

2008 ACADEMY OF MANAGEMENT WORKSHOP:

Creating New Access to Business Data using Virtual Organizations

INSIDE THIS ISSUE:

NIST/TIP	2
RESEARCH UPDATE	
USDA ERS/NASS RESEARCH UPDATE	2
KAUFFMAN RESEARCH UPDATE	3
AOM LEAD STORY CONTINUED...	4
UPDATE FROM NORC	5
UPCOMING PRESENTATIONS	5

SPECIAL POINTS OF INTEREST:

- Data Enclave Version II in December '08
- USDA ERS NASS Pilot Testing Remote Training Platform
- Kauffman, NIST and Data Enclave at 2008 AOM
- Metadata Workshop

A professional development workshop was held at the 2008 AOM meetings to engage the management research community and to get their scientific advice on the structure and content of the Data Enclave. Researchers also were asked to identify existing datasets that could be targeted to include in remote access facilities.

Workshop organizers included data producers (Kauffman Foundation and NIST/TIP), enclave managers (NORC/University of Chicago) and researchers (Alicia Robb and Julia Lane).

Attendees had a number of ideas for **enclave structure**. They felt it should be more than a clearinghouse for researchers to access datasets. The enclave should be flexible enough to meet the needs of different users, both senior and junior researchers, and could be structured in two ways to meet this need: one for datasets that require strict security and one for less sensitive, public use type data.

The less security intensive, public use-type platform could focus on building collaborative functionality and open source software. This large repository could also serve as a one-stop shop for web-based remote access to well organized datasets, real peer-to-peer collaboration and small group-to-group collaboration. This less security intensive, public use-type data platform could potentially provide a mechanism for hundreds of

Ph.D. students per year to define their research questions and complete their dissertation work in a cooperative environment.

The AOM attendees also had a number of suggestions about **incentives** that could be put in place to encourage researchers to collaborate on content and building a data infrastructure. These included the creation of citation standards for people who provide data and code (see, e.g. <http://five.dartmouth.edu/datasets/conditions>).

They recommended having regular conferences to promote collaboration among researchers as well as identifying and celebrating "master researchers" who provided new ideas and guidance. They suggested creating geographical/industry/subject foci on the portal to connect researchers studying similar topics and expanding the listserv to provide information about targeted issues.

Another suggestion was to create an online journal, like STATA journal, to which researchers can contribute and get publications – and have people review contributions (use a star rating system like Amazon.com). There was a vigorous debate about the merits of providing rebates to reduce costs if people contribute, and creating a marketplace to exchange datasets or add/enhance existing datasets. Finally, they recommended developing tagging tools to index datasets and provide links (with permission) to who tagged the data (*cont. on page 4*)

DATA ENCLAVE VERSION II: DECEMBER '08 ROLLOUT

In December, NORC will roll out the second generation version of the enclave. This will include a number of enhancements intended to improve the current researcher environment. For example, researchers will now have the ability to interact with one another in a secure manner while in the enclave via peer-to-peer and small group-to-group instant messaging. Researchers also will be able to share files with one another through the soon-to-be implemented Office Communication System (OCS).

The new platform will provide functionality for broad based collaboration, such as group presentations and webinars. Researchers may also see one another by using personal webcams. Another major

change will involve replacing the existing technical support mechanism with a dynamic Help Desk functionality that will improve the work flow process considerably.

Another major enhancement involves a redeployment of the existing mechanism for importing and exporting files from the enclave. Among other utilities, this new mechanism will improve the existing disclosure review process and make it more efficient and transparent to researchers. For example, researchers will be able to track the progress of their Help Desk requests and import and export requests in real time and get a clear sense of when their requests may be ready for export.

NATIONAL INSTITUTE OF STANDARDS & TECHNOLOGY: TECHNOLOGY INNOVATION PROGRAM

DATA MATCHING EXERCISE: ALL COMPANY DATA IN THE BRS, PPS, 3 APPLICANT SURVEYS, & JV SURVEY

As a result of TIP's new information collection requirements to implement more streamlined instruments for impact assessment, in the future a greater emphasis will be placed upon matching relevant information from outside data sources for impact assessment. NORC will work with TIP to perform a matching exercise for existing data at the Data Enclave to provide researchers a richer dataset while protecting the confidentiality of past awardees' data.

