricultural, development, energy, environmental, resource and other areas of applied economics. It currently contains over 37,000 documents in 8 languages from 215 organizations in 40 countries.

Organizations that contribute materials include academic departments, professional societies, government agencies, and non-government organizations. AgEcon Search is housed at the University of Minnesota, and co-sponsored by the Agricultural and Applied Economics Association, which is based in the United States. Strong support also comes from the other large agricultural economics professional societies, including those representing Australia, Canada, and Europe, as well as the profession's international association (Letnes and Kelly, 2002).

While freely available, most of the material in AgEcon Search is not open access in the strict sense. Authors usually retain the rights to working papers and conference papers, and the publishers of the journals, mostly small societies, hold the copyright on the journal articles. For some journals, this is the only electronic access available, and for those that have another e-version, it is typically an organizational Web site with links to PDF versions of individual papers. A small number of the 36 journals also make some of their material available through commercial databases that include full text.

The project began as a local solution for both economists and librarians. Economists have a long tradition of writing working papers as the first step in their publication cycle. These are often sponsored by an academic department. In the days before the Internet, copies would be mailed to other departments or interested individuals, and there was little bibliographic control. If a department was lucky enough to have its own library, working papers might have been retained there, but they seldom made their way to the main campus library or its card catalog. Authors would cite working papers, but locating a copy even a few months after its publication could be very challenging.

With the introduction of the Internet and Gopher (the text-only predecessor to the Web) on campuses in the early 1990s, two librarians at the University of Minnesota (Pat Rodkewich, now retired, and Louise Letnes) had an idea about making these elusive working papers available to researchers. In 1995, working with local IT staff and colleagues in the neighboring states of Wisconsin and North Dakota, they began distributing Word Perfect versions of working papers via a Gopher server (Rodkewich and Letnes, 1995).

AgEcon Search is currently running on DSpace, which is its 4th software platform, including Gopher. It is operated in conjunction with University Digital Conservancy, <http://conservancy.umn.edu>, which is the University of Minnesota's institutional repository.

AgEcon Search is run, to use the language of economics, as a distributed network. Each participating group contributes to the effort by either uploading all of their own papers, or providing the funds to have a student worker at the hub at the University of Minnesota do the task for them. This keeps the overhead very low. Several small grants have been obtained to do special projects, often involving digitizing of older materials or uploading large numbers of papers from groups without the personnel to accomplish such a task.

Connections with other like-minded groups have helped the project grow and develop over the years. Since 2007, AgEcon Search has been a SPARC Partner, in the category of SPARC Scientific Communities. SPARC, <http://www.arl.org/sparc/> the Scholarly Publishing and Academic Resources Coalition, is an international alliance of academic and research libraries working on issues of scholarly publishing.

In late 2008, AgEcon Search began supplying citation information and links to full text to RePEc, <http://repec.org/> Research Papers in Economics. RePEc is a large database of citations to working papers, conference papers and journal articles covering all aspects of economics. When available, RePEc links to the full text of materials. They in turn cooperate with EconLit, a commercial database produced by the American Economic Association.

Aspects that are easier in a subject repository

Both authors have some involvement in the operation of the University Digital Conservancy, the University of Minnesota's institutional repository. Working on both projects, the differences have become more apparent to the authors, and noting them may serve to help others who work in one arena or the other.

A number of factors have made creating and sustaining a subject repository an easier undertaking than doing the same for an institutional repository. An obvious one is the fact that AgEcon Search covers just one subdiscipline, although a variety of work is done under the umbrella of agricultural and applied economics. Papers cover diverse topics, including farm management in Kenya, energy costs in the tourism industry, and environmental policy in Costa Rica. Research reports may contain a multitude of formulas, or results from a qualitative survey. But researchers, policy makers and students in this field share many understandings, including publication practices that lend themselves to the use of a subject repository.

Economists have a pre-print culture, which is to say that prior to the publication of an article in a peer-reviewed journal, the results of a research project are very likely to have appeared in a more preliminary but full-length form as either a working paper, research report, or conference paper. The most widely known example of a discipline with a pre-print culture is probably physics, with its subject repository arXiv, <http://arxiv.org/>. Till (2001) describes the concept and how it exists in some fields and not others.

