Expanding Access to the Agricultural Resource Management Survey: USDA’s Experience with Remote Data Analysis Using the NORC Data Enclave

2014 Federal Committee on Statistical Methodology (FCSM) Statistical Policy Seminar

“Virtual Data Access” for Statistical and Research Purposes
December 15-16, 2014

Mitch Morehart– Economic Research Service, USDA
What is the Agricultural Resource Management Survey?

The ARMS is critical to the research and analysis mission of the Economic Research Service, and is a key input to official estimates across the Department of Agriculture and in other federal agencies.

It is a valued and unique resource, since it is the only national survey from which observations of field level farm practices, the economics of the farm business and the characteristics of the household operating the farm, are all collected annually in a representative sample.
Wide Variety of Data Users/Uses Reflect Importance of ARMS

Major Data Users Include:

- Office of the Secretary of Agriculture
  Chief Economist
- USDA Agencies
  Farm Services Agency
  Natural Resources Conservation Service
- Government Agencies Outside USDA
  Bureau of Economic Analysis (Commerce)
- Congressional Committees, Agencies, and Staff
  Agriculture Committees
  Congressional Research Service, Congressional Budget Office
- Customers Outside Government
  Farm Credit, Farm Groups (such as Farm Bureau)
- Academic Researchers

According to a Google Scholar search (October 2014) using “USDA ARMS data”, 23,800 documents appeared.
External Peer Review Drives Development

“ARMS is an invaluable source of information on the current state of American agriculture, as well as the sole source of some important information on the linkages between fields, farms, and families that serves to illuminate the challenges faced by agriculture policy makers and farm families. Because the survey is so critical to understanding agriculture, it carries a special burden. Its methods, practices, and procedures must be designed to yield data of impeccable quality in view of their uses, and the data must be made available to the research community both inside and outside the federal government in order to generate the improved analytical knowledge the data makes possible.”

Committee on National Statistics (CNSTAT)
“The Advisory Committee recommends conducting a pilot study of making the ARMS micro data available through the National Institutes of Standards and Technology and National Opinion Research Center (NIST-NORC) data enclave”
Prior to participation in the enclave there were two primary ways to get ARMS data:

- **Access summary data on the ERS website**
  ...variety of information updated twice annually, interactive ARMS data tool provides custom tables

- **Sign-up & travel to ERS or NASS Remote Data Center**
  ... **Approved researchers and projects gain access to micro data files**
PILOT PROJECT TO EXPAND ARMS DATA ACCESS

In 2006 ERS entered into a cooperative agreement with NORC to “improve the ability of internal and external stakeholders to access and understand ARMS data and products while protecting the underlying microdata.”

- The agreement was entered into in the spirit of a congressional mandate to preserve and enhance the unique partnership that which exists in food and agricultural research, extension, and instruction between the Federal Government, the governments of several States, institutions and other organizations.

- The overarching objective was to improve the efficiency and effectiveness with which sensitive ARMS data are managed and accessed. The project involved providing geographically dispersed researchers secure remote access to ARMS data and technical support services.
What is a Data Enclave?

• Secure environment for accessing sensitive data
  • **Access**: remote desktop, encryption, audit logs
  • **Security**: controlled information flow, group isolation
  • **Facility**: software, tools, collaborative space, technical support
ERS & NASS Data Dissemination Goals Consistent with NORC Data Enclave Mission

- Promote access to sensitive microdata
- Protect confidentiality (portfolio approach)
- Archive, index and curate micro data
- Encourage researcher collaboration
- Leverage cutting edge technologies (e.g., metadata preparation & documentation best practices)
Enclave Fosters Multi-layered Protection Approach

Utility Confidentiality

Technological Protection
Statistical Protection
Operational Protection
Legal Protection
Partnership with NORC has encouraged innovation
Technical Protection

- Encrypted connection with the data enclave using virtual private network (VPN) technology. VPN technology prevents outsiders from reading the data transmitted between the researcher’s computer and NORC’s network.
- Users access the data enclave from a static or pre-defined narrow range of IP addresses.
- Citrix Web-based security interface.
- All applications and data run on the server at the data enclave.
- Data enclave can prevent the user from transferring any data from data enclave to a local computer.
- Data files cannot be downloaded from the remote server to the user’s local PC.
- User cannot use the “cut and paste” feature in Windows to move data from the Citrix session.
- User is prevented from printing the data on a local computer.
- Audit logs and audit trails
Researcher Collaboration Incentives