NORC will assist in distilling recommendations and best practices from the matching exercise which will inform TIP collection efforts for future data matching and collaboration with other federal agencies and organizations addressing impact assessment. The data matching exercise will at once maintain anonymity of data and include counts of papers, patents, etc., as well as citation counts and patent classes. The matching will be for all companies in the Business Reporting Survey (BRS) and Post Project Surveys (PPS), all companies in the three applicant surveys (awardees and non-awardees), all companies in the Joint Venture (JV) survey, and person-level matching from the key personnel listed in the

latest BRS version.

OUTREACH TO EVALUATION & SCIENTIFIC POLICY COMMUNITY

NORC will work with TIP to develop and facilitate a mechanism for TIP's Impact Assessment Office (IAO) to collaborate with other agencies and organizations and their impact assessment efforts. This effort is intended to provide a gateway for soliciting input for program development and engaging the larger Science & Technology policy community. Outcomes and impacts are inseparable from the scientific challenges of a given critical national need (CNN). TIP needs to develop appropriate methodologies to assess outcomes and impacts across divers CNNs.

Outputs will include:

- (1) Matched data for legacy projects & recommendations to assessing the needs for future matching efforts
- (2) Outreach (e.g., conference) to larger evaluation and scientific policy community and
- (3) Summary of findings from outreach to be in-



USDA NATIONAL AGRICULTURAL STATISTICS SERVICE & ECONOMIC RESEARCH SERVICE UPDATE

NEW DEVELOPMENTS: USDA DATA ENCLAVE

New Remote Training Platform. USDA/NASS/ERS recently piloted the first remote enclave training session with a group of four researchers, including Ani Katchova, Ph.D., Assistant Professor, Agricultural Economics Department, University of Kentucky; Ashok Mishra, Ph.D., Associate Professor Department of Agricultural Economics and Agribusiness and Editor of the Journal of Agricultural and Applied Economics; Helen Jensen, Ph.D., Professor Department of Economics, Iowa State University Center for Agricultural & Rural Development; and Fengxia Dong, Ph.D., Associate Scientist, Iowa State University Center for Agricultural & Rural Development. The remote trainees accessed training materials from a secure NORC FTP site and participated in the remote training at the State field offices in Kentucky, Louisiana, and Iowa, respectively, where they watched pre-recorded training videos. Thanks very much to the researchers for pilot testing the new training platform and to the NASS State field staff at each of the sites for their assistance in making the training sessions such a great success.

University-based Data Enclave Secure Workstation. USDA/NASS/ERS is in the process of pilot testing the development and implementation of a Univer-

sity-based Data Enclave secure workstation at Oregon State University's (OSU) Department of Agricultural & Resource Economics. NORC and ERS/NASS are collaborating with Christian Langpap, Ph.D., Assistant Professor who will oversee the effort in concert with Oregon State Field Office Director Chris Mertz. A secure room has been identified and the workstation is in the process of being configured. Kandice Kleiber, Ph.D. student and Graduate Teaching Assistant in the Department of Economics at OSU, traveled to NORC's Bethesda Office to be trained and certified as an authorized enclave researcher and will serve as the student trainee throughout the duration of the pilot program.

Coming in November: New (2007) & Revised (all other years) ARMS Data

New ARMS data will be placed in the Data Enclave and will be available for use where authorized in individual projects agreements. The specific changes are:

- * 2007 data for farm finances, households, and costs of production and
- * Updated data files from previous years.



UPDATE FROM THE EWING MARION KAUFFMAN FOUNDATION

Ewing Marion
KAUFFMAN
Foundation

KAUFFMAN HOSTS RECEPTION AT 2008 AOM MEETING IN ANAHEIM

The Ewing Marion Kauffman Foundation hosted a reception at the Hilton Anaheim on August 8 from 5-8pm. Newly trained KFS enclave researchers, Kauffman staff, NORC Data Enclave staff, Professional Development Workshop attendees, and other AOM program attendees participated.

KAUFFMAN HOLDS DATA ENCLAVE TRAINING AT 2008 AOM MEETING

On August 8, four new KFS scholars traveled to Anaheim California and participated in a 1-day Data Enclave training session conducted by Julia Lane, Tim Mulcahy, and Alicia Robb at the 2008 AOM meeting in Anaheim, CA. New KFS researchers include Charlie Braymen (Kansas State Univ.), Harold Fried (Union College) and Loren Tauer (Cornell Univ.), David Robinson (Duke Univ.), and Robert Scott (Monmouth Univ.).