As with other disciplines, reports of some research projects never make it to the journal article stage, and the only record of the work is the pre-print version. But unlike some other disciplines, pre-prints in economics are viewed as credible accounts of current research. The peer review standards are not the double blind ones employed by the respected journals, but working papers, research reports, and conference papers all undergo some peer evaluation process. Since these papers and reports precede the associated journal article by months or years, pre-prints are the source for the most recent research in the field.

Some subject repositories struggle with the policy about who is allowed to contribute material, and how it is reviewed. Since AgEcon Search only accepts papers that are already part of a series or journal, it is not a concern.

In economics, the characteristics of the publication cycle have lead to the establishment of two very large free Web resources covering the field as a whole. RePEc is one, and the other is Economists Online, <http://www.economistsonline.org/>, a newly established portal that highlights the work of scholars at Europe's top economics departments while also including citations and links to full text for thousands of other works. It is run by the Nereus consortium, <http://www.nereus4economics.info/>.

Although economists find working papers and conference papers useful, obtaining copies of them used to be quite problematic. They could be a frustration to librarians, too, so the existence of a subject repository that captured the pre-prints in agricultural and applied economics helped to solve an ongoing problem for both groups. As Hardesty (1995) notes, one key to success in projects involving the library and research faculty is the idea that librarians are helping to solve an existing problem.

Much to the consternation of some college and university administrators, researchers often have more loyalty to their discipline than to their institution. In the repository world, this may translate into more interest in subject repositories than institutional ones. The interest and ownership that some researchers have shown in AgEcon Search is not often seen in the world of institutional repositories. Journal editors are anxious for their publication to join with others of similar topics, and researchers seemed increasingly interested in AgEcon Search once there was a critical mass of papers. Students seem especially drawn to the free access to full text that will continue to be available after they have graduated. The same is true for researchers and policy makers in the developing world, where access to the journal literature can be quite limited.

On occasion, a friendly competition between groups has helped in recruiting efforts. Solicitations to potential participants always include download statistics from a comparable group, whether it is directed to the head of an academic department, the editor of a journal, or the president of a professional society.

Beyond their publication practices, economists have a professional interest in intellectual property, and some study the economics of publishing in general (Getz 2005) and open access in particular (Houghton and Sheehan 2009).

Communicating with current and potential submitters is relatively easy when many of them belong to the same professional organizations, attend the same meetings, and subscribe to the same listservs. Items are regularly submitted to the newsletters of the 4 or 5 biggest professional organizations, announcing new features, encouraging use, and recruiting new groups to submit.

AgEcon Search's narrow subject focus allows communication in ways that would be much less fruitful in a setting covering many unrelated topics. A weekly e-mail alert is distributed, listing 15-20 new papers, with a link to the full text, reaching about 3000 people. Occasionally a short news item is included at the beginning ("Stop at the booth at the upcoming meeting in Ghent" or "Download statistics now available for each paper"). A news blog is maintained, at <http://blog.lib.umn.edu/aesearch/ageconsearch/> and blogging from the professional meetings is a new feature.

The most effective method for getting new groups to add their papers to AgEcon Search is to have an exhibit at the professional meetings of agricultural economists, talk to individuals about the possibility, and follow up with e-mails or phone calls after the meeting. This type of activity is just not possible for an institutional repository covering a myriad of subjects.

The professional organizations have been invaluable to the growth of AgEcon Search in another way – by either passing along the full papers associated with talks presented at their meetings, or requiring that authors submit their papers beforehand. In some cases, authors are told that if they do not submit the paper by the deadline, they will be dropped from the program. Again, there is no comparable activity in the institutional repository world.

Agricultural economists, especially those whose professional societies actively support AgEcon Search, seem to be quite connected to the project. They introduce students and young colleagues to it and send unsolicited comments and notes of thanks. They appear to see submission to AgEcon Search as a way to contribute to their profession. Thomas and McDonald (2007) studied participation patterns of authors in several repositories, including AgEcon Search, and found there to be many authors, although few (about 10%) with more than 5 papers in the repository. While they know that the co-coordinators of the project are both librarians, AgEcon Search is not viewed as a library project.

Being focused on one subject area has allowed recruitment of members for a small advisory board who have turned out to be great champions of the project. Since the profession is rather small, all the board members know each other and have worked together in the past. On the library side, there is a small professional organization, the Agricultural Economics Reference Organization, whose members have been invaluable in providing input and outreach for AgEcon Search over the years.