• Post announcements on the portal homepage
• Postings to the calendar
• ARMS discussions
• Library additions
• Post FAQs
• Contributions to the Wiki
• Add/refine existing new documentation
• Identify new (open source) software
• Assist fellow researcher in code / statistical analysis / software
• IM / chat / sharing attachments, code, other knowledge, help desk questions
• Share imported datasets, macros, derived/constructed variables

• Code sharing
• Database enhancements such as variable documentation and other meta data enhancements.
• Post/update researcher bios/project descriptions for collaborative and website
• Journal articles/papers on DE data analyses
• Webinar presentation (to fellow enclave researchers /larger USDA + audience)
• Conference presentations on DE data analyses
• Other work disseminated from the enclave

Reduce fees based on amount of collaborative activities
Enclave Scholarships

Promote wider access to institutions and researchers who otherwise could not afford to utilize the data.

The Economic Research Service (ERS) and National Agricultural Statistical Service (NASS) are pleased to announce the availability of the 2010-2011 Agricultural Resource Management Survey (ARMS) Scholarship Program, the recipients of which will receive a one year scholarship ($5,200 value) to be applied toward researcher access to the NORC Data Enclave, a confidential, protected remote access platform within which authorized researchers can access ARMS data. NORC Data Enclave already provides secure access to the ARMS data for professors and other senior analysts working on approved research projects. More information about NORC Data Enclave can be obtained from http://www.norc.org/DataEnclave/.

The deadline to apply for this scholarship opportunity is November 15, 2010, by 5:00 p.m. Eastern Standard Time. Applications are to be evaluated on the quality and relevance of the research proposal and are geared towards institutions that have not previously received a USDA scholarship to access the Enclave. Please include a brief project description and objective, a detailed description of the data needed, budget, dissemination plan, and the planned uses of the data.
The USDA Experience

Instead of creating additional costly brick and mortar RDCs, USDA can now roll out a virtual RDC to any university in the country.

Geographically dispersed researchers travel to secure Research Data Centers (RDCs)

Thin client terminals are installed in secure locations at researchers’ universities
# Current USDA Scholars