PROFESSIONAL DEVELOPMENT WORKSHOPS AT AOM 2008

Also on August 8th, Julia Lane and Alicia Robb held a professional development workshop on *Creating New Access to Business Data using Virtual Organizations*. The workshop included discussion on new ways that are being developed to promote researcher access to business microdata – particularly researchers who are geographically dispersed.

New modalities are being developed both nationally and internationally to promote researcher access to business microdata – particularly remote access enclaves. Two recent workshops, one NSF funded and in Ann Ar-

bor Michigan, and one in Nuremburg Germany brought together data producers and enclave managers to discuss the best ways to promote such access. This workshop engaged the research community to find out what works best for them in terms of structure and content, as well as what kind of datasets exist that could be targeted to include in remote access facilities (see lead story).

Workshop organizers included data producers (Kauffman Foundation and NIST/TIP), enclave managers (NORC/University of Chicago and IAB/Germany), researchers (Alicia Robb and Julia Lane).

FOUR ADDITIONAL KFS ENCLAVE RESEARCHER SCHOLARSHIPS AWARDED

The Kauffman Foundation made four new awards last quarter. Recipients included:

Susan Coleman, Ph.D., the Ansley Professor of Finance at the University of Hartford's Barney School of Business;

Kristie Briggs, Ph.D., Assistant Professor of Economics at Creighton University;

Andrea Asoni, Ph.D. student in Economics at the University of Chicago; and

Roni Kisin, Ph.D. student in Finance in the Graduate School of Business at the University of Chicago.

2008 KAUFFMAN DATA SYMPOSIUM ON ENTREPRENEURSHIP AND INNOVATION DATA NOVEMBER 21, 2008

(BY INVITATION ONLY)

Kauffman Symposia on Entrepreneurship and Innovation Data are a multi-year series of workshops focused on the important and growing body of data collected on entrepreneurship and innovation. The 2008 Symposium theme is "Proposals to Advance Measurement of Entrepreneurship and Innovation in the United States." Twenty presentations will be made highlighting different proposals for advancing our understanding of entrepreneurship and/or innovation through modifications to existing data collection tools and methodologies. Although the Data Symposium has reached full capacity, proceedings from the meetings will be posted for those who are interested at <http://www.kauffman.org>.

KAUFFMAN HOSTS WEBINAR TO SHOWCASE PRELIMINARY RESULTS FROM KFS ENCLAVE RESEARCH

On Wednesday, October 8th, 2008, the Kauffman Foundation conducted a webinar moderated by Kauffman Fellow and KFS Principal Investigator Alicia Robb and E.J. Reedy, Kauffman's Research and Policy Manager. The webinar showcased preliminary results from five KFS researchers, including:

Winston Smith (Temple University): "*New Firm Innovation*"

David Robinson (Duke University): "*A New Pecking Order*"

Susan Coleman (University of Hartford): "*Gender and Financing*"

Ethan Lewis (Dartmouth College): "*Labor Market Endowments and Firm Performance*" and

Jared DeLisle (Florida State University): "*Capital Structure of New*



CREATING NEW ACCESS TO BUSINESS DATA USING VIRTUAL ORGANIZATIONS (CONTINUED FROM P1)

The role of the enclave as a **data disseminator** was the focus of the third group. That group suggested that the enclave should create a clearinghouse web listing of all datasets available. They suggested that the enclave get information about membership groups & professional organizations since they have detailed information on their members. The group recommended the enclave get access to a wide variety of business data, like venture capital datasets, microfinance organizations data, Small Business Association (SBA) 8a and SBA Loan data, bankruptcy data and patent data. The enclave should also work to get researcher access to confidential versions of public use datasets, like the Survey of Small Business Finances, which include detailed geography codes and banking details for their suppliers; Census and other government datasets like the Longitudinal Database of Establishments (LBD) Covered by State Unemployment Insurance Programs; Survey of Business Owners (SBO) and unemployment insurance (UI), etc. from Census and BLS. The enclave could perform a useful public service by disseminating codebooks and other documentation for each dataset that will help users. We have found all of these ideas very useful as we expand the use and functionality of the enclave. If you have additional suggestions, please email us at dataenclave@norc.org.

REMOTE WEB-BASED DATA ENCLAVE TRAINING IN NOVEMBER

NORC will conduct a web based training (TBA) in November 2008. This 1-day remote training session will include new enclave researchers as well as current researchers who need to conduct their annual booster training.