In AgEcon Search, the narrow focus of topics covered has made it easier to pursue grant funding for a few special projects. The subjects covered are a close match with the mission of the Farm Foundation, which funded a few small projects, and funds for two small digitization endeavors came from the AgNIC, <http://www.agnic.org/> and the National Agricultural Library.

While there is neither time nor personnel to include multiple, narrowly focused subject headings in the metadata for each paper, one or more of the 30 AgEcon Search "broad

categories” is added to each record. Examples are agribusiness, environmental economics and policy, and land economics/use, and submitters are required to select at least one as part of the submission process.

Aspects that are harder in a subject repository

Lest readers think everything is better in a subject repository versus an institutional one, here are some of the areas that may be more challenging.

First, there is the logistics of working with 215 different groups who are not part of the same administrative structure, and do not necessarily speak the same language or have the same reasons for contributing. Some have a person in charge who simply decides that they will begin adding their papers, while others must get the approval of the editorial board or several government officials, or the entire membership has to take a vote. The main contacts move on to other positions, which means orienting a new person to the system. And as additional contributors are recruited, each new one adds to the complexity.

Authors not only move from institution to institution, but in any given year, they may submit papers from their academic department and the meetings of two different professional organizations, plus two journals might add papers which the person has authored. Authority control in this environment is difficult, and it taxes the limits of software that was built for a one-institution setting.

Managers of institutional repositories often lament that fact that self-archiving is much less common than mediated submission (Swan and Brown, 2005), but having over 1300 registered submitters means fielding many questions via e-mail and phone about the idiosyncrasies of the system. And the system and e-mail traffic can be huge on a day before a deadline for a meeting with over 400 papers due. Again, the software was not built with this in mind.

Currently, all documents in AgEcon Search are in PDF, so having papers in multiple languages and alphabets does not cause much problem. It can be a challenge to be sure that instructions, FAQs, and submission forms are clear and designed with all readers and submitters in mind. If the submitters were all on campus and an in-person meeting could easily be scheduled, less attention might need to be paid to this area.

Conducting follow-up conversations via e-mail is usually quite satisfactory, but as noted earlier, making initial contacts in person seems to be the most effective method to get new groups to begin submitting their papers. The narrow window – just a few days once or twice a year at a professional meeting – is very different than being on the same campus 5 days a week. And securing travel funds is a constant challenge.

When recruiting new materials, there is no legion of department liaisons, perfectly positioned to recruit. Nor are there dedicated technical services staff to work on ingest or metadata issues.

Journals, which have become a more and more important component of AgEcon Search, are challenging in many ways. Typical open source software used for institutional repositories is not designed to deal easily with volumes, issues, and displaying individual articles in page number order. Customization of the software is a possibility, but that often comes at a cost, both in the long and short runs. Some journals have embargos of 1 or 2 years which need to be tracked. Journals that are with commercial publishers have to try and negotiate for the release of backfiles, and groups have found out that they did not have certain rights that they thought they did.

Opportunities and future plans

With all its challenges, there are also many opportunities ahead in the development of AgEcon Search and other subject repositories. They include capitalizing on the mandates of funders to make research results available, working with journals, reaching out more to researchers in the developing world, expanding the categories of material included, adding the capability for communication and community activities, and continuing to cooperate with other large databases.

More funding agencies are considering the example of the National Institutes of Health in the US and the Wellcome Trust in the UK in requiring articles resulting from projects they support to be freely available in a matter of months after their publication. The rapid growth and heavy usage of PubMed Central, <http://www.ncbi.nlm.nih.gov/pmc/>, could be the future for subject repositories in other subject areas, should new mandates become reality.

During the last year, the largest growth in AgEcon Search has been in the area of small journals, mainly from outside the United States. Some were actively recruited, but others were not. Subject repositories may be a good fit for small journals looking for an alternative to going with a large commercial publisher. Pinfield (2009) recently suggested that open access journals and open access archives should consider more avenues to collaborate.

Taking a print-only journal into the electronic world can be a costly and labor-intensive task. AgEcon Search can offer a relatively simple solution, especially now that creating PDFs and even scanning are no longer huge barriers. In the developing world, the printing of each issue of a journal may be dependent on an individual grant, so having a free venue to upload papers can make a huge difference in the dissemination and even viability of a journal.