<table>
<thead>
<tr>
<th>Project Leader</th>
<th>Organization</th>
<th>Co-researchers</th>
<th>Location</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>George Frisvold</td>
<td>Univ. of Arizona</td>
<td>N/A</td>
<td>NORC</td>
<td>Glyphosate resistant weeds effects on corn &amp; cotton production</td>
</tr>
<tr>
<td>Jennifer Ifft</td>
<td>Cornell</td>
<td>N/A</td>
<td>NORC</td>
<td>Farm financial management and government policies</td>
</tr>
<tr>
<td>Barry Goodwin</td>
<td>NC State</td>
<td>N/A</td>
<td>NORC</td>
<td>Measuring the impact of crop diversification on farm risk</td>
</tr>
<tr>
<td>Mitch Renkow</td>
<td>NC State</td>
<td>Robert Dinterman</td>
<td>NORC</td>
<td>Fed. broadband loans to internet providers effects on farmers</td>
</tr>
<tr>
<td>Azzeddine Azzam</td>
<td>Univ. Nebraska-Lincoln</td>
<td>Kate Brooks, Brian Mills</td>
<td>Lincoln</td>
<td>Examine vertical economies of scope in cow-calf operations</td>
</tr>
<tr>
<td>Dan Black</td>
<td>Univ. Chicago</td>
<td>Sam Seo</td>
<td>NORC</td>
<td>Investigate farmer welfare effects from contracting</td>
</tr>
<tr>
<td>Nattai Bergman</td>
<td>MIT</td>
<td>Richard Thakor</td>
<td>NORC</td>
<td>Temporary shocks, long-run effects, role of credit and collateral</td>
</tr>
<tr>
<td>Paul Mitchell</td>
<td>Univ. Wisconsin</td>
<td>Fengxia Dong</td>
<td>NORC</td>
<td>Impact of sustainable practices adoption on profitability</td>
</tr>
<tr>
<td>Richard Shumway</td>
<td>Washington State</td>
<td>Michael Brady</td>
<td>NORC</td>
<td>Emerging technologies adoption and producer decisions</td>
</tr>
<tr>
<td>Paul Adler</td>
<td>USDA/ARS Harrisburg</td>
<td>Hernan Bejarno, Chapman College</td>
<td>NORC&amp;ERS</td>
<td>Incentives for double-crop adoption for corn-bean rotations</td>
</tr>
<tr>
<td>Terrance Hurley</td>
<td>Univ. Minnesota</td>
<td>N/A</td>
<td>NORC</td>
<td>Food security and consequences of herbicide resistant weeds.</td>
</tr>
<tr>
<td>Kohei Ueda</td>
<td>Univ. Maryland</td>
<td>Erik Lichtenberg</td>
<td>ERS</td>
<td>Biotechnology, herbicide demand, and stream runoff</td>
</tr>
<tr>
<td>Lee Schulz</td>
<td>Iowa State</td>
<td>Joleen Hadrich-Colorado State</td>
<td>NORC</td>
<td>Composition of weight gain in pig production</td>
</tr>
<tr>
<td>Walter N. Thurman</td>
<td>NC State</td>
<td>Dallas Woods</td>
<td>NORC</td>
<td>Economic factors leading to herbicide resistant weeds</td>
</tr>
<tr>
<td>Lurleen Walters</td>
<td>Mississippi State</td>
<td>N/A</td>
<td>NORC</td>
<td>Community-focused Agriculture in the Southeast</td>
</tr>
<tr>
<td>Brad Barham</td>
<td>Univ. Wisconsin</td>
<td>JeanPaul Chavas</td>
<td>NORC</td>
<td>Heartland family farm diversification strategies</td>
</tr>
<tr>
<td>Michael Black</td>
<td>Univ Texas Austin</td>
<td>N/A</td>
<td>NORC</td>
<td>Environmental life-cycle assessment of organic crops</td>
</tr>
<tr>
<td>Barrett Kirwan</td>
<td>Univ. Illinois</td>
<td>N/A</td>
<td>NORC</td>
<td>Credit Demand and Changing Farm Policy</td>
</tr>
<tr>
<td>Mario J. Miranda</td>
<td>Ohio State</td>
<td>Meng Fen Yen</td>
<td>NORC</td>
<td>Credit, savings, and insurance in risk management</td>
</tr>
<tr>
<td>Keri Jacobs</td>
<td>Iowa State</td>
<td>John Miranowski</td>
<td>NORC</td>
<td>Intensive corn management practices in a dynamic setting</td>
</tr>
<tr>
<td>Jeff Gillespie</td>
<td>Louisianna</td>
<td>N/A</td>
<td>NORC</td>
<td>Competitiveness of US beef, poultry, and dairy industries</td>
</tr>
<tr>
<td>Carol Barford</td>
<td>Univ. Wisconsin</td>
<td>N/A</td>
<td>Wisc. Office</td>
<td>Energy intensity of crop and dairy farming in Wisconsin</td>
</tr>
<tr>
<td>Valentina Hartarska</td>
<td>Auburn</td>
<td>Denis Nadolnyak</td>
<td>NORC</td>
<td>Climate change risk and financing constraints</td>
</tr>
<tr>
<td>Bruce Ahrendsden</td>
<td>Univ. of Arkansas</td>
<td>Bruce Dixon, Charles Dodson</td>
<td>NORC</td>
<td>Borrower Characteristics of FSA Guaranteed Loans</td>
</tr>
<tr>
<td>Joleen Hadrich</td>
<td>Colorado State Univ.</td>
<td>Kamina Johnson, APHIS</td>
<td>NORC</td>
<td>Risk diversification on dairy farms</td>
</tr>
</tbody>
</table>

Active ARMS MOUs and project agreements, Sept. 2014
"The enclave is easy to work with and convenient for researchers...Using the ARMS data gave me the opportunity to conduct a study that otherwise would be been impossible."
Next Steps

• Continue to apply practical lessons learned / feedback for continuous improvement
• Continue to innovate
• Identify and implement new collaborative functionality
• Continue to pilot test ARMS Data Enclave remote training
• Explore possibilities for Data Enclave Regional Data Research Centers for graduate students
• Add more microdata to the enclave that enhances access to unique and valuable ERS and NASS resources
Summary

• Goal: To promote access to sensitive micro data while also protecting confidentiality

• Benefits:
  • Secure, low-cost approach to data sharing
  • Applicable and customizable to agency needs and requirements
  • Engagement of broad community of researchers

• Challenges:
  • Building trust among data providers
  • Developing a user community
  • Keeping up with IT evolution
Contact Information

Mitch Morehart
morehart@ers.usda.gov