For information on the date and time, please contact Tim Mulcahy at mulcahy-tim@norc.uchicago.org

NEW STATISTICAL SOFTWARE CURRENTLY BEING ROLLED OUT IN THE DATA ENCLAVE

At the request of many researchers, the Data Enclave has purchased licenses for Matlab, version 9 of LIMDEP, SAS version 9.2, R (<http://www.r-project.org/>) the open source statistical computing and graphics environment, and Stata version 10.

We are always open to new ideas and suggestions, so please let us know what software and analytical tools are of most interest to you in your research.



CONTACT: TIM MULCAHY

National Opinion Research Center
at the University of Chicago
4350 East-West Highway
Bethesda, MD 20814
Phone: 301-634-9352
Fax: 301-634-9301
E-mail: mulcahy-tim@norc.uchicago.edu

UK DATA ARCHIVE ADOPTS NORC DATA ENCLAVE MODEL FOR SECURE REMOTE DATA ACCESS

The UK Data Archive recently adopted NORC's Data Enclave model for providing secure remote access to micro data as part of its Secure Data Service (SDS). The UK Data Archive was awarded funds by the Economic and Social Research Council to build the SDS to support the development of the National Strategy for Data Resources for Research in the Social Sciences. According to the tender specifications, this service will have a number of functions. First and foremost, it will assist researchers who wish to make use of administrative data resources for research purposes to do so ethically and within a clear legal framework. Second, it will become a focal centre for knowledge about the availability of appropriate administrative data, their suitability for specific research purposes and the procedures required to gain access to and use such data. Third, it will work in tandem with government departments and agencies, seeking to develop and improve the use of administrative data resources for research purposes. The service will not hold or store such data, but will act as an intermediary between researchers and the organization(s) providing access to administrative data for research purposes.

The service will provide controlled access to sensitive and/or disclosive personal or organizational information which cannot be released for research purposes under End User License or Special License conditions. Access to data placed within the SDS will be managed and operated by the UKDA, according to the rules and conditions established by the individuals or organizations with responsibility for the guardianship of the data.

PRESENTATIONS, BACKGROUND REPORTS, PUBLICATIONS**SEPTEMBER 17, 2008**

Tim Mulcahy presented on the Data Enclave to the NORC at the University of Chicago's Board of Trustees.

SEPTEMBER 22 , 2008

Chet Bowie and Tim Mulcahy presented at the Washington Statistical Society (WSS) on recent developments on the Data Enclave.

RECENT DATA ENCLAVE PUBLICATIONS

"Data Access in a Cyber World: Making Use of Cyberinfrastructure" Julia Lane, Pascal Heus and Timothy Mulcahy, *Transactions in Data Privacy*, 2008

"Informing Public Policy on Science and Innovation: The Advanced Technology Program's Experience" Stephanie Shipp, Stephen Campbell, Timothy Mulcahy and Ted Allen, *Journal of Technology Transfer*, 2008 (in press)

UPCOMING EVENTS**DECEMBER 2, 2008**

The Next Frontier in Documenting Survey Data: What Data Producers and Archivists Need to Know about DDI, XML and SDMX

This free, 1-day workshop, sponsored by NORC at the University of Chicago, is intended for a non-technical audience and is targeted at data producers and archivists. The objectives of the workshop are twofold:

To provide an introduction to the basics of some emerging data documentation standards, such as the Data Documenta-

tion Initiative (DDI), eXtensible Markup Language (XML), and Statistical Data and Metadata Exchange Standard (SDMX) and To encourage adoption of these emerging data documentation standards.

This metadata workshop will cover the following topics:

- * Introduction to metadata and XML
- * The case for metadata standards: benefits, challenges, and the future of metadata
- * Introduction to the Data Documentation Initiative (DDI) for the documentation of survey and administrative micro data
- * Essentials of the Statistical Data and Metadata Exchange Standard (SDMX) for the documentation of aggregate and time series data
- * DDI & SDMX interoperability and other important metadata specifications
- * Major contributors, users, tools, ongoing activities, next steps

The workshop will be held at NORC's Bethesda, MD office at 4350 East-West Highway, Suite 800, from 9:30am to 3:30 pm.

RSVP by November 17, 2008 to

mulcahy-tim@norc.uchicago.edu

For more information, contact Tim Mulcahy at 301-634-9352