The use of AgEcon Search is very high in the developing world, especially as access to the Internet becomes more widely available and reliable. Expanding use and recruiting more material produced by researchers and students in developing countries is a

priority for the project. Scholars in the developing world are anxious for their work to be more widely available, and scholars in the developed world are anxious to have better access to it. The situation is likely the same in many disciplines.

In trying to solve the initial local problems of access and bibliographic control, it was not clear that these solutions would benefit others far outside the region. The first solid evidence about use outside the United States came from a user survey in 2005, seeking input to help inform plans for a software upgrade. Many respondents were from Africa, Russian, and Eastern Europe, and that trend has been borne out by more recent download and page hit information. A recurring idea is to establish an AgEcon Search "hub" in Africa, Eastern Europe, or Asia, to promote use of AgEcon Search and to recruit more material from the surrounding area.

Inspired by both an interest in developing world documents and the growing movement to make documents as freely available as possible, AgEcon Search has just begin soliciting dissertations and theses, particularly from outside North America, and pre-prints and post-prints of journal articles, to the extent that they are allowed. A pilot project is underway to include the Masters papers from the Collaborative Masters Program in Agricultural and Applied Economics (<http://www.agriculturaleconomics.net/home/home.asp?00=1>), which involves several universities in sub-Saharan Africa.

The expansion of institutional repositories in many parts of the world may mean that subject repositories may be able to harvest dissertations and theses as well as other documents from them, creating a critical mass of easily-searchable documents on a particular topic.

Including data sets in AgEcon Search is a topic that continues to come up, and possibilities in that area are being monitored. The current software platform, DSpace, was not designed with data as its main file type, so data-related issues will be carefully investigated before the migration to a new platform.

Subject repositories with a regular following of researchers and students in a particular discipline seem like ideal venues to experiment with online communication tools. The development of EthicShare, <https://www.ethicshare.org/> which is also based at the University of Minnesota, is being closely watched. EthicShare is a Web space that incorporates public and private shared workspace with citations to the literature, as well as news updates and other communication options.

The collaboration between AgEcon Search and RePEc has been beneficial to both projects, and through RePEc, links to full text are being supplied to EconLit. In the future, relationships with other large databases in the area of economics will be investigated and pursued.

Literature cited

Getz, M. (2005) "Open Scholarship and research universities" *Vanderbilt University, Department of Economics Working Paper* No. 05-W17, May, 21p.
<http://www.vanderbilt.edu/Econ/wparchive/workpaper/vu05-w17.pdf>

Hardesty, L (1995) "Faculty Culture and Bibliographic Instruction: An Exploratory Analysis" *Library Trends*, Vol. 44, No. 2, Fall, 339-367

Houghton, J. and Sheehan, P. (2009) "Estimating the Potential Impacts of Open Access to Research Findings", *Economic Analysis and Policy*, Vol. 39 No. 1, March, 127-142. <http://www.eap-journal.com.au/download.php?file=696>

Letnes, L. and Kelly, J. (2002) "AgEcon Search: Partners Build a Web Resource." *Issues in Science and Technology Librarianship*, no.34, Spring, <http://www.istl.org/02-spring/article3.html>

Pinfield, S. (2009) "Journals and repositories: an evolving relationship?" *Learned Publishing*, Vol. 22, No. 3, July. 165-175. <http://eprints.nottingham.ac.uk/1092/>

Rodkewich, P. and Letnes, L. (1995) "AgEcon Search: Research in Agricultural Economics, Working Papers on the Internet." *Journal of Agricultural and Food Information*, Vol. 3, No. 2, 1995, 23-29.

Swan, A. and Brown, S. (2005) *Open access self-archiving: An author study* (Key Perspectives Limited, Cornwall, UK), <http://cogprints.org/4385/1/jisc2.pdf>

Thomas, C. and McDonald, R. (2007) "Measuring and Comparing Participation in Digital Repositories" *D-Lib Magazine* Vol. 13, No. 9/10.
<http://www.dlib.org/dlib/september07/mcdonald/09mcdonald.html>.

James E. Till. (2001) "Predecessors of preprint servers" *Learned Publishing*, Vol. 14, No. 1, 7-13. <http://arxiv.org/abs/physics/0102